Proceedings Paper:
Song, L. and Lu, H. (2016) Proper Inner Product with Mean Displacement for Gaussian Noise Invariant ICA. In: Proceedings of The 8th Asian Conference on Machine Learning. The 8th Asian Conference on Machine Learning, November 16-18 2016, The University of Waikato, Hamilton, New Zealand. , pp. 398-413.

JMLR: Workshop and Conference Proceedings 63:398–413, 2016 ACML 2016

Proper Inner Product with Mean Displacement for Gaussian Noise Invariant ICA

Liyan Song LXS189@CS.BHAM.AC.UK
Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China
School of Computer Science, The University of Birmingham, Birmingham, UK

Haiping Lu HPLU@IEEE.ORG
Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China

Editors: Robert J. Durrant and Kee-Eung Kim

Abstract

Independent Component Analysis (ICA) is a classical method for Blind Source Separation (BSS). In this paper, we are interested in ICA in the presence of noise, i.e., the noisy ICA problem. Pseudo-Euclidean Gradient Iteration (PEGI) is a recent cumulant-based method that defines a pseudo Euclidean inner product to replace a quasi-whitening step in Gaussian noise invariant ICA. However, PEGI has two major limitations: 1) the pseudo Euclidean inner product is improper because it violates the positive definiteness of inner product; 2) the inner product matrix is orthogonal by design but it has gross errors or imperfections due to sample-based estimation. This paper proposes a new cumulant-based ICA method named as PIMD to address these two problems. We first define a Proper Inner product (PI) with proved positive definiteness and then relax the centering preprocessing step to a mean displacement (MD) step. Both PI and MD aim to improve the orthogonality of inner product matrix and the recovery of independent components (ICs) in sample-based estimation. We adopt a gradient iteration step to find the ICs for PIMD. Experiments on both synthetic and real data show the respective effectiveness of PI and MD as well as the superiority of PIMD over competing ICA methods. Moreover, MD can improve the performance of other ICA methods as well.

Keywords: Blind Source Separation, Noisy Independent Component Analysis, Cumulants, Inner Product, Pseudo-whitening

1. Introduction

Blind Source Separation (BSS) assumes that the observed data are drawn from unknown latent sources, and aims to recover these sources with the observations only. Independent Component Analysis (ICA) is a classical method for BSS (Hyvärinen and Oja, 2000; Hyvärinen, 2013). It assumes that the sources are mutually independent and they are linearly mixed. Standard ICA methods typically have three steps: (1) centering that shifts the sample mean to the origin, (2) whitening that removes the second-order statistics from the data, and (3) Independent Component (IC) estimation that recovers the latent sources. Most methods differ only at the last step. Conventionally, the observations are assumed to be noise-free. However, real-world observations often contain noise that deteriorates the performance of ICA (Voss et al., 2013, 2015). Thus, Hyvärinen et al. (2001) suggested the *noisy ICA* model that includes a noise term.

The noisy ICA problem is more challenging. In particular, the conventional whitening step fails for noisy ICA. In noise-free ICA where $\mathbf{x} = \mathbf{A}\mathbf{s}$, the *whitening* step employs the covariance matrix of the observations, denoted as $cov(\mathbf{x})$, to orthogonalize the mixing matrix $\mathbf{A}$ with its exact information via $cov(\mathbf{x}) = \mathbf{A}\mathbf{A}^T$. However, this covariance-based *whitening* does not work well in the presence of noise. In this case, $cov(\mathbf{x})$ contains not only the information from the sources but also that from the noise, i.e. $cov(\mathbf{x}) = \mathbf{A}\mathbf{A}^T + cov(\boldsymbol{\eta})$, posing difficulties on noise-free ICA methods. Therefore, *noisy ICA methods* have been developed to take noise into account.

The studies of *noisy ICA* can be dated back to the 1990s. One approach aims to remove or reduce the bias caused by noise so that slight modification would be enough for noise-free ICA methods (Hyvärinen, 1999a). Another approach uses probabilistic models with likelihood maximization (Hyvärinen, 1998) or posterior estimation (Koivunen et al., 2001). A third and more popular *noisy ICA* approach uses higher-order *cumulants* that are invariant to Gaussian noise. This paper focuses on this (third) *cumulant-based Gaussian noise invariant ICA* approach.

Earlier *cumulant-based noisy ICA* methods (Cardoso, 1991; Albera et al., 2004) are limited in their applicability. Recently, Arora et al. (2012) re-examined the second step of ICA, i.e. *whitening*, and proposed to replace the second-order covariance matrix with a higher-order cumulant tensor and introduced a *quasi-whitening* step via the Hessian of a kurtosis-related function, provided that the values of source kurtosis are all of the same sign. *Quasi-whitening* is less restrictive than *whitening* because it does not force the converted observation variables to have identity covariance matrix, where diagonal covariance matrix is good enough. Belkin et al. (2013) further lifted the restriction on the sign of source kurtosis with a new *quasi-whitening* method, which was computational heavy though. Next, Voss et al. (2013) developed a more efficient *quasi-whitening* method using the Hessian of a cumulant-based function, followed by a new deflationary algorithm named as Gradient Iteration (GI), and thus their method was called GI-ICA. Nonetheless, the *quasi-whitening* step is complex and error-prone. Voss et al. (2015) proposed a *pseudo Euclidean inner product* so that the mixing matrix are orthogonal in columns in the pseudo Euclidean inner product space and the GI algorithm can be applied to estimate ICs, leading to the Pseudo-Euclidean Gradient Iteration (PEGI) method that further improves GI-ICA.

However, there are two limitations on PEGI: (1) As Voss et al. (2015) pointed out in their paper, the pseudo Euclidean inner product is *improper* because it is not *positive definite*. This violates the definition of inner product and poses difficulties in further analysis and development. (2) Although the pseudo Euclidean inner product matrix is supposed to be orthogonal by the design of PEGI, we found that due to sample-based estimation, there were gross errors (imperfections) in such orthogonality (to be shown later). In addition, PEGI was evaluated only on synthetic data, while its performance on real-world data is not clear yet (neither was that for GI-ICA).

This paper aims to address the above limitations of PEGI with two contributions:

1. We design a *Proper Inner product* (PI) based on the fourth-order cumulants, and prove its positive definiteness. This new inner product brings improved orthogonality on sample-based estimations and the GI algorithm can be adopted for IC estimation.

2. We propose a *Mean Displacement (MD)* preprocessing step via investigating the orthogonality of the inner product matrix in sample-based estimation using a newly defined metric. With gross errors or imperfections of such orthogonality found in PEGI, we re-examine the first step of ICA and pose the following question: *can we move the data along certain direction to gain better orthogonality?* Subsequently, we introduce MD that moves the data along the direction of their mean to relax the centering step, leading to better orthogonality of the inner product matrix and better IC recovery in turn.

We name our method as *Proper Inner product with Mean Displacement (PIMD)* for Gaussian noise invariant ICA. Besides the studies on synthetic data to show the effectiveness of PIMD, we also conduct real-world data experiments on audio signal separation to show its superiority over other ICA methods.

