# Reprint of paper published in "Signal Processing"

Paper reference:
N. Charkani, Y. Deville, "Self-adaptive separation of convolutively mixed signals with a recursive structure. Part II: Theoretical extensions and application to synthetic and real signals", Signal Processing, vol. 75, no. 2, pp. 117-140, 1999.

# Self-Adaptive Separation of Convolutively Mixed Signals with a Recursive Structure - Part II: Theoretical Extensions and Application to Synthetic and Real Signals

Nabil CHARKANI (1)(2) Yannick DEVILLE (1)†

(1) Laboratoires d'Electronique Philips S.A.S. (LEP)
22, Avenue Descartes, BP 15, 94453 Limeil-Brévannes Cedex, FRANCE
(2) Laboratoire de Traitement d'Images et Reconnaissance de Formes (LTIRF - INPG)
46, Avenue Félix Viallet, 38031 Grenoble Cedex, FRANCE

e-mail: charkani@lep.research.philips.com
e-mail: ydeville@lep.research.philips.com

† Author to whom correspondence about this paper should be sent.

Short title: Self-adaptive separation of convolutively mixed signals - Part II

**Keywords:** asymptotic behaviour optimization, blind source separation, coloured signals, convolutive mixtures, non-linear algorithms, recurrent network, self-adaptive algorithms, speech processing, stability analysis, sub-optimum non-linear functions.

Number of typewritten pages of the paper: 41

Number of tables of the paper: 4

Number of figures of the paper: 18

**Figure captions:**

Fig. 1: Basic mixture model for source separation.
Fig. 2: Recurrent structure for the separating system.
Fig. 3: Recurrent separating system for coloured signals.
Fig. 4: Generalized Gaussian family p.d.f for several values of $\beta$.
Fig. 5: Efficiency vs parameter $k$ of the separating function $f_i$. Each plot corresponds to a specific value of the parameter $\beta$ of the sources: —: $\beta = 1$, ....: $\beta = 2$, -.-.: $\beta = 7$.
Fig. 6: $SNRI/SNRI_{max}$ vs parameter $k$ of the separating function $f_i$. Each plot corresponds to a specific value of the parameter $\beta$ of the sources: —: $\beta = 1$, ....: $\beta = 2$, -.-.: $\beta = 7$.
Fig. 7: Efficiency vs parameter $k$ of the separating function $g_i$. Each plot corresponds to a specific value of the parameter $\beta$ of the sources: —: $\beta = 1$, ....: $\beta = 2$, -.-.: $\beta = 7$.
Fig. 8: $SNRI/SNRI_{max}$ vs parameter $k$ of the separating function $g_i$. Each plot corresponds to a specific value of the parameter $\beta$ of the sources: —: $\beta = 1$, ....: $\beta = 2$, -.-.: $\beta = 7$.
Fig. 9: Error variance vs adaptation gain $\mu$. Each plot corresponds to a specific separating function $f_i(x)$ (while $g_i(x) = x$): —: $f_i(x) = sign(x)$, -.-.-: $f_i(x) = x$, - - -: $f_i(x) = x^3$.
Fig. 10: SNRI (dB) vs adaptation gain $\mu$. Each plot corresponds to a specific separating function $f_i(x)$ (while $g_i(x) = x$): —: $f_i(x) = sign(x)$, -.-.-: $f_i(x) = x$, - - -: $f_i(x) = x^3$.
Fig. 11: Error variance vs adaptation gain $\mu$. Each plot corresponds to a specific separating function $f_i(x)$ (while $g_i(x) = x$): —: $f_i(x) = sign(x)$, -.-.-: $f_i(x) = h^{(2D)}(x)$, - - -: $f_i(x) = h^{(3D)}(x)$.
Fig. 12: SNRI (dB) vs adaptation gain $\mu$. Each plot corresponds to a specific separating function $f_i(x)$ (while $g_i(x) = x$): —: $f_i(x) = sign(x)$, -.-.-: $f_i(x) = h^{(2D)}(x)$, - - -: $f_i(x) = h^{(3D)}(x)$.
Fig. 13: Error variance vs adaptation gain $\mu$. Each plot corresponds to a specific separating function $f_i(x)$ (while $g_i(x) = x$): —: $f_i(x) = x$, -.-.-: $f_i(x) = h^{(2D)}(x)$,

- - -: $f_i(x) = h^{(3D)}(x)$.

Fig. 14: SNRI (dB) vs adaptation gain $\mu$. Each plot corresponds to a specific separating function $f_i(x)$ (while $g_i(x) = x$): —: $f_i(x) = x$, -.-.-: $f_i(x) = h^{(2D)}(x)$, - - -: $f_i(x) = h^{(3D)}(x)$.

Fig. 15: SNRI vs adaptation gain $\mu$ for the mixing matrix based on $(A^1_{12}(z), A^1_{21}(z))$. Each plot corresponds to (NW1) with a specific separating function $f_i(x)$ (while $g_i(x) = x$): —: $f_i(x) = sign(x)$, -.-.-: $f_i(x) = x$, - - -: $f_i(x) = x^3$.

Fig. 16: SNRI vs adaptation gain $\mu$ for the mixing matrix based on $(A^2_{12}(z), A^2_{21}(z))$. Each plot corresponds to (NW1) with to a specific separating function $f_i(x)$ (while $g_i(x) = x$): —: $f_i(x) = sign(x)$, -.-.-: $f_i(x) = x$, - - -: $f_i(x) = x^3$.

Fig. 17: SNRI vs adaptation gain $\mu$ for the mixing matrix based on $(A^1_{12}(z), A^1_{21}(z))$. Each plot corresponds to (NW1) with a specific separating function $f_i(x)$ (while $g_i(x) = x$): —: $f_i(x) = sign(x)$, -.-.-: $f_i(x) = h^{(2D)}(x)$, - - -: $f_i(x) = h^{(3D)}(x)$.

Fig. 18: SNRI vs adaptation gain $\mu$ for the mixing matrix based on $(A^2_{12}(z), A^2_{21}(z))$. Each plot corresponds to (NW1) with to a specific separating function $f_i(x)$ (while $g_i(x) = x$): —: $f_i(x) = sign(x)$, -.-.-: $f_i(x) = h^{(2D)}(x)$, - - -: $f_i(x) = h^{(3D)}(x)$.

**Abstract:** This paper deals with the separation of two convolutively mixed signals. The proposed approach uses a recurrent structure adapted by generic rules involving arbitrary separating functions. While the basic versions of this approach were defined and analyzed in our companion paper [3], two extensions are considered here. The first one is intended for possibly-coloured signals. In addition, the second one may be used even when the probability density functions of the sources are unknown. We first analyze the convergence properties of these extended approaches at the separating state, i.e. we derive their equilibrium and stability conditions and their asymptotic error variance. We then determine the separating functions which minimize this error variance. We also report experimental results obtained in various conditions, ranging from synthetic data to mixtures of speech signals measured in real situations. These results confirm the validity of the proposed approaches and show that they significantly outperform classical source separation methods in the considered conditions.

**Résumé:** Cet article concerne la séparation de deux signaux mélangés de manière convolutive. L'approche proposée utilise une structure récurrente adaptée par des règles génériques mettant en jeu des fonctions séparatrices arbitraires. Les versions de base de cette approche ayant été définies et analysées dans notre article associé [3], deux extensions sont considérées ici. La première concerne des signaux pouvant être colorés, tandis que la seconde peut être utilisée même si les densités de probabilité des sources sont inconnues. Après avoir introduit les principes de ces approches étendues, nous analysons leurs propriétés de convergence au point de séparation, en déterminant leurs conditions d'équilibre et de stabilité ainsi que leur variance d'erreur asymptotique. Nous en déduisons ensuite les fonctions séparatrices qui minimisent cette variance d'erreur. Des résultats expérimentaux sont présentés. Ils ont été obtenus dans diverses conditions, comprenant aussi bien des données synthétiques que des mélanges de signaux de parole mesurés dans des situations réelles. Ces résultats confirment la validité des approches proposées et montrent qu'elles fournissent des performances nettement meilleures que les méthodes classiques de séparation de sources dans les conditions considérées.

# 1 Introduction

Multichannel blind source separation is a relatively new signal processing technique. It is based on a model inside which statistically independent signals (sources) are linearly mixed through an unknown invertible medium. Blind source separation then consists in recovering the sources using only several observations of the mixtures. Many methods for achieving source separation were proposed in the literature in the instantaneous case, which corresponds to media without memory (see [3] and references therein for more details). On the contrary, only a few algorithms were introduced in the convolutive domain, corresponding to media with memory [3]. Moreover, their behaviour was almost not analyzed in the literature. This topic was addressed in detail for white sources in our companion paper [3]: we presented a stability analysis and an asymptotic behaviour characterization for a large class of adaptive algorithms involving non-linear separating functions. We also derived the optimum choice for the separating functions, i.e the choice that yields the best quadratic matching of the mixing filters.

In this article, we first briefly redefine the source separation system that we proposed in [3] and we summarize the results that we derived about its behaviour (see Section 2). We then present an extension of this system and associated analysis (see Section 3). This extension concerns the case of possibly-coloured sources, and especially AR processes. An analysis of the stability of this system and of its asymptotic properties is provided. Its optimum separating functions are also derived. They are shown to be related to the probability density functions (p.d.f) of the whitened versions of the sources. As these p.d.f are often unknown in practice, we then study in Section 4 a sub-optimum approach, that performs a self-adaptive projection of the optimum separating functions on a predefined set of classical functions. Section 5 presents experimental results obtained with these approaches. The conclusions drawn from this investigation are eventually provided in Section 6.

# 2 Problem statement and available results

In this section, we briefly present some definitions and notations already used in our companion paper [3]. Some important results derived in [3] are also recalled.

## 2.1 Definitions and notations

Let us consider two unknown zero-mean and mutually independent sequences $x_1(n)$ and $x_2(n)$ that are mixed through an unknown bi-dimensional linear system (see Fig. 1). The corresponding observations are respectively denoted $y_1(n)$ and $y_2(n)$. The source-observation relationship of this model can be expressed as follows in the Z-domain:

$$\begin{pmatrix} Y_1(z) \\ Y_2(z) \end{pmatrix} = A(z) \begin{pmatrix} X_1(z) \\ X_2(z) \end{pmatrix} = \begin{pmatrix} 1 & A_{12}(z) \\ A_{21}(z) & 1 \end{pmatrix} \begin{pmatrix} X_1(z) \\ X_2(z) \end{pmatrix} \tag{1}$$

where $A(z)$ denotes *the mixing matrix*.
The aim of the blind source separation technique is to recover the original sources $x_1(n)$ and $x_2(n)$. This can be achieved by using the recurrent structure shown in Figure 2, so as to derive an estimate $G(z)$ of the inverse of the mixing matrix (see [3] for more details). $G(z)$ is called the *separating matrix*. The associated relationship between the observations and the outputs of the separating system reads as follows in the Z-domain:

$$\begin{pmatrix} S_1(z) \\ S_2(z) \end{pmatrix} = G(z) \begin{pmatrix} Y_1(z) \\ Y_2(z) \end{pmatrix} = \frac{1}{1 - C_{12}(z)C_{21}(z)} \begin{pmatrix} 1 & -C_{12}(z) \\ -C_{21}(z) & 1 \end{pmatrix} \begin{pmatrix} Y_1(z) \\ Y_2(z) \end{pmatrix}. \tag{2}$$

Assuming that both the mixing and the separating filters, $A_{ij}(z)$ and $C_{ij}(z)$, are causal and have a finite memory, only the solution

$$C_{ij}(z) = A_{ij}(z) \quad \forall i \neq j \in \{1, 2\} \tag{3}$$

ensures source separation [3]. (3) is also equivalent to

$$s_i(n) = x_i(n) \quad \forall i \in \{1, 2\}. \tag{4}$$

Source separation then becomes equivalent to the identification of the mixing filters. Unfortunately, $A_{ij}(z)$ are unknown so that (3) cannot be used as a practical method for recovering the sources. Therefore, criteria for assigning $C_{ij}(z)$ are needed. A solution consists in adapting their coefficients, i.e. $(c_{ij}(k))_{k\geq 0}$, so that the outputs of the source separation structure become statistically independent. In practice, only partial statistical independence tests are tractable. Several such tests were proposed in the literature. An attractive solution, presented by H.L. Nguyen Thi and C. Jutten[1] [8], is the stochastic algorithm that reads:

$$c_{ij}(n+1, k) = c_{ij}(n, k) + \mu_n f(s_i(n)) g(s_j(n-k)) \quad i \neq j \in \{1, 2\}, k \in [0, M] \tag{5}$$

where

- $c_{ij}(n, k)$ is the $k^{th}$ coefficient of $C_{ij}(z)$ at the $n^{th}$ iteration,
- $\mu_n$ is a "small" positive adaptation gain,
- $f()$ and $g()$ are odd nonlinear functions.

## 2.2 Summary of fundamental results

In our companion paper [3], we studied a generalization of (5), still applied to the recurrent structure of Figure 2. This extended algorithm is denoted (N0) and reads:

$$c_{ij}(n+1, k) = c_{ij}(n, k) + \mu_n f_i(s_i(n)) g_i(s_j(n-k)) \quad i \neq j \in \{1, 2\}, k \in [0, M] \tag{6}$$

where $f_i()$ and $g_i()$ are arbitrary functions that are not necessarily odd and that are called the "separating functions". Furthermore, we investigated normalized versions of (N0):

- normalization scheme (N1):

$$c_{ij}(n+1, k) = c_{ij}(n, k) + \mu_n \frac{f_i(s_i(n))}{\sqrt{E[f_i^2(s_i)]}} \frac{g_i(s_j(n-k))}{\sqrt{E[g_i^2(s_j)]}} \quad i \neq j \in \{1, 2\}, k \in [0, M], \tag{7}$$

[1] The Nguyen-Jutten algorithm is an extension of the Hérault-Jutten rule [6].

