

# Separating Mixed Signals in a Noisy Environment Using Global Optimization

George Hou <sup>\*</sup>

January 15, 2015

## Abstract

In this paper, we propose and analyze a class of blind source separation (BSS) methods to recover mixed signals in a noisy environment. Blind source separation aims at recovering source signals from their mixtures without detailed knowledge of the mixing process. Motivated by the work presented in [23], we propose a new optimization method based on second order statistics that considers the impact of Gaussian noise. By treating the Gaussian noise as a separate source signal and using an extra measurement of the mixed signals, we formulate the source separation problem as a global optimization problem that minimizes the cross-correlation of the recovered signals. In the case when the cross-correlation of the source signals is exactly zero, we give precise solvability conditions and prove that our global optimization method gives an exact recovery of the original source signals up to a scaling and permutation. In the case when the cross-correlation is small but nonzero, we perform stability and error analysis to show that the global optimization method still gives an accurate recovery with a small error. We also analyze the solvability for the two-signal case when the mixing matrix is degenerate. To the best of our knowledge, this is the first error analysis of BSS methods. The numerical results using realistic signals confirm our theoretical findings and demonstrate the robustness and accuracy of our methods.

## 1 Introduction

Often, in our daily encounters, situations require the need to listen to a particular speaker in a complex auditory scene consisting of multiple speakers in an environment with a noisy background. Amazingly, humans have the ability to separate the background noise (or surrounding speakers' speech) from that of the targeted speaker. Attempts to understand this amazing human ability inspire the development of Blind Source Separation (BSS). This method aims at recovering source signals from their mixtures without detailed knowledge of the mixing process.

---

<sup>\*</sup>Email: georgehou23@yahoo.com, Arcadia High School, CA 91006. This work was under the mentorship of Professor Jack Xin from University of California at Irvine. Email: jxin@math.uci.edu

---

## 1.1 A Brief Review

Broadly speaking, there are three approaches used to perform blind source separation of time-dependent linear mixtures. The first approach uses a time domain method based on de-correlation or the independence hypothesis of source signals. The idea is to estimate the de-mixing matrix from the resulting statistical equations involving moments of mixture signals and their time translations. The source signals are assumed to be stationary in a short time scale (up to a few hundred milliseconds for sound signals), and non-stationary in longer time scales. The approach can be formulated for both instantaneous and convolutive mixtures. We refer to [15, 13] for methods that use second order statistics and to [5, 6, 3, 1, 8, 22, 10, 23], which use higher order statistics.

The second approach uses a frequency domain method for convolutive mixtures. First, it takes the discrete Fourier transform to convert the time nonlocal de-mixing problem into many instantaneous de-mixing problems; then it applies one of the instantaneous methods; see e.g. [17, 11, 16]. Due to the scaling and permutation uncertainties in the instantaneous methods, the frequency domain approach has a complication in sorting de-mixed results from each frequency bin before the inverse Fourier transform can produce final time domain results. Also, different segments of Fourier transform have to be glued together nicely to avoid artifacts.

The third approach proceeds in the time-frequency domain where the working hypothesis is that the different source signals do not overlap (in the time-frequency domain or time dependent Fourier spectra). The separation is based on feature vectors and clustering of time-dependent Fourier spectra of the mixtures. The feature vectors can be geometrically based (related to arrival times and directions of the incoming sounds) or perceptually based (pitch, timbre, speech onset-offset, amplitude, frequency modulations, etc. of the source sounds). The direction of arrival-based approach is also known as the beam-forming method, which is discussed in detail in [18]; see also [24, 2, 20, 19] for more discussions of related works.

Nearly all methods encounter difficulties with long convolution (strongly reverberant rooms) or source signals coming from the same direction (e.g. one person standing in front of the other). The latter problem corresponds to the ill-conditioned mixing cases. Another difficulty encountered by existing methods is how to effectively separate mixed signals in a noisy environment, especially when the noise level is large. This is precisely the main focus of this paper.

## 1.2 Summary of the Main Contributions

One of the main contributions of our work is to develop and analyze a class of blind source decomposition methods that can be used to separate mixed signals in a noisy environment. In view of past research [23], we use second order statistics and a global optimization approach. Most existing methods (Chapter 5 of [23]) do not explicitly consider the impact of noise. In fact, direct implementation of the second order statistic method to mixed signals polluted with noise does not yield a satisfactory recovery of the source signals. In this paper, we consider Gaussian noise and treat the noise as a separate source signal. With an extra measurement of the mixed signals, we formulate

---

the source separation problem as a global optimization problem that minimizes the cross-correlation of the recovered signals.

We develop rigorous error and stability analysis of the proposed methods. Our analysis reveals the desirable properties of the measurements and source signals under which one can guarantee a stable and accurate recovery of the mixed signals. The analysis applies to both clean and noisy mixtures with a large signal-to-noise ratio. In the case when the cross-correlation of the source signals is exactly zero, we give precise solvability conditions and prove that our global optimization method gives an exact recovery of the original source signals up to a scaling and permutation. Changing the scale of the recovered signal corresponds to changing the volume of the signal without affecting the actual content of the signal. And, multiplication by a permutation matrix only affects the order of the recovered signals. Thus recovering the original signals up to a scaling and permutation can be considered as an exact recovery.

The exact recovery of the original signals can be expressed in terms of the sparsity property of matrix  $P$  which maps the source signals  $s$  to the recovered signals  $v$  by  $v = Ps$ . If we denote  $x$  as the given mixture and  $A$  as the mixing matrix, we have  $x = As$ . Note that both  $A$  and  $s$  are unknown. We would like to construct a demixing matrix  $B$  that approximates  $A^{-1}$  in some sense so that  $v = Bx$  gives an accurate recovery of  $s$  up to a scaling and permutation. The matrix  $P$  can be expressed as  $P = BA$ . The exact recovery of the signals corresponds to requiring that  $P$  be written as a product of a permutation matrix and a diagonal matrix, which implies that each column of  $P$  lies on one of the Cartesian coordinates. In some sense, the exact recovery of the signals is equivalent to minimizing the  $L^0$  norm of each column of  $P$ , which is known to be extremely difficult. Surprisingly, we prove that minimizing the cross-correlation of the recovered signals in effect solves this  $L^0$  minimization problem for the columns of  $P$ .

In practice, the cross-correlation of the source signals is small but not zero. Solving the global optimization problem numerically also incurs a numerical error. In order to show that one can still obtain an accurate recovery in this case, we perform stability analysis of our global optimization method. We can prove that when the cross-correlation is small but nonzero and there is a small error in the global optimization problem, the global optimization method still gives an accurate recovery with a small error. We also analyze the solvability for the two-signal case when the mixing matrix is degenerate. To the best of our knowledge, this is the first error analysis of BSS methods.

A new difficulty associated with noisy mixtures is that the self-correlation of a Gaussian noise signal with any non-zero shift is essentially as small as the cross-correlation, leading to a degenerate recovery problem. Guided by our analysis, we propose a modified method that includes the cross-correlation with no-shift ( $n = 0$ ) as part of the energy objective functional. With this modification, we can show that the resulting recovery problem is non-degenerate for noisy mixtures. The new method works just as effectively for clean and noisy mixtures. Separating noise from real signals is known to be a challenging problem. Most de-noising methods tend to remove the high frequency information from the targeted signals, resulting in an unsatisfactory recovery. Surprisingly, our optimization method based on second order statistics can accurately separate mixed signals in a noisy environment even with a large signal-to-noise ratio.

---

We also study the challenging problem that occurs when the mixed matrix becomes degenerate. By carefully estimating the interaction between the degenerate self-correlation terms and the degenerate cross-correlation terms, we prove that we can still recover the original source signals essentially independent of the degeneracy of the measurements. Our theoretical results are confirmed by the numerical experiments.

### 1.3 Organization of the Paper

The rest of this paper is organized as follows. Section 2 is devoted to the study of instantaneous models. We begin with the simplified case, where we assume that the cross-correlation of the signals is exactly zero and that there is no error in the optimization method. We then transition into the more realistic case, where the cross-correlation of the signals is small, but not zero and there is an optimization error. Stability analysis is performed to demonstrate the robustness of the proposed method. We also consider the recovery of mixed signals in a noisy environment at the end of this section. Section 3 is devoted to convolutive mixtures. By using the Kronecker product, we can develop a parallel theory for the convolutive case. In Section 4, we consider the instantaneous two-signal mixture with degenerate measurements. Finally, in Section 5, we present several numerical results that confirm our theoretical results and demonstrate the accuracy and robustness of the proposed methods.

## 2 Recovery of Noisy Mixtures: the Instantaneous Case

In this section, we consider the simpler instantaneous mixture problems with three or more signals in a noisy environment. This case will help one to gain better understanding of the main ideas before considering the more complicated convolutive case. The stability analysis is also more easily done for the instantaneous case than that for the convolutive case.

### 2.1 Problem Setting: the Instantaneous Case

Often, our ears perceive multiple different signals mixed together. Whether it be while walking on the busy city streets or having a conversation with multiple people, we encounter these mixed signals daily. However, there are times when we desire to hear only one signal. For example, it may be tough to hear a person's voice in a busy restaurant when there is heavy background noise. Cases similar to these produce the following mathematical problem. Suppose  $s_i$  ( $i = 1, 2, 3$ ) are the three original sources, and  $x_i$  ( $i = 1, 2, 3$ ) are the three receivers (linear combinations of the sources), i.e.

$$x = As. \quad (1)$$

Our goal is to recover the original sources  $s$  from the mixture  $x$ . If  $A$  is known and nonsingular, then the problem is trivial, leading to the equation  $s = A^{-1}x$ . However, we do not know the exact value of matrix  $A$ . We are only given that it is a constant

---

matrix. Therefore, we are going to look for a matrix  $B$  such that

$$v = Bx \quad (2)$$

is an accurate recovery of the signals. An accurate recovery is a recovered signal  $v$  that is equal to the original signal  $s$  up to a scaling and permutation. This would be reasonable because the scaling would only vary the volume of the signals, and the permutation would only interchange the order of the original signals during the recovery. To find an accurate  $B$ , we need to make reasonable assumptions and design algorithms correspondingly.

In this paper, we first study the second order de-correlation method, see [21, 13, 15, 23] among others. This method is based on the assumption that the cross-correlations between original signals are 0, i.e.

**Assumption 2.1 (ZeroCorrelation).** *For a finite number of delay,  $|n| \leq N$  and  $i \neq j$ , we have*

$$\mathbb{E}[s_i(t)s_j(t-n)] = 0, \quad (3)$$

where  $\mathbb{E}$  is expectation, which is approximated by sample average using data in the frames.

Applying this non-correlation assumption to  $v$ , we may solve for  $B$ . In the case where we have two signals, the recovery matrix  $B$  can be constructed analytically, as was done in chapter 5 of [23]. However, when we have three or more signals to be separated from their mixtures, we cannot recover  $B$  analytically. we will formulate the problem to recover  $B$  as an optimization problem by minimizing the cross-correlation of  $v$ . The optimization formulation offers a more flexible approach to handle multiple signal recoveries, and this is the approach that we will adopt in this paper. And, as long as the global minimizer gives zero cross-correlation, we can prove that the optimization method yields an exact recovery under some non-degeneracy assumption on the mixing matrix  $A$ . However, for practical signals, the ZeroCorrelation assumption is not exactly satisfied; the cross-correlation is small, but never zero. One cannot get a minimizer that gives the cross-correlation to be zero because the cross-correlation is not zero and there will be error from the optimization algorithm. In this case, we can obtain a solution  $B$  that gives a small cross-correlation. Under several reasonable assumptions, we will prove that we can still recover the original source signals with good accuracy with an error depending on the size of the original signals' cross-correlation and the optimization algorithm's error.

## 2.2 Optimization: the Instantaneous Case

Since the volume of the recovered signals  $v$  does not matter, we may normalize each row of  $B$  to be a unit vector, i.e. for each  $i = 1, 2, 3$

$$\sum_{j=1}^3 b_{ij}^2 = 1. \quad (4)$$

---

**Remark 2.2.** There are other methods for normalizing  $B$ . For example, row normalization by  $L^1$  norm has been used in [23] (see page 164). One can also force the diagonal entries to be 1. Recall that the normalization of  $B$  must occur on the rows, because normalizing the rows of  $B$  is equivalent to rescaling the volume of the recovered signals  $v$ .

We propose to solve the following optimization problem to get an estimate of  $B$ .

$$\begin{aligned} \text{minimize } & F(B) := \sum_{i \neq j} \sum_{n \in \mathcal{I}} (\mathbb{E}[v_i(t)v_j(t-n)])^2 \\ \text{subject to: } & v = Bx, \\ & \sum_{j=1}^3 b_{ij}^2 = 1, \quad i = 1, 2, 3, \\ & |\det B| \geq b_0, \end{aligned} \tag{5}$$

where  $b_0$  is a positive parameter to be specified later. The index set  $\mathcal{I}$  is the set of shifts in which one wants to minimize the correlation. Let

$$\mathcal{I} = \{0, \pm 1, \dots, \pm N\}, \tag{6}$$

where  $N$  is the maximum shift. The constraint

$$|\det B| \geq b_0, \tag{7}$$

is to ensure that  $B$  is invertible and that the recovered signals are not degenerate.