2. Background

2.1. The Noisy ICA Model

The noisy ICA model is formulated as

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \boldsymbol{\eta} = \sum_{i=1}^{d} \mathbf{A}_i s_i + \boldsymbol{\eta}, \tag{1}$$

where $\mathbf{x} = [x_1, \cdots, x_d]^T$ are the observation variables, $\mathbf{s} = [s_1, \cdots, s_d]^T$ are the independent components (ICs), d is the number of sources or observations (usually assumed to be equal), and $\mathbf{A} = [\mathbf{A}_1, \cdots, \mathbf{A}_d]$ is the unknown constant matrix, namely *mixing matrix*. Noise is denoted by $\boldsymbol{\eta}$, and is assumed to be Gaussian distributed with zero mean. Noisy ICA methods aim to estimate sources $\mathbf{s}$ and the mixing matrix $\mathbf{A}$ simultaneously using observations $\mathbf{x}$ only with the corruption of the unknown Gaussian noise $\boldsymbol{\eta}$.

2.2. Cumulants and Their Properties

Cumulants are defined as the coefficients of the *Taylor expansion* of the second characteristic function. The r-th order cumulants are conventionally denoted by $\kappa_r(x)$ for random variable x, and $(\mathcal{Q}_\mathbf{x})_{i_1,\cdots,i_r}$ or $cum(x_{i_1}, \cdots, x_{i_r})$ for random vector $\mathbf{x}$ where $i_1, \cdots, i_r$ are the indices. Note that $\kappa_r(x_i) = cum(x_i, \cdots, x_i)$ (r times) is the r-th univariate cumulant of x_i.

Cumulants have polynomial expansions of the moments of fewer and the same orders. In particular, the *third* and *fourth* cumulants of a zero-mean random variable are the *skewness* and *kurtosis* respectively. Nevertheless, cumulants are preferable to moments for the following properties that are commonly not shared by moments:

(1) ***symmetry***: $(\mathcal{Q}_\mathbf{x})_{i_1,\cdots,i_r} = (\mathcal{Q}_\mathbf{x})_{i_{\sigma(1)},\cdots,i_{\sigma(r)}}$ for any permutation σ.

(2) ***multi-linearity***: for any random variable y and scalar α, we have $cum(x_1, \cdots, x_i + y, \cdots, \alpha x_j, \cdots, x_r) = \alpha[cum(x_1, \cdots, x_r) + cum(x_1, \cdots, y, \cdots, x_r)]$.

(3) ***independence***: if $\exists p, q \in \{1, \cdots, r\}$ where x_{i_p} and x_{i_q} are independent, we have $(\mathcal{Q}_\mathbf{x})_{i_1,\cdots,i_r} = 0$. Combined with the multi-linear property, this implies that if random variables $y_1, \cdots, y_r$ are independent of $x_1, \cdots, x_r$, we have $cum(x_1 + y_1, \cdots, x_r + y_r) = cum(x_1, \cdots, x_r) + cum(y_1, \cdots, y_n)$.

(4) ***vanishing Gaussian***: if $\mathbf{x}$ is Gaussian, $(\mathcal{Q}_\mathbf{x})_{i_1,\cdots,i_r} = 0$ for any order $r \geq 3$.

For the univariate case of x, these properties become:

(1) ***homogeneity***: $\kappa_r(\alpha x) = \alpha^r \kappa_r(x)$ for $\forall \alpha \in \mathbb{R}$.

(2) ***additivity***: if random variables x and y are independent, $\kappa_r(x+y) = \kappa_r(x) + \kappa_r(y)$.

(3) ***vanishing Gaussian***: if random variable x is Gaussian, $\kappa_r = 0$ for $\forall r \geq 3$.

2.3. Inner Product and its Matrix

Definition 1 *An **inner product** of a real vector space $\mathbb{V}$ is an assignment that, for any two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{V}$, there is a real number $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ satisfying the following three properties:*

1. ***linearity:*** $\langle a\boldsymbol{u} + b\boldsymbol{v}, \boldsymbol{w} \rangle = a\langle \boldsymbol{u}, \boldsymbol{w} \rangle + b\langle \boldsymbol{v}, \boldsymbol{w} \rangle$, *where* $\forall \boldsymbol{w} \in \mathbb{V}$ *and* $\forall a, b \in \mathbb{R}$.
2. ***symmetry:*** $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle \boldsymbol{v}, \boldsymbol{u} \rangle$.
3. ***positive definiteness:*** $\langle \boldsymbol{u}, \boldsymbol{u} \rangle \geq 0$; *and* $\langle \boldsymbol{u}, \boldsymbol{u} \rangle = 0$ *if and only if* $\boldsymbol{u} = \boldsymbol{0}$.

*The vector space $\mathbb{V}$ with an inner product is called an **inner product space** (Strang, 2016).*

Note that positive definiteness ensures that the length of a vector $\mathbf{u}$ is well defined as the square root of the self inner product, i.e. $||\mathbf{u}|| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$.

Definition 2 *Let $\mathcal{B} = \{\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_d\}$ be a basis of an d-dimensional inner product space $\mathbb{V}$. We call the $d \times d$ matrix*

$$\boldsymbol{C} = \begin{bmatrix} \langle \boldsymbol{\beta}_1, \boldsymbol{\beta}_1 \rangle, \langle \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \rangle, \cdots, \langle \boldsymbol{\beta}_1, \boldsymbol{\beta}_d \rangle \\ \langle \boldsymbol{\beta}_2, \boldsymbol{\beta}_1 \rangle, \langle \boldsymbol{\beta}_2, \boldsymbol{\beta}_2 \rangle, \cdots, \langle \boldsymbol{\beta}_2, \boldsymbol{\beta}_d \rangle \\ \cdots \\ \langle \boldsymbol{\beta}_d, \boldsymbol{\beta}_1 \rangle, \langle \boldsymbol{\beta}_d, \boldsymbol{\beta}_2 \rangle, \cdots, \langle \boldsymbol{\beta}_d, \boldsymbol{\beta}_d \rangle \end{bmatrix}$$

*the **inner product matrix** relative to the basis $\mathcal{B}$ (Strang, 2016).*

Definition 3 *For arbitrary vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{V}$, where $\boldsymbol{u} = x_1\boldsymbol{\beta}_1 + \cdots + x_d\boldsymbol{\beta}_d$ and $\boldsymbol{v} = y_1\boldsymbol{\beta}_1 + \cdots + y_d\boldsymbol{\beta}_d$, by linearity of inner product, we know that*

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \sum_{i=1}^{d} \sum_{j=1}^{d} x_i y_j \langle \boldsymbol{\beta}_i, \boldsymbol{\beta}_j \rangle. \tag{2}$$

Therefore, we can have an inner product $\langle \cdot, \cdot \rangle_{\boldsymbol{C}}$ such that

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle_{\boldsymbol{C}} = \boldsymbol{x}^T \boldsymbol{C} \boldsymbol{y}.$$

*Vectors $\boldsymbol{x} = [x_1, \cdots, x_d]$ and $\boldsymbol{y} = [y_1, \cdots, y_d]$ are called **coordinates** of vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ (Strang, 2016).*

Theorem 1 *Let $\boldsymbol{C}$ be an $d \times d$ real symmetric matrix. Define a functional $\langle \boldsymbol{u}, \boldsymbol{v} \rangle_{\boldsymbol{C}} = \boldsymbol{u}^T \boldsymbol{C} \boldsymbol{v}$ for any $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^d$. Then $\langle \cdot, \cdot \rangle_{\boldsymbol{C}}$ is an inner product if and only if the matrix $\boldsymbol{C}$ is positive definite (Strang, 2016).*

3. Proper Inner Product with Mean Displacement (PIMD)

3.1. Limitations of Pseudo Euclidean Inner Product in PEGI

To solve noisy ICA problem, a fast and simple deflationary approach, namely *Gradient Iteration (GI-ICA)*, is proposed in (Voss et al., 2013). However, GI-ICA requires that mixing matrix must be orthogonal in columns, which needs to orthogonalize the observation. This prerequisite is, though, less restrictive compared with *whitening* as it does not force the converted observation variables to have identity covariance matrix (diagonal matrix is fine), a complicated *quasi-whitening* step is required.