- normalization scheme (N2):

$$c_{ij}(n+1,k) = c_{ij}(n,k) + \mu_n \frac{f_i(s_i(n))}{E[f_i'(s_i)]\sqrt{E[s_i^2]}} \frac{g_i(s_j(n-k))}{\sqrt{E[g_i^2(s_j)]}} \qquad i \neq j \in \{1,2\}, k \in [0,M]. \tag{8}$$

The algorithms (N0), (N1) and (N2) were investigated in detail in [3] under the asssumption that both sources are independent identically distributed (i.i.d) sequences. More precisely, we studied:

- **The equilibrium conditions**. They consist of the conditions on the source statistics and on the separating functions for the separating solution (3) to be an equilibrium state of the algorithm. For all the algorithms (N0), (N1) and (N2), the equilibrium conditions are[2]:

$$E[f_i(x_i)]E[g_i(x_j)] = 0, \qquad i \neq j \in \{1,2\}. \tag{9}$$

- **The stability conditions**. Using the so-called Ordinary Differential Equation technique (ODE), we derived the conditions under which the algorithms (N0), (N1) and (N2) are asymptotically stable in the vicinity of the separating solution. The stability conditions are shown to be related to the sources statistics, the separating functions and the mixing filters (see details in [3]).

- **The steady state (or asymptotic) behaviour**. This refers to the algorithm properties once convergence is achieved. This investigation was carried out for strictly causal mixing and separating filters. The convergence accuracy was measured by the variance $\sigma_\infty$ of the asymptotic estimation error defined as:

$$\sigma_\infty = \lim_{n\to\infty} E\left[\sum_{k=1}^{M} (c_{12}(n,k) - a_{12}(k))^2 + (c_{21}(n,k) - a_{21}(k))^2\right]. \tag{10}$$

We investigated the minimization of $\sigma_\infty$ with respect to the separating functions. We thus first showed that (N0) suffers from an ambiguity due to the lack of normalization. We then proved that (N1) and (N2) overcome this problem. The corresponding optimum separating functions (i.e. the functions which minimize $\sigma_\infty$) were then shown to be for both algorithms:

$$f_{iopt}(x) = -\nu_{i1} \frac{p_{x_i}'(x)}{p_{x_i}(x)}, \tag{11}$$

$$g_{iopt}(x) = \nu_{i2}\, x, \tag{12}$$

where $p_{x_i}$ is the p.d.f of the source $x_i$ and $(\nu_{i1}, \nu_{i2})$ is a couple of arbitrary real constants suitably chosen to ensure the algorithm stability.

[2] (9) assumes that the sources are stationary.

# 3 Extension to coloured signals

In the companion paper [3], all the theoretical results were derived using the whiteness of the sources as a key assumption. Most often, this hypothesis is used to ease mathematical calculus more than to take into account physical models, as most of the signals used in real applications are not white. Our goal hereafter is therefore to define a modified version of our approach, which applies to possibly-coloured source signals and which takes advantage of the results already derived in [3]. To this end, we propose a new separating structure based on whitened versions of the separating module outputs. We also present the associated adaptation rules. A detailed analysis of these algorithms (equilibrium states, stability and asymptotic properties) is then provided.

## 3.1 Models and proposed structure

In this section, we still consider the two-dimensional mixing system of Figure 1. All the previous assumptions on the mixing filters and on the sources are still made except that the sources $x_1(n)$ and $x_2(n)$ are no more required to be i.i.d sequences but only possibly-coloured versions of i.i.d sequences (denoted respectively $\tilde{x}_1(n)$ and $\tilde{x}_2(n)$). Furthermore, we assume that the relationship between $x_i(n)$ and $\tilde{x}_i(n)$ for $i \in \{1,2\}$ reads in the Z-domain:

$$X_i(z) = \frac{\tilde{X}_i(z)}{\tilde{B}_i(z)}, \qquad i \in \{1,2\}, \tag{13}$$

where $\tilde{B}_i(z)$ for $i \in \{1,2\}$ represents the Z-transform of a causal and minimum-phased $q_i$th-order filter, i.e. $\tilde{B}_i(z) = \sum_{k=0}^{q_i} \tilde{b}_i(k) z^{-k}$. We also assume that each source $x_i$ has a non-singular correlation matrix $R_{x_i}$. The sources $x_i(n)$ for $i \in \{1,2\}$ are then AR processes. This assumption does not allow to recover all coloured signals but it makes it possible to handle a large class of processes (especially speech signals, as shown in Section 5). It can also be seen as a good approximation of the larger class of ARMA processes. In the following, we assume that $\tilde{b}_i(0) = 1$. This assumption, which is not restrictive, enables to define $\tilde{x}_i$ as the normalized innovation process of $x_i$ [9].
In the context of such coloured sources, one could think of still using the recurrent structure of Fig. 2 and the adaptive rule (6) that we proposed for white sources. However, since sources are no more white sequences, the results presented in Subsection 2.2 do not hold anymore here and the analysis of the properties of the rule (6) becomes tedious.
In order to overcome this problem and to take advantage of the previous results, we propose to use a new separating structure (Fig. 3) which consists of two stages:

- A *separating module* that implements an inverse matrix, and that derives the outputs $s_i(n)$ from the observations $y_i(n)$ and separating filters $C_{ij}(z)$. This stage is the same as the one previously used for white signals.

- A *whitening module* that drives the separating filter adaptation, according to a rule detailed in the next subsection. Filters are thus adaptively estimated using whitened versions of the separating module outputs.

### 3.2 Formulation of the separation algorithm

Three adaptation rules for the above structure are presented hereafter. They are denoted (NW0), (NW1) and (NW2) as respective extensions of (N0), (N1) and (N2) defined in [3] in the case of i.i.d sources:

- normalization scheme (NW0):

$$c_{ij}(n+1,k) = c_{ij}(n,k) + \mu_n f_i(v_i(n)) g_i(v_j(n-k)) \qquad i \neq j \in \{1,2\}, k \in [0,M], \quad (14)$$

- normalization scheme (NW1):

$$c_{ij}(n+1,k) = c_{ij}(n,k) + \mu_n \frac{f_i(v_i(n))}{\sqrt{E[f_i^2(v_i)]}} \frac{g_i(v_j(n-k))}{\sqrt{E[g_i^2(v_j)]}} \qquad i \neq j \in \{1,2\}, k \in [0,M], \quad (15)$$

- normalization scheme (NW2):

$$c_{ij}(n+1,k) = c_{ij}(n,k) + \mu_n \frac{f_i(v_i(n))}{E[f_i'(v_i)]\sqrt{E[v_i^2]}} \frac{g_i(v_j(n-k))}{\sqrt{E[g_i^2(v_j)]}} \qquad i \neq j \in \{1,2\}, k \in [0,M], \quad (16)$$

where $v_1(n)$ and $v_2(n)$ are adaptively whitened versions of the outputs $s_1(n)$ and $s_2(n)$ of the separating module (i.e. $v_i(n)$ is the estimated innovation process of $s_i(n)$). The order of each whitening filter $B_i(z)$ is assumed to be equal to the order of the corresponding colouring filter $\tilde{B}_i(z)$, which is not a restrictive assumption[1]. Furthermore, we assume that the zero-lag coefficient of each filter $B_i(z)$ (i.e. $b_i(0)$) is equal to[3] 1. The whitening algorithm then becomes:

$$b_i(n+1,k) = b_i(n,k) - \gamma_n v_i(n) s_i(n-k) \qquad i \in \{1,2\},\ k \in [1,q_i] \quad (17)$$

where $s_i$ and $v_i$ are given by:

$$\begin{cases} s_i(n) &= \dfrac{1}{1 - c_{12}(n,0)c_{21}(n,0)} \left(y_i(n) - \displaystyle\sum_{k=1}^{M} c_{ij}(n,k) s_j(n-k)\right) \qquad i \in \{1,2\} \\ v_i(n) &= \displaystyle\sum_{k=0}^{q_i} b_i(n,k) s_i(n-k) \qquad i \in \{1,2\} \end{cases} \quad (18)$$

and $\gamma_n$ is a "small" positive adaptation step. For clarity, we assume in the next sections that $\gamma_n = \mu_n \ \forall n$.

[1] More generally speaking, it is only required that $Order(B_i) \geq Order(\tilde{B}_i)$ for $i \in \{1,2\}$. Equality is assumed here only to ease the presentation of the results.

[3] $b_i(0)$ can be fixed to any non null value. The value of $b_i(0)$ does not influence the quality of the whitening algorithm since it is only related to signal amplitude.

## 3.3 Analysis of the separation algorithm

### 3.3.1 Principles

The normalization schemes (NW0), (NW1) and (NW2) have to be analyzed in association with the whitening rule (17), as the coefficients $b_i(n,k)$ of the latter rule are also part of the overall set of adaptive parameters of the considered system. Furthermore, (NW1) and (NW2) use normalizing terms, i.e. $E[f_i'(v_i)]$, $\sqrt{E[v_i^2]}$, $\sqrt{E[f_i^2(v_i)]}$ and $\sqrt{E[g_i^2(v_j)]}$, that are not available in practice. These terms have then to be adaptively estimated and therefore the adaptive parameters of their estimation rules also have to be incorporated in the algorithm analysis. The analysis of the resulting overall algorithms is then based on an approach similar to the one already used in [3] to analyze the algorithms (N0), (N1) and (N2). The details of this analysis are therefore skipped in the current paper, and only its major steps are presented (a detailed presentation may be found in [2] however). Some of these steps are only provided for (NW0). The corresponding results for (NW1) and (NW2) may be derived from those related to (NW0), using respectively the following transforms[4]:

- for (NW1)

$$\begin{aligned} f_i(x) \longrightarrow F_i(x) &= \frac{f_i(x)}{\sqrt{E[f_i^2(x)]}}, \\ g_i(x) \longrightarrow G_i(x) &= \frac{g_i(x)}{\sqrt{E[g_i^2(x)]}}, \end{aligned} \tag{19}$$

- for (NW2)

$$\begin{aligned} f_i(x) \longrightarrow F_i(x) &= \frac{f_i(x)}{E[f_i'(x)]\sqrt{E[x^2]}}, \\ g_i(x) \longrightarrow G_i(x) &= \frac{g_i(x)}{\sqrt{E[g_i^2(x)]}}. \end{aligned} \tag{20}$$

### 3.3.2 Equilibrium states

To analyze (NW0), we consider hereafter a global reformulation of (14) and (17) as:

$$\Theta_{n+1} = \Theta_n + \mu_n H(\Theta_n, \xi_{n+1}), \tag{21}$$

where $\Theta_n$, $\xi_{n+1}$ and $H(\Theta_n, \xi_{n+1})$ are the following column vectors:

$$\Theta_n = [\theta_{1n}, \theta_{2n}]^T, \tag{22}$$

$$H(\Theta_n, \xi_{n+1}) = [H_1, H_2]^T, \tag{23}$$

$$\xi_{n+1} = [\xi_{1,n+1}, \xi_{2,n+1}]^T, \tag{24}$$

where

$$\theta_{1n} = [c_{12}(n,0), \ldots, c_{12}(n,M), c_{21}(n,0), \ldots, c_{21}(n,M)], \tag{25}$$

$$\theta_{2n} = [b_1(n,1), \ldots, b_1(n,q_1), b_2(n,1), \ldots, b_2(n,q_2)], \tag{26}$$

$$\begin{aligned} H_1 = [&f_1(v_1(n))g_1(v_2(n)), \ldots, f_1(v_1(n))g_1(v_2(n-M)), \\ &f_2(v_2(n))g_2(v_1(n)), \ldots, f_2(v_2(n))g_2(v_1(n-M))], \end{aligned} \tag{27}$$

$$H_2 = -[v_1(n)s_1(n-1), \ldots, v_1(n)s_1(n-q_1), v_2(n)s_2(n-1), \ldots, v_2(n)s_2(n-q_2)], \tag{28}$$

[4] The validity of these transforms is proved in [2].

$$\begin{aligned}\xi_{1,n+1} &= [y_1(n), y_2(n), s_1(n-1), \ldots, s_1(n-max(M,q_1)), \\ &\quad s_2(n-1), \ldots, s_2(n-max(M,q_2))], \quad (29)\\ \xi_{2,n+1} &= [v_1(n-1), \ldots, v_1(n-M), v_2(n-1), \ldots, v_2(n-M)]. \quad (30)\end{aligned}$$

The equilibrium states of (21) are then the states $\Theta^*$ that meet:

$$E_{\Theta^*}[H(\Theta^*, \xi_{n+1})] = 0, \quad (31)$$

where $E_{\Theta^*}[]$ denotes the mathematical expectation associated to the asymptotic probability law of the vector $\xi_{n+1}$ for a given vector $\Theta^*$. In the next sections, we are interested in the properties of the separating state, i.e. the vector $\Theta^s$ that corresponds to the identification of the mixing matrix filters and of the colouring filters (see motivations in [3]). $\Theta^s$ is defined by:

$$\Theta^s = [\theta_1^s, \theta_2^s]^T, \quad (32)$$

with:

$$\begin{aligned}\theta_1^s &= [a_{12}(0), \ldots, a_{12}(M), a_{21}(0), \ldots, a_{21}(M)], \quad (33)\\ \theta_2^s &= [\tilde{b}_1(1), \ldots, \tilde{b}_1(q_1), \tilde{b}_2(1), \ldots, \tilde{b}_2(q_2)]. \quad (34)\end{aligned}$$