In this paper, we use the global optimization code “fminsearch” provided by Matlab to solve the above optimization problem. Since we normalize the rows of matrix  $B$  by  $L^2$  norm, the parameters in our objective functional have finite range. To take advantage of this finite range of parameters in our global optimization problem, we use a slightly modified version of fminsearch provided by John D’Errico <sup>1</sup> in our study. The only change in D’Errico’s code from fminsearch is a rescaling from finite intervals to the whole space so that one can call fminsearch which is designed for the whole space. In our implementation, we first disregard the determinant constraint (7) and solve the simplified optimization problem. Upon getting the minimizer of  $B^*$ , we determine whether this minimizer is desirable. If it satisfies (7), we accept it; otherwise, we discard the output and continue to solve the same optimization problem using a different initial guess until we have obtained a minimizer of  $B^*$  that satisfies (7). From our experience, it does not take more than two trials before an accurate solution is attained.

### 2.3 Physical Assumptions: the Instantaneous Case

In order to guarantee a successful recovery of the original signals, first, we must ensure that the original signals  $s$  and the mixing matrix  $A$  have some good properties. Furthermore, the 3 measurements  $x$  should be independent enough to represent the 3 original signals  $s$ . Mathematically, it is

---

<sup>1</sup><http://www.mathworks.com/matlabcentral/fileexchange/8277-fminsearchbnd-fminsearchcon>

---

**Assumption 2.3 (Asigma).** *There exists a positive constant  $C_{cond}$  such that*

$$cond(A) := \frac{\sigma_M}{\sigma_m} \leq C_{cond}, \quad (8)$$

where  $\sigma_M$  and  $\sigma_m$  are the largest and smallest singular values of  $A$  respectively.

Secondly, the original signals  $s$  should have some intrinsic properties. We denote the second order statistics of a vector signal  $x$  as:

$$c_{n,ij}^x := \mathbb{E}[x_i(t)x_j(t-n)], \quad |n| \leq N, \quad i, j \in \{1, 2, 3\} \quad (9)$$

and the covariance matrix

$$C_n^x := (c_{n,ij}^x)_{i,j \in \{1,2,3\}}, \quad |n| \leq N. \quad (10)$$

The following intrinsic property of the original signals  $s$  will play an essential role in the recovery.

**Assumption 2.4 (Cinvertible).** *There exists three shifts  $n_1, n_2$  and  $n_3$  such that*

$$C_{n_1,n_2,n_3}^s := \begin{bmatrix} c_{n_1,11}^s & c_{n_1,22}^s & c_{n_1,33}^s \\ c_{n_2,11}^s & c_{n_2,22}^s & c_{n_2,33}^s \\ c_{n_3,11}^s & c_{n_3,22}^s & c_{n_3,33}^s \end{bmatrix}_{3 \times 3} \quad (11)$$

is invertible.

To prove stability, we need a stronger condition for  $C_{n_1,n_2,n_3}^s$ .

**Assumption 2.5 (Cstable).** *There exists three shifts  $n_1, n_2, n_3$  and a positive constant  $C_s$  such that  $C_{n_1,n_2,n_3}^s$  is invertible and*

$$\|(C_{n_1,n_2,n_3}^s)^{-1}\|_{L^2} \leq C_s. \quad (12)$$

The reason why we require these assumptions will become clear later in our analysis.

## 2.4 Exact Recovery: the Instantaneous Case

In this subsection, we assume that the ZeroCorrelation assumption holds true. In a sense, the following theorem shows that  $A^{-1}$  is one of the global minimizers to the optimization method (5).

**Theorem 2.6.** *Suppose Assumption Asigma (8) holds true and let*

$$b_0 = \frac{1}{C_{cond}^3}. \quad (13)$$

*Then there exists a feasible  $B^*$  for the optimization problem (5) and it gives the global minimum to be 0.*

---

*Proof.* First, we construct  $B^*$  by normalizing each row of  $A^{-1}$ . Suppose that the matrix  $A$  has the following singular value decomposition (SVD):

$$A = U\Sigma V^T = U \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \sigma_3 \end{bmatrix}_{3 \times 3} V^T, \quad (14)$$

then

$$A^{-1} = V\Sigma^{-1}U^T = V \begin{bmatrix} 1/\sigma_1 & & \\ & 1/\sigma_2 & \\ & & 1/\sigma_3 \end{bmatrix}_{3 \times 3} U^T. \quad (15)$$

Write  $V$  as

$$V = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}_{3 \times 3},$$

then

$$A^{-1} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}_{3 \times 3} = V\Sigma^{-1}U^T = \begin{bmatrix} v_1\Sigma^{-1}U^T \\ v_2\Sigma^{-1}U^T \\ v_3\Sigma^{-1}U^T \end{bmatrix}_{3 \times 3}. \quad (16)$$

Since  $U$  and  $V$  are unitary, we have  $\|v_i\|_2 = 1$  and

$$\|a_i\|_2 := \|v_i\Sigma^{-1}U^T\|_2 = \|v_i\Sigma^{-1}\|_2 \leq \|\Sigma^{-1}\|_2 = 1/\sigma_m, \quad (17)$$

where  $\sigma_m$  is the smallest singular value of  $A$ .

Define

$$B^* = \begin{bmatrix} a_1/\|a_1\|_2 \\ a_2/\|a_2\|_2 \\ a_3/\|a_3\|_2 \end{bmatrix}_{3 \times 3}, \quad (18)$$

then  $B^*$  is row-normalized, and the recovery

$$v = B^*x = B^*As = \begin{bmatrix} s_1/\|a_1\|_2 \\ s_2/\|a_2\|_2 \\ s_3/\|a_3\|_2 \end{bmatrix}_{3 \times 3}$$

gives an exact recovery up to a scaling and the objective function value will be 0, which is the global minimum. Now we only need to show that  $B^*$  satisfies the determinant constraint (7). Specifically, we have

$$|\det B^*| = \frac{|\det A^{-1}|}{\|a_1\|_2\|a_2\|_2\|a_3\|_2} = \frac{1}{\prod_{i=1}^3 \sigma_i \|a_i\|_2}.$$

Using estimate (17), recall that  $\sigma_M$  is the largest singular value of  $A$ . Therefore, we have

$$|\det B^*| \geq \frac{1}{\prod_{i=1}^3 \sigma_M/\sigma_m} = \frac{1}{\text{cond}(A)^3} \geq \frac{1}{C_{\text{cond}}^3},$$

where Assumption Asigma (8) is used in the last inequality. Combined with condition (13), we get

$$|\det B^*| \geq \frac{1}{C_{\text{cond}}^3} = b_0,$$

which proves that  $B^*$  satisfies the determinant condition.  $\square$

The following theorem shows that if  $B$  is a solution of the optimization method, then  $v = Bx$  will result in an exact recovery up to a scaling and permutation.

**Theorem 2.7.** *Suppose the ZeroCorrelation assumption (3) and the Cinvertible assumption (11) hold true. If there is an invertible matrix  $P$  such that*

$$v = Ps, \quad (19)$$

and for delay  $|n| \leq N$  and  $i \neq j$ ,

$$\mathbb{E}[v_i(t)v_j(t-n)] = 0, \quad (20)$$

then there exists a 3 by 3 permutation matrix  $\Pi$  and nonzero scalars  $\lambda_i$  ( $i = 1, 2, 3$ ) such that

$$P = \Pi \Lambda := \Pi \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix}_{3 \times 3}. \quad (21)$$

*Proof.* From the bilinear property of covariance, we know for any shift  $|n| \leq N$ , we have

$$C_n^v = PC_n^s P^T. \quad (22)$$

Writing  $P$  as

$$P = [p_1 \ p_2 \ p_3]_{3 \times 3},$$

we then get

$$\begin{aligned} C_n^v &= [p_1 \ p_2 \ p_3]_{3 \times 3} \begin{bmatrix} c_{n,11}^s & c_{n,12}^s & c_{n,13}^s \\ c_{n,21}^s & c_{n,22}^s & c_{n,23}^s \\ c_{n,31}^s & c_{n,32}^s & c_{n,33}^s \end{bmatrix}_{3 \times 3} \begin{bmatrix} p_1^T \\ p_2^T \\ p_3^T \end{bmatrix}_{3 \times 3} \\ &= [p_1 \ p_2 \ p_3]_{3 \times 3} \begin{bmatrix} c_{n,11}^s & 0 & 0 \\ 0 & c_{n,22}^s & 0 \\ 0 & 0 & c_{n,33}^s \end{bmatrix}_{3 \times 3} \begin{bmatrix} p_1^T \\ p_2^T \\ p_3^T \end{bmatrix}_{3 \times 3} \\ &\quad + [p_1 \ p_2 \ p_3]_{3 \times 3} \begin{bmatrix} 0 & c_{n,12}^s & c_{n,13}^s \\ c_{n,21}^s & 0 & c_{n,23}^s \\ c_{n,31}^s & c_{n,32}^s & 0 \end{bmatrix}_{3 \times 3} \begin{bmatrix} p_1^T \\ p_2^T \\ p_3^T \end{bmatrix}_{3 \times 3} \\ &= \sum_{i=1}^3 c_{n,ii}^s p_i p_i^T + \sum_{i \neq j} c_{n,ij}^s p_i p_j^T. \end{aligned} \quad (23)$$

Using the ZeroCorrelation assumption (3), we have

$$C_n^v = \sum_{i=1}^3 c_{n,ii}^s p_i p_i^T. \quad (24)$$

Using the ZeroCorrelation condition (20) of  $v$ , we have

$$\begin{aligned} \sum_{i=1}^3 c_{n,ii}^s p_i(1)p_i(2) &= 0, \\ \sum_{i=1}^3 c_{n,ii}^s p_i(1)p_i(3) &= 0, \\ \sum_{i=1}^3 c_{n,ii}^s p_i(2)p_i(3) &= 0. \end{aligned} \quad (25)$$

Taking  $n = n_1, n_2, n_3$  in the Cinvertible assumption (11), we obtain:

$$C_{n_1, n_2, n_3}^s \begin{bmatrix} p_1(1)p_1(2) \\ p_2(1)p_2(2) \\ p_3(1)p_3(2) \end{bmatrix}_{3 \times 1} = 0. \quad (26)$$

The Cinvertible assumption (11) implies that

$$\begin{bmatrix} p_1(1)p_1(2) \\ p_2(1)p_2(2) \\ p_3(1)p_3(2) \end{bmatrix}_{3 \times 1} = 0. \quad (27)$$

Similarly, we have

$$\begin{bmatrix} p_1(1)p_1(3) \\ p_2(1)p_2(3) \\ p_3(1)p_3(3) \end{bmatrix}_{3 \times 1} = 0, \quad \begin{bmatrix} p_1(2)p_1(3) \\ p_2(2)p_2(3) \\ p_3(2)p_3(3) \end{bmatrix}_{3 \times 1} = 0. \quad (28)$$

From all the equations involving  $p_1$ , we get:

$$[p_1(1)p_1(2) \ p_1(1)p_1(3) \ p_1(2)p_1(3)]_{1 \times 3} = 0. \quad (29)$$

Therefore, there exist  $i_1 \in \{1, 2, 3\}$  and a scalar  $\lambda_1$  such that

$$p_1 = \lambda_1 e_{i_1}, \quad (30)$$

where  $e_{i_1}$  is the unit vector along the  $i_1$ 's coordinate. Similarly, we have

$$p_2 = \lambda_2 e_{i_2}, \quad p_3 = \lambda_3 e_{i_3}. \quad (31)$$

As a result, we obtain (21), i.e.

$$P = [\lambda_1 e_{i_1} \ \lambda_2 e_{i_2} \ \lambda_3 e_{i_3}]_{3 \times 3} = \Pi \Lambda$$

where we denote

$$\Pi := [e_{i_1} \ e_{i_2} \ e_{i_3}]_{3 \times 3}.$$

Taking the determinant of  $P$ , we have

$$\det P = \det \Pi \det \Lambda = \epsilon_{i_1, i_2, i_3} \prod_{i=1}^3 \lambda_i,$$

where  $\epsilon_{i_1, i_2, i_3}$  is the Levi-Civita symbol, which is defined as follows:  $\epsilon_{ijk} = 1$  if  $(i, j, k)$  is  $(1, 2, 3)$ , or  $(2, 3, 1)$  or  $(3, 1, 2)$ ;  $\epsilon_{ijk} = -1$  if  $(i, j, k)$  is  $(3, 2, 1)$ , or  $(2, 1, 3)$  or  $(1, 3, 2)$ ;  $\epsilon_{ijk} = 0$  if  $i = j$ , or  $j = k$ , or  $k = i$ . By assumption,  $P$  is invertible. Therefore  $i_1, i_2$  and  $i_3$  cannot repeat and  $\lambda_i (i = 1, 2, 3)$  must be nonzero.  $\square$

The following corollary guarantees that the solution given from the optimization method will yield an exact recovery.

**Corollary 2.8.** *Suppose Assumption Asigma (8) holds true. Further, we assume that the parameter  $b_0$  satisfies condition (13). Then every solution of the optimization method will yield an accurate recovery. More precisely, if  $B$  is a global minimizer, and if we define*

$$P = BA, \tag{32}$$

*then there exists nonzero scalars  $\lambda_i (i = 1, 2, 3)$  and a 3 by 3 permutation matrix  $\Pi$  such that*

$$P = \Pi \Lambda := \Pi \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix}_{3 \times 3}. \tag{33}$$

**Remark 2.9.** Observe that each column of  $P$  lies on a Cartesian coordinate axis. Thus exact recovery is equivalent to imposing sparsity on each column of  $P$ . It is interesting to note that although we did not impose sparsity on  $B$  or  $P$  explicitly in our global optimization problem, the exact recovery of the source signals up to a scaling and permutation corresponds to finding  $B$  that minimizes the  $L^0$  norm of the columns of  $P$ .