To avoid the complex and error-prone *quasi-whitening* step of GI-ICA, Voss et al. (2015) propose a *pseudo Euclidean inner product*, so that the mixing matrix is orthogonal in columns in this *pseudo Euclidean inner product space*. Their *pseudo Euclidean inner product* is defined as:

$$\langle \mathbf{u}, \mathbf{v} \rangle_{\widetilde{\mathbf{C}}} := \mathbf{u}^T \widetilde{\mathbf{C}}\, \mathbf{v}, \tag{3}$$

where the pseudo Euclidean inner product matrix $\widetilde{\mathbf{C}}$ has the following form:

$$\widetilde{\mathbf{C}} = \mathbf{A}^{-T} \widetilde{\mathbf{D}} \mathbf{A}^{-1}. \tag{4}$$

Here, $\mathbf{A}$ is the mixing matrix, $\mathbf{A}^{-T}$ denotes the transpose of $\mathbf{A}^{-1}$, and $\widetilde{\mathbf{D}}$ is diagonal with non-zero entries approximated as follows:

$$\widetilde{D}_{ii} = (12\kappa_4(s_i)||\mathbf{A}_i^T \mathbf{u}||)^{-1}, \tag{5}$$

where $\mathbf{u} \in \mathbb{R}^d$ and has unit norm. Therefore, in this pseudo Euclidean inner product space, we have:

$$\langle \mathbf{A}_i, \mathbf{A}_j \rangle_{\widetilde{\mathbf{C}}} = \mathbf{A}_i^T \widetilde{\mathbf{C}} \mathbf{A}_j = \mathbf{e}_i^T \widetilde{\mathbf{D}} \mathbf{e}_j = \begin{cases} 0, & \text{for } i \neq j \\ (12\kappa_4(s_i)||\mathbf{A}_i^T \mathbf{u}||)^{-1}, & \text{for } i = j \end{cases} \tag{6}$$

i.e. mixing matrix $\mathbf{A}$ is orthogonal in columns in the pseudo Euclidean inner product space.

However, there are two issues for this Pseudo Euclidean inner product (PE). First, we can see from Eq. (6) that PE depends on the kurtosis of the sources, so it is not guaranteed to be *positive definite*, which violates the definition of inner product according to Theorem 1. Second, we find (in our studies below) that although columns of $\mathbf{A}$ should be orthogonal to each other by definition, such orthogonality suffers from **gross errors** because the pseudo Euclidean inner product has to be estimated from limited number of samples in practice. We will tackle the two problems individually.

3.2. A Proper Inner Product (PI)

Inspired by the construction of *quasi-whitening* procedure of (Belkin et al., 2013), we propose a new Proper Inner product (PI) based on the Gaussian-invariant cumulants to tackle the first problem in PEGI, and prove its positive definiteness.

Definition 4 ***Linear operator*** $\mathscr{F}$ *with respect to a random vector* $\boldsymbol{x} \in \mathbb{R}^d$ *is defined by the fourth-order cumulant tensor of* $\boldsymbol{x}$ *in the following way (Hyvärinen et al., 2001):*

$$\begin{array}{rcl} \mathscr{F}_{\boldsymbol{x}} : \mathbb{R}^{d\times d} & \to & \mathbb{R}^{d\times d} \\ \boldsymbol{M} & \to & \mathscr{F}_{\boldsymbol{x}}(\boldsymbol{M}), [\mathscr{F}_{\boldsymbol{x}}(\boldsymbol{M})]_{ij} := \sum_{k,l} (\mathscr{Q}_{\boldsymbol{x}})_{ijkl} \cdot M_{kl}, \end{array} \tag{7}$$

where $(\mathscr{Q}_{\boldsymbol{x}})_{ijkl}$ *denotes the* (i, j, k, l) *entry of the fourth-order cumulant tensor for random vector* $\boldsymbol{x}$*, and* $\{M_{kl}\}$ *are the elements of the matrix* $\boldsymbol{M}$*.*

Lemma 1 *The cumulant operator $\mathscr{F}$ has the following properties (Belkin et al., 2013):*

(1) It is linear and symmetric.

(2) It maps any symmetric matrix to another symmetric matrix.

(3) With noisy ICA model (1), for $\forall \boldsymbol{M} \in \mathbb{R}^{d\times d}$, $\mathscr{F}_{\boldsymbol{x}}(\boldsymbol{M})$ can be decomposed into $\boldsymbol{A}\boldsymbol{D}\boldsymbol{A}^T$, where $\boldsymbol{D}$ is diagonal and $D_{ii} = \kappa_4(s_i)\langle \boldsymbol{A}_i\boldsymbol{A}_i^T, \boldsymbol{M}\rangle = \kappa_4(s_i)\boldsymbol{A}i^T\boldsymbol{M}\boldsymbol{A}_i$.

Theorem 2 *Given noisy ICA model (1), we construct the matrix $\boldsymbol{C}$ with the following steps: (1) compute matrix $\boldsymbol{M}_1 = [\mathscr{F}_{\boldsymbol{x}}(\boldsymbol{I})]^{-1}$; (2) compute matrix $\boldsymbol{M}_2 = \mathscr{F}_{\boldsymbol{x}}(\boldsymbol{M}_1)$; (3) set matrix $\boldsymbol{C} = (\boldsymbol{M}_2)^{-1}$. Then, $\boldsymbol{C}$ is symmetric and positive definite.*

Proof By Lemma 1, $\mathscr{F}_{\mathbf{x}}(\mathbf{I}) = \mathbf{A}\mathbf{D}_1\mathbf{A}^T$, where $\mathbf{D}_1$ is diagonal and $(D_1)_{ii} = \kappa_4(s_i)\mathbf{A}_i^T\mathbf{A}_i = \kappa_4(s_i)||\mathbf{A}_i||^2$. So $\mathbf{M}_1 = \mathbf{A}^{-T}\mathbf{D}_1^{-1}\mathbf{A}^{-1}$. Apply Lemma 1 a second time, and we have $\mathbf{M}_2 = \mathbf{A}\mathbf{D}_2\mathbf{A}^T$ where $\mathbf{D}_2$ is diagonal with $(D_2)_{ii} = \kappa_4(s_i)\mathbf{A}_i^T\mathbf{M}_1\mathbf{A}_i$. Substitute $\mathbf{M}_1$, and we know

$$\begin{aligned}(D_2)_{ii} &= \kappa_4(s_i)\mathbf{A}_i^T(\mathbf{A}^{-T}\mathbf{D}_1^{-1}\mathbf{A}^{-1})\mathbf{A}_i = \kappa_4(s_i)\mathbf{e}_i^T\mathbf{D}_1^{-1}\mathbf{e}_i \\ &= \kappa_4(s_i)/(D_1)_{ii} = \kappa_4(s_i)/[\kappa_4(s_i)||\mathbf{A}_i||^2] = ||\mathbf{A}_i||^{-2}.\end{aligned} \tag{8}$$