The condition for the separating state $\Theta^s$ to be an equilibrium state of the algorithm (21) is then:

$$E[f_i(\tilde{x}_i(n))g_i(\tilde{x}_j(n-k))] = 0 \quad k \in [0, M],\ i \neq j \in \{1, 2\}, \quad (35)$$

or equivalently,

$$E[f_i(\tilde{x}_i)]E[g_i(\tilde{x}_j)] = 0 \quad i \neq j \in \{1, 2\} \quad (36)$$

since the sources are assumed to be independent and stationary. (36) sets restrictions on the separating functions $f_i$ and $g_i$. Many choices can then be made depending on the available information on the statistical properties of the sources. In order not to constrain both $f_i$ and $g_i$, we assume that:

$$E[g_i(\tilde{x}_j)] = 0. \quad (37)$$

### 3.3.3 Stability conditions

The Jacobian matrix $J(\Theta)$ of (NW0) at the state $\Theta$ is composed of the elements:

$$J_{ij}(\Theta) = \frac{\partial(E_\Theta[H(\Theta, \xi_{n+1})]^{(i)})}{\partial\Theta^{(j)}}, \quad (38)$$

where $E_\Theta[H(\Theta, \xi_{n+1})]^{(i)}$ is the $i^{th}$ component of $E_\Theta[H(\Theta, \xi_{n+1})]$ and $\Theta^{(j)}$ is the $j^{th}$ component of vector $\Theta$. The condition for a given equilibrium state $\Theta^*$ to be locally stable is then: all the eigenvalues of $J(\Theta^*)$ have negative real parts [3]. Using (22) and (23), this Jacobian matrix can be written as a 2x2 block matrix that reads:

$$J(\Theta) = \begin{pmatrix} \frac{\partial(E(H_1))}{\partial\theta_1} & \frac{\partial(E(H_1))}{\partial\theta_2} \\ \frac{\partial(E(H_2))}{\partial\theta_1} & \frac{\partial(E(H_2))}{\partial\theta_2} \end{pmatrix} \quad (39)$$

where $\frac{\partial(E(H_i))}{\partial\theta_j}$ for $i,j \in \{1,2\}$ are partial derivative matrices. In the following, we focus on the properties of $J$ at the separating state $\Theta^s$ considered as an equilibrium state. It can then be shown that:

$$\frac{\partial(E(H_1))}{\partial\theta_2}|_{\Theta=\Theta^s} = 0_{q_1+q_2}, \tag{40}$$

$$\frac{\partial(E(H_2))}{\partial\theta_2}|_{\Theta=\Theta^s} = -\begin{pmatrix} R_{x_1} & 0 \\ 0 & R_{x_2} \end{pmatrix}, \tag{41}$$

where $0_{q_1+q_2}$ and $R_{x_i}$ for $i \in \{1,2\}$ denote respectively the $(q_1+q_2)\times(q_1+q_2)$ null matrix and the correlation matrix of the source $x_i$. $J(\Theta^s)$ then becomes lower block-diagonal and its eigenvalues are equal to those of $\frac{\partial(E(H_1))}{\partial\theta_1}$, $-R_{x_1}$ and $-R_{x_2}$. Besides, the matrices $(R_{x_i})_{i\in\{1,2\}}$ are assumed to be invertible. Their eigenvalues are then strictly positive[1] and so those of $(-R_{x_i})_{i\in\{1,2\}}$ are strictly negative. The above-defined stability condition is then only related to the eigenvalues of $\frac{\partial(E(H_1))}{\partial\theta_1}$. Using the same type of approach as in [3] (especially the calculus presented in its Appendix A), the stability condition can be formulated as:

- if $\Delta > 0$

$$\begin{cases} \tilde{\alpha}_1\tilde{w}(0) > 0 \\ \tilde{\alpha}_2\tilde{w}(0) > 0 \\ (\tilde{\alpha}_1 - \tilde{\varphi}_1)(\tilde{\alpha}_2 - \tilde{\varphi}_2) > \tilde{\beta}_1\tilde{\beta}_2 \\ (\tilde{\alpha}_1 + \tilde{\alpha}_2)\tilde{w}(0) > (\tilde{\varphi}_1 + \tilde{\varphi}_2)\tilde{w}(0) \end{cases} \tag{42}$$

- if $\Delta \leq 0$

$$\begin{cases} \tilde{\alpha}_1\tilde{w}(0) > 0 \\ \tilde{\alpha}_2\tilde{w}(0) > 0 \\ (\tilde{\alpha}_1 + \tilde{\alpha}_2)\tilde{w}(0) > (\tilde{\varphi}_1 + \tilde{\varphi}_2)\tilde{w}(0) \end{cases} \tag{43}$$

where

$$\Delta = (\tilde{\alpha}_1 - \tilde{\varphi}_1 - \tilde{\alpha}_2 + \tilde{\varphi}_2)^2 + 4\tilde{\beta}_1\tilde{\beta}_2, \tag{44}$$

and

$$\begin{cases} \tilde{\alpha}_i = E[f_i'(\tilde{x}_i)]E[\tilde{x}_j g_i(\tilde{x}_j))] \\ \tilde{\beta}_i = E[\tilde{x}_i f_i(\tilde{x}_i)]E[g_i'(\tilde{x}_j))] \\ \tilde{\varphi}_i = a_{ji}(0)E[f_i(\tilde{x}_i)]E[\tilde{x}_j g_i'(\tilde{x}_j))] \\ \tilde{w}(0) = \frac{1}{1-a_{12}(0)a_{21}(0)} \end{cases} \tag{45}$$

We consider hereafter two particular cases:

1. **Strictly causal filters**[1]: the Jacobian matrix $\frac{\partial(E(H_1))}{\partial\theta_1}$ then becomes block-diagonal with lower triangular diagonal blocks. We also have $\varphi_i = 0$ for $i \in \{1,2\}$. The stability condition then reads:

[1] $R_{x_i}$ is a priori a semi-definite matrix. Its eigenvalues are then positive or null. The invertibility assumption implies that no eigenvalues are null.

[1] $a_{12}(0) = a_{21}(0) = c_{12}(0) = c_{21}(0) = 0$.

$$\begin{cases} \tilde{\alpha}_1 > 0 \\ \tilde{\alpha}_2 > 0 \end{cases} \tag{46}$$

2. **Instantaneous mixtures**[2]: the stability condition then reads:

   - if $\Delta > 0$

$$\begin{cases} (\tilde{\alpha}_1 - \tilde{\varphi}_1)(\tilde{\alpha}_2 - \tilde{\varphi}_2) > \tilde{\beta}_1\tilde{\beta}_2 \\ (\tilde{\alpha}_1 + \tilde{\alpha}_2)\tilde{w}(0) > (\tilde{\varphi}_1 + \tilde{\varphi}_2)\tilde{w}(0) \end{cases} \tag{47}$$

   - if $\Delta \leq 0$

$$(\tilde{\alpha}_1 + \tilde{\alpha}_2)\tilde{w}(0) > (\tilde{\varphi}_1 + \tilde{\varphi}_2)\tilde{w}(0) \tag{48}$$

### 3.3.4 Asymptotic behaviour

The asymptotic behaviour of the algorithms (NW0), (NW1) and (NW2) is investigated hereafter. Here again, we assume that:

- The mixing and the separating filters are strictly causal.
- The adaptation gain $\mu_n$ is such as $\mu_n = \mu > 0$, $\forall n \geq 0$.

Under these assumptions, the stability condition presented in (46) is still valid. The asymptotic error variances associated to the considered algorithms become [2]:

- for (NW0):

$$\sigma_\infty = \mu \sum_{i \neq j=1}^{2} \frac{E[f_i^2(\tilde{x}_i)]}{E[f_i'(\tilde{x}_i)]} \frac{E[g_i^2(\tilde{x}_j)]}{E[\tilde{x}_j g_i(\tilde{x}_j)]} (q_{i2} + q_{i1}a_i), \tag{49}$$

- for (NW1):

$$\sigma_\infty = \mu \sum_{i \neq j=1}^{2} \frac{\sqrt{E[f_i^2(\tilde{x}_i)]}}{E[f_i'(\tilde{x}_i)]} \frac{\sqrt{E[g_i^2(\tilde{x}_j)]}}{E[\tilde{x}_j g_i(\tilde{x}_j)]} (q_{i2} + q_{i1}a_i), \tag{50}$$

- for (NW2):

$$\sigma_\infty = \mu \sum_{i \neq j=1}^{2} \frac{E[f_i^2(\tilde{x}_i)]}{E^2[f_i'(\tilde{x}_i)]} \frac{\sqrt{E[g_i^2(\tilde{x}_j)]}}{E[\tilde{x}_j g_i(x_j)]} \frac{1}{\sqrt{E[\tilde{x}_i^2]}} (q_{i2} + q_{i1}a_i), \tag{51}$$

where $a_i = \dfrac{E^2[f_i(\tilde{x}_i)]}{E[f_i^2(\tilde{x}_i)]}$ and $q_{ij}$ for $i, j \in \{1, 2\}$ are real values that depend only on the mixing matrix.

Following the same approach as in [3], the optimality for (NW1) and (NW2) can be shown to be reached for[5]:

[2] $a_{ij}(k) = c_{ij}(k) = 0$ for $i \neq j \in \{1, 2\}$ and $k \geq 1$.

[5] As for the algorithm (N0) studied in [3], the optimum choice for the separating functions cannot be defined for (NW0) because the adaptation gain and the separating functions are coupled terms of the algorithm.

$$f_{iopt}(x) = -\nu_{i1}\frac{p'_{\tilde{x}_i}(x)}{p_{\tilde{x}_i}(x)} \quad (52)$$

$$g_{iopt}(x) = \nu_{i2}x \quad (53)$$

where $p_{\tilde{x}_i}$ is the p.d.f of the innovation process $\tilde{x_i}$ of $x_i$ and $(\nu_{i1}, \nu_{i2})$ is a couple of real constants that must be chosen in order to meet the stability condition (46). Using (19)-(20), the optimum normalized separating functions are then:

- for (NW1):

$$F_{iopt}(x) = -\frac{\frac{p'_{\tilde{x}_i}(x)}{p_{\tilde{x}_i}(x)}}{\sqrt{E[(\frac{p'_{\tilde{x}_i}(x)}{p_{\tilde{x}_i}(x)})^2]}}, \quad (54)$$

$$G_{iopt}(x) = \frac{x}{\sqrt{E[x^2]}}, \quad (55)$$

- for (NW2):

$$F_{iopt}(x) = \frac{\frac{p'_{\tilde{x}_i}(x)}{p_{\tilde{x}_i}(x)}}{E[(\frac{p'_{\tilde{x}_i}(x)}{p_{\tilde{x}_i}(x)})']\sqrt{E[x^2]}}, \quad (56)$$

$$G_{iopt}(x) = \frac{x}{\sqrt{E[x^2]}}. \quad (57)$$

# 4 Sub-optimum separating functions

The implementation of the optimum separating functions $f_{iopt}$, or equivalently $F_{iopt}$, requires the p.d.f $p_{\tilde{x}_i}$ of the whitened versions of the sources to be known. Unfortunately those functions are often not available. A natural solution to this problem is the estimation of these p.d.f. However, this approach is computationally expensive, hard to use in real-time applications and difficult to extend to the case of non-stationary sources. An alternative solution consists in determining a sub-optimum estimate of these optimum separating functions (see also [4]). Especially, this can be achieved by deriving the projection of the optimum functions on a set of classical functions (polynomial functions for example). Hence, each separating function $f_i$ or $F_i$ has the form:

$$h(x) = \sum_{k=1}^{L} \omega_k \psi_k(x), \quad (58)$$

where:

- $(\psi_k(x))_{k\in[1,L]}$ is a set of continuously derivable functions that span the projection space,
- $(\omega_k)_{k\in[1,L]}$ is a set of scalar coefficients associated to $(\psi_k(x))_{k\in[1,L]}$.

The expansion coefficients $(\omega_k)_{k\in[1,L]}$ depend on the optimum function that is projected, on the set of projection functions and also on the constraints put on $h(x)$. Many such constraints can be used, leading to more or less complicated solutions that may depend on the mixture characteristics. One way to achieve an easy presentation of the results is to constrain $h(x)$ to be zero mean. This assumption makes sense since the optimum function is zero mean[6], but this does not ensure that the sub-optimum function thus obtained is very close to the optimum. This approach is therefore a trade-off between complexity and optimality. It is illustrated hereafter for the two normalization schemes introduced above and for strictly causal mixtures.
It should be noted that a sub-optimum approach is only required for the functions $F_i$. On the contrary, as for the functions $G_i$, the optimum is proportional to the identity function, which is easily implemented and therefore used hereafter.
All the results presented below are also valid in the case when the sources are white signals. One then just has to replace $\tilde{x}_i$ by $x_i$ and to set the whitening filters to 1 in the final results.