*Proof.* From Theorem 2.6, we know that the global minimum for the global optimization problem must be zero. Therefore, if  $B$  is a global minimizer, then  $v = Bx$  will satisfy equation (20)

$$\mathbb{E}[v_i(t)v_j(t-n)] = 0.$$

The Assumption Asigma (8) and condition (13) imply that  $A$  and  $B$  are both invertible. Thus, this leads to the conclusion that  $P$  is invertible. Therefore, we can apply Theorem 2.7 to prove our corollary.  $\square$

## 2.5 Stability Analysis: the Instantaneous Case

In more realistic circumstances, the cross-correlation is “small,” but the ZeroCorrelation assumption (3) is not exactly satisfied. Therefore, we need to modify (3) accordingly to the following “small” cross-correlation assumption.

---

**Assumption 2.10 (EpsCorrelation).** *There exists a small number  $\epsilon$  such that for delay  $|n| \leq N$  and  $i \neq j$*

$$\mathbb{E}[s_i(t)s_j(t-n)] \leq \epsilon. \quad (34)$$

Under Assumption 2.10, we have a theorem similar to Theorem 2.6. However, now the global minimum is not guaranteed to be zero anymore.

**Theorem 2.11.** *Suppose Assumption 2.10 and Assumption Asigma (8) hold true. If the parameter  $b_0$  satisfies the condition (13), then there exists a feasible  $B^*$  for the optimization problem (5) and*

$$F(B^*) \leq C\sigma_M^4\epsilon^2, \quad (35)$$

where  $\sigma_M$  is the largest singular value of  $A$  and  $C$  is a constant depending linearly on  $N$ .

*Proof.* We construct  $B^*$  similar to that in Theorem 2.6. As a result, we will recover the signals

$$v = B^*x = B^*As = \begin{bmatrix} s_1/\|a_1\|_2 \\ s_2/\|a_2\|_2 \\ s_3/\|a_3\|_2 \end{bmatrix}_{3 \times 3}.$$

Thus, we get

$$F(B^*) = \sum_{i \neq j} \sum_{n \in \mathcal{I}} \left( \frac{\mathbb{E}[s_i(t)s_j(t-n)]}{\|a_i\|_2\|a_j\|_2} \right)^2.$$

As in Equation (17), we have

$$\|a_i\|_2 := \|v_i\Sigma^{-1}U^T\|_2 = \|v_i\Sigma^{-1}\|_2 \geq \|v_i\|_2/\sigma_M = 1/\sigma_M. \quad (36)$$

Therefore, we get:

$$F(B^*) \leq \sigma_M^4 \sum_{i \neq j} \sum_{n \in \mathcal{I}} \epsilon^2 \leq C\sigma_M^4\epsilon^2.$$

□

In practice, we cannot expect that the recovered signals  $v$  have zero cross-correlation (20), so instead, we assume that the objective function  $F(B)$  is small. We assume that  $B$  is a global minimizer of the optimization problem and that there exists a small positive number  $\delta$  such that

$$F(B) \leq \delta^2. \quad (37)$$

Equation (37) implies that the recovered signals have small cross-correlation, i.e. for all  $|n| \leq N$  and  $i \neq j$ ,

$$|\mathbb{E}[v_i(t)v_j(t-n)]| \leq \delta. \quad (38)$$

Now we want to show that the recovered signals  $v = Bx = BAs := Ps$  is an accurate recovery within an error tolerance depending on  $\epsilon$  and  $\delta$ . However, before we state

the theorem, we would like to state the following inequality. We write  $P = BA$  in row vectors:

$$\begin{bmatrix} p^1 \\ p^2 \\ p^3 \end{bmatrix}_{3 \times 3} = \begin{bmatrix} b^1 \\ b^2 \\ b^3 \end{bmatrix}_{3 \times 3} A.$$

Recall that when we normalize  $\|b^i\|_2$  to be 1, we get an estimate of the row vector of  $P$ :

$$\|p^i\|_2 \leq \|b^i\|_2 \|A\|_2 = \sigma_M. \quad (39)$$

The following theorem guarantees that the optimization method is robust with respect to a slight perturbation of the ZeroCorrelation Assumption (3).

**Theorem 2.12.** *Suppose Assumption 2.10, Assumption Asigma (8) and Assumption Cstable (12) hold true. Because we have already set the parameter  $b_0$  to satisfy condition (13), if  $B$  is a global minimizer of the optimization problem and if there exists a small positive number  $\delta$  such that (37) holds true, then*

$$P = BA = \Pi \Lambda + \sqrt{C_d C_s} (\sigma_M^2 \epsilon + \delta)^{1/2}, \quad (40)$$

where  $C_d$  is a constant that only depends on the number of signals  $d$  (we have  $d = 3$ ) and  $C_s$  is the constant in Assumption Cstable (12),

$$\Pi := [e_{i_1} \ e_{i_2} \ e_{i_3}]_{3 \times 3}, \quad i_1, i_2, i_3 \in \{1, 2, 3\}$$

and

$$\Lambda := \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix}_{3 \times 3}.$$

Moreover, if  $\epsilon$  and  $\delta$  are small enough, then  $\Pi$  is a permutation matrix and  $\lambda_i (i = 1, 2, 3)$  are nonzero scalars.

*Proof.* From Equation (22) and (23), we have

$$\begin{aligned} C_n^v &= [p_1 \ p_2 \ p_3]_{3 \times 3} \begin{bmatrix} c_{n,11}^s & 0 & 0 \\ 0 & c_{n,22}^s & 0 \\ 0 & 0 & c_{n,33}^s \end{bmatrix}_{3 \times 3} \begin{bmatrix} p_1^T \\ p_2^T \\ p_3^T \end{bmatrix}_{3 \times 3} \\ &\quad + \begin{bmatrix} p^1 \\ p^2 \\ p^3 \end{bmatrix}_{3 \times 3} \begin{bmatrix} 0 & c_{n,12}^s & c_{n,13}^s \\ c_{n,21}^s & 0 & c_{n,23}^s \\ c_{n,31}^s & c_{n,32}^s & 0 \end{bmatrix}_{3 \times 3} [(p^1)^T \ (p^2)^T \ (p^3)^T]_{3 \times 3} \\ &= \sum_{i=1}^3 c_{n,ii}^s p_i p_i^T + [p^i C^{off}(p^j)^T]_{i,j=1,2,3}, \end{aligned} \quad (41)$$

where

$$C^{off} := \begin{bmatrix} 0 & c_{n,12}^s & c_{n,13}^s \\ c_{n,21}^s & 0 & c_{n,23}^s \\ c_{n,31}^s & c_{n,32}^s & 0 \end{bmatrix}_{3 \times 3}. \quad (42)$$

From Assumption 2.10, we get

$$\|C^{off}\|_2 \leq C_d\epsilon. \quad (43)$$

From the Cauchy Schwarz inequality we get for any  $i, j = 1, 2, 3$ :

$$\|p^i C^{off}(p^j)^T\|_2 \leq \|p^i\|_2 \|C^{off}(p^j)^T\|_2 \leq \|p^i\|_2 \|C^{off}\|_2 \|(p^j)^T\|_2.$$

Using Equations (39) and (43), we have

$$\|p^i C^{off}(p^j)^T\|_2 \leq C_d\sigma_M^2\epsilon. \quad (44)$$

Using Equations (38) and (44), we obtain

$$\begin{aligned} \left| \sum_{i=1}^3 c_{n,ii}^s p_i(1)p_i(2) \right| &\leq C_d\sigma_M^2\epsilon + \delta, \\ \left| \sum_{i=1}^3 c_{n,ii}^s p_i(1)p_i(3) \right| &\leq C_d\sigma_M^2\epsilon + \delta, \\ \left| \sum_{i=1}^3 c_{n,ii}^s p_i(2)p_i(3) \right| &\leq C_d\sigma_M^2\epsilon + \delta. \end{aligned} \quad (45)$$

Similar to how we obtain Equation (26), we get the following linear system

$$C_{n_1,n_2,n_3}^s \begin{bmatrix} p_1(1)p_1(2) \\ p_2(1)p_2(2) \\ p_3(1)p_3(2) \end{bmatrix}_{3 \times 1} = C_d\sigma_M^2\epsilon + \delta. \quad (46)$$

Thus, we get

$$\begin{aligned} \left\| \begin{bmatrix} p_1(1)p_1(2) \\ p_2(1)p_2(2) \\ p_3(1)p_3(2) \end{bmatrix}_{3 \times 1} \right\|_2 &\leq C_d \|(C_{n_1,n_2,n_3}^s)^{-1}\|_2 (\sigma_M^2\epsilon + \delta) \\ &\leq C_d C_s (\sigma_M^2\epsilon + \delta), \end{aligned} \quad (47)$$

where we have used Assumption Cstable (12) in the last inequality. Similarly, we obtain

$$\left\| \begin{bmatrix} p_1(1)p_1(3) \\ p_2(1)p_2(3) \\ p_3(1)p_3(3) \end{bmatrix}_{3 \times 1} \right\|_2 \leq C_d C_s (\sigma_M^2\epsilon + \delta), \quad \left\| \begin{bmatrix} p_1(2)p_1(3) \\ p_2(2)p_2(3) \\ p_3(2)p_3(3) \end{bmatrix}_{3 \times 1} \right\|_2 \leq C_d C_s (\sigma_M^2\epsilon + \delta). \quad (48)$$

From all of the above estimates involving  $p_1$ , for  $i \neq j$ , we get:

$$[p_1(1)p_1(2) \quad p_1(1)p_1(3) \quad p_1(2)p_1(3)]_{1 \times 3} = C_d C_s (\sigma_M^2\epsilon + \delta). \quad (49)$$

---

Obviously, there cannot be two entries of  $p_1$  whose absolute values are larger than  $\sqrt{C_d C_s (\sigma_M^2 \epsilon + \delta)}$  at the same time. Therefore, we obtain

$$p_1 = \lambda_1 e_{i_1} + \sqrt{C_d C_s} (\sigma_M^2 \epsilon + \delta)^{1/2}, \quad (50)$$

where  $e_{i_1}$  is the unit vector along the  $i_1$ 's coordinate. Similarly, we have

$$\begin{aligned} p_2 &= \lambda_2 e_{i_2} + \sqrt{C_d C_s} (\sigma_M^2 \epsilon + \delta)^{1/2}, \\ p_3 &= \lambda_3 e_{i_3} + \sqrt{C_d C_s} (\sigma_M^2 \epsilon + \delta)^{1/2}. \end{aligned} \quad (51)$$

As a consequence, we have proved (40)

$$P = BA = \Pi \Lambda + \sqrt{C_d C_s} (\sigma_M^2 \epsilon + \delta)^{1/2}.$$

Taking the determinant of  $P$ , we have

$$|\det P| = |\det B||\det A| \geq b_0 \prod_{i=1}^3 \sigma_i.$$

Furthermore, since determinant,  $\det$ , is a differentiable function, we use Taylor expansion around  $\Pi \Lambda$  and get

$$|\det P| = |\det(\Pi \Lambda) + \mathcal{O}((\sigma_M^2 \epsilon + \delta)^{1/2})| = |\epsilon_{i_1, i_2, i_3} \prod_{i=1}^3 |\lambda_i| + \mathcal{O}((\sigma_M^2 \epsilon + \delta)^{1/2})|.$$

Therefore, if  $\epsilon$  and  $\delta$  are small enough,  $|\epsilon_{i_1, i_2, i_3} \prod_{i=1}^3 |\lambda_i|$  must be nonzero, implying that  $\Pi$  is a permutation matrix and  $\lambda_i (i = 1, 2, 3)$  are nonzero scalars.  $\square$

**Remark 2.13.** In order to ensure a successful recovery of the original signals within an error of tolerance, we need to prove that

$$|\lambda_i| \gg \sqrt{C_d C_s} (\sigma_M^2 \epsilon + \delta)^{1/2}. \quad (52)$$

We can show this by using the fact that singular values are continuously (differentially) dependent on the matrix entries. In the case when a matrix has different eigenvalues, all these eigenvalues are smoothly ( $\mathcal{C}^\infty$ ) dependent on the entries of the matrix [7].

Consider the matrix

$$\Pi \Lambda = P - \sqrt{C_d C_s} (\sigma_M^2 \epsilon + \delta)^{1/2}.$$

We assume that  $P$  has three different singular values,  $\sigma_m^P$  being the smallest singular value of  $P$ . Note that the smallest singular value of  $\Pi \Lambda$  is  $\min_{i=1,2,3} |\lambda_i|$ . Using the theorem above, we get:

$$\min_{i=1,2,3} |\lambda_i| = \sigma_m^P + \mathcal{O}(\sigma_M^2 \epsilon + \delta)^{1/2}.$$

Notice that

$$\sigma_m^P \geq \sigma_m^B \sigma_m^A = \sigma_m^B \sigma_m.$$

---

From the definition of the matrix  $L^2$  norm, we have

$$\begin{aligned}
\|B\|_2 &= \sup_{x \neq 0} \frac{\|Bx\|_2}{\|x\|_2} \\
&= \sup_{x \neq 0} \left\| \begin{bmatrix} b^1 x \\ b^2 x \\ b^3 x \end{bmatrix}_{3 \times 1} \right\|_2 / \|x\|_2 \\
&= \sup_{x \neq 0} \frac{\sqrt{(b^1 x)^2 + (b^2 x)^2 + (b^3 x)^2}}{\|x\|_2} \\
(\text{Cauchy Schwarz}) &\leq \sup_{x \neq 0} \frac{\sqrt{(\|b^1\|_2 \|x\|_2)^2 + (\|b^2\|_2 \|x\|_2)^2 + (\|b^3\|_2 \|x\|_2)^2}}{\|x\|_2} \\
(\text{use (4)}) &= \sqrt{3}.
\end{aligned}$$

Thus  $\sigma_M^B \leq \sqrt{3}$  and we have

$$\sigma_m^B \geq \frac{\prod_{i=1}^3 \sigma_i^B}{(\sigma_M^B)^2} = \frac{|\det B|}{(\sigma_M^B)^2} \geq b_0/3.$$

Therefore, we get:

$$\min_{i=1,2,3} |\lambda_i| \geq b_0 \sigma_m / 3 + \mathcal{O}(\sigma_M^2 \epsilon + \delta)^{1/2}.$$

We can set  $b_0 = 1/\text{cond}(A)^3$ . As long as  $\text{cond}(A)$  and  $\sigma_m$  are both of order 1, Equation (52) is satisfied. Consequently, we will have an accurate recovery within an error of tolerance.