Then, we have $\mathbf{C} = (\mathbf{M}_2)^{-1} = \mathbf{A}^{-T}\mathbf{D}\mathbf{A}^{-1}$ and $\mathbf{D} = (\mathbf{D}_2)^{-1}$, for which

$$D_{ii} = |\mathbf{A}_i||^2 > 0. \tag{9}$$

Therefore, we have proven that $\mathbf{C}$ is symmetric and positive definite. ■

Theorem 3 *Define a functional for any vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^d$ as*

$$\langle \boldsymbol{u}, \boldsymbol{v}\rangle_{\boldsymbol{C}} := \boldsymbol{u}^T\boldsymbol{C}\boldsymbol{v}, \tag{10}$$

*where the matrix $\boldsymbol{C}$ is constructed following Theorem 2, then $\langle\cdot,\cdot\rangle_{\boldsymbol{C}}$ is **a (proper) inner product**, and $(\mathbb{V}, \langle\cdot,\cdot\rangle)$ is **a (proper) inner product space**.*

Proof The proof is straightforward by combining Theorem 1 and Theorem 2. ■

Theorem 4 *Given noisy ICA model (1), mixing matrix $\boldsymbol{A}$ is orthogonal in columns with the inner product $\langle\cdot,\cdot\rangle_{\boldsymbol{C}}$.*

Proof For $\forall i, j \in \{1, \cdots, d\}$, we have

$$\langle \mathbf{A}_i, \mathbf{A}_j\rangle_{\mathbf{C}} = \mathbf{A}_i^T\mathbf{C}\mathbf{A}_j = \mathbf{A}_i^T(\mathbf{A}^T)^{-1}\mathbf{D}\mathbf{A}^{-1}\mathbf{A}_j = \mathbf{e}_i^T\mathbf{D}\mathbf{e}_j = \mathbf{D}_{ij} = \begin{cases} 0, & i \neq j, \\ ||\mathbf{A}_i||_2^2, & i = j,\end{cases} \tag{11}$$

which means different columns of mixing matrix $\mathbf{A}$ are orthogonal with each other. ■

3.3. Mean Displacement (MD) to Relax Centering for Better Orthogonality

To tackle the second problem of PEGI, we investigate the orthogonality property of its inner product in sample-based estimation. In the following presentation, the orthogonality of inner product refers to that of its sample-based estimation for convenience. We first define a metric for the orthogonality error of an inner product. Next, we introduce a mean displacement step to seek for better orthogonality, which relaxes the conventional *centering*.

3.3.1. Orthogonality of the Pseudo Euclidean Inner Product

For a specific inner product $\langle \cdot, \cdot \rangle_{\boldsymbol{C}}$, we define the following metric to measure the orthogonality of this inner product.

Definition 5 *We define the **orthogonality of inner product** $\langle \cdot, \cdot \rangle_{\boldsymbol{C}}$ for the mixing matrix $\boldsymbol{A}$ of the noisy ICA model (1) as*

$$\delta_{\boldsymbol{C}} = \frac{\sum_i \sum_{j \neq i} |\Delta_{ij}|}{\sum_i |\Delta_{ii}|}, \tag{12}$$

where $\Delta_{ij} = \langle \boldsymbol{A}_i, \boldsymbol{A}_j \rangle_{\boldsymbol{C}}$. It measures the orthogonality of the inner product $\langle \cdot, \cdot \rangle_{\boldsymbol{C}}$ in sample-based estimation. $\delta_{\boldsymbol{C}}$ can be considered as an error *measurement so the smaller, the better.*

Using the above metric, we study the orthogonality of the pseudo Euclidean inner product with respect to the number of samples in Fig. 1 (PE - the blue dash line). We can see that its orthogonality improves with the growth of sample sizes at the beginning but becomes almost saturated at certain point, without converging to zero. This indicates that although by design, the columns of the mixing matrix $\mathbf{A}$ in PE should be orthogonal with each other, this is not the case in practical with finite samples. Such violation of the orthogonality property could have an adverse effect on the performance.

3.3.2. Mean Displacement as a Relaxation of Centering

Motivated by the imperfect orthogonality of PEGI in Fig. 1 and the replacement of the *whitening* step by *quasi-whitening* in (Arora et al., 2012), we go one step further (back) to re-examine the first step of centering in ICA.

Centering is basically a displacement step to move the mean to the origin (along the direction of the mean), which has a positive effect on simplifying the theoretical analysis and algorithm derivation for noise-free ICA. However, in noisy ICA, *the presence of noise corrupts sample-based estimations of mean and covariance*, especially when the number of samples is small. Existing noisy ICA methods address the influence of noise on sample covariance matrix by relaxing the conventional *whitening* to *quasi-whitening* (Arora et al., 2012; Voss et al., 2013). However, the *adverse effect on the mean* has not been taken into consideration. Although lacking theoretical results, we suspect such corruption on the mean estimation has negative effects as well, e.g., it could affect the orthogonality of an inner product in noisy ICA.

Given noisy ICA model (1), we relax *centering* to *Mean Displacement* (MD) that displaces the data along the direction of its mean. Therefore, we define MD as follows:

$$\tilde{\mathbf{x}} = \mathbf{x} - p\mathbb{E}[\mathbf{x}], \tag{13}$$

where $\mathbb{E}[\mathbf{x}]$ is the expectation of $\mathbf{x}$, and parameter p controls the amount of the displacement. It is noteworthy that centering is a special case of MD when $p = 1$. Another special case is MD with $p = 0$, which means keeping the original data.

As a *relaxation of centering*, MD has some *regularization* effect and adds model flexibility to achieve better orthogonality for noisy ICA problem. Figure 1 demonstrates the improvement of orthogonality in PE by mean displacement (PEGI+MD$_{p=0}$ – the yellow dot line). We can see that MD largely improves the orthogonality of the pseudo Euclidean inner product of PEGI. Comparing the orthogonality of the proposed proper inner product (PI – the brown dash line) with PIMD (PIMD$_{p=0}$ – the black solid line), we can see that

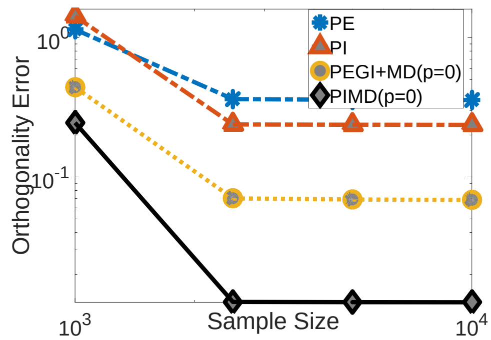


Figure 1: *Orthogonality* of inner products with or without mean displacement (MD). The results are obtained from a 4-source noisy ICA task with noise level 0.1. Each point is averaged over 100 repetitions. We show the orthogonality errors between 0 and 0.8 for clarity.

