This is the published version:


Available from Deakin Research Online:

http://hdl.handle.net/10536/DRO/DU:30059312

Reproduced with the kind permission of the copyright owner.

Copyright: 2008, IEEE
Nonnegative Matrix Factorization Applied to Nonlinear Speech and Image Cryptosystems

Shengli Xie, Senior Member, IEEE, Zuyuan Yang, and Yuli Fu

Abstract—Nonnegative matrix factorization (NMF) is widely used in signal separation and image compression. Motivated by its successful applications, we propose a new cryptosystem based on NMF, where the nonlinear mixing (NLM) model with a strong noise is introduced for encryption and NMF is used for decryption. The security of the cryptosystem relies on following two facts: 1) the constructed multivariable nonlinear function is not invertible; 2) the process of NMF is unilateral, if the inverse matrix of the constructed linear mixing matrix is not nonnegative. Comparing with Lin’s method (2006) that is a theoretical scheme using one-time Padding in the cryptosystem, our cipher can be used repeatedly for the practical request, i.e., mutlitime padding is used in our cryptosystem. Also, there is no restriction on statistical characteristics of the ciphers and the plaintexts. Thus, more signals can be processed (successfully encrypted and decrypted), no matter they are correlative, sparse, or Gaussian. Furthermore, instead of the number of zero-crossing-based method that is often unstable in encryption and decryption, an improved method based on the kurtosis of the signals is introduced to solve permutation ambiguities in waveform reconstruction. Simulations are given to illustrate security and availability of our cryptosystem.

Index Terms—Nonnegative matrix factorization (NMF), blind source separation (BSS), signal encryption.

I. INTRODUCTION

WITH the fast development of the information technology, the communication between persons becomes more and more convenient, meantime, the security has been attracting more and more attention. As a result, the encryption for information or signals is a natural choice, and multifarious methods have been utilized, such as the traditional data encryption standard (DES) and Rivest, Shamir, and Adleman (RSA) [1]–[3]. The analog encryption is another technique used in speech communication [4]. Besides, the method based on chaotic system and some other methods have also been developed [5]–[8]. In fact, for speech and image sources, the original plaintexts are composed of partly distorted samples. So, in the decryption, an efficient estimation needs to be used to recover the plaintext, instead of the unbiased estimation. It is acceptable to recover the plaintext approximately and not analytically. To this end, a new technique based on blind source separation (BSS) is applied to image and speech encryption [9]–[11]. Considering the cryptosystem, it adopts the linear encryption scheme and its security relies on the difficulty of solving the underdetermined BSS problem, instead of the traditional apparently intractability of the computational problem. It is an attractive symmetric-key cryptosystem, where the encryption is quite laconic and the decryption is also very convenient using independent component analysis (ICA) algorithms [12]–[17]. However, the following implicit assumptions limit its practical application: 1) the plaintext signals are mutually independent, and there is at most one Gaussian signal; 2) the cipher signals are mutually independent and one cipher should be used only once, i.e., one-time padding. The limits of this cryptosystem are based on the reasons that many real signals are correlative or dependent, such as image signals, and it is quite troublesome to use one-time padding since the high cost of frequent transmission of the keys or the secret seeds in a symmetric-key encryption scheme, especially for online encryption where the signal is encrypted frame by frame [11]. Usually, speech and image signals are sparse (or sparse in a transform domain). The sparsity is often successfully used to solve underdetermined BSS [18]–[22]. From paper [22], one can find that the quality of the separation for underdetermined BSS would be improved when the sparsity of the sources goes higher. Therefore, for this cryptosystem, the security may be destroyed by the sparsity of the sources. In fact, it is nearly inevitable when the cipher is used repeatedly (i.e., mutlitime padding). This phenomenon will be verified in Section V.

From above analysis, one can see that the assumption of one-time padding is necessary in the cryptosystem based on BSS/ICA, and it cannot be extended to mutlitime padding. Furthermore, the ICA algorithms may fail for the separation of dependent signals since the independence of the plaintext is required for a better decryption. So, it is an unsolved problem to design a more secure and practical cryptosystem using the BSS technology, especially for sparse and dependent signals.

Generally speaking, the decryption process is equivalent to matrix factorization, including the decryption algorithms based on ICA and second order statistics [23]. Nonnegative matrix factorization (NMF) is one of the directions in matrix factorization [24]. The convergence of the algorithm is analyzed in [25]. Furthermore, some relative mature algorithms are also developed [26]–[29] and work well especially for sparse signal. Since NMF algorithms do not require statistical characteristics of the original signals, it can be used for the decryption of the correlative or dependent signals, Gaussian signals et al.