## 2.6 Separating Mixed Signals in a Noisy Environment

It is not easy to separate a clean signal from Gaussian noise if one only has a single measurement. One may try to use an adaptive filter, such as wavelet filter to separate the noise from the signal; however, such filtering tends to damage the high frequency component of the signal. In this paper, we propose to treat the noise background as a separate signal. With one extra measurement, we can solve the optimization problem (5) to separate the noise from the mixed signals. As can be seen from our analysis above, the key component of this approach is to carefully choose the index set  $\mathcal{I}$  to make the Assumption (12) valid. In this section, we will restrict ourselves to the simpler instantaneous case.

In this paper, we will consider the noise as white noise  $W$ , which has the following correlation property:

$$C_n^W = \mathbb{E}[W(t)W(t-n)] = \begin{cases} \|W\|_2^2 & \text{for } n = 0, \\ 0 & \text{otherwise.} \end{cases} \quad (53)$$

**Remark 2.14.** For a sample of white noise  $W$ , if we calculate  $\mathbb{E}[W(t)W(t-n)]$  by average in time, the self-correlation may not be zero for a nonzero shift as in (53); however,

---

the self-correlation with a nonzero shift will definitely be very small. To simplify our analysis, we will assume that (53) holds true in the following analysis. The case where the self-correlation is small can be analyzed similarly.

In the case where we have one clean signal with white noise and two different measurements of its mixtures, it falls in the two-signal instantaneous case. To make sure that our optimization method can produce an accurate recovery, we require that the clean signal and noise satisfy the two-dimensional Cstable Assumption, i.e. there exist two shifts  $n_1$  and  $n_2$  such that

$$\left\| \left( \begin{bmatrix} c_{n_1,11}^s & c_{n_1,22}^s \\ c_{n_2,11}^s & c_{n_2,22}^s \end{bmatrix} \right)^{-1} \right\|_2 \leq C_s. \quad (54)$$

Since the second signal is white noise, one of  $n_1$  and  $n_2$  must be zero in order to satisfy the above condition. Without loss of generality, we make  $n_2 = 0$ , then

$$C_{n_1,0}^s = \begin{bmatrix} c_{n_1,11}^s & 0, \\ c_{0,11}^s & \|W\|_2^2 \end{bmatrix}. \quad (55)$$

Based on the stability analysis in Section 4, we conclude the following: for one clean signal mixed with a white noise signal, if we choose our index set  $\mathcal{I}$  to contain 0, then even in cases where the measurement  $A$  is nearly degenerate, we can still yield an accurate recovery of the clean signal!

**Remark 2.15.** For a signal,  $c_0^s = \|s\|_2^2$  represents its energy or volume. Therefore, if we assume that the clean signal has order one volume, then  $c_{0,11}^s$  is of order one. For the other case where  $c_{n_1,11}^s$  is too small for every  $n_1 \neq 0$ , the clean signal  $s_1$  degenerates to a nearly white noise signal, which is not an interesting case for us to consider. Therefore, for a regular clean signal mixed with white noise, our optimization problem always produces a satisfactory recovery.

Similarly, for the three-dimensional instantaneous case with  $s_3$  to be white noise, we must contain shift 0 in our index set  $\mathcal{I}$ . We will make  $n_3 = 0$ . Then we have

$$C_{n_1,n_2,0}^s := \begin{bmatrix} c_{n_1,11}^s & c_{n_1,22}^s & 0 \\ c_{n_2,11}^s & c_{n_2,22}^s & 0 \\ c_{n_3,11}^s & c_{n_3,22}^s & \|W\|_2^2 \end{bmatrix}_{3 \times 3}. \quad (56)$$

In this case, the Cinvertible Assumption is reduced to the corresponding two-dimensional invertible assumption, which requires that there exist two non-zero shifts  $n_1$  and  $n_2$  such that  $\begin{bmatrix} c_{n_1,11}^s & c_{n_1,22}^s \\ c_{n_2,11}^s & c_{n_2,22}^s \end{bmatrix}$  is invertible. The Cstable Assumption can also be reduced to stability of the same  $2 \times 2$  matrix.

Another interesting case is when the noise for each measurement is independent. In this case, we cannot recover the original source signals by introducing additional measurements. In this case, the problem can be formulated as follows:

$$x = As + \eta, \quad (57)$$

---

where  $\eta = (\eta_1, \eta_2, \eta_3)^T$  and  $\eta_i$  are independent Gaussian noise. Since  $A$  is assumed to be invertible, we can write  $\eta = A\xi$ , where  $\xi = A^{-1}\eta$  is also Gaussian. Then we can reformulate the above problem as

$$x = A\tilde{s}, \quad (58)$$

where  $\tilde{s} = s + \xi$ . If we treat  $\tilde{s}$  as our new source signals to be recovered from the measurements  $x$ , then we can consider the above problem as a standard Blind Source Separation problem with no noise. Our previous analysis can be applied to show that we can recover  $\tilde{s}$  accurately using global optimization. The good news of this approach is that we can still separate different signals from each other. The bad news is that the recovered signals still contain noise  $\xi$ . To further reduce the effect of noise in the recovered signal, one can apply some effective adaptive filtering method to  $\tilde{s}$  [9]. De-noising  $\tilde{s}$  is much easier than de-noising the original mixed measurements  $x$  since  $\tilde{s}_i = s_i + \xi_i$ , i.e.  $\tilde{s}_i$  does not contain any other source signal  $s_j$  for  $j \neq i$ .

We have performed numerical experiments to test the above finding, and our numerical results confirm the above analysis. Indeed, the additional post-processing filtering further reduces the noise, resulting a cleaner recovery. Additionally, even for one source signal plus noise, it is hard to eliminate the noise completely. Filtering tends to damage the high frequency components of the signal. In comparison to other methods, we found that other existing BSS methods, such as the Info-Max method [3, 1, 8, 22] and the Jade method [5, 6] (both use higher order statistics), and the analytical de-correlation method (section 5.2 of [23]) fail to separate the signals from each other when the mixtures are polluted with independent Gaussian noise for each measurement.

### 3 Recovery of Noisy Mixtures: the Convulsive Case

In this section, we will consider the more realistic case: the convulsive case, specifically, convulsive mixtures in a noisy environment. We will show that the analysis that we have developed in the previous section can still apply to the convulsive case – however, just with more complicated, involved algebra manipulation.

#### 3.1 Problem Setting: the Convulsive Case

In an enclosed environment, sound mixing is not instantaneous (or local in time), because sound tends to reflect from multiple surfaces; therefore, the signals mix non-locally in time. This leads to a convulsive mixing process and may be viewed as a discrete version of Green's formula to acoustic wave equations. The convulsive mixing model of two sources  $s_i (i = 1, 2)$  with mixtures recorded by two receivers is [23]:

$$\begin{aligned} x_1(t) &= \sum_{k=1}^q a_k^{11} s_1(t+1-k) + \sum_{k=1}^q a_k^{12} s_2(t+1-k), \\ x_2(t) &= \sum_{k=1}^q a_k^{21} s_1(t+1-k) + \sum_{k=1}^q a_k^{22} s_2(t+1-k). \end{aligned} \quad (59)$$

The model is an extension of (1) and can be written as:

$$x = As \quad (60)$$

with

$$A = \begin{bmatrix} a^{11}* & a^{12}* \\ a^{21}* & a^{22}* \end{bmatrix}, \quad (61)$$

where  $*$  is the abbreviation of linear convolution in (59) and  $a^{ij}$  is extended to the length of  $s_i$  and  $x_i$  by zero padding. In the following analysis, we will treat the signals  $s_i$  and  $x_i$  as periodic signals and calculate their shifted correlations. Because we will only use correlations with a small shift relative to their length, the periodically shifted correlations are nearly the same as the truncated-shifted correlations.<sup>2</sup> The collection of all the finite sequences forms a semi-group with the binary operation  $*$  being associative and commutative, namely,

$$(f * g) * h = f * (g * h) \quad f * g = g * f. \quad (62)$$

Similar to that of the instantaneous case, we would like to find a convolutive recovery of the form:

$$\begin{aligned} v_1(t) &= \sum_{k=1}^r b_k^{11} x_1(t+1-k) + \sum_{k=1}^r b_k^{12} x_2(t+1-k), \\ v_2(t) &= \sum_{k=1}^r b_k^{21} x_1(t+1-k) + \sum_{k=1}^r b_k^{22} x_2(t+1-k). \end{aligned} \quad (63)$$

And, this can be written as:

$$v = Bx \quad (64)$$

with

$$B = \begin{bmatrix} b^{11}* & b^{12}* \\ b^{21}* & b^{22}* \end{bmatrix}. \quad (65)$$

If the mechanism of mixing is known, i.e.,  $A$  is known, then one can consider choosing  $B$  as  $A$ 's adjoint matrix. To this end, let us consider:

$$\begin{aligned} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} &= \begin{bmatrix} a^{22}* & -a^{12}* \\ -a^{21}* & a^{11}* \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} a^{22}* & -a^{12}* \\ -a^{21}* & a^{11}* \end{bmatrix} \begin{bmatrix} a^{11}* & a^{12}* \\ a^{21}* & a^{22}* \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \\ &= \begin{bmatrix} a^{22} * a^{11} * -a^{12} * a^{21} * & a^{22} * a^{12} * -a^{12} * a^{22} * \\ -a^{21} * a^{11} * +a^{11} * a^{21} * & -a^{21} * a^{12} * +a^{11} * a^{22} * \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \\ &= \begin{bmatrix} a^{11} * a^{22} * -a^{12} * a^{21} * & s_1 \\ a^{11} * a^{22} * -a^{12} * a^{21} * & s_2 \end{bmatrix}. \end{aligned} \quad (66)$$

---

<sup>2</sup>Periodic shifted correlation and truncated shifted correlation are two ways to calculate the shifted correlation (9). The former will be used in the following analysis, while the latter is used in [23].

---

One can regard this as an accurate recovery because the recovered signal  $v_i(i = 1, 2)$  only depends on  $s_i(i = 1, 2)$ .

However, in realistic problems,  $A$  is unknown. We look for an accurate recovery matrix  $B$  to separate the mixed signals. To quantify the condition of an accurate recovery, we define:

$$P = BA = \begin{bmatrix} b^{11} * a^{11} * + b^{12} * a^{21} * & b^{11} * a^{12} * + b^{12} * a^{22} * \\ b^{21} * a^{11} * + b^{22} * a^{21} * & b^{21} * a^{12} * + b^{22} * a^{22} * \end{bmatrix} \quad (67)$$

and denote it as:

$$P := \begin{bmatrix} p^{11}* & p^{12}* \\ p^{21}* & p^{22}* \end{bmatrix}. \quad (68)$$

When either  $P$  or  $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} P$  is diagonal with nonzero diagonal entries, we call such  $B$  an accurate recovery matrix. In other words, it indicates that the recovered signal  $v_i(i = 1, 2)$  depends on one of the source signals  $s_i(i = 1, 2)$  only.

Since  $p^{ij}$  contains at most  $q + r - 1$  nonzero entries, we will make use of this sparsity property in separating the mixed signals.

### 3.2 Physical Assumptions and Optimization: the Convulsive Case

For the convulsive mixing model (59), it is natural to assume that the absolute value of  $a^{ij}$ 's entries decrease for every pair of  $(i, j)$ , since the energy decays after every reflection. Therefore, for each pair of  $(i, j)$ , we set the following condition:

$$|a^{ij}(k)| \leq |a^{ij}(1)|, \quad k = 1, 2, \dots, q. \quad (69)$$

From (66), we know that there will be an accurate recovery  $B$  which has the same length  $q$  and satisfies the same decay condition, i.e. for each pair of  $(i, j)$

$$|b^{ij}(k)| \leq |b^{ij}(1)|, \quad k = 1, 2, \dots, q. \quad (70)$$

To yield an accurate recovery, we still need an accurate measurement condition like (8). There are many ways to define a similar condition for the convulsive case. Here, we will just use the first entry, which contains the most energy. We define it as:

$$\text{cond}(A) := \text{cond}(A_1) = \frac{\sigma_M}{\sigma_m} \leq C_{\text{cond}}, \quad (71)$$

where  $\sigma_M$  and  $\sigma_m$  are respectively the largest and smallest singular values of  $A_1 = \begin{bmatrix} a_1^{11} & a_1^{12} \\ a_1^{21} & a_1^{22} \end{bmatrix}$ .