MD also greatly improve the orthogonality of PI. Besides, PI or PIMD outperforms PE or PEGI+MD in orthogonality respectively, showing the positive effects of PI as well.

Thus, to achieve positive definiteness and better orthogonality than PEGI, PIMD can be viewed as replacing the Pseudo Euclidean inner product (PE) with the Proper Inner product (PI), and relax the conventional centering step to Mean Displacement (MD), leading to a better noisy ICA method.

3.4. Gradient Iteration in the Proper Inner Product Space

We now adopt the GI algorithm (Voss et al., 2013) to solve our PIMD noisy ICA problem. Given noisy ICA model (1), a kurtosis-based function $f(\mathbf{u}) = \kappa_4(\langle \mathbf{x}, \mathbf{u}\rangle_\mathbf{C})$ is defined, where $\mathbf{u}$ is an arbitrary vector in $\mathbb{R}^d$ space, and $\langle \cdot, \cdot \rangle_\mathbf{C}$ is the inner product in Theorem 2. According to the properties of inner product and kurtosis, we have

$$\begin{aligned} f(\mathbf{u}) &= \kappa_4(\langle \mathbf{A}\mathbf{s}, \mathbf{u}\rangle_\mathbf{C} + \langle \boldsymbol{\eta}, \mathbf{u}\rangle_\mathbf{C}) = \kappa_4(\langle \textstyle\sum_{i=1}^d \mathbf{A}_i s_i, \mathbf{u}\rangle_\mathbf{C} + \langle \boldsymbol{\eta}, \mathbf{u}\rangle_\mathbf{C}) \\ &= \kappa_4(\textstyle\sum_{i=1}^d s_i \langle \mathbf{A}_i, \mathbf{u}\rangle_\mathbf{C} + \langle \boldsymbol{\eta}, \mathbf{u}\rangle_\mathbf{C}) = \textstyle\sum_{i=1}^d (\langle \mathbf{A}_i, \mathbf{u}\rangle_\mathbf{C})^4 \kappa_4(s_i). \end{aligned} \tag{14}$$

Take first-order derivative with respect to vector $\mathbf{u}$, we have

$$\nabla f(\mathbf{u}) = 4\sum_{i=1}^d (\langle \mathbf{A}_i, \mathbf{u}\rangle_\mathbf{C})^3 \kappa_4(s_i)\mathbf{A}_i, \tag{15}$$

which shows the form of a generalized *power iteration*. Thus, we can adopt gradient iteration to estimate one column of mixing matrix $\mathbf{A}$ iteratively by

$$\nabla f(\mathbf{u})/||\nabla f(\mathbf{u})|| \to \mathbf{A}_i, \text{ for } \forall i \in \{1, \cdots, d\}. \tag{16}$$

3.5. SINR-optimal Source Estimation

Following (Voss et al., 2015), given access to the mixing matrix $\mathbf{A}$, we adopt

$$\mathbf{W} = \mathbf{A}^T cov(\mathbf{x})^{-1} \tag{17}$$

as the optimal demixing matrix. (Koldovský and Tichavský, 2006) shows that $\mathbf{W}$ can achieve the greatest *Signal to Interference-Plus-Noise Ratio* (SINR) for each component.

Algorithm 1 PIMD for noisy ICA.

Input: Observation matrix $\mathbf{X}$ with multiple samples aligned in columns.
Mean displacement
  Decide a proper p to move observations according to Eq. (13) ($p = 0$ by default).
Inner product matrix C (Theorem 2)
  (1) Compute $\mathbf{M}_1 = [\mathscr{F}_{\mathbf{x}}(\mathbf{I})]^{-1}$.
  (2) Compute $\mathbf{M}_2 = \mathscr{F}_{\mathbf{x}}(\mathbf{M}_1)$.
  (3) Inner product matrix $\mathbf{C} = \mathbf{M}_2^{-1}$.
Gradient iteration with inner product $\langle \cdot, \cdot \rangle_{\mathbf{C}}$
  For $i = 1$ until d
    $i \leftarrow 1$
    repeat
      $\mathbf{u}_i \leftarrow \nabla f(\mathbf{u}_{i-1}) / ||\nabla f(\mathbf{u}_{i-1})||$ based on Eq. (15) and (16)
    until Convergence
    $\mathbf{A}_i \leftarrow \mathbf{u}_i$
  end FOR
SINR-optimal source estimation
  (1) SINR-optimal demixing matrix $\mathbf{W} = \mathbf{A}^T cov^{-1}(\mathbf{x})$ from Eq. (17).
  (2) Source estimation $\mathbf{S} = \mathbf{W}\mathbf{X}$.
Output: Mixing matrix $\mathbf{A}$ and sources $\mathbf{S}$.

Therefore, after estimating the mixing matrix $\mathbf{A}$ in Sec. 3.4, we calculate the SINR-optimal demixing matrix $\mathbf{W}$ and estimate the latent components by

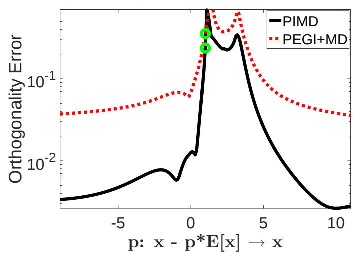
$$\mathbf{s} = \mathbf{W}\mathbf{x}. \tag{18}$$

Algorithm 1 summarizes the GI-based algorithm of PIMD for noisy ICA, where d is the number of sources or observations (assumed to be equal by convention).

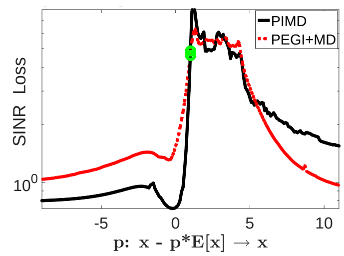
4. Experiments

In this section, we evaluate PIMD against the following three groups of ICA methods on mixtures of both synthetic and real data:

1. Representative noise-free ICA methods: FastICA (Hyvärinen and Oja, 1997), its robust version rFastICA (Hyvärinen, 1999b), JADE (Cardoso and Souloumiac, 1993), and Infomax (Bell and Sejnowski, 1995).
2. State-of-the-art noisy ICA methods: 1FICA (Koldovský and Tichavský, 2007) as a variation of FastICA with the *tanh* contrast function designed to have low bias for performing SINR-optimal ICA in the presence of Gaussian noise, GI-ICA (Voss et al., 2013) as a Gaussian noise invariant ICA which generalizes the conventional *whitening* to *quasi-whitening* based on higher-order cumulants, and PEGI (Voss et al., 2015) which introduces a pseudo Euclidean inner product and circumvents the noise-susceptible *whitening* step.
3. The MD versions of the above methods: FastICA+MD, rFastICA+MD, JADE+MD, Infomax+MD, 1FICA+MD, GI-ICA+MD, and PEGI+MD. Note that the methods in the first two groups are special cases of their MD versions with $p = 1$.



(a) Orthogonality vs. p.



(b) SINR Loss vs. p.

Figure 2: The effects of mean displacement (MD) with respect to p (see Eq. (13)). The values with green circles at $p = 1$ denote the orthogonality and SINR Loss of pseudo and proper inner product *without MD*, which correspond to the conventional *centering*. The ICA problem has 4 sources with sample size 10,000 and noise level $\sigma^2 = 0.1$. Each point is the average over 100 repetitions.