### 4.1 Sub-optimum solution for normalization scheme (NW1)

When setting $G_i$ to its optimum value (55) and $\mu_n = \mu$, the normalization scheme (NW1) becomes:

$$c_{ij}(n+1,k) = c_{ij}(n,k) + \mu F_i(v_i(n)) \frac{v_j(n-k)}{\sqrt{E[v_j^2]}} \qquad i \neq j \in \{1,2\}, k \in [1,M]. \tag{59}$$

Using the sub-optimum approach in (59), $F_i(x)$ is made equal to $h_i(x)$ defined according to (58), i.e:

$$h_i(x) = \sum_{k=1}^{L} \omega_{ik}\psi_{ik}(x) = \frac{f_i(x)}{\sqrt{E[f_i^2(x)]}}. \tag{60}$$

At the separating equilibrium state, the stability condition derived from (46) according to (19) and from (55) reads:

$$E[h_i^{'}(\tilde{x}_i)] > 0. \tag{61}$$

Combining (50), (53) and (60), the asymptotic error variance becomes:

$$\sigma_\infty = \mu \sum_{i=1,i\neq j}^{i,j=2} \frac{(q_{i2}E[h_i^2(\tilde{x}_i)] + q_{i1}E^2[h_i(\tilde{x}_i)])}{\sqrt{E[h_i^2(\tilde{x}_i)]}E[h_i^{'}(\tilde{x}_i)]} \frac{1}{\sqrt{E[\tilde{x}_j^2]}}. \tag{62}$$

Furthermore, $h_i(\tilde{x}_i)$ has to meet two constraints: $E[h_i^2(\tilde{x}_i)] = 1$ and $E[h_i(\tilde{x}_i)] = 0$ corresponding respectively to (60) and to the zero-mean assumption made on $h_i(x)$. Under those constraints, the error variance becomes:

$$\sigma_\infty = \mu \sum_{i,j=1,i\neq j}^{i,j=2} \frac{q_{i2}}{E[h_i^{'}(\tilde{x}_i)]\sqrt{E[\tilde{x}_j^2]}}. \tag{63}$$

[6] (54) or (56), combined with the assumption $\lim_{|x|\to+\infty} p_{\tilde{x}_i}(x) = 0$ (made in [3]), imply $E[F_{iopt}(x)] = 0$.

The minimization of $\sigma_\infty$ as defined in (63) then corresponds to the constrained maximization of $E[h_i'(\tilde{x}_i)]$ which leads to the following set of coefficients $(\omega_{ik})_{k\geq 1}$ (see [2]):

$$\left\{ \begin{array}{rcl} \omega_{i1} & = & \dfrac{sign(\sum_{k=1}^{L} \frac{d_{ik}}{d_{i1}} E[\psi_{ik}'(\tilde{x}_i)])}{\sqrt{\sum_{k,l=1}^{L} \frac{d_{ik}d_{il}}{d_{i1}^2} E[\psi_{ik}(\tilde{x}_i)\psi_{il}(\tilde{x}_i)]}} \\ \omega_{ik} & = & \omega_{i1}\dfrac{d_{ik}}{d_{i1}} \quad k \in [2, L] \end{array} \right. \tag{64}$$

where $d_{ik}, k \in [1, L]$, are the entries of the vector $D_i$ defined by:

$$\begin{array}{rcll} D_i & = & \Psi_i^{-1}\Gamma_i, & if\ V_i = 0, \\ D_i & = & \Psi_i^{-1}\Gamma_i - (\frac{V_i^T\Psi_i^{-1}\Gamma_i}{V_i^T\Psi_i^{-1}V_i})\Psi_i^{-1}V_i, & if\ V_i \neq 0, \end{array} \tag{65}$$

where

$$\Psi_i = \begin{pmatrix} E[\psi_{i1}(\tilde{x}_i)\psi_{i1}(\tilde{x}_i)] & \dots & E[\psi_{i1}(\tilde{x}_i)\psi_{iL}(\tilde{x}_i)] \\ \vdots & \vdots & \vdots \\ \vdots & E[\psi_{ik}(\tilde{x}_i)\psi_{il}(\tilde{x}_i)] & \vdots \\ \vdots & \vdots & \vdots \\ E[\psi_{iL}(\tilde{x}_i)\psi_{i1}(\tilde{x}_i)] & \dots & E[\psi_{iL}(\tilde{x}_i)\psi_{iL}(\tilde{x}_i)] \end{pmatrix}, \tag{66}$$

$$\Gamma_i^T = \left[E[\psi_{i1}'(\tilde{x}_i)], \dots, E[\psi_{iL}'(\tilde{x}_i)]\right]^T, \tag{67}$$

and

$$V_i^T = [E[\psi_{i1}(\tilde{x}_i)], \dots, E[\psi_{iL}(\tilde{x}_i)]]^T. \tag{68}$$

Note that $d_{i1}$ was assumed to be non null above. If this assumption does not hold, a permutation of indices in the vector $D$ allows to fulfill this requirement.

### 4.2 Sub-optimum solution for normalization scheme (NW2)

Here again, the function $G_i$ is set to its optimum value (57), which corresponds to the new form of (NW2):

$$c_{ij}(n+1,k) = c_{ij}(n,k) + \mu F_i(v_i(n))\ \frac{v_j(n-k)}{\sqrt{E[v_j^2]}} \quad i \neq j \in \{1,2\}, k \in [1, M]. \tag{69}$$

Using the sub-optimum approach in (69), $F_i(x)$ is made equal to $h_i(x)$, so that:

$$h_i(x) = \sum_{k=1}^{L} \omega_{ik}\psi_{ik}(x) = \frac{f_i(x)}{E[f_i'(x)]\sqrt{E[x^2]}}. \tag{70}$$

The stability condition derived from (46) according to (20), (57) and (70) becomes:

$$E[\tilde{x}_i^2] > 0 \quad i \in \{1,2\}, \tag{71}$$

which is always met, i.e. independently from the properties of the separating functions $f_i$. Combining (51), (53) and (70), the asymptotic error variance is then:

$$\sigma_\infty = \mu \sum_{i=1, i\neq j}^{2} \frac{1}{E^2[h_i'(\tilde{x}_i)]}(q_{i2}E[h_i^2(\tilde{x}_i)] + q_{i1}E^2[h_i(\tilde{x}_i)])\frac{1}{\sqrt{E[\tilde{x}_i^2]E[\tilde{x}_j^2]}}. \quad (72)$$

Moreover, $h_i(\tilde{x}_i)$ has to meet two constraints: $E[h_i'(\tilde{x}_i)] = \frac{1}{\sqrt{E[\tilde{x}_i^2]}}$ which results from (70) and the zero-mean constraint $E[h_i(\tilde{x}_i)] = 0$. The asymptotic error variance becomes then:

$$\sigma_\infty = \mu \sum_{i,j=1, i\neq j}^{i,j=2} q_{i2}E[h_i^2(\tilde{x}_i)]\sqrt{\frac{E[\tilde{x}_i^2]}{E[\tilde{x}_j^2]}}. \quad (73)$$

The constrained minimization of $\sigma_\infty$ leads to the set of coefficients $(\omega_{ik})_{k\geq 1}$ defined by (see [2]):

$$\begin{cases} \omega_{i1} &= \dfrac{1}{\sqrt{E[\tilde{x}_i^2]}\displaystyle\sum_{k=1}^{L}\frac{d_{ik}}{d_{i1}}E[\psi_{ik}'(\tilde{x}_i)]} \\ \omega_{ik} &= \omega_{i1}\dfrac{d_{ik}}{d_{i1}} \quad k \in [2, L] \end{cases} \quad (74)$$

where $[d_{i1}, \ldots, d_{iL}]^T$ is the solution of the linear system given in (65).

# 5 Experimental results

This section is devoted to the experimental validation of the analysis performed in the previous sections. Three basic conditions are considered, depending on the nature of the sources and of the mixing matrix: *i)* synthetic sources and synthetic mixing system, *ii)* real sources and synthetic mixing system, *iii)* real sources and real mixing system. The assumptions made in the theoretical analysis can be rigorously checked only in the first case. However, the most interesting case from a practical point of view is the most realistic one, i.e. the third one. The validation of the theoretical results for the adaptation rules (N1), (N2), (NW1) and (NW2) is fully detailed in [2]. In seek for brevity, only the experimental results associated to (NW1) are presented hereafter.

## 5.1 Definitions

**Definition 1** *The efficiency of a separating function $f_i$, denoted $eff(f_i)$, is the ratio of the minimal asymptotic error variance (associated to the optimum function $f_{iopt}$) to the asymptotic error variance achieved by this separating function $f_i$.*

An equivalent definition holds for the separating functions $g_i$.

**Definition 2** *The Signal to Noise Ratio Improvement at output $i$ of the separating system is defined as:*

$$SNRI_i = 10\ log_{10}\left(\frac{E[(y_i(n) - x_i(n))^2]}{E[(s_i(n) - x_i(n))^2]}\right), \quad for\ i \in \{1, 2\}. \quad (75)$$

*$SNRI_i$ measures the logarithmic difference between the signal to noise ratio at output $i$ and the signal to noise ratio at observation $i$ (i.e. at input $i$ of the separating system). In*

*other words, $SNRI_i$ measures the reduction of the crosstalk achieved by the source separation sysem for source $i$. In practice, $SNRI_i$ will be estimated by*[7]*:*

$$SNRI_i = 10\ log_{10}\left(\frac{\sum_{n=L_1}^{L_2}(y_i(n)-x_i(n))^2}{\sum_{n=L_1}^{L_2}(s_i(n)-x_i(n))^2}\right), \quad for\ i \in \{1,2\} \tag{76}$$

*where $L_2$ is the record length and $L_1$ is such that the convergence is achieved for $n \geq L_1$.*

**Definition 3** *$SNRI$ denotes the average Signal to Noise Ratio Improvement defined by:*

$$SNRI = \frac{SNRI_1 + SNRI_2}{2}. \tag{77}$$

**Definition 4** *The* **Generalized Gaussian Family** *(GGF) is the set of p.d.f which depend on a parameter $\beta$ and are expressed as:*

$$p_\beta(x) = K_\beta\ exp(-\frac{|x|^\beta}{\lambda_\beta}) \tag{78}$$

*where*

$$\beta > 0, \tag{79}$$

$$K_\beta = \frac{\beta}{2\sigma}\left(\frac{\Gamma(\frac{3}{\beta})}{\Gamma^3(\frac{1}{\beta})}\right)^{\frac{1}{2}}, \tag{80}$$

$$\lambda_\beta = \left(\frac{\Gamma(\frac{1}{\beta})}{\Gamma(\frac{3}{\beta})}\right)^{\beta/2}\sigma^\beta, \tag{81}$$

$$\sigma = \sqrt{E[x^2]}, \tag{82}$$

*and $\Gamma$ denotes the Gamma function given by $\Gamma(x) = \int_0^{+\infty} e^{-t}t^{x-1}dt$.*

The GGF includes some classical p.d.f such as the Laplace p.d.f ($\beta = 1$), Gaussian p.d.f ($\beta = 2$) and uniform p.d.f ($\beta = +\infty$) (see Fig. 4). It also allows to approximate a large set of unimodal symmetric p.d.f.

## 5.2 Synthetic sources and synthetic mixing matrix

### 5.2.1 Test strategy and conditions

The experimental validation of the algorithm (NW1) requires to vary the separating functions $f_1$, $f_2$, $g_1$ and $g_2$ in order to determine the experimentally optimum functions. Such

[7] $SNRI_i$ can be computed only if the contributions of each source in the observed mixed signals are available. In experimental tests, these contributions may be obtained by emitting a single source signal at a time and recording the corresponding partial observed signals. Each overall observed mixed signal is then obtained by adding the above two contributions successively measured for the two sources on the considered sensor. We used this method in the context of real signals and real mixing systems reported below.

a validation would therefore ideally consist of an infinite number of tests. In this paper however, the complexity of this validation is reduced as explained hereafter. As shown in (53), the optimality of each function $g_i$ is reached when this function is the identity function[8], which is independent from the source statistics and which is the same for $g_1$ and $g_2$. A degree of freedom is therefore removed by restricting the investigation to the case when $g_1 = g_2$. In order to further reduce the complexity of the simulations, the considered sources have the same p.d.f hereafter. According to the theoretical result (52), this assumption leads to the same optimum choice for the separating functions $f_1$ and $f_2$ and therefore allows to obtain relevant results by considering only the case $f_1 = f_2$. With the approach defined at this stage, the tests are performed in a two-dimensional space spanned by $f_1 = f_2 = f$ and $g_1 = g_2 = g$. This space is eventually restricted to one dimension by investigating the optimum choice for $f_i$ while $g_i$ is fixed to its theoretical optimum, and then vice versa. More precisely, $f_i$ or $g_i$ is varied as follows. In each series of tests, the p.d.f of both sources belongs to the above-defined GGF, with a fixed value $\beta$. The optimum separating functions (52)-(53) may then be shown to become:

$$f_{iopt}(x) \propto sign(x)|x|^{\beta-1}, \quad i \in \{1,2\} \tag{83}$$

$$g_{iopt}(x) \propto x, \quad i \in \{1,2\}. \tag{84}$$

where $\propto$ stands for the proportionality symbol. Now consider the family of separating functions $\mathcal{F} = \{f\rangle\}\backslash(\S)|\S|^{\|}, \ \| \geq \prime\}$, where each function corresponds to a specific value of the parameter $k$. The optimum function $f_{iopt}$ defined in (83) belongs to this family and corresponds to $k = \beta - 1$. The optimality of this function can therefore be checked experimentally by successively performing simulations with functions $f_i$ belonging to the family $\mathcal{F}$ and corresponding to various values $k$ (while $g_i(x) = g_{iopt}(x) = x$), and by checking that the minimum experimental error variance is achieved for $k = \beta - 1$. Similarly, the optimality of $g_{iopt}$ defined in (84) is checked by considering functions $g_i$ belonging to the family $\mathcal{F}$ (while $f_i(x) = f_{iopt}(x) = sign(x)|x|^{\beta-1}$), and by checking that the minimum experimental error variance is achieved for $k = 1$ (for which $g_i$ is the identity function). This test procedure is applied hereafter (and additional tests are defined in Subsection 5.2.4). To this end, the following experimental conditions are used:

- *Sources*: coloured versions of i.i.d sequences with GGF p.d.f and with unit power. The tests are successively performed for $\beta = 1, 2$ and 7.
- *Colouring filters* (used to obtain the sources): $b_1 = [1, -0.4, 0.5, -0.7]$ and $b_2 = [1, 0.8, 0.4, -0.2]$.
- *Mixing filters*: we use strictly causal MA(4) filters defined by:

$$\begin{aligned} A_{12}^{(1)} &= [0.0, -0.38077, 0.136010, 0.080955], \\ A_{21}^{(1)} &= [0.0, -0.3268, -0.184256, 0.02734]. \end{aligned} \tag{85}$$

- *Separating filters*: we use strictly causal MA(4) filters.