Similar to that of the instantaneous case, we normalize our variable  $B$  with  $L^2$  norm, but only on the first entry, i.e. for each  $i = 1, 2$

$$\sum_{j=1}^2 (b_1^{ij})^2 = 1. \quad (72)$$

Therefore, we propose the following optimization problem to get an estimate of  $B$ .

$$\begin{aligned}
& \text{minimize} \quad F(B) := \sum_{i \neq j} \sum_{n \in \mathcal{I}} (\mathbb{E}[v_i(t)v_j(t-n)])^2 \\
& \text{subject to: } v = Bx, \\
& \quad B = \begin{bmatrix} b^{11}* & b^{12}* \\ b^{21}* & b^{22}* \end{bmatrix} \text{with length } q, \\
& \quad \sum_{j=1}^2 (b_1^{ij})^2 = 1, \quad i = 1, 2, \\
& \quad |b^{ij}(k)| \leq |b^{ij}(1)|, \quad k = 1, 2, \dots, q \\
& \quad |\det B_1| \geq b_0,
\end{aligned} \tag{73}$$

where  $b_0$  is a positive parameter to be specified later. The function of the constraint

$$|\det B_1| \geq b_0 \tag{74}$$

has the same meaning as constraint (7) in the instantaneous case.

### 3.3 Conditions on $C^s$ : the Convulsive Case

We need to estimate the correlation of  $v$  given by  $v = Ps$ , where convolution is involved. In the optimization problem (69), we specify the length of  $B$  to be  $q$ , and as a result, the length of  $P$  to be  $2q - 1$ . We denote  $\tau = 2q - 1$ . From the following expression

$$\begin{aligned}
v_1(t) &= \sum_{k=1}^{\tau} p_k^{11} s_1(t+1-k) + \sum_{k=1}^{\tau} p_k^{12} s_2(t+1-k), \\
v_2(t) &= \sum_{k=1}^{\tau} p_k^{21} s_1(t+1-k) + \sum_{k=1}^{\tau} p_k^{22} s_2(t+1-k),
\end{aligned} \tag{75}$$

we get:

$$C_{n,12}^v = (p^{11})^T C_{n,11}^{s,\tau} p^{21} + (p^{12})^T C_{n,22}^{s,\tau} p^{22} + (p^{12})^T C_{n,21}^{s,\tau} p^{21} + (p^{11})^T C_{n,12}^{s,\tau} p^{22}, \tag{76}$$

where the  $\tau \times \tau$  dimensional matrix  $C_{n,ij}^{s,\tau}$  is defined as:

$$C_{n,ij}^{s,\tau}(k_1, k_2) = C_{n+k_2-k_1, ij}^s, \quad k_1, k_2 = 1, 2, \dots, \tau. \tag{77}$$

Note that in the two-source case, we have

$$F(B) := \sum_{n \in \mathcal{I}} (\mathbb{E}[v_1(t)v_2(t-n)])^2 = \sum_{n \in \mathcal{I}} (C_{n,12}^v)^2. \tag{78}$$

If we reshape  $C_{n,ij}^{s,\tau}$  to be a row vector  $c_{n,ij}^{s,\tau}$  with dimension  $1 \times \tau^2$ , by padding row after row, (76) can be written as:

$$c_{n,11}^{s,\tau}(p^{11} \otimes p^{21}) + c_{n,22}^{s,\tau}(p^{12} \otimes p^{22}) = C_{n,12}^v - (p^{12})^T C_{n,21}^{s,\tau} p^{21} - (p^{11})^T C_{n,12}^{s,\tau} p^{22}, \tag{79}$$

---

where  $\otimes$  is the standard Kronecker product [4]. In the special case where  $a = (a_1, a_2, \dots, a_r)$  and  $b = (b_1, b_2, \dots, b_r)$  are two row vectors of length  $r$ ,  $a \otimes b$  is a row vector of length  $r^2$  given as follows:

$$a \otimes b = (a_1 b_1, a_1 b_2, \dots, a_1 b_r, a_2 b_1, a_2 b_2, \dots, a_2 b_r, \dots, a_r b_1, a_r b_2, \dots, a_r b_r).$$

Because the right hand side of the equation is expected to be small, we can assume that  $p^{11} \otimes p^{21}$  and  $p^{12} \otimes p^{22}$  are nearly zero, implying that the recovery is accurate. To make this argument valid, we make the following assumption:

**Assumption 3.1 (Cinvertible2).** *There exist  $2\tau^2$  shifts  $n_i$  such that*

$$C^s := \begin{bmatrix} c_{n_1,11}^{s,\tau} & c_{n_1,22}^{s,\tau} \\ c_{n_2,11}^{s,\tau} & c_{n_2,22}^{s,\tau} \\ \vdots & \vdots \\ c_{n_{2\tau^2},11}^{s,\tau} & c_{n_{2\tau^2},22}^{s,\tau} \end{bmatrix}_{2\tau^2 \times 2\tau^2} \quad (80)$$

is invertible.

To prove stability, we need a stronger condition for  $C^s$ :

**Assumption 3.2 (Cstable2).** *There exists  $2\tau^2$  shifts  $n_i$  and a positive constant  $C_s$  such that  $C^s$  is invertible and*

$$\|(C^s)^{-1}\|_{L^2} \leq C_s. \quad (81)$$

### 3.4 Exact Recovery: the Convulsive Case

In this section, we assume that the ZeroCorrelation assumption (3) holds true. Similar to Theorem 2.6, the following theorem shows that the adjoint matrix of  $A$  is one of the global minimizers of our optimization method (73) after normalization.

**Theorem 3.3.** *Suppose Assumption Asigma (71) holds true and let*

$$b_0 = \frac{1}{C_{cond}^2}, \quad (82)$$

*then there exists a feasible  $B^*$  for the optimization problem (73) which gives the global minimum to be 0.*

*Proof.* We first construct  $B^*$  from normalizing  $\begin{bmatrix} a^{22*} & -a^{12*} \\ -a^{21*} & a^{11*} \end{bmatrix}$  with rule (72). This proof is the same as the proof of Theorem 2.6, except, in this case, we have two dimensions.  $\square$

We can also get the following results that are parallel to those of the instantaneous case.

---

**Theorem 3.4.** Suppose the ZeroCorrelation assumption (3) and the Cinvertible assumption (80) hold true. If there is a convolutive matrix  $P$  such that  $P_1$  is invertible, and

$$v = Ps, \quad (83)$$

and for delay  $|n| \leq N$ , and  $i \neq j$ ,

$$\mathbb{E}[v_i(t)v_j(t-n)] = 0, \quad (84)$$

then either  $P$  or  $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} P$  is diagonal with nonzero diagonal entries.

*Proof.* Under these assumptions, the right hand side of equation (79) is zero. Combined with the Cinvertible assumption (80), we conclude that  $p^{11} \otimes p^{21}$  and  $p^{12} \otimes p^{22}$  must be zero, which implies that either  $P$  or  $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} P$  is diagonal.

Since  $P_1$  is invertible, the corresponding diagonal entries are not zero.  $\square$

The following corollary guarantees that the solution given by our optimization method will yield an accurate recovery.

**Corollary 3.5.** Suppose Assumption A sigma (71) holds true and the parameter  $b_0$  satisfies condition (82). Then, every solution of the optimization method will yield an accurate recovery. Specifically, if  $B$  is a global minimizer, and we define

$$P = BA, \quad (85)$$

then either  $P$  or  $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} P$  is diagonal with nonzero diagonal entries.

*Proof.* Making use of the fact  $P_1 = B_1 A_1$ , the rest of the proof is exactly the same as the proof of Corollary 2.8.  $\square$

### 3.5 Stability Analysis: the Convulsive Case

Before we perform the stability analysis, we state an important inequality. Suppose vector  $a^{ij}$  and  $b^{ij}$  enjoy decay properties (69) and (70) respectively; then, we have:

$$\|b^{ij} * a^{ij}\|_2 \leq |a^{ij}(1)| |b^{ij}(1)| \sqrt{\frac{2q^3 + q}{3}} =: |a^{ij}(1)| |b^{ij}(1)| C_{1,q}. \quad (86)$$

We can prove the above inequality by simply replacing  $a^{ij}$  and  $b^{ij}$  with positive constant vectors.

Under Assumption 2.10, we have a theorem similar to Theorem 3.3. However, now, the global minimum is no longer guaranteed to be zero.

---

**Theorem 3.6.** Suppose Assumption 2.10 and Assumption Asigma (71) hold true. If the parameter  $b_0$  satisfies condition (82), then there exists a feasible  $B^*$  for our optimization problem (73) and

$$F(B^*) \leq C_q \sigma_M^4 \epsilon^2, \quad (87)$$

where  $\sigma_M$  is the largest singular value of  $A_1$ , and  $C_q$  is a constant dependent on  $q$  and  $N$ .

*Proof.* We construct  $B^*$  by normalizing  $\begin{bmatrix} a^{22}* & -a^{12}* \\ -a^{21}* & a^{11}* \end{bmatrix}$  with rule (72). Denoting  $\hat{a} = a^{11} * a^{22} - a^{12} * a^{21}$  and using equation (66), we get:

$$C_{n,12}^v = \frac{\hat{a}^T C_{n,12}^{s,2q-1} \hat{a}}{\sqrt{(a^{22}(1))^2 + (a^{12}(1))^2} \sqrt{(a^{21}(1))^2 + (a^{11}(1))^2}}. \quad (88)$$

By inequality (86), we get:

$$\begin{aligned} \|\hat{a}\|_2 &\leq \|a^{11} * a^{22}\|_2 + \|a^{12} * a^{21}\|_2 \\ &\leq C_{1,q}(|a^{11}(1)||a^{22}(1)| + |a^{12}(1)||a^{21}(1)|) \\ &\leq C_{1,q} \sqrt{(a^{22}(1))^2 + (a^{12}(1))^2} \sqrt{(a^{21}(1))^2 + (a^{11}(1))^2}. \end{aligned} \quad (89)$$

Since every entry of  $C_{n,12}^{s,2q-1}$  is smaller than  $\epsilon$ , there exists a constant  $C_{2,q}$  dependent on  $q$  such that

$$\|C_{n,12}^{s,2q-1}\|_2 \leq \epsilon C_{2,q}. \quad (90)$$

Then we have

$$\begin{aligned} C_{n,12}^v &\leq \frac{\|C_{n,12}^{s,2q-1}\|_2 \|\hat{a}\|_2^2}{\sqrt{(a^{22}(1))^2 + (a^{12}(1))^2} \sqrt{(a^{21}(1))^2 + (a^{11}(1))^2}} \\ &\leq \epsilon C_{2,q} C_{1,q}^2 \sqrt{(a^{22}(1))^2 + (a^{12}(1))^2} \sqrt{(a^{21}(1))^2 + (a^{11}(1))^2} \\ &\leq \epsilon C_{2,q} C_{1,q}^2 \sigma_M^2, \end{aligned} \quad (91)$$

where  $\sigma_M$  is the largest singular value of  $A_1$ . In the last inequality, we have used the fact that the  $L^2$  norm of each column or row of a matrix is not larger than the  $L^2$  norm of the full matrix.

Summing  $n$  over the index set  $\mathcal{I}$  proves the theorem.  $\square$

The following theorem guarantees that our optimization method is stable with respect to a small perturbation of the ZeroCorrelation Assumption (3).

**Theorem 3.7.** Suppose Assumption 2.10, Assumption Asigma (71), and Assumption Cstable (81) hold true, and the parameter  $b_0$  satisfies condition (82). If  $B$  is a feasible minimizer for the optimization problem and there exists a small positive number  $\delta$  such that (37) holds true, then

$$\left\| \begin{bmatrix} p^{11} \otimes p^{21} \\ p^{12} \otimes p^{22} \end{bmatrix} \right\|_2 \leq C_q C_s (C_{2,q} C_{1,q}^2 \sigma_M^2 \epsilon + \delta), \quad (92)$$

---

where  $C_q$  is a constant that only depends on  $q$ ,  $C_s$  is the constant in Assumption Cstable (12), and  $C_{1,q}$  and  $C_{2,q}$  are the constants in Theorem 3.6.

Moreover, if  $\epsilon$  and  $\delta$  are small enough, then either  $P$  or  $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} P$  has  $\mathcal{O}(1)$  diagonal entries and  $\mathcal{O}(\sqrt{C_q C_s} (C_{2,q} C_{1,q}^2 \sigma_M^2 \epsilon + \delta)^{1/2})$  off-diagonal entries.

*Proof.* First, we note that

$$\begin{aligned} \|p^{11}\|_2 &\leq \|b^{11} * a^{11} + b^{12} * a^{21}\|_2 \\ &\leq C_{1,q}(|b^{11}(1)| |a^{11}(1)| + |b^{12}(1)| |a^{21}(1)|) \\ &\leq C_{1,q} \sqrt{(b^{11}(1))^2 + (b^{12}(1))^2} \sqrt{(a^{11}(1))^2 + (a^{21}(1))^2} \\ &\leq C_{1,q} \sigma_M, \end{aligned} \tag{93}$$

using the normalization in (72). Similarly, we have:

$$\|p^{ij}\|_2 \leq C_{1,q} \sigma_M. \tag{94}$$

Using (79), we get:

$$|c_{n,11}^{s,\tau}(p^{11} \otimes p^{21}) + c_{n,22}^{s,\tau}(p^{12} \otimes p^{22})| \leq \delta + C_{2,q} C_{1,q}^2 \sigma_M^2 \epsilon. \tag{95}$$

Combined with Cstable assumption (81), we obtain (92).