To guarantee the invertibility of mixing matrices, they are generated in three steps: (i) uniformly generate a $d \times d$ matrix with each entry between zero and one, where d is the number of sources; (ii) normalize the generated matrix by column; (iii) add an identity matrix to the one in (ii). The Gaussian noise is assumed to follow

$$\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}), \tag{19}$$

where $\sigma^2 \in [0, 1]$ controls the *noise level*. Each source component of both synthetic and real-world data is normalized to have unit variance.

SINR Loss is used to measure the performance of noisy ICA methods in noisy setting, which is defined as (Koldovský and Tichavský, 2007):

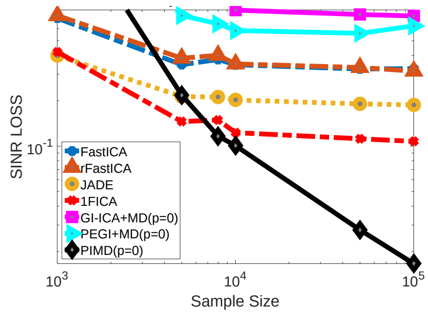
$$\text{SINR Loss} = \text{Optimal SINR - Achieved SINR}, \tag{20}$$

where *Optimal SINR* uses the oracle demixing by Eq. (17) to achieve SINR-optimal result.

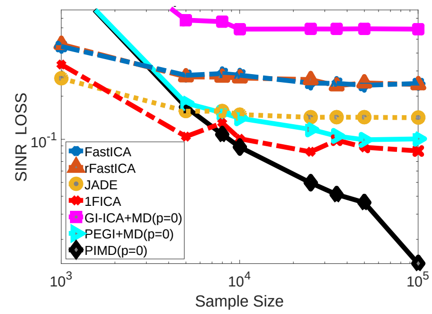
4.1. Studies on Mean Displacement (MD)

Figure 2(a) shows the orthogonality of Pseudo Euclidean inner product (PE) in PEGI+MD and that of Proper Inner product (PI) in PIMD for p ranging from -9 to 11. The original PEGI is simply PEGI+MD with $p = 1$. The values highlighted with green circles denote the orthogonality of PE and PI without MD, i.e., $p = 1$, which corresponds to centering. We can see that MD can substantially improve the orthogonality of PE for PEGI and PI for PIMD. It also shows that PI of PIMD can further improve the orthogonality over PE of PEGI. These studies confirm that MD is a simple yet powerful technique to improve the orthogonality of inner products in noisy ICA.

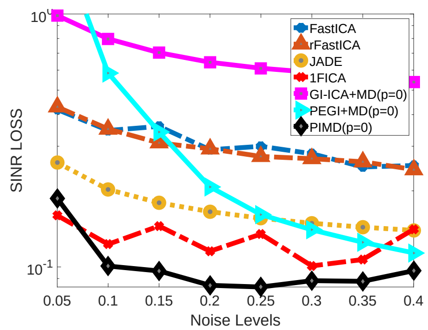
Figure 2(b) shows the IC recovery performance of PIMD and PEGI+MD (including the original PEGI as a special case with $p = 1$). We can see that good values of p in MD can largely improve the performance of both PIMD and PEGI in terms of SINR Loss. We can also see that Fig. 2(a) and Fig. 2(b) follow similar trends mostly, indicating that better orthogonality can lead to better performance in general. Moreover, PEGI+MD outperforms PEGI ($p = 1$) for most p, and PIMD outperforms PEGI+MD mostly except for $p > 5$.



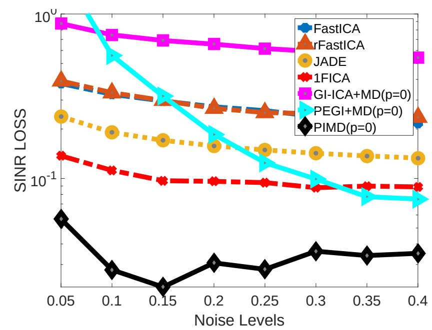
(a) Error with different sample sizes. Noise level is 0.1.



(b) Error with different sample sizes. Noise level is 0.3.



(c) Error under additive Gaussian noise. Sample size is 10,000.



(d) Error under additive Gaussian noise. Sample size is 50,000.

Figure 3: Sensitivity studies for 4-source synthetic noisy BSS. We do not investigate the effects of MD here and thus its p is set to zero for PIMD, GI-ICA, and PEGI. We use the without MD version of all the other methods. Each point is the average value over 100 repetitions. We show errors below 0.8 for clarity.

4.2. Blind Source Separation on Noisy Synthetic Data

We further study BSS with four synthetic sources: $s_1(t) = sin(2\pi 800t)$, $s_2(t) = sin(2\pi 90t)$, $s_3(t) = sin(2\pi 300t - 6cos(2\pi 60t))$, and $s_4(t) = sin(2\pi 9t)sin(2\pi 300t)$ (Tang and Wang, 2012). In this set of experiments, Infomax, GI-ICA, and PEGI perform poorly so their results are not included for clarity. Here, we focus on a particular case of MD with $p = 0$ for demonstration[1] although other values of p could lead to better performance.

In Fig. 3(a) and Fig. 3(b), we compare methods at various sample sizes.[2] For each sample size, 100 square mixing matrices are randomly generated to get 100 sets of mixtures (four mixtures per set) and then each mixing matrix is estimated from its four mixtures by ICA methods. The mean SINR Loss of each ICA method is reported. We can see that the performance of PIMD improves at a greater rate as the sample size increases. Given large numbers of samples (more than 10^4), PIMD outperforms all the other methods. 1FICA gives the second best results on the whole. In particular, when the sample size is small, e.g. smaller than 10^4, 1FICA obtains the best results.

1. The default setting in Algorithm 1, i.e., no centering and keeping the original data.
2. Note that the term 'sample' refers to samples of random variables in noisy ICA model rather than training examples or observations that are commonly referred to as 'sample' in machine learning.

There are two main reasons that noisy ICA methods such as PEGI or PIMD need larger sample size for good performance than classical ICA methods such as FastICA or JADE. One is that the 'whitening' (via PE or PI) in PEGI or PIMD needs the fourth-order statistics while the classical whitening in FastICA or JADE involves only the second-order statistics.[3] In general, higher-order statistics need more samples to estimate well (Hyvärinen et al., 2001). The other reason is that the noisy ICA model is more complex than the classical noiseless ICA model. A complex model considering noise explicitly needs more samples to estimate well than a simpler noiseless model. Today's real-world data often have high sampling rate or resolution, and large sample size in turn.

Figures 3(*c*) and 3(*d*) show the performance of ICA methods for different noise levels. Each of the four sources has at least 10,000 or 50,000 samples. Each point is the mean SINR Loss over 100 repetitions. We can see that on the whole, PIMD outperforms the others and 1FICA is the second-best method.

Note that the relative performance of PEGI improves drastically with higher noise level. PEGI achieves the third best when the noise level is higher and the sample size is larger (right half of Figs. 3(*b*), 3(*c*), and 3(*d*)) and it can even outperform 1FICA when the noise level is high enough and the sample size is large enough (the right most points of Figs. 3(*c*), and 3(*d*) at noise level of 0.4), which is consistent with Fig. 1(a) of (Voss et al., 2015).