[8] More precisely, optimality corresponds to $g_i$ proportional to the identity function. However, we only consider the algorithm (NW1) here. This algorithm is such that proportionality coefficients on separating functions have no influence on the separation quality, since they disappear in this algorithm, as can be seen in (15). There is therefore no need considering proportionality factors on $g_i$ here. This also holds for $f_i$ and for the family of functions $\mathcal{F}$ defined below.

### 5.2.2 Optimal function $f_i$

The evolution of the experimental efficiency $eff(f_i)$ with respect to the parameter $k$ of the function $f_i$ varied in the family $\mathcal{F}$ is represented in Fig. 5 for various values of $\beta$. The optimum separating function (i.e. the one that maximizes the efficiency) corresponds to $k = 0$, 1 and 6 respectively for $\beta = 1$, 2 and 7. This is in full agreement with the above-mentioned theoretical result, i.e. $k_{opt} = \beta - 1$. In addition, Fig. 6 represents the normalized SNRI[2] for each considered value of $\beta$. It shows clearly that the SNRI is maximized for $k = 0$, 1 and 6 respectively when $\beta = 1$, 2, and 7. This confirms the optimality of the separating function $f_i(x) = sign(x)|x|^{\beta-1}$ for GGF sources and the validity of the separating approach based on the whitening module presented in Section 3.

### 5.2.3 Optimal function $g_i$

Figure 7 represents the evolution of the experimental efficiency $eff(g_i)$ with respect to the parameter $k$ of the function $g_i$ varied in the family $\mathcal{F}$. Each plot corresponds to a specific value of $\beta$. This figure shows that the experimentally optimum separating function $g_i$ is the identity function (corresponding to $k = 1$) whatever $\beta$, which is the expected result. The measured normalized SNRI (see Fig. 8) leads to the same conclusion.

### 5.2.4 Validation of the sub-optimum approach and of the robustness of the (sub-)optimality with respect to the adaptation gain $\mu$

Two topics are addressed in this subsection: the validation of the sub-optimum approach developed in Section 4 and the test of the robustness of the (sub-)optimality with respect to the adaptation gain parameter $\mu$. We deal with these two items in one global approach by studying the error variance and the SNRI evolutions vs $\mu$ for different separating functions $f_i$ (including the optimum function and its projections). The experimental conditions are the same as the ones we used in the previous subsections except that only the case $\beta = 1$ is considered (i.e. the p.d.f of the innovation processes of the sources are fixed to the Laplace law, as this corresponds to the real signals considered in the subsequent subsections). The separating functions are still all associated to the normalization scheme (NW1), and a comparison is made between:

- The classical separating functions, i.e. the functions which are commonly used in the literature (but which are then applied to classical adaptation rules, as opposed to the rule (NW1) in which they are used here). These functions are: $((f_i(x), g_i(x)) \in \{(x, x), (x^3, x)\})$.
- The original separating functions which result from our approach, i.e: $((f_i(x), g_i(x)) \in \{(sign(x), x), (h^{(2D)}(x), x), (h^{(3D)}(x), x)\})$. Among these functions $f_i(x)$, $sign(x)$ is the optimum function for the considered sources, while $h^{(2D)}(x)$ and $h^{(3D)}(x)$ denote respectively the projection of the optimum function $f_{iopt}$ on the sub-space spanned by $(x, x^3)$ and $(sign(x)\sqrt{|x|}, x, x^3)$.

Figures 9, 11 and 13 represent the evolution of the error variance vs the adaptation gain $\mu$ for the rules corresponding to the considered functions. All separating functions yield the same trend: for small values of $\mu$, the error variance is relatively high because convergence

[2]The normalized SNRI is defined as $SNRI/SNRI_{max}$ where $SNRI_{max}$ is the maximum value of SNRI.

is not reached yet at the end of the considered records. For sufficiently high values of $\mu$, the error variance becomes an increasing function of $\mu$. This phenomenon corresponds to the classical trade-off between convergence speed and accuracy.
The SNRI evolution (see Fig. 10, 12 and 14) yields the same trend (taking into account that when the error variance reduces SNRI grows and vice versa). Fig. 11 and Fig. 12 confirm that the theoretically optimum rule yields in practice a better separating accuracy than 2D and 3D projections. These figures also show that the 3D projection performs better than the 2D projection. Besides, Figures 13 and 14 show that the 2D and the 3D projections outperform the classical separating functions $(f_i(x), g_i(x)) = (x, x)$.
To sum up, also taking into account Fig. 9 and 10 and using the notation $(f_i^{(1)}(x), g_i^{(1)}(x)) > (f_i^{(2)}(x), g_i^{(2)}(x))$ if $(f_i^{(1)}(x), g_i^{(1)}(x))$ yields better separation performance than $(f_i^{(2)}(x), g_i^{(2)}(x))$, the investigation of the evolution of the separating rule performance vs the adaptation gain $\mu$ shows that $(sign(x), x) > (h^{(3D)}(x), x) > (h^{(2D)}, x) > (x, x) > (x^3, x)$ in the considered conditions. This yields two fundamental results:

- The validity of the optimum and sub-optimum approaches, i.e. the fact that the associated rules outperform the rules corresponding to classical functions.
- The robustness of the optimum and sub-optimum approaches with respect to the adaptation gain.

## 5.3 Real sources and synthetic mixing matrix

### 5.3.1 Laplace modelling of speech signals

Speech signals are non-stationary sequences (they can be considered as stationary only over short time periods, i.e. often less than 30 ms). However, it is often convenient to use a statistical model of speech. We showed in [2] that speech may be approximated by an AR process. In this model, the order of the AR filters typically ranges from 8 to 20 and the innovation processes of the signals have a Laplace p.d.f (which belongs to the GGF and corresponds to $\beta = 1$, as stated above)[9]. Therefore, speech belongs to the class of source signals that can be handled by the source separation approach defined in Sections 3 and 4. Moreover, as the sources are modelled with a Laplace p.d.f here, like in the previous simulations, the corresponding optimum separating functions of our source separation system are again $(f_i(x), g_i(x)) = (sign(x), x)$.

### 5.3.2 Test conditions and results

The sources used here are two speech signals. Two mixing matrices are successively considered: the first one is based on the same filters $A_{12}^{(1)}$ and $A_{21}^{(1)}$ as in Subsection 5.2. The second one is based on 13th-order strictly causal filters $A_{12}^{(2)}$ and $A_{21}^{(2)}$ defined by:

$$\begin{aligned} A_{12}^{(2)} &= [0, -0.05268, -0.104256, 0.32734, 0.5997, 0.242796, -0.051458, \\ &\quad -0.0159489, 0.010203, 0.00012, 0.002, 0.12, 0.019], \\ A_{21}^{(2)} &= [0, -0.018077, 0.03601, 0.280955, 0.46475, 0.320127, -0.077205, \\ &\quad -0.024521, 0.012375, -0.000011, 0.001, -0.32, 0.009]. \end{aligned}$$

[9] This is to be contrasted with classical speech models [10], in which the AR filters are excited by non i.i.d signals, i.e. periodic signals for voiced speech or coloured noise for unvoiced speech.

The separating matrix is based on strictly causal MA filters having the same orders as the corresponding mixing filters. The whitening filters are MA(20) with zero-lag coefficients $b_i(0)$ set to 1. The separating rules that we test are:

- Algorithm (N1) associated to the separating functions $(f(x), g(x)) \in \{(x, x), (x^3, x)\}$. These approaches are considered because they provide a good idea of the performance that can be achieved by classical methods reported in the literature (the only improvement added here to the classical methods is the normalization scheme (N1)).

- The approaches already defined in Subsection 5.2.4, i.e. algorithm (NW1) associated to the separating functions:
  $(f(x), g(x)) \in \{(x, x), (x^3, x), (sign(x), x), (h^{(2D)}(x), x), (h^{(3D)}(x), x)\}$.

The comparison of these approaches is performed for different values of the adaptation gain $\mu$. This again enables us to examine the dependency of their performance vs $\mu$ and to check that the robustness of the (sub-)optimum approach is valid for real signals. Figures 15 to 18 represent the SNRI evolution vs the adaptation gain $\mu$ for different rules and for both mixing matrices. In seek for clarity, only the approaches based on (NW1) are represented at this stage. For small values of $\mu$, the SNRI is relatively low because the convergence is not reached yet at the end of the considered records. For sufficiently high values of $\mu$, the SNRI becomes a decreasing function of $\mu$ (this again corresponds to the trade-off between convergence speed and accuracy mentioned in Subsection 5.2.4). According to these figures, the maximum SNRI is achieved for the separation rule $(f_i(x), g_i(x)) = (sign(x), x)$ for both mixing matrices[10].
In Tables 1 and 2, we provide the simulation results for all the algorithms that we tested. These results correspond to the best performance achieved for each rule[11]. For both mixing matrices, the optimum and the sub-optimum solutions yield a significant improvement both in crosstalk reduction (or equivalently SNRI) and in convergence accuracy. More precisely, the optimum separating functions outperform the classical approaches in terms of SNRI by about 5 dB for the mixing matrix $A^{(1)}(z)$ and 9 dB for the mixing matrix $A^{(2)}(z)$. In addition, Tables 1 and 2 show that the whitening module slightly improves the separation accuracy and that 2D projection performs better than the classical separating functions. However, this 2D projection yields lower performance than the 3D projection and the optimum approach (see also Fig. 17 and 18). Therefore, the 2D projection is not considered in the next section.

## 5.4 Real sources and real mixing matrix

In this part, we consider the realistic situation when both the sources and the mixing filters are real: the sources are two speech signals considered in the telephone band [300 Hz - 3400 Hz]. These signals are emitted by two loudspeakers and picked up by an antenna made up of 8 microphones, with the following configuration:

- the inter-source distance is 40 cm,

[10] As the Laplace model is only an approximation of the actual p.d.f. of the innovation processes of the considered real sources, $f_i(x) = sign(x)$ is only an approximation of the optimum separating function $f_{iopt}$ for these sources. This is the reason why the functions $f_i(x) = h^{(2D)}(x)$ and $f_i(x) = h^{(3D)}(x)$ slightly outperform $f_i(x) = sign(x)$ in a few cases, although they are sub-optimum for the considered source model.

[11] For a given separation rule, the best performance is the best SNRI measured for different values of the adaptation gain $\mu$.

- the source-antenna distance is 40 cm,
- the distance between adjacent microphones is 5 cm.

The underlying motivation for considering different microphone distances $d$ is to investigate the potential degradation of the performance of the source separation algorithms in the case of ill-conditioned mixtures, which corresponds to small inter-microphone distances. The measurements were performed inside two different rooms (Room 1 and Room 2) corresponding to different volumes. This enables to investigate the influence of the acoustic channel complexity on the separation performance. The separating filters are strictly causal MA(30) for Room 1 and MA(100) for Room 2. The whitening filters are again MA(20) with zero-lag coefficients $b_i(0)$ set to 1. Several adaptation rules are considered. They again correspond to (NW1) with $f_i(x) \in \{sign(x), h^{(3D)}(x), x, x^3\}$ and $g_i(x) = x$.

Since the mixing filters are unknown, one cannot measure the estimated error variance defined in the previous sections. Therefore, only the SNRI of the separating system is considered hereafter (as explained in Subsection 5.1). Tables 3 and 4 provide the best performance achieved by each rule. They show clearly that the separating function $f_i(x) = sign(x)$ and the 3D-projection $f_i(x) = h^{(3D)}(x)$ yield better performance than classical separating functions in the two experimental situations. The extra improvement is about 5 dB in Room 1 and about 3 dB in Room 2. These values represent a significant improvement with respect to the moderate SNRI achieved by classical rules. It should also be noted that the SNRI decreases when the inter-microphone distances $d$ become very small (i.e. about 5 cm). This results from the fact that the contributions of each source at the two microphones then become very similar, which yields a badly conditioned mixing matrix.