As in the proof of Theorem 2.12 and Remark 2.13, we can perform similar analysis to the first entry of  $P_1$ , making use of  $P_1 = B_1 A_1$ . Then we can get the estimate that either  $P$  or  $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} P$  has  $\mathcal{O}(1)$  diagonal entries and  $\mathcal{O}(\sqrt{C_q C_s} (C_{2,q} C_{1,q}^2 \sigma_M^2 \epsilon + \delta)^{1/2})$  off-diagonal entries.  $\square$

## 4 Analysis for Two-Dimensional Degenerate Case

In this section, we consider the two-signal case in the instantaneous setting where the mixing matrix is degenerate. The analysis that we perform for the three-signal instantaneous case can be generalized to any finite number of signals. However, for the simpler case  $D = 2$ , the problem has a better structure which allows us to obtain sharper stability estimates.

For the two-signal case, as in [23], we can represent the mixing matrix  $A$  as follows:

$$A = \begin{bmatrix} \sin(\varphi) & \sin(\theta) \\ \cos(\varphi) & \cos(\theta) \end{bmatrix}. \tag{96}$$

Note: based on this form of  $A$ , the original sources  $s_1$  and  $s_2$  may have different amplitudes, which may result in failing to satisfy the Cstable Assumption (12).

From our numerical experiments, we found that even when  $A$  was moderately ill-conditioned, i.e.  $\varphi$  and  $\theta$  are close to each other, which violates the Asigma Assumption (8), we can still obtain an accurate recovery. This suggests that our method works

---

better than the stability analysis suggests. For the two-signal case, we will provide a more refined stability analysis to justify that this is indeed correct.

In the two-signal case, we use the  $L^2$  normalization (4) of  $B$  by writing  $B$  as follows:

$$B = \begin{bmatrix} \cos(\hat{\theta}) & -\sin(\hat{\theta}) \\ -\cos(\hat{\varphi}) & \sin(\hat{\varphi}) \end{bmatrix}, \quad (97)$$

and then

$$P = BA = \begin{bmatrix} \sin(\varphi - \hat{\theta}) & \sin(\theta - \hat{\theta}) \\ \sin(\hat{\varphi} - \varphi) & \sin(\hat{\varphi} - \theta) \end{bmatrix}. \quad (98)$$

Suppose that Assumption 2.10 holds true. We denote  $\delta = |\varphi - \theta|$ . By choosing  $\hat{\varphi} = \varphi$  and  $\hat{\theta} = \theta$ , we obtain

$$|\mathbb{E}[v_i(t)v_j(t-n)]| \leq \sin^2(\varphi - \theta)\epsilon \curvearrowright \mathcal{O}(\delta^2\epsilon), \quad |n| \leq N. \quad (99)$$

Note that the above estimate (99) is much sharper than what we obtained in Theorem 35. This sharp estimate will enable us to prove the following theorem of an accurate recovery when  $A$  is ill-conditioned.

**Theorem 4.1.** Suppose that Assumption 2.10 holds true, and that  $A = \begin{bmatrix} \sin(\varphi) & \sin(\theta) \\ \cos(\varphi) & \cos(\theta) \end{bmatrix}$  with  $|\varphi - \theta| = \delta$ . If  $B = \begin{bmatrix} \cos(\hat{\theta}) & -\sin(\hat{\theta}) \\ -\cos(\hat{\varphi}) & \sin(\hat{\varphi}) \end{bmatrix}$  is a global minimum for our optimization problem (5) with the shift set  $\mathcal{I} = \{n_1, n_2\}$  which satisfies the Assumption Cstable (12), then we have either

$$\begin{aligned} |\hat{\varphi} - \varphi| &= \mathcal{O}(\sqrt{\epsilon}\delta), \\ |\hat{\theta} - \theta| &= \mathcal{O}(\sqrt{\epsilon}\delta), \end{aligned} \quad (100)$$

or

$$\begin{aligned} |\hat{\varphi} - \theta| &= \mathcal{O}(\sqrt{\epsilon}\delta), \\ |\hat{\theta} - \varphi| &= \mathcal{O}(\sqrt{\epsilon}\delta). \end{aligned} \quad (101)$$

These two estimates will guarantee an accurate recovery with or without swapping.

*Proof.* As in the three-signal case, we can get the following linear system from the off-diagonal part of  $C_n^v = PC_n^sP^T$ :

$$\begin{bmatrix} c_{n_1,11}^s & c_{n_1,22}^s \\ c_{n_2,11}^s & c_{n_2,22}^s \end{bmatrix} \begin{bmatrix} \sin(\varphi - \hat{\theta}) \sin(\hat{\varphi} - \varphi) \\ \sin(\theta - \hat{\theta}) \sin(\hat{\varphi} - \theta) \end{bmatrix} = \begin{bmatrix} c_{n_1,12}^v \\ c_{n_2,12}^v \end{bmatrix} - \begin{bmatrix} c_{n_1,12}^s & c_{n_1,21}^s \\ c_{n_2,12}^s & c_{n_2,21}^s \end{bmatrix} \begin{bmatrix} \sin(\varphi - \hat{\theta}) \sin(\hat{\varphi} - \theta) \\ \sin(\theta - \hat{\theta}) \sin(\hat{\varphi} - \varphi) \end{bmatrix}. \quad (102)$$

Since  $B$  is a global minimum, we have from (99) that

$$\begin{bmatrix} c_{n_1,12}^v \\ c_{n_2,12}^v \end{bmatrix} \curvearrowright \mathcal{O}(\delta^2\epsilon). \quad (103)$$

From Assumption 2.10, we have

$$\begin{bmatrix} c_{n_1,12}^s & c_{n_1,21}^s \\ c_{n_2,12}^s & c_{n_2,21}^s \end{bmatrix} \backsim \mathcal{O}(\epsilon). \quad (104)$$

Combining the above estimate with the Assumption Cstable (12), we have

$$\begin{aligned} & \sin(\varphi - \hat{\theta}) \sin(\hat{\varphi} - \varphi) \\ &= \epsilon(c_{11} \sin(\varphi - \hat{\theta}) \sin(\hat{\varphi} - \theta) + c_{12} \sin(\theta - \hat{\theta}) \sin(\hat{\varphi} - \varphi)) + c_{13} \epsilon \delta^2, \end{aligned} \quad (105)$$

$$\begin{aligned} & \sin(\theta - \hat{\theta}) \sin(\hat{\varphi} - \theta) \\ &= \epsilon(c_{21} \sin(\varphi - \hat{\theta}) \sin(\hat{\varphi} - \theta) + c_{22} \sin(\theta - \hat{\theta}) \sin(\hat{\varphi} - \varphi)) + c_{23} \epsilon \delta^2, \end{aligned} \quad (106)$$

with bounded constants  $c_{ij}$  depending on  $C_s$  only. We will consider two cases:

Case 1:  $|\sin(\varphi - \hat{\theta}) \sin(\hat{\varphi} - \theta)| > |\sin(\theta - \hat{\theta}) \sin(\hat{\varphi} - \varphi)|$ .

From (105), we have

$$|\sin(\varphi - \hat{\theta}) \sin(\hat{\varphi} - \theta)| \leq \epsilon(|c_{11}| + |c_{12}|) |\sin(\varphi - \hat{\theta}) \sin(\hat{\varphi} - \theta)| + c_{13} \epsilon \delta^2. \quad (107)$$

If  $|\sin(\varphi - \hat{\theta})| \leq \sqrt{\epsilon} \delta$ , then we have

$$|\hat{\theta} - \varphi| = \mathcal{O}(\sqrt{\epsilon} \delta). \quad (108)$$

If  $|\sin(\varphi - \hat{\theta})| > \sqrt{\epsilon} \delta$ , then we divide both sides of equation (107) by  $|\sin(\varphi - \hat{\theta})|$  to get

$$\begin{aligned} |\sin(\hat{\varphi} - \varphi)| &\leq \epsilon(|c_{11}| + |c_{12}|) |\sin(\hat{\varphi} - \theta)| + \mathcal{O}(\sqrt{\epsilon} \delta) \\ &\leq (|c_{11}| + |c_{12}|) \epsilon (|\hat{\varphi} - \varphi| + |\varphi - \theta|) + \mathcal{O}(\sqrt{\epsilon} \delta). \end{aligned} \quad (109)$$

Note: from  $\frac{2}{\pi} |\hat{\varphi} - \varphi| \leq |\sin(\hat{\varphi} - \varphi)|$  and  $|\varphi - \theta| = \delta$ , we have

$$\begin{aligned} & \left( \frac{2}{\pi} - \epsilon(|c_{11}| + |c_{12}|) \right) |\hat{\varphi} - \varphi| \\ & \leq \epsilon(|c_{11}| + |c_{12}|) \delta + \mathcal{O}(\sqrt{\epsilon} \delta). \end{aligned} \quad (110)$$

Then, we get:

$$|\hat{\varphi} - \varphi| = \mathcal{O}(\sqrt{\epsilon} \delta). \quad (111)$$

Similarly, from (106), we will have either

$$|\hat{\varphi} - \theta| = \mathcal{O}(\sqrt{\epsilon} \delta), \quad (112)$$

or

$$|\hat{\theta} - \theta| = \mathcal{O}(\sqrt{\epsilon} \delta). \quad (113)$$

Recall that  $|\varphi - \theta| = \delta$ . We can only have the following two combinations among (108), (111), (112) and (113):

$$\begin{aligned} & |\hat{\varphi} - \varphi| = \mathcal{O}(\sqrt{\epsilon} \delta), \\ & |\hat{\theta} - \theta| = \mathcal{O}(\sqrt{\epsilon} \delta), \end{aligned}$$

---

or

$$|\hat{\varphi} - \theta| = \mathcal{O}(\sqrt{\epsilon}\delta), \\ |\hat{\theta} - \varphi| = \mathcal{O}(\sqrt{\epsilon}\delta).$$

All other combinations would lead to  $\varphi - \theta = O(\sqrt{\epsilon}\delta)$ , which contradicts  $\delta = |\varphi - \theta|$ .

Case 2:  $|\sin(\varphi - \hat{\theta}) \sin(\hat{\varphi} - \theta)| \leq |\sin(\theta - \hat{\theta}) \sin(\hat{\varphi} - \varphi)|$ . Following the same argument as above, we will obtain either (100) or (101).  $\square$

We believe that under some appropriate assumption on the degeneracy of the mixing matrix  $A$ , we will be able to extend Theorem 4.1 to any finite dimensional case. To achieve this, we need to obtain a sharper estimate on every term in equation (23) and take into account all possible cancellations among the degenerate terms. This will be reported in a future work.

## 5 Numerical Results

In this section, we present several numerical results to demonstrate the accuracy and robustness of our method. As our stability analysis suggests, it is important to select an appropriate index set so the conditions listed in our stability analysis are satisfied. Since we use an optimization approach, we can take a larger shift index set for the shifts than the minimum number of shifts that is required by our stability analysis. Thus, even though the stability constant in our stability theorem may be large for minimum shift index set (e.g.  $N = d$  for the  $d$ -signal instantaneous case), one can still get a reasonably accurate recovery. In our numerical implementation, we choose a shift index set to be  $\mathcal{I} = \{n \mid 0 \leq n \leq 10\}$  in the instantaneous case and  $\mathcal{I} = \{n \mid -200 \leq n \leq 200\}$  in the convolutive case with  $q = 5$  and  $\tau = 9$ .

We also compare the performance of our method with that of the analytical method (section 5.2 of [23]) based on second order statistics, as well as with the Info-Max method that uses information of speech probability distribution. We do not compare our method with the Jade method [5, 6] since this method is limited to the instantaneous mixtures only. This comparison shows that our method can indeed give better performance than that of the analytical de-correlation method (section 5.2, [23]) and the Info-Max method, especially when the mixing matrix is degenerate. The Info-Max method has been proposed and investigated by several authors, see, e.g., [3, 1, 8, 22]. The code that we use for the comparison is based on the work of [22]. Our numerical results confirm our theoretical findings. We consider both the instantaneous model and the convolutive model. In the instantaneous case, we consider both the 2-signal case and the 3-signal case. Of particular interest is the case when the mixing matrix is degenerate. In each of the following examples, we mix together sound signals, specifically that of speech. We use diverse speech signals to ensure that our method can accommodate with all types of signals.

In each test example, the output results include SIRI, which measures the signal-to-interference ratio improvement (see page 145 of [23] for its definition), and the matrix

---

$P$ , which confirms Theorem 2.7 and demonstrates the relative accuracy of the recovered signals through the relationship between  $v$  and  $s$ . A larger SIRI indicates greater improvement in the signal. Furthermore, we introduce a quantity  $\text{sigma}P$  that measures the relative error between  $P$  (or its permutation) and the diagonal matrix – a bigger  $\text{sigma}P$  translates to a better recovery. Roughly speaking,  $\text{sigma}P = \min_i \text{sigma}P_i$ , where  $\text{sigma}P_i$  is defined as the ratio between the largest element in absolute value and the second largest element in absolute value of the  $i$ th row of  $P$ . In all examples, the signal-to-noise ratio  $\text{SNR} = \frac{P_{\text{signal}}}{P_{\text{noise}}}$  is about  $\text{SNR} = 2$ .