4.3. Noisy Blind Audio Separation

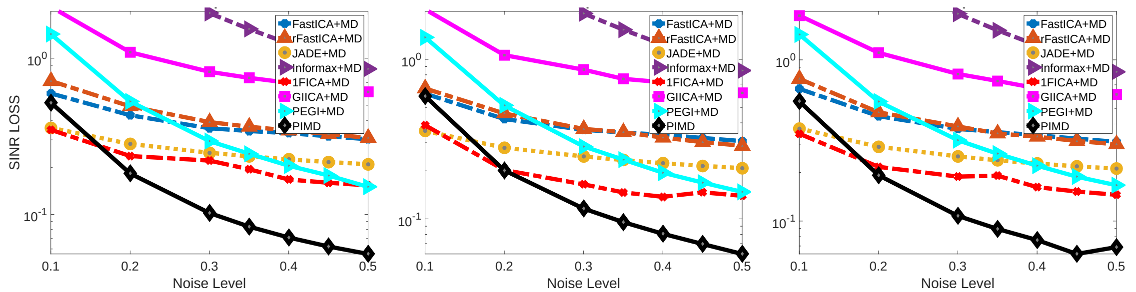
With experiments on synthetic data, we have shown the superiority of PIMD over other ICA methods in noisy ICA problem. Now we examine PIMD on real data against others. We study BSS of audio mixtures.

Audio data: The source audio signal data are collected from a public venue "*Stereo Audio Source Separation Evaluation Campaign*" (Vincent et al., 2007). We use the single-channel audio sources of the development data. They are generated from four sets of source signals with 10 second duration sampled at 16 kHz, which contain 4 male speeches, 4 female speeches, 3 non-percussive pieces of music sources, and 3 music sources including drums. Each of these 14 audio sources is standardized to have 160,000 samples and unit variance. We then inject different levels of Gaussian noise to each audio source.

Experimental settings: We randomly select four audio sources, and mix them with a random square mixing matrix with injected Gaussian noise of a chosen level to get four mixtures. We then separate these four mixtures by ICA methods. We repeat this BSS experiment 50 times and report the average SINR Loss. We study PIMD from two perspectives. First, we evaluate PIMD against other methods at different noise levels. Second, we investigate the effects of mean displacement on PIMD and the sensitivity to parameter p on real data. In addition, we study the MD versions of all the other ICA methods where the original versions correspond to $p = 1$.

Audio separation performance: Figures 4(*a*), 4(*b*), and 4(*c*) show the results for different noise levels. PIMD gives the best results mostly for noise levels above 0.2. PIMD outperforms the other methods by a greater amount at higher noise levels. 1FICA is the second best in most cases. PEGI+MD gives the third best results for noise levels above

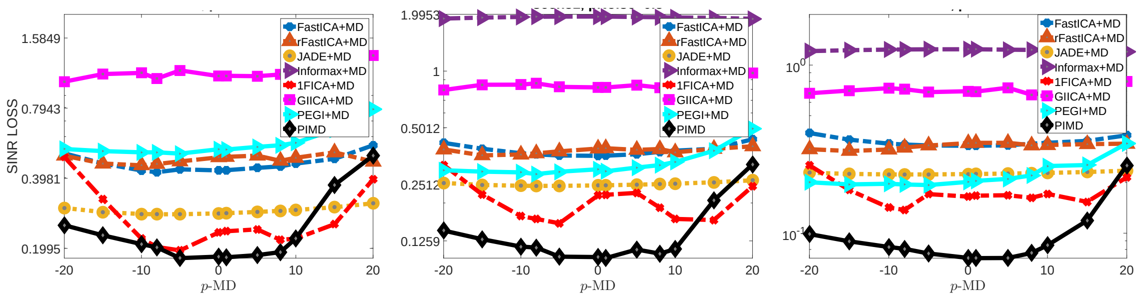
3. Note FastICA/JADE uses the fourth-order statistics in the IC estimation step not the whitening step.



(a) The p in MD is 1, i.e. the conventional *centering*.

(b) The p in MD is -8.

(c) The p in MD is 8.



(d) Noise level is 0.2.

(e) Noise level is 0.3.

(f) Noise level is 0.4.

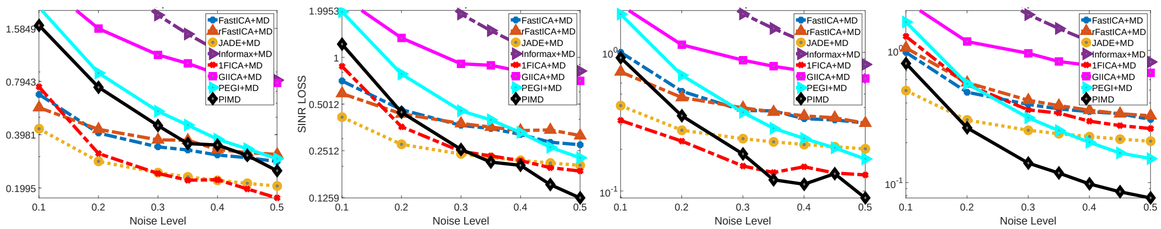
Figure 4: Experiments on noisy audio data with 100% of the whole samples (i.e. 160,000). The top three graphs demonstrate the performance of ICA methods for noisy audio separation with respect to different noise levels, and the bottom three graphs demonstrate the effects of mean displacement (MD) with respect to p on all ICA methods, where $p = 1$ indicates the conventional *centering* step. This audio separation problem has 4 sources. Each point is the average over 50 repetitions. We show SINR Loss below 2 for more details.

0.35. An interesting observation is that the conventional noise-free ICA methods such as JADE and FastICA perform reasonably well in the noisy setting, which was first observed in (Koldovský and Tichavský, 2006) and then in (Voss et al., 2015). The robust version of FastICA can be similar or slightly better than FastICA for higher noise levels, though for lower noise levels, FastICA outperforms its robust version.

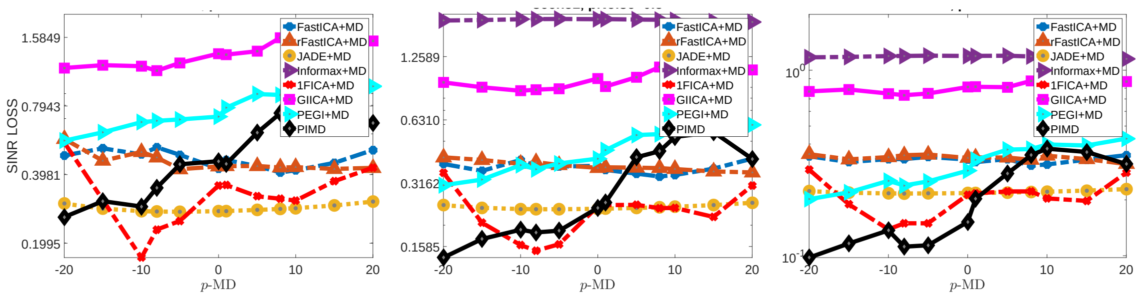
Effects of mean displacement: Figures 4(d), 4(e), and 4(f) demonstrate the effects of MD on all ICA methods studied. We can see that MD has little effect on standard ICA methods including FastICA, its robust version rFastICA, JADE, and Infomax. In contrast, better values of p in MD will largely benefit PIMD and 1FICA+MD. For these audio source data, the performance of PIMD is less sensitive to the values of p for those around $p = 1$. In contrast, 1FICA is more sensitive to the values of p.