# 6 Conclusion

This paper deals with the separation of two convolutively mixed signals. The proposed approach is based on a recurrent separation structure adapted by generic rules involving arbitrary separating functions. While the basic versions of this approach were defined and analyzed in our companion paper [3], two extensions were considered here. The first extended approach that we defined is intended for possibly-coloured signals (while only white signals were considered in [3]). This approach applies to many real signals, including mixtures of speech sources. The second extension yields a system which may be operated even when the p.d.f of the sources are unknown and which uses sub-optimum separating functions.

After defining these extended approaches, we analyzed various aspects of their convergence properties at the separating state, i.e. equilibrium and stability conditions and the asymptotic error variance at this state. We also determined the optimum separating functions, i.e. the functions which minimize the asymptotic error variance. We then reported experimental results obtained in various conditions, ranging from synthetic data to mixtures of speech signals measured in real situations. These results confirm the validity of the proposed approach and illustrate its practical performance. They clearly demonstrate the attractiveness of the (sub-)optimization of the separating functions developed in this paper, since both the proposed optimum and sub-optimum approaches significantly outperform classical source separation methods in the considered conditions.

## References


[1] A. Benveniste, M. Metivier and P. Priouret, "Adaptive algorithms and stochastic approximations," *Applications of Mathematics*, Springer-Verlag, Vol. 22, 1990.

[2] N. Charkani, *Séparation auto-adaptative de sources pour les mélanges convolutifs. Application à la téléphonie mains-libres dans les voitures*, PhD thesis, INP Grenoble, France, 1996.

[3] N. Charkani and Y. Deville "Self-adaptive separation of convolutively mixed signals with a recursive structure - Part I: Stability Analysis and Optimization of Asymptotic Behaviour", to appear in Signal Processing, vol. 73, no. 3.

[4] Ph. Garat, *Approche statistique pour la séparation aveugle de sources*, PhD thesis, UJF Grenoble, France, 1994.

[5] J. Hérault, C. Jutten and B. Ans, "Détection des grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé," *Proc. Grestsi'85*, Nice, France, May 1995, pp. 1017-1022.

[6] C. Jutten and J. Hérault, "Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture," *Signal Processing*, Vol. 24, No. 1, July 1991, pp. 1-10.

[7] H.L. Nguyen Thi, *Séparation aveugle de sources à large bande dans un mélange convolutif*, PhD thesis, INP Grenoble, France, 1993.

[8] H.L. Nguyen Thi and C. Jutten, "Blind source separation for convolutive mixtures", *Signal Processing*, Vol. 45, No. 2, March 1996, pp. 209-229.

[9] A. Papoulis, *Probability, random variables and stochastic processes*, McGraw-Hill, New York, 1984.

[10] L.R. Rabiner and R.W. Schafer, *Digital processing of speech signals,* Prentice-Hall, Englewood Cliffs, New Jersey, 1978.

## Tables

| Class of approaches | Algorithm | $(f(x), g(x))$ | SNRI (dB) | $\sigma_\infty \times 10^4$ |
|---|---|---|---|---|
| Classical functions $(f, g)$ without whitening | (N1) | $(x, x)$ | 32.3 | 1.45 |
| | (N1) | $(x^3, x)$ | 29.8 | 1.54 |
| Classical functions $(f, g)$ with whitening | (NW1) | $(x, x)$ | 33.6 | 1.16 |
| | (NW1) | $(x^3, x)$ | 31.2 | 1.09 |
| New functions $(f, g)$ with whitening | (NW1) | $(h^{(2D)}(x), x)$ | 33.7 | 0.91 |
| | (NW1) | $(h^{(3D)}(x), x)$ | 35.2 | 0.76 |
| | (NW1) | $(sign(x), x)$ | 37.1 | 0.42 |

Table 1: Comparative performance for the mixing matrix based on $A_{12}^{(1)}$ and $A_{21}^{(1)}$.

| Class of approaches | Algorithm | $(f(x), g(x))$ | SNRI (dB) | $\sigma_\infty \times 10^4$ |
|---|---|---|---|---|
| Classical functions $(f, g)$ without whitening | (N1) | $(x, x)$ | 17.3 | 7.84 |
| | (N1) | $(x^3, x)$ | 17.0 | 7.14 |
| Classical functions $(f, g)$ with whitening | (NW1) | $(x, x)$ | 17.9 | 5.30 |
| | (NW1) | $(x^3, x)$ | 16.4 | 6.83 |
| New functions $(f, g)$ with whitening | (NW1) | $(h^{(2D)}(x), x)$ | 19.1 | 3.09 |
| | (NW1) | $(h^{(3D)}(x), x)$ | 25.2 | 0.99 |
| | (NW1) | $(sign(x), x)$ | 25.8 | 0.93 |

Table 2: Comparative performance for the mixing matrix based on $A_{12}^{(2)}$ and $A_{21}^{(2)}$.

| $d$ (cm) | SNRI (dB) | | | |
|---|---|---|---|---|
| | $(x, x)$ | $(x^3, x)$ | $(h^{(3D)}(x), x)$ | $(sign(x), x)$ |
| 5 | 3.7 | 4.3 | 8.0 | 7.4 |
| 10 | 5.2 | 5.1 | 9.1 | 9.1 |
| 15 | 4.9 | 4.6 | 10.3 | 10.0 |
| 20 | 4.8 | 5.7 | 10.0 | 9.6 |
| 25 | 5.0 | 6.3 | 10.0 | 10.0 |
| 30 | 5.4 | 6.4 | 10.0 | 9.7 |

Table 3: Comparative performance for Room 1.
Each considered couple of separating functions $(f, g)$ is associated to the adaptation rule (NW1). $d$ is the inter-microphone distance.

| $d$ (cm) | SNRI (dB) | | | |
|---|---|---|---|---|
| | $(x, x)$ | $(x^3, x)$ | $(h^{(3D)}(x), x)$ | $(sign(x), x)$ |
| 5 | 3.0 | 1.8 | 3.2 | 3.4 |
| 10 | 3.9 | 2.4 | 6.2 | 5.8 |
| 15 | 4.0 | 4.0 | 7.8 | 7.8 |
| 20 | 4.9 | 5.2 | 7.6 | 7.6 |
| 25 | 5.2 | 5.6 | 7.2 | 7.2 |
| 30 | 6.9 | 7.5 | 9.5 | 9.5 |

Table 4: Comparative performance for Room 2.
Each considered couple of separating functions $(f, g)$ is associated to the adaptation rule (NW1). $d$ is the inter-microphone distance.

# Figures

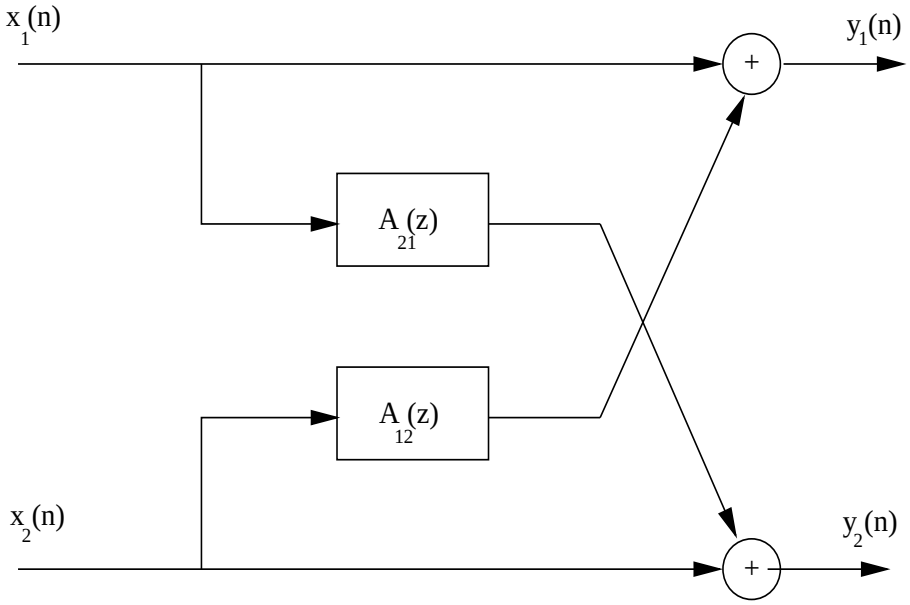


Figure 1: Basic mixture model for source separation.

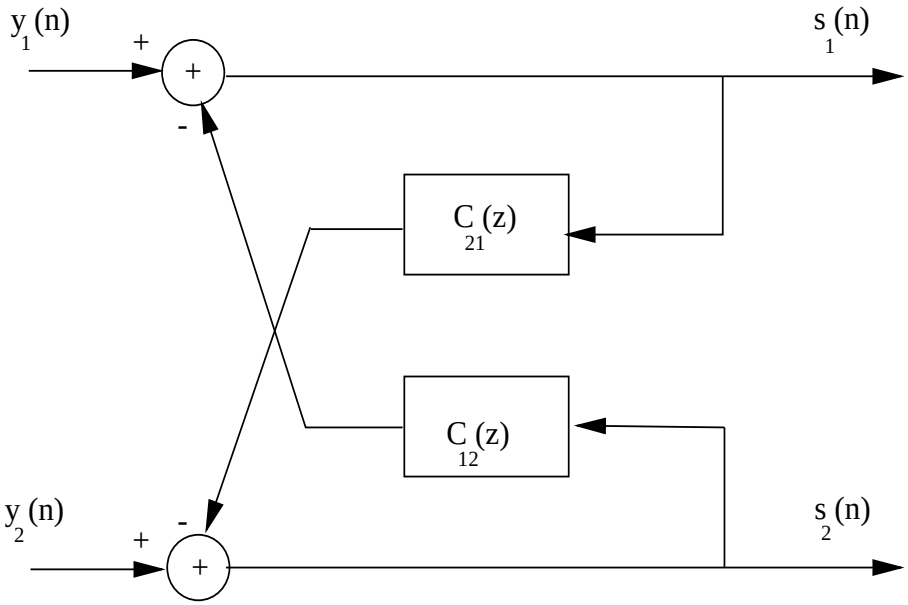


Figure 2: Recurrent structure for the separating system.

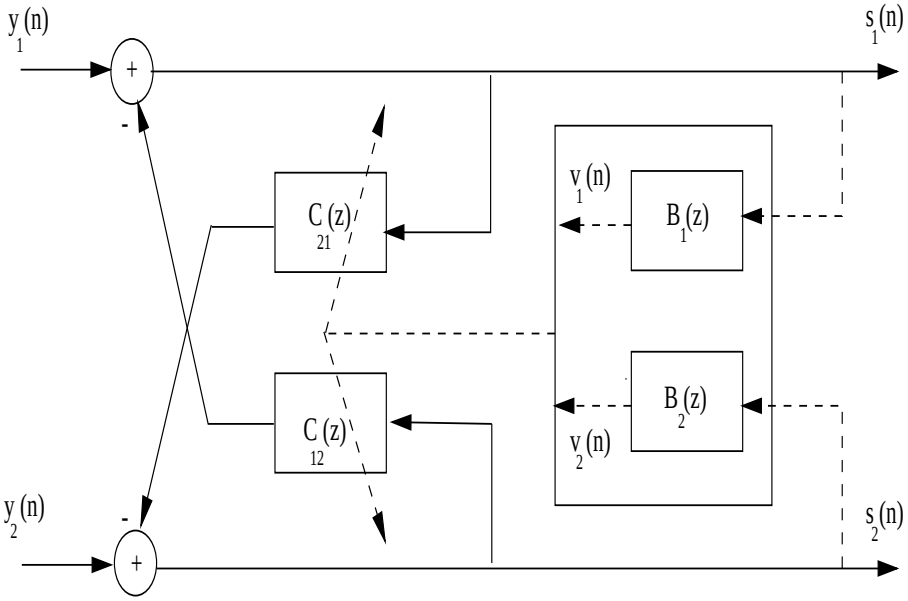


Figure 3: Recurrent separating system for coloured signals.

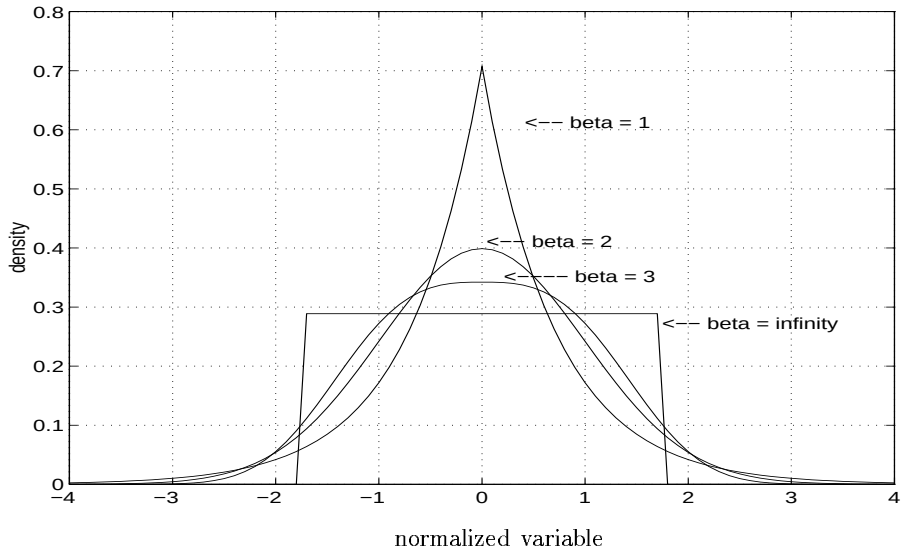


Figure 4: Generalized Gaussian family p.d.f for several values of $\beta$.