**Example 1.** In the first example, we consider the mixtures of one speech signal with Gaussian noise in the instantaneous case. The results obtained by our optimization method are presented in Figure 1. The top row of Figure 1 plots the two mixed signals that are generated using the following mixing matrix

$$A = \begin{bmatrix} 0.1191 & 0.8377 \\ 0.9929 & 0.5461 \end{bmatrix}. \quad (114)$$

The recovered signals are plotted in the second row, while the original source signals are plotted in the third row of Figure 1. In our optimization method, the shift of  $n$  is taken from 0 to 10. Our method gives SIRI = 52.2411, which indicates that our method gives significant enhancement over the original mixed signals. This is also confirmed by the large value of  $\text{sigma}P = 382.9470$ . Finally, the fact that matrix

$$P = \begin{bmatrix} -0.6877 & -0.0016 \\ -0.0018 & -0.6879 \end{bmatrix} \quad (115)$$

is nearly diagonal gives another confirmation that our optimization method gives an excellent recovery of the original mixed signals without swapping.

We also compare the performance of our method with that of the analytical de-correlation method presented in section 5.2 of [23] (page 142-143). As we mentioned earlier, this method of [23] with  $n = 1, 2$  does not take into account explicitly the impact of noise. Although the analytical method in [23] still gives a reasonable enhancement of the mixed signals with SIRI = 18.8522, one can still hear the background noise from the recovered signal. This can be explained by the  $P$  matrix:

$$P = \begin{bmatrix} 1.0000 & 0.0916 \\ 0.0000 & 0.9438 \end{bmatrix}. \quad (116)$$

One can see that the second component of the first row of  $P$  is not small, about 9.16%, indicating that 9.16% of the noise is contained in the recovered speech signal. The value,  $\text{sigma}P = 10.3055$ , also confirms this finding. In comparison, our optimization method gives a significantly larger  $\text{sigma}P$ , 382.9470.

Finally, we compare our optimization method with the Info-Max method, which approximates the speech probability distribution function. Using higher order statistics, we observe a considerable improvement over the analytical method of [23] that uses second order statistics. In particular, the Info-Max method gives SIRI = 19.4455,

---

$\sigma_P = 160.31$ , and

$$P = \begin{bmatrix} -0.0048 & 0.7695 \\ 0.9149 & -0.0027 \end{bmatrix}. \quad (117)$$

It is interesting to note that the SIRI value does not give an adequate description of the improvement of the Info-Max method over the second order analytical method [23]. The  $\sigma_P$  value seems to be a better indicator. Note that our optimization method still outperforms the Info-Max method in this case by a factor of 2.4.

**Example 2.** In this example, we consider the 3-signal mixtures in the instantaneous case. We use two speech signals mixed with a Gaussian noise signal. The mixing matrix  $A$  is given by

$$A = \begin{bmatrix} 2.0795 & -1.2925 & -0.3978 \\ 0.2554 & -0.5062 & 1.1136 \\ -0.953 & 1.8890 & -0.2922 \end{bmatrix}. \quad (118)$$

We compare the performance of our optimization method with that of the Info-Max method. The results obtained by our optimization method are given in Figure 2. Again, we plot the mixtures in the first row, the recovered signals in the second row, and the original source signals in the third row; which are in Figure 2. Our method gives SIRI = 20.6428,  $\sigma_P = 43.4990$ , and

$$P = \begin{bmatrix} 0.9825 & 0.0226 & -0.0054 \\ -0.0125 & 1.0119 & -0.0011 \\ -0.0037 & 0.0073 & 0.9979 \end{bmatrix}. \quad (119)$$

Although the results are not as accurate as those of the 2-signal case, the recovered signals still have reasonable accuracy.

We tend to believe that a main factor in the decrease of accuracy of the 3-signal case as well as the convolutive case is due to the performance of the global optimization algorithm in the Matlab code. In the 3-signal instantaneous case and the convolutive case, the number of unknowns increases significantly, making it extremely difficult to find the global minimum of our nonconvex energy functional. In many cases, the algorithm only returns a local minimum instead of a global minimum. The ability to design an even more effective global optimization method will be the key to our proposed approach, a subject that we will investigate in the future.

We have also compared this method with the Info-Max method using the same mixing matrix and data. The Info-Max method gives SIRI = 17.4613,  $\sigma_P = 23.9777$ , and

$$P = \begin{bmatrix} 1.3979 & 0.0583 & -0.0008 \\ 0.0003 & 0.0092 & 1.0318 \\ 0.0143 & 1.1425 & 0.0083 \end{bmatrix}. \quad (120)$$

As in the two-signal case, we can see that our optimization method still outperforms the Info-Max method for the 3-signal case by about a factor of 2 if we compare the values of  $\sigma_P$  obtained by the two approaches.

This numerical experiment demonstrates that our global optimization method can, with reasonable accuracy, recover the two source signals mixed with noise. Not only do

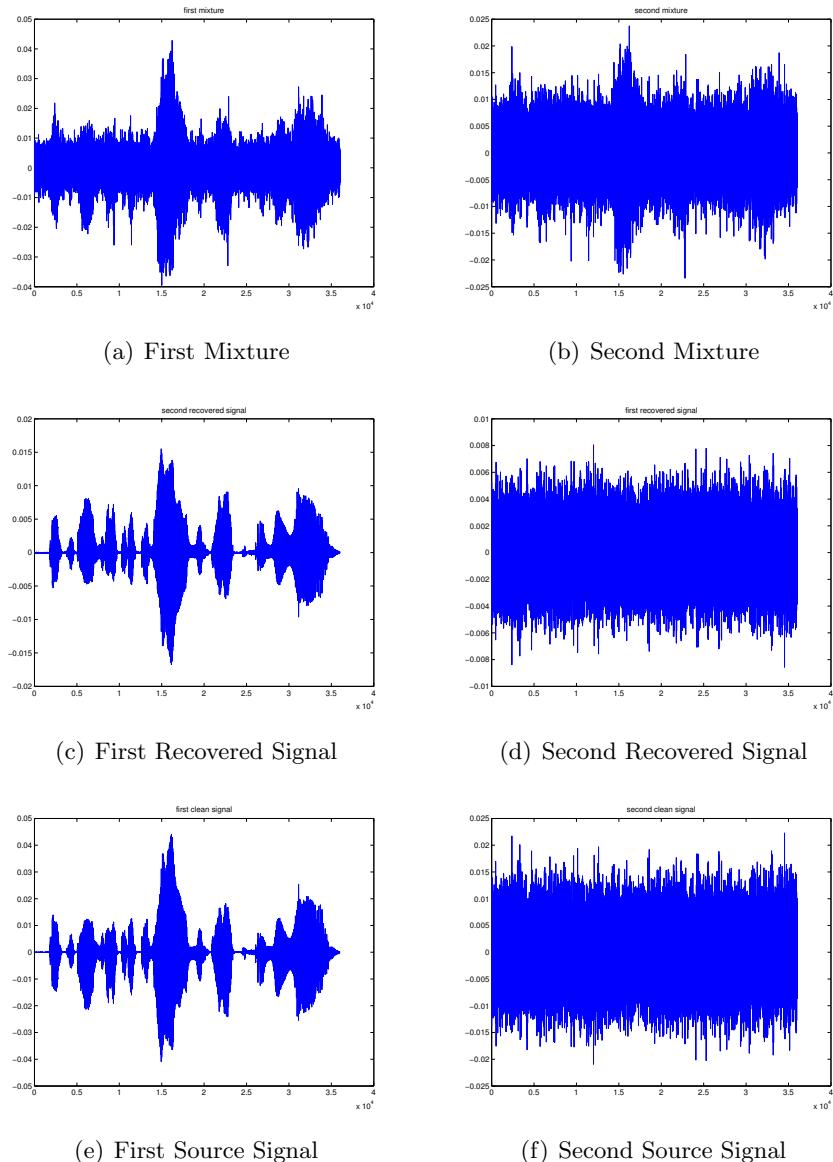


Figure 1: This is the instantaneous case where we have a regular signal and a signal composed of random noise. The computation is performed using our optimization method with shifts  $n$  taken from 0 to 10. Our method gives SIRI = 52.2411 and  $\sigma_P = 382.9470$ .

our results show an excellent recovery of generic sound signals, but we also demonstrate that we can handle the impact of Gaussian noise. The small values of the off-diagonals in the  $P$  matrix solidify our claims that our three-signal global optimization method can accurately recover source signals.

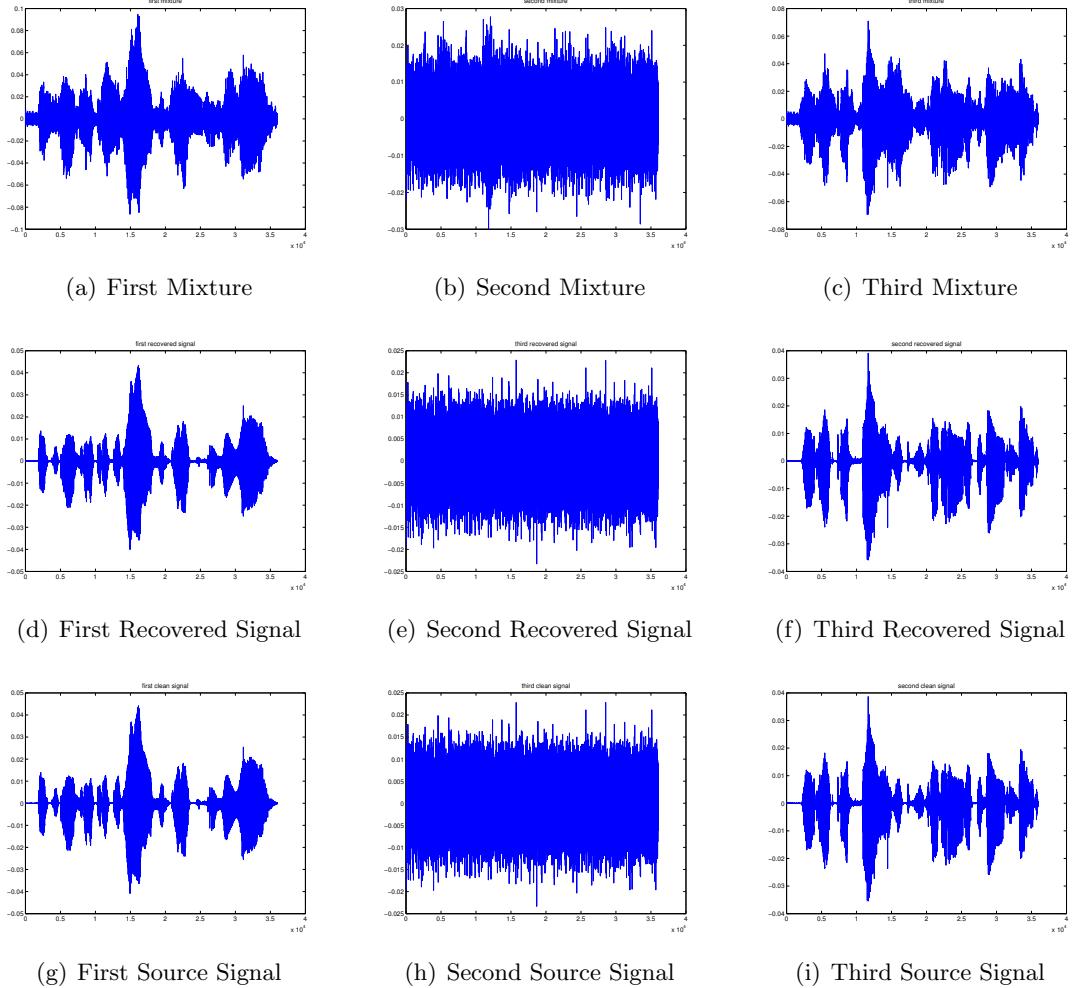


Figure 2: This is the instantaneous case where we explore the 3-signal case, where two sources are regular speech signals and one source is a random noise signal. To recover the signals, we use our global optimization method. Our method gives SIRI = 20.6428 and  $\sigma_P = 43.4990$ .

**Example 3.** Next, we consider the two-signal instantaneous case with a degenerate mixing matrix  $A$ . Again, we consider the case where we have one regular speech signal mixed with random noise. Our numerical results will confirm the theoretical results obtained in Section 4 – that we can still obtain accurate recovery of the mixed signals even when the mixing matrix  $A$  is degenerate.

In Figure 3, we plot the results obtained by our our optimization method. The

---

layout is exactly the same as that of Figure 1. The degenerate matrix  $A$  is given by

$$A = \begin{bmatrix} 0.7071 & 0.7078 \\ 0.7071 & 0.7064 \end{bmatrix}. \quad (121)$$

The degree of degeneracy can be measured by its determinant,  $\det(A) = 0.00607406$ . Our optimization method gives  $SIRI = 22.1174$ ,  $sigmaP = 41.1483$ , and

$$P = \begin{bmatrix} 1.0031 & 0.0179 \\ -0.0031 & 0.9821 \end{bmatrix}. \quad (122)$$

Given the nature of the degenerate measurement, these results are quite impressive.

We compare our optimization method with the analytical method of [23] with shift  $n$  taken from 1 to 2. This method of [23] gives  $SIRI = 10.3856$ ,  $sigmaP = 9.9176$ , and

$$P = \begin{bmatrix} 1.0000 & 0.0916 \\ 0.0000 & 0.9084 \end{bmatrix}. \quad (123)$$

As one can infer from matrix  $P$ , there is about 9.16% of mixture of the noise background in the first recovered source signal. When we play the recovered signal, the effect of the noise is quite noticeable.