As mentioned above, there exist several drawbacks for the BSS/ICA-based cryptosystem in practice. In this study, the encryption is implemented via two parts of ciphers. The first part and the second part are generated by two key signal generators.
(KSGs) driven by secret seeds $I_0$ and $I_1$, respectively. Usually, the first part is set to be static while the second part is set to be dynamic for different encryption. The decryption will be implemented using NMF and the static part of the same ciphers [30]. A nonlinear mixing (NLM) model with a strong noise is introduced for encryption, and a NMF algorithm is utilized for decryption. The security of our cryptosystem is proved for some usual attacks, such as ciphertext-only attack, known-plaintext attack, and chosen-plaintext attack.

The rest of this paper is organized as follows. In Section II, we present the NMF model and NLM model with a strong noise. Section III describes the structure of the proposed cryptosystem. Section IV introduces the construction of the nonlinear function (NLF) and the linear mixing matrix. Simulations and discussions are introduced in Section V. Finally, conclusions are presented in Section VI.

II. NMF Model and NLM Model

Before introducing the models, two definitions are given by:

**Definition 1:** For multivaluable function $y = f(x_1, \ldots, x_r)$ and the fixed $p, (p < r)$ variables of it, a $p$-conditional function is defined as $y = f^p_i(x_1, \ldots, x_r)$, where $i \in \Omega_1, f \in \Omega_2, \Omega_1 \cap \Omega_2 = \emptyset, \Omega_1 \cup \Omega_2 = \{1, \ldots, r\}, \forall i, \forall \Omega_2, \mathcal{C}(\Omega_2) = p - r, \mathcal{C}(\Omega_1) = r - p$, and $\mathcal{C}(\cdot)$ means cardinality of a set.

**Definition 2:** For the $p$-conditional function $y = f^p_i(x_1, \ldots, x_r)$, when $p = r - 1$, if there exists $j \in \Omega_1$ such that $x_i = f^p_i(\{x_j\})$, then $f$ is $p$-conditionally invertible with respect to $x_i$. And $f_i$ is the corresponding $p$-conditional inverse function of $f$ with respect to $x_i$.

**Remarks:** If $p < r - 1$, $f$ is not $p$-conditionally invertible to $x_i$, since at least two of the rest variables are free. If $p = r - 1$, although there is only one free variable, $f$ may be still invertible (e.g., $f$ is a transcendental function, see model (14)).

A. NMF Model

For nonnegative data matrix $Y^n = [y^n(1), \ldots, y^n(T)] \in \mathbb{R}^{d \times T}$, the aim of the matrix factorization is to find two nonnegative matrix $A^n \in \mathbb{R}^{d \times m}$ and $S^n = [s^n(1), \ldots, s^n(T)] \in \mathbb{R}^{m \times T}$, such that $Y^n = A^nS^n$. The following approximate decomposition is generally considered [27], [29]:

$$Y^n = A^n S^n + V$$  \hspace{1cm} (1)

where $V \in \mathbb{R}^{d \times T}$ is the error matrix, $y^n(t) = [y^n_1(t), \ldots, y^n_d(t)]^T$ is a vector of the observed signals at the discrete time instants $t$, and $s^n(t) = [s^n_1(t), \ldots, s^n_m(t)]^T$ is a vector of the components or source signals at the same time instant, generally, $T \gg d \geq m$. The scalar form of the model (1) is: $y^n_i(t) = y^n_d = \sum_{j=1}^{d} a^n_{dj} s^n_j(t) + v^n_i(t), i = 1, \ldots, d, t = 1, \ldots, T$. The following notations are used: $s^n_i(t) = s^n_d, y^n_i(t) = y^n_d$, and $[A^nS^n]_{dt}$ means the $dt$-element of the matrix $A^nS^n$.

B. NLM Model and Demixing Model

Given the signal matrix $S^n = [s^n(1), \ldots, s^n(T)] \in \mathbb{R}^{m \times T}$ and the strong noise matrix $U = [u(1), \ldots, u(T)] \in \mathbb{R}^{m \times T}$, then the NLM model respect to $S^n$ with a strong noise is

$$X^n = G(U, S^n)$$  \hspace{1cm} (2)

where $X^n \in \mathbb{R}^{m \times T}$, or equivalently in a scalar form as $x^n_i(t) = g_{it}(u_i(t), s^n_i(t))$, $i = 1, \ldots, m, t = 1, \ldots, T$ and $g_{it} = [G]_{it}$ is a NLF. If $U$ is known, the demixing model is

$$S^n \approx G^{-1}(X^n | U)$$  \hspace{1cm} (3)

or equivalently in a scalar form as $s^n_i(t) \approx g_{it}^{-1}(x^n_i(t)|u_i(t)), i = 1, \ldots, m, t = 1, \ldots, T$, where $g_{it}^{-1} = [G]^{-1}_{it}$ is the $(p = 1)$-conditional inverse function of $g_{it}$ with respect to $s^n_i(t)$. Note that the noise in the NLM model could be very strong.