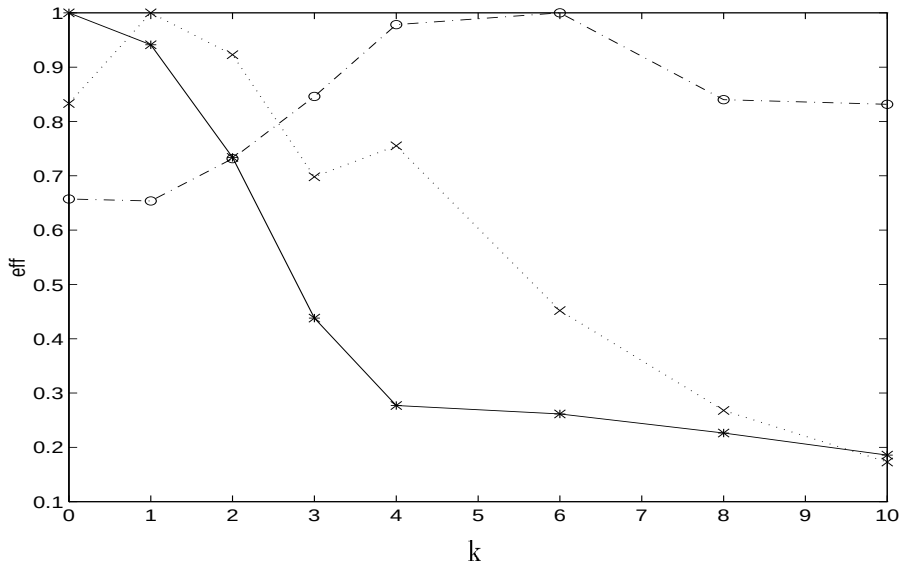


Figure 5: Efficiency vs parameter $k$ of the separating function $f_i$. Each plot corresponds to a specific value of the parameter $\beta$ of the sources: —: $\beta = 1$, ....: $\beta = 2$, -.-.: $\beta = 7$.

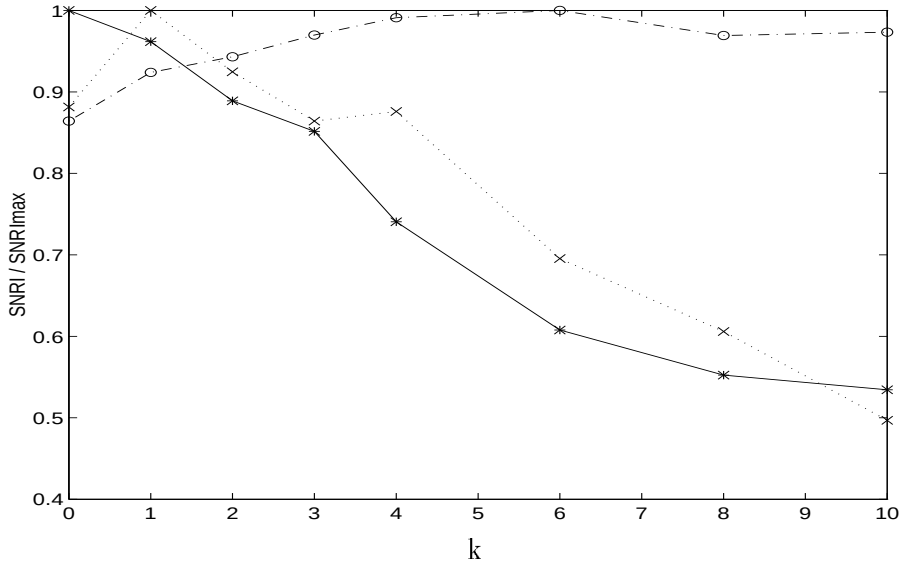


Figure 6: $SNRI/SNRI_{max}$ vs parameter $k$ of the separating function $f_i$. Each plot corresponds to a specific value of the parameter $\beta$ of the sources: —: $\beta = 1$, ....: $\beta = 2$, -.-.: $\beta = 7$.

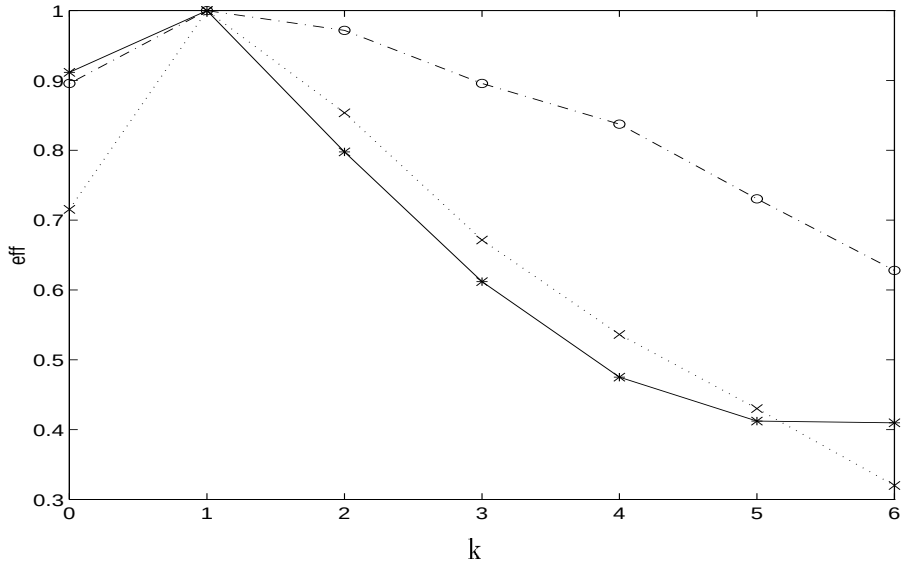


Figure 7: Efficiency vs parameter $k$ of the separating function $g_i$. Each plot corresponds to a specific value of the parameter $\beta$ of the sources: —: $\beta = 1$, ....: $\beta = 2$, -.-.: $\beta = 7$.

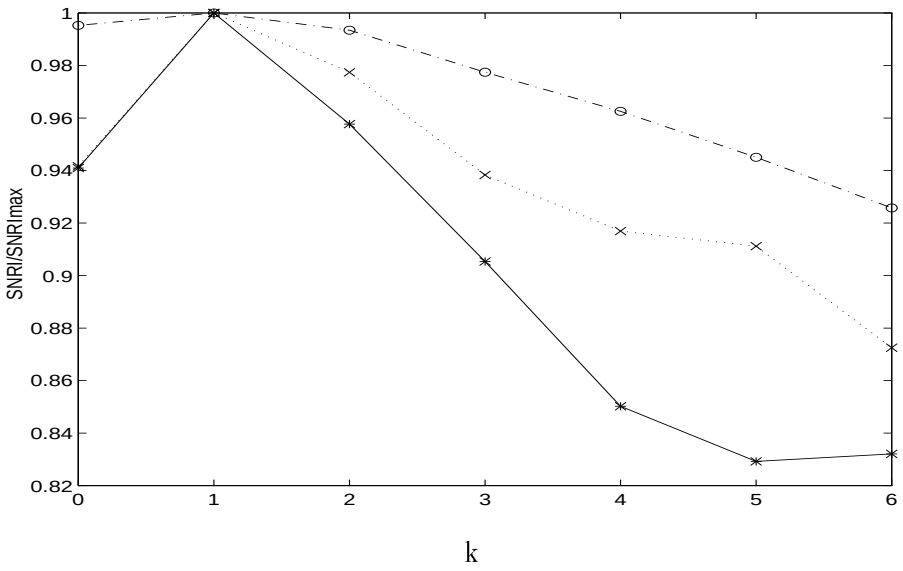


Figure 8: $SNRI/SNRI_{max}$ vs parameter $k$ of the separating function $g_i$. Each plot corresponds to a specific value of the parameter $\beta$ of the sources: —: $\beta = 1$, ....: $\beta = 2$, -.-.: $\beta = 7$.

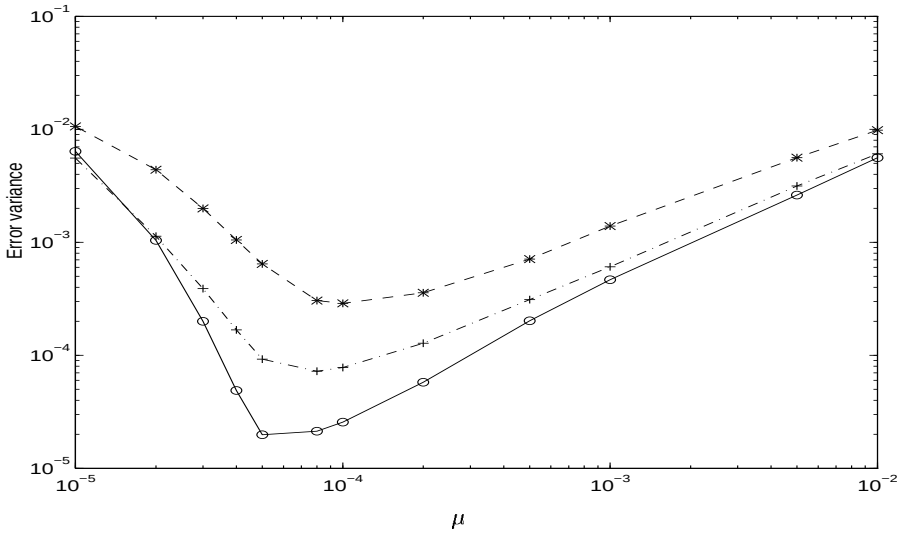


Figure 9: Error variance vs adaptation gain $\mu$. Each plot corresponds to a specific separating function $f_i(x)$ (while $g_i(x) = x$): —: $f_i(x) = sign(x)$, -.-.-: $f_i(x) = x$, - - -: $f_i(x) = x^3$.

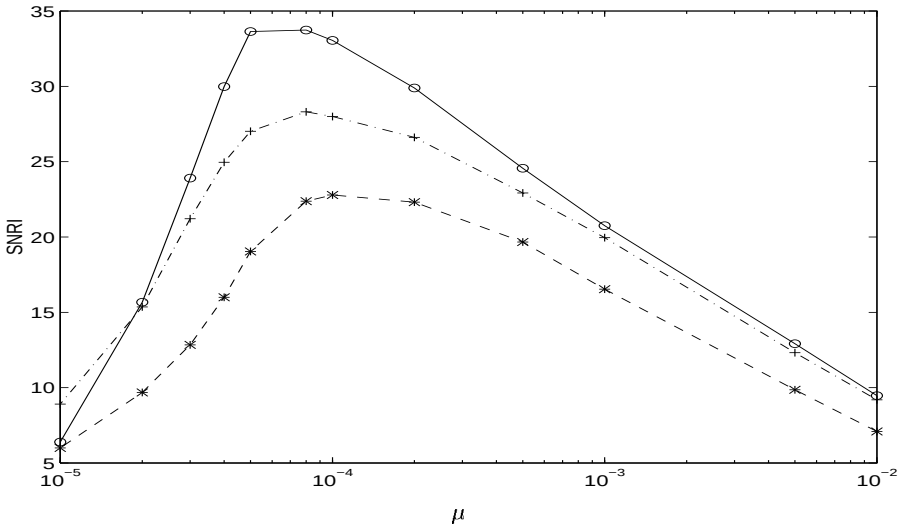


Figure 10: SNRI (dB) vs adaptation gain $\mu$. Each plot corresponds to a specific separating function $f_i(x)$ (while $g_i(x) = x$): —: $f_i(x) = sign(x)$, -.-.-: $f_i(x) = x$, - - -: $f_i(x) = x^3$.

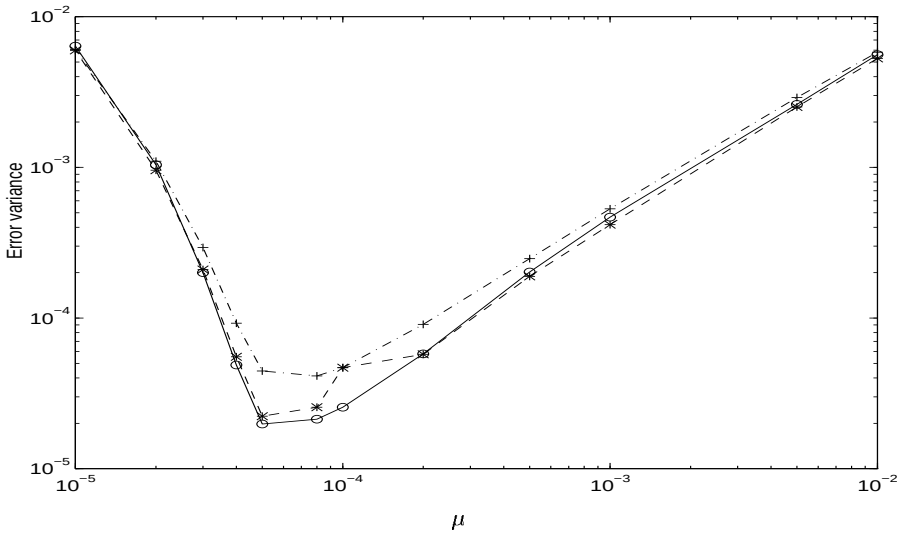


Figure 11: Error variance vs adaptation gain $\mu$. Each plot corresponds to a specific separating function $f_i(x)$ (while $g_i(x) = x$): —: $f_i(x) = sign(x)$, -.-.-: $f_i(x) = h^{(2D)}(x)$, - - -: $f_i(x) = h^{(3D)}(x)$.