Note here that the sample size (160,000) is quite large so the effects of MD are smaller compared to the cases of smaller sample sizes, as shown below.

Smaller sample size and values of p: Figure 5 shows the same audio separation experiments with only 75% (i.e. 120,000 out of 160,000) of the available samples. We can see that the relative performance of PIMD deteriorates with smaller sample size, becoming inferior to 1FICA and JADE in most cases when $p = 1$ and $p = 8$ though keeping superior



(a) The p in MD is 8. (b) The p in MD is 1, i.e. conventional *centering*. (c) The p in MD is -8. (d) The p in MD is -20.



(e) Noise level is 0.2. (f) Noise level is 0.3. (g) Noise level is 0.4.

Figure 5: Experiments on noisy audio data with 75% of the whole samples (i.e. 120,000). The top four graphs demonstrate the performance of ICA methods for noisy audio separation with respect to different noise levels, and the bottom three graphs demonstrate the effects of mean displacement (MD) with respect to p on all ICA methods, where $p = 1$ indicates the conventional *centering* step. This audio separation problem has 4 sources. Each point is the average over 50 repetitions. We show SINR Loss below 2 for more details.

to PEGI and GI-ICA consistently. We can also see that the relative performance of PIMD becomes better with higher noise levels, which is consistent with previous observations with all available samples. Thus, PIMD is still advantageous to 1FICA for highly noisy ICA with a properly determined p. On the other hand, comparing Fig. 5(e)∼5(g) with Fig. 4(d)∼4(f), we can see that MD (or the value of p) has larger effects on PIMD for this smaller sample size case. Positive large values of p lead to substantially worse performance, while negative large values of p continue to improve the performance and can outperform all other methods. Similar observations can be made from Fig. 5(a)∼5(d).

Viewing MD from the perspective of regularization, a larger value of p indicates a stronger regularization. It is expected that smaller sample size needs stronger regularization (larger p), while the sign of p or the direction of displacement for an effective regularization could be data dependent (e.g., depending on the mean).

Based on the studies above, we have the following suggestions on setting the values of p. For large sample size, we should set p to a (small) value near $p = 1$ where $p = 0$ is a reasonably good choice. For small sample size, we should set a larger p for a stronger regularization, where the sign of p could be determined based on some hypothesis testing, cross validation procedures, or data-dependent knowledge.

5. Conclusion

This paper addresses two limitations of a state-of-the-art noisy ICA method, namely Pseudo Euclidean Gradient Iteration (PEGI). First, the Pseudo Euclidean inner product (PE) is not positive definite, which violates the definition of inner product. To address this problem, we define a *Proper* Inner product (PI) and prove its positive definiteness. Second, we examine the orthogonality of PE in practical finite sample case. We define an orthogonality metric of inner product and introduce a relaxed version of *centering*, namely *Mean Displacement (MD)*, to improve the orthogonality of inner product and the recovery performance. We name our method as Proper Inner product with Mean Displacement (PIMD).

PIMD is evaluated on both synthetic and real-world data compared against other ICA methods as well as their MD versions. Experiments demonstrate that PIMD presents better orthogonality and better recovery performance in noisy settings given sufficient sample size. In particular, its superiority over other methods becomes greater for higher levels of noise. Furthermore, MD is effective for other methods including PEGI, GI-ICA, and 1FICA.

Acknowledgments

This work was supported by Research Grants Council of the Hong Kong SAR (Grant 12200915). Liyan is partially financially supported by FRG2/13-14/073. We thank Dr. James Voss for providing their codes and many helpful discussions.

References

L. Albera, A. Ferréol, P. Comon, and P. Chevalier. Blind identification of overcomplete mixtures of sources (biome). *Linear Algebra and Its Applications*, 391:3–30, 2004.

S. Arora, R. Ge, A. Moitra, and S. Sachdeva. ProvableICAwith unknown gaussian noise, and implications for gaussian mixtures and autoencoders. In *Neural Information Processing Systems (NIPS)*, pages 1284–2392, 2012.

M. Belkin, L. Rademacher, and J. Voss. Blind signal separation in the presence of gaussian noise. In *Conf. on Learning Theory (COLT)*, pages 270–287, 2013.

A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

J. F. Cardoso. Super-symmetric decomposition of the fourth-order cumulant tensor: Blind identification of more sources than sensors. In *Int. Conf. on Acoustics, Speech, and Signal Processing*, pages 3109–3112, 1991.

J. F. Cardoso and A. Souloumiac. Blind beamforming for non-gaussian signals. *IEE Processings F (Radar and Signal Processing)*, 140(6):362–370, 1993.

A. Hyvärinen. Independent component analysis in the presence of gaussian noise by maximizing joint likelihood. *Neurocomputing*, 22(13):49–67, 1998.

A. Hyvärinen. Fast ICA for noisy data using gaussian moments. In *IEEE Int. Symposium on Circuits and Systems*, volume 5, pages 57–61, 1999a.

A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. on Neural Networks*, 10:626–634, 1999b.

A. Hyvärinen. Independent component analysis: Recent advances. *Philosophical Trans. of the Royal Society of London: Mathematical, Physical and Engineering Sciences*, 371 (1984), 2013.

A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9:1483–1492, 1997.

A. Hyvärinen and E. Oja. Independent component analysis: Algorithms and applications. *IEEE Trans. on Neural Networks*, 13(4-5):411–430, 2000.

A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley-Interscience, 2001. ISBN 0-471-22131-7.

V. Koivunen, M. Enescu, and E. Oja. Adaptive algorithm for blind separation from noisy time-varying mixtures. *Neural Computation*, 13(10):2339–2357, 2001.

Z. Koldovský and P. Tichavský. Methods of fair comparison of performance of linear ICA techniques in presence of additive noise. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, volume 5, pages 873–876, 2006.

Z. Koldovský and P. Tichavský. Blind instantaneous noisy mixture separation with best interference-plus-noise rejection. In *Int. Conf. on Independent Component Analysis and Signal Separation*, pages 730–737, 2007.

Gilbert Strang. *Introduction to Linear Algebra (Fifth Edition)*. Wellesley Cambridge, 2016. ISBN 978-09802327-7-6.

H. Tang and S. Wang. Noisy blind source separation based on adaptive noise removal. In *World Congress on Intelligent Control and Automation*, pages 4255–4257, 2012.

E. Vincent, H. Sawada, P. Bofill, Makino S., and J. P. Rosca. First stereo audio source separation evaluation campaign: data, algorithms and results. In *Independent Component Analysis and Signal Separation*, pages 552–559, 2007.

J. Voss, L. Rademacher, and M. Belkin. Fast algorithms for gaussian noise invariant ica. In *Neural Information Processing Systems (NIPS)*, pages 2544–2552, 2013.

J. Voss, M. Belkin, and L. Rademacher. A pseudo-euclidean iteration for optimal recovery in noisy ICA. In *Neural Information Processing Systems (NIPS)*, pages 2872–2880, 2015.