III. PROPOSED CRYPTO SYSTEM

First of all, the general assumptions of a useful cryptosystem are considered as 1) validity: the signals can be decrypted efficiently after the encryption; 2) security: the signals cannot be decrypted without the cipher, meantime, the cipher cannot be obtained from the public plaintext when the Multi-Time padding method is adopted. And the corresponding typical attacks to a cryptosystem are:

**Attack 1:** ciphertext-only attack, the attacker knows only the ciphertext and aims at the plaintext.

**Attack 2:** known-plaintext attack, the attacker knows parts of the pairs of plaintext and ciphertext, and aims at the cipher.

**Attack 3:** chosen-plaintext attack, the attacker can select some special pairs of plaintext and ciphertext, and aims at the cipher.

Considering the BSS/ICA-based cryptosystem in [11], we have two conclusions as follows.

**Theorem 1:** The BSS/ICA-based cryptosystem fails for dependent or Gaussian sources, i.e., the cryptosystem cannot work in the case of more than 2 dependent or Gaussian plaintexts.

It is easy to prove the theorem, since the validity of the cryptosystem is violated and at least two plaintexts cannot be encrypted.

**Theorem 2:** To encrypt the sparse signals, the one-time padding method cannot be extended into the multi-time padding, i.e., the BSS/ICA-based cryptosystem is not secure when the multtime padding method is adopted.

The proof is given in Appendix I.

**Remarks:** Theorem 1 and Theorem 2 express the restrictions on the cryptosystem in [11], and the trap for sparse signal encryption exists widely in the linear encryption scheme. The phenomena are verified in Section V. Furthermore, with the increase of the prior information, the attack becomes easier. In fact, according to the proof of the Theorem 2, no matter the
source signal is sparse or not, the linear BSS/ICA-based cryptosystem using the multitime padding cannot protect the plaintexts from the Attack 2 and the Attack 3. Also, the proof supplies a method to decrypt the plaintext without the cipher.

The nonlinear encryption scheme will be introduced. The block diagram of the proposed cryptosystem is shown in Fig. 1, mainly including the preprocessing, the encryption, the decryption, and the waveform reconstruction. The full structure will be introduced in detail in this Section.

A. Preprocessing

For a source signal \(s(t)\), at first, using segment splitter in [11], it is divided into \(N\) frames, then the each frame is put into \(P\) segments with the unify length \(T\). The samples in every segment are transformed into nonnegative ones by an invertible regularized transformation (RT). Also, after dividing by maximal norm of the signals, the nonnegative signals are restricted in the interval \([0, 1]\). The waveform information of every segment is stored in a definite format, including kurtosis (kurt), maximum (\(s_{\text{max}}\)), and minimum (\(s_{\text{min}}\)). It is inserted into the head data of the encrypted signal in a definite format for transmission, together with the parameters \(P, T\).

The sign function is utilized to define the RT as

\[
\text{RT}(s^n_t) = \text{sgn}(s^n_t) \cdot s^n_t(t)
\]

where \(s^n_t(t)\) is the \(i\)th segment of the \(n\)th frame \(t = 1, \ldots, T\).

\[
\text{sgn}(s^n_t) = \begin{cases} 
1, & \text{if } s^n_t(t) \geq 0 \\
-1, & \text{else}
\end{cases}
\]

B. Encryption

Assumptions: 1) The data matrix of the nth frame is \(S^n = [s^n(1), s^n(2), \ldots, s^n(T)] \in \mathbb{R}^{P \times T}\). 2) The key signal matrix generated by \(I_0\) is \(U = [u(1), u(2), \ldots, u(T)] \in \mathbb{R}^{P \times T}\). 3) The NLF matrix is \(G = [g_1, g_2, \ldots, g_T]\), and the ciphertext data matrix is \(X^n = [x^n(1), x^n(2), \ldots, x^n(T)] \in \mathbb{R}^{P \times T}\), where \(x^n(t) = [s^n(t), \ldots, s^n(t)]^T,\ u(t) = [u_1(t), \ldots, u_P(t)]^T,\ g(t) = [g_{1t}, \ldots, g_{Pt}]^T,\ \text{and } x^n(t) = [x^n_1(t), \ldots, x^n_P(t)]^T\), for \(t = 1, \ldots, T\). Then the encryption is described by

\[
X^n = G(U, S^n, A^n)
\]

where \(A^n \in \mathbb{R}^{P \times P}\) is nonnegative matrix generated by the seed \(I_1\) and \(u_i(t)\) distributes uniformly between \(-1\) and \(1\).