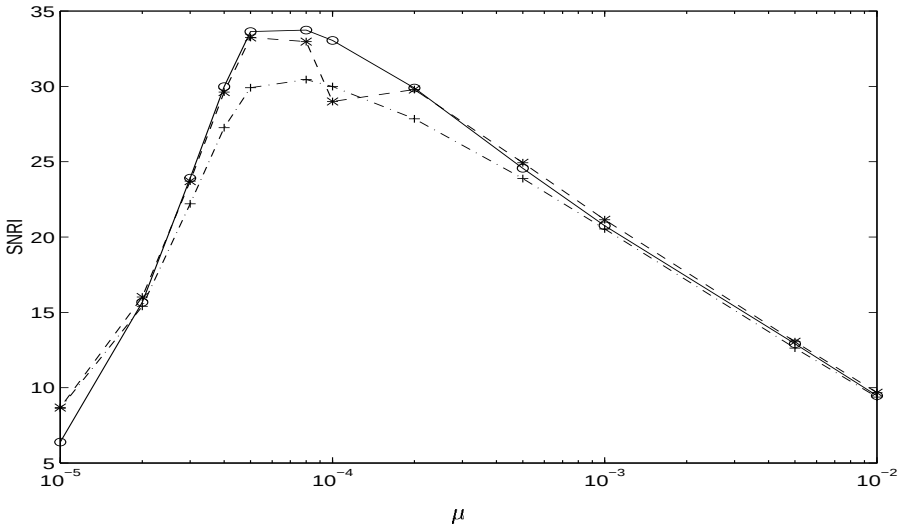


Figure 12: SNRI (dB) vs adaptation gain $\mu$. Each plot corresponds to a specific separating function $f_i(x)$ (while $g_i(x) = x$): —: $f_i(x) = sign(x)$, -.-.-: $f_i(x) = h^{(2D)}(x)$, - - -: $f_i(x) = h^{(3D)}(x)$.

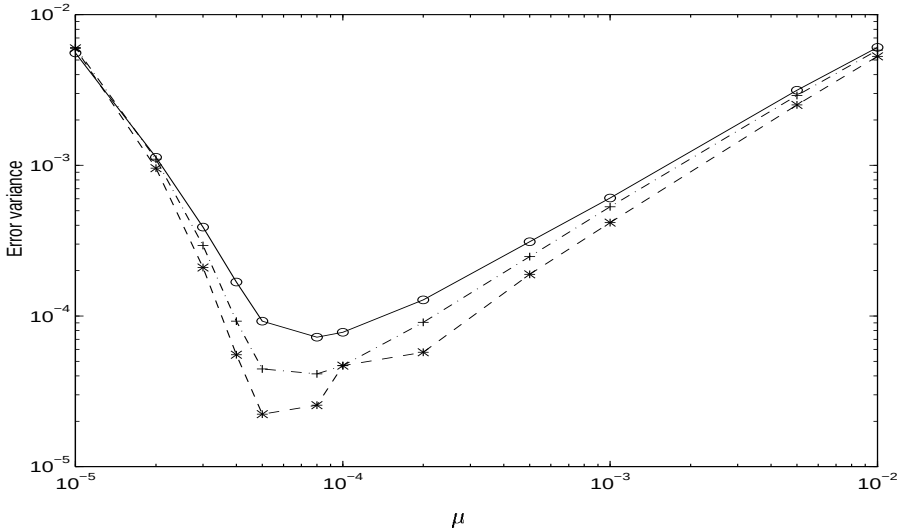


Figure 13: Error variance vs adaptation gain $\mu$. Each plot corresponds to a specific separating function $f_i(x)$ (while $g_i(x) = x$): —: $f_i(x) = x$, -.-.-: $f_i(x) = h^{(2D)}(x)$, - - -: $f_i(x) = h^{(3D)}(x)$.

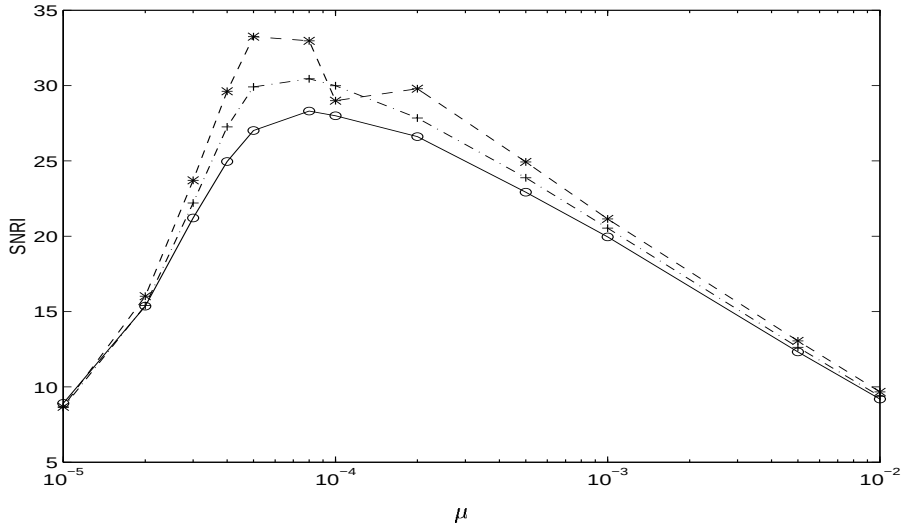


Figure 14: SNRI (dB) vs adaptation gain $\mu$. Each plot corresponds to a specific separating function $f_i(x)$ (while $g_i(x) = x$): —: $f_i(x) = x$, -.-.-: $f_i(x) = h^{(2D)}(x)$, - - -: $f_i(x) = h^{(3D)}(x)$.

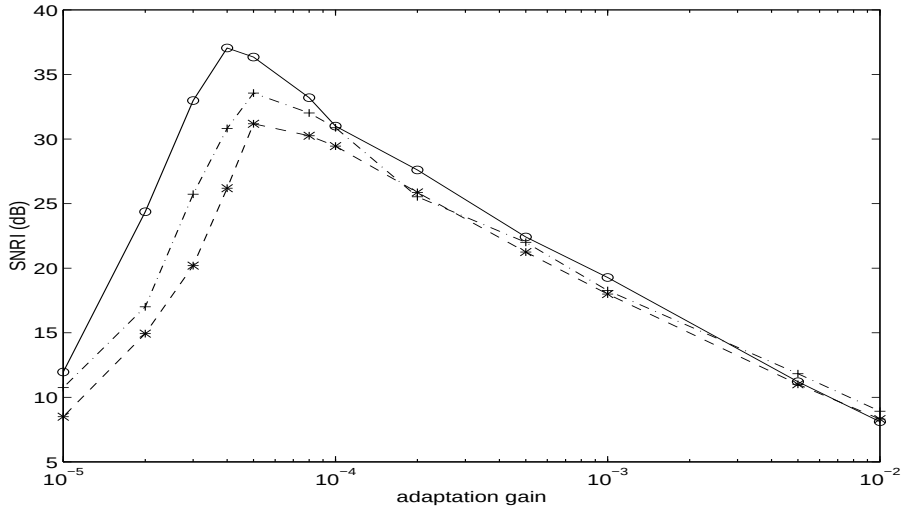


Figure 15: SNRI vs adaptation gain $\mu$ for the mixing matrix based on $(A_{12}^1(z), A_{21}^1(z))$. Each plot corresponds to (NW1) with a specific separating function $f_i(x)$ (while $g_i(x) = x$): —: $f_i(x) = sign(x)$, -.-.-: $f_i(x) = x$, - - -: $f_i(x) = x^3$.

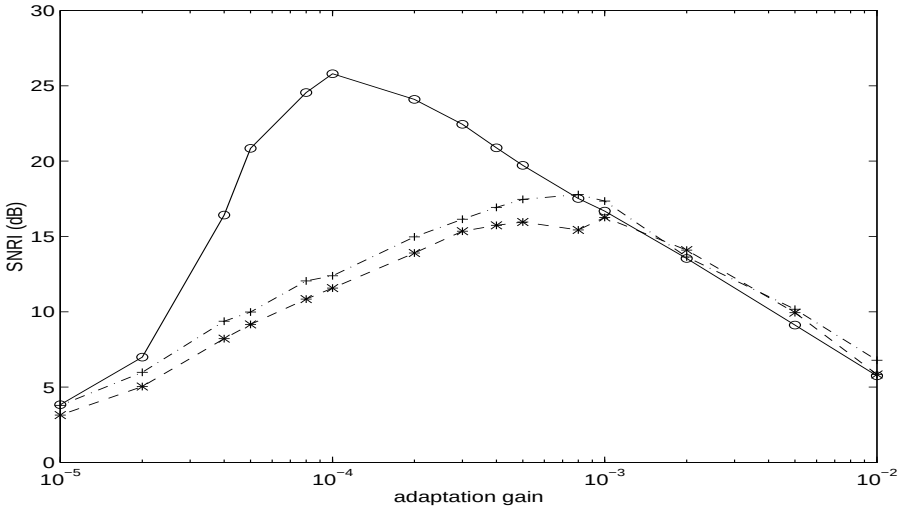


Figure 16: SNRI vs adaptation gain $\mu$ for the mixing matrix based on $(A_{12}^2(z), A_{21}^2(z))$. Each plot corresponds to (NW1) with to a specific separating function $f_i(x)$ (while $g_i(x) = x$): —: $f_i(x) = sign(x)$, -.-.-: $f_i(x) = x$, - - -: $f_i(x) = x^3$.

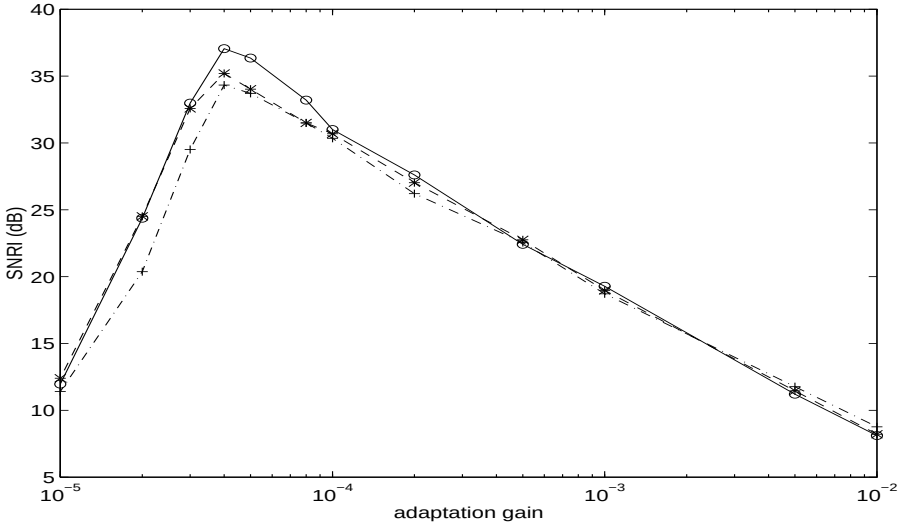


Figure 17: SNRI vs adaptation gain $\mu$ for the mixing matrix based on $(A_{12}^1(z), A_{21}^1(z))$. Each plot corresponds to (NW1) with a specific separating function $f_i(x)$ (while $g_i(x) = x$): —: $f_i(x) = sign(x)$, -.-.-: $f_i(x) = h^{(2D)}(x)$, - - -: $f_i(x) = h^{(3D)}(x)$.

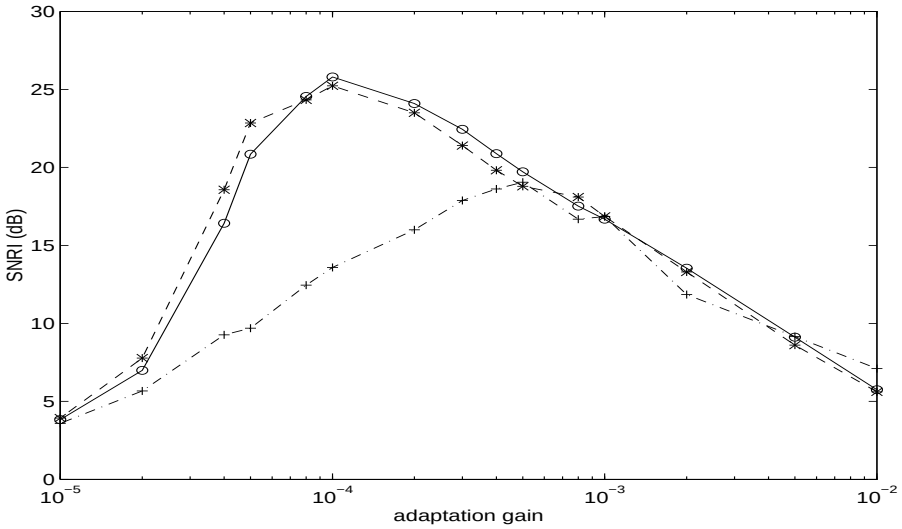


Figure 18: SNRI vs adaptation gain $\mu$ for the mixing matrix based on $(A_{12}^2(z), A_{21}^2(z))$. Each plot corresponds to (NW1) with to a specific separating function $f_i(x)$ (while $g_i(x) = x$): —: $f_i(x) = sign(x)$, -.-.-: $f_i(x) = h^{(2D)}(x)$, - - -: $f_i(x) = h^{(3D)}(x)$.