We also apply the Info-Max method to this degenerate measurement case. The Info-Max method gives  $SIRI = 5.2362$ ,  $sigmaP = 3.9571$ , and

$$P = \begin{bmatrix} 0.2530 & -0.0202 \\ 0.1096 & 0.4337 \end{bmatrix}. \quad (124)$$

As we can see from matrix  $P$ , about 23% of the first source signal is still mixed with the second source signal in the recovered second signal. When we play the recovered signals, the mixing of the two signals is quite strong. We have also applied the Info-Max method to many other signals with a degenerate measurement. We found that the Info-Max method always gives very poor recovery when  $A$  is degenerate.

**Example 4.** Finally, we test the performance of our optimization method for the more realistic convolutive case. We use the same signals and their mixtures as those used in [23]. The results obtained by our global optimization method are presented in Figure 4. The layout is the same as before. Again, the mixed signals are plotted in the first row, the recovered signals in the second row, and the original source signals in the third row. Because of complications in calculating the SIRI for the convolutive case, we use the  $P$  matrix to determine the accuracy of our signal recoveries for convolutive cases.

Our method gives  $sigmaP = 24.7826$ , which is calculated by averaging the relative error of each component (or its permutation) of  $P$  to the diagonal matrix. In particular, the first component of  $P$ , denoted as  $P_1$ , contains the largest energy. To illustrate the effectiveness of our recovery, we present  $P_1$  below.

$$P_1 = \begin{bmatrix} 1.0000 & 0.0007 \\ -0.0063 & 1.0000 \end{bmatrix}. \quad (125)$$

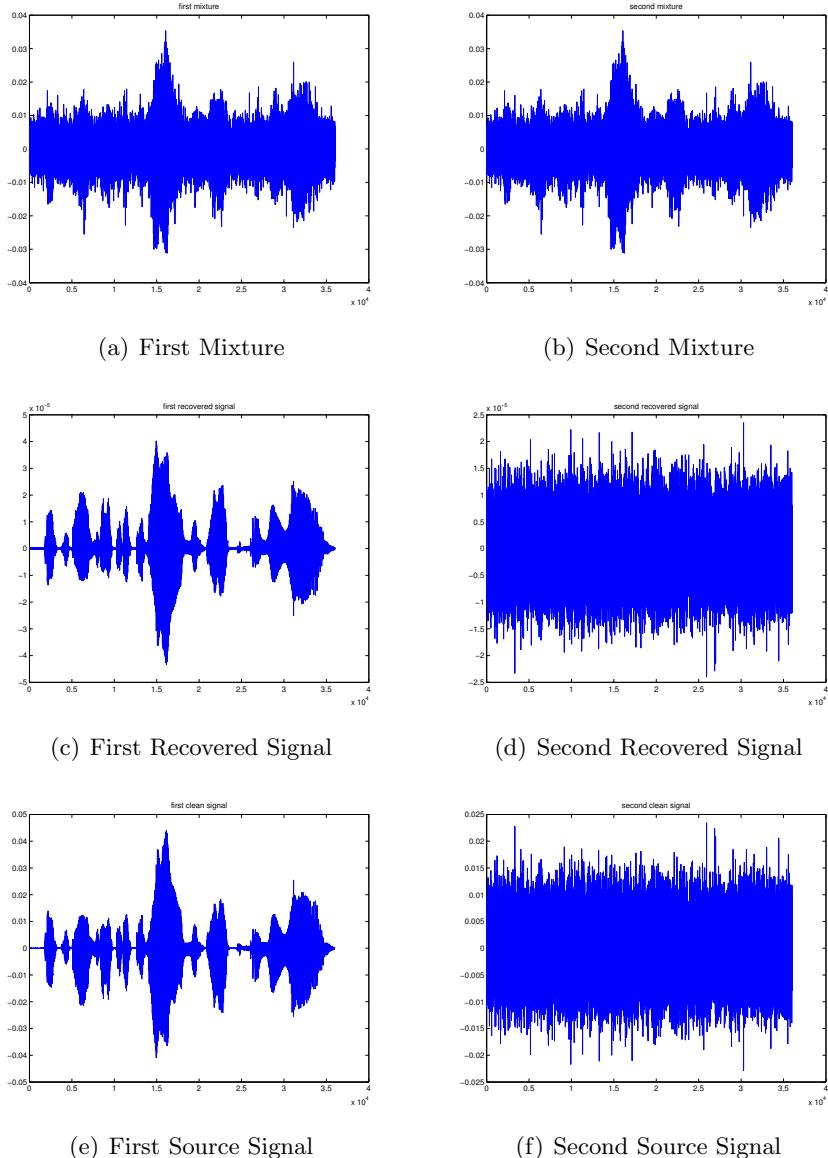


Figure 3: This is the instantaneous case where the  $A$  matrix is degenerate. We have one regular signal mixed with random noise in the instantaneous setting. Our optimization method gives  $SIRI = 22.1174$  and  $\sigma_{AP} = 41.1483$ .

As we can see, this first component of  $P$  is very close to a diagonal matrix.

We also compare the performance of our global optimization method with that of the convulsive decorrelation method [13] (see section 5.5 of [23], pp. 160-164) which

---

gives  $\sigma_P = 8.1993$  with

$$P_1 = \begin{bmatrix} 0.1199 & 0.0036 \\ 0.0011 & 0.1533 \end{bmatrix}. \quad (126)$$

Both  $\sigma_P$  and the  $P_1$  matrix indicate that even in the convolutive cases, we are still able to recover the signals. Using our global optimization method, we observe that our global optimization method gives a better recovery than that of the de-correlation method of [13, 23], which seeks a local minimum in its optimization step. Both  $P_1$  matrices' off-diagonals and measurements of  $\sigma_P$  further confirm our conclusion. In particular, the  $\sigma_P$  value for our global optimization method is 3 times that of a local minimum from the de-correlation method of [13, 23].

## References

- [1] S. Amari, A. Cichocki and H-H. Yang, “A new learning algorithm for blind signal separation”, *Adv. Neural Information Processing System*, 8 (1996), 757-763.
- [2] S. Araki, H. Sawada and S. Makino, “K-means Based Underdetermined Blind Speech Separation”, Chapter 9 in “Blind Speech Separation”, S. Makino, T.W. Lee, H. Sawada, eds, Springer 2007.
- [3] A. Bell and T. Sejnowski, “An Information-Maximization Approach to Blind Separation and Blind Deconvolution”, *Neural Computation*, 7(1995), pp 1129-1159.
- [4] J. W. Brewer, “A Note on Kronecker Matrix Products and Matrix Equation Systems”, *SIAM Journal on Applied Mathematics*, 17(3) (1969), pp. 603-606.
- [5] J-F. Cardoso and A. Souloumiac, “Blind beamforming for non-Gaussian signals”, *IEEE-Proceedings-F (Radar and Signal Processing)*, 140(6) (1993), pp. 362-370.
- [6] J-F. Cardoso and A. Souloumiac, “Jacobi angles for simultaneous diagonalization”, *SIAM Journal on Matrix Analysis and Applications*, 17 (1996), pp. 161-164.
- [7] M. Do Carmo, “Differential Geometry of Curves and Surfaces”, 1976, Weierstrasss Publication.
- [8] S.C. Douglas, M. Gupta, “Scaled Natural Gradient Algorithms for Instantaneous and Convulsive Blind Source Separation”, *Proceedings of ICASSP*, II637-II640, 2007.
- [9] S. Haykin, “Adaptive Filter Theory”. Prentice Hall, 2002. ISBN 0-13-048434-2.
- [10] Q. He, J. Xin, “A Randomly Perturbed InfoMax Algorithm for Blind Source Separation”, *Proceedings of the 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, pp. 3218-3222, 2013.

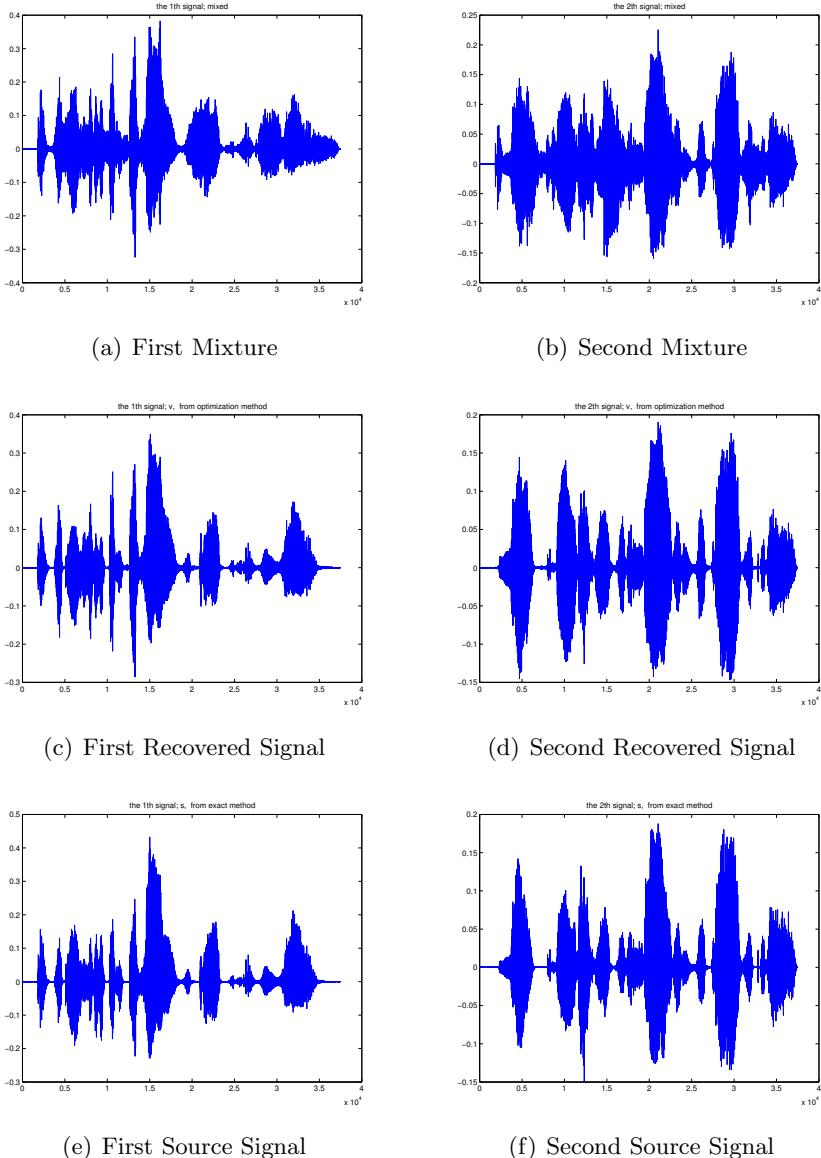


Figure 4: This is the convulsive case using our global optimization method. Our global optimization method gives  $\sigma_P = 24.7826$ .

- [11] J. Liu, J. Xin, Y. Y Qi, “A Dynamic Algorithm for Blind Separation of Convulsive Sound Mixtures”, Neurocomputing, 72(2008), pp 521-532.
- [12] J. Liu, J. Xin, Y-Y. Qi, “A Soft-Constrained Dynamic Iterative Method of Blind Source Separation”, SIAM J. Multiscale Modeling Simulations, 7(4), 2009, pp 1795-1810.

- 
- [13] J. Liu, J. Xin, Y. Qi, F-G. Zeng, “A Time Domain Algorithm for Blind Separation of Convulsive Sound Mixtures and L-1 Constrained Minimization of Cross Correlations”, Comm. Math Sciences, 7(11), 2009, pp. 109-128.
  - [14] N. Murata, S. Ikeda, and A. Ziehe, “An approach to blind source separation based on temporal structure of speech signals”, Neurocomputing, 41 (2001), pp. 1-24.
  - [15] L. Parra and C. Spence, “Convulsive blind separation of non-stationary sources”, IEEE Trans. Speech Audio Processing, 8(3) (2000), pp. 320-327.
  - [16] H. Sawada, S. Araki, S. Makino, “Frequency-domain Blind Source Separation”, chapter 2 in “Blind Speech Separation”, S. Makino, T.W. Lee, H. Sawada, eds, Springer 2007.
  - [17] P. Smaragdis, “Blind separation of convolved mixtures in the frequency domain”, Neurocomputing, 22(1-3) (1998), pp. 21-34.
  - [18] I. Tashev, “Sound Capture and Processing: Practical Approaches”, John Wiley and Sons, 2009.
  - [19] D. Wang, G.J. Brown, “Computational Auditory Scene Analysis: Principles, Algorithms, and Applications”, IEEE Press, Wiley, 2006.
  - [20] Y. Wang, O. Yilmaz and Z. Zhou, “Phase Aliasing Correction for Robust Blind Source Separation Using DUET”, Applied and Computational Harmonic Analysis, 35(2), pp. 341-349, 2013.
  - [21] E. Weinstein, M. Feder and A.V. Oppenheim, “Multi-channel signal separation by decorrelation”, IEEE Trans. on Speech and Audio Processing, 1(4), 405-413, 1993.
  - [22] J. Xin, M. Yu, Y-Y Qi, H-I Yang, F-G Zeng, “Nonlocally Weighted Soft-Constrained Natural Gradient Algorithm for Blind Source Separation of Reverberant Speech”, Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), pp 81-84, 2009, New Paltz, NY, eds. Jacob Benesty and Tomas Gaensler.
  - [23] J. Xin and Y. Qi, “Mathematical Modeling and Signal Processing in Speech and Hearing Sciences”, Modeling, Simulation & Applications, Vol. 10, Springer publication, 2014.
  - [24] O. Yilmaz and S.Rickard, “Blind separation of speech mixtures via time-frequency masking”, IEEE Trans. Signal Processing, 52(7) (2004), 1830-1847.