There exists a function matrix \(G\), such that the model (6) can be rewritten as follows (see also the model (16))

\[
X^n = G(U, A^nS^n) = G(U, Y^n)
\]

where \(Y^n = [y^n(1), y^n(2), \ldots, y^n(T)] = A^nS^n\), i.e.,

\[
x^n(t) = g_t(u(t), A^n s^n(t)) = g_t(u(t), y^n(t))
\]
or equivalently in a scalar form as \(x^n(t) = g_t(u(t), y^n(t)), i = 1, \ldots, P, t = 1, \ldots, T\), and \(y^n(t) = [y^n_1(t), \ldots, y^n_P(t)]^T\). Suppose that the key signal \(U\) is the noise, then the model (7) is an NLN model respect to \(Y^n\) with a strong noise, and it can also be regarded as a post-NLM model respect to \(S^n\).

According to the model (7), the NLF matrix \(G\) plays a key role in the security of the cryptosystem, and the construction of it will be exploited in the next Section. The source signal \(s(t)\) is encrypted frame by frame with the same method, where the ciphers \(U\) and the nonlinear functions can be used repeatedly. The styles of the nonlinear functions are public, and their parameters can be inserted into the head data of the encrypted signal in a definite format for transmission.

C. Decryption

After receiving the \(nth\) encrypted frame (i.e., the ciphertext), combining with the cipher regenerated by \(I_0\) and the public NLF (see Section IV), we can obtain a linear mixture of the plaintext, according to the models (3), (7), and (8), i.e.,

\[
y^n(t) = g_t(x^n(t)u(t))
\]

where \(y^n(t) = [y_{1n}(t), \ldots, y_{Pn}(t)]^T = A^n s^n(t), t = 1, \ldots, T,\ g_t = [g_{1n}, \ldots, g_{Pn}]^T,\ \text{and } g_t\ is the (p = 1)-conditional inverse function of g_d respect to y_d(t).

Note that \(y^n(t)\) is calculated analytically, both the mixing matrix \(A^n\) and the plaintext \(s^n(t)\) (or \(S^n\)) are nonnegative, then \(y^n(t)\) (or \(Y^n\)) is nonnegative. Based on the model (1), where \(m = d = P\), the plaintext signals can be decrypted from \(Y^n\) using the NMF algorithm, except for the indeterminacy of permutation and scale.
The following typical optimization problem based on Euclidean distance is often used for solving the model (1): Minimize
\[
\begin{align*}
D(A^n, S^n) &= \|Y^n - A^nS^n\|^2 \\
S_{i,0}a^n_{ij} &\geq 0, s^n_{jk} \geq 0, \forall i, j, k
\end{align*}
\]
and the corresponding algorithm is as follows [25, Th. 1].

**NMF Algorithm:**

1. **Initialization:** generate nonnegative matrices \(A^n, S^n\) randomly, then, calculate the best \(S^n\) by iterative loop. For \(\text{iter}=1: \max \text{num}\)

   \[
   s^n_{jk} \leftarrow s^n_{jk} \frac{\left[ (A^n)^T Y^n \right]_{jk}}{\left[ (A^n)^T A^n S^n \right]_{jk}} \quad (11)
   \]

   \[
   a^n_{ij} \leftarrow a^n_{ij} \frac{\left[ Y^n (S^n)^T \right]_{ij}}{\left[ A^n S^n (S^n)^T \right]_{ij}} \quad (12)
   \]

   \[
   a^n_{ij} \leftarrow \frac{a^n_{ij}}{\sum_i a^n_{ij}} \quad (13)
   \]

   end

where \(\max \text{num}\) is the maximum iterative number, and the added model (13) is used to fix the scale of the mixing matrix.

This is just one of the NMF algorithms, and the accuracy of which will improve when the linear mixing matrix and the original signal matrix are sparse [26].

**D. Reconstruction**

Just like decryption method based on ICA, the output of the model (10) is the matrix composed of \(P\) segments with ambiguities of permutation and scale. To solve the permutation problem, a simple method based on the number of zero-crossings (nzc) is used in [11]. However, the index nzc is often unstable when we aim at efficient estimations of the original segments. One may find that \(\text{kurt}\) index seems more robust. In Table I, a comparison is made for these two indexes before the encryption and after the decryption, where the source comes from the first frame of the speech in [11]. The test is made for three partitions \((P = 4, 5, 6)\), respectively.

In the proposed cryptosystem, the \(\text{kurt}\) information of the original segments is used to solve the indeterminacy of the permutation problem. The maximum \(s_{\max}\) and the minimum \(s_{\min}\) of the waveforms are still utilized to solve the scale ambiguity problem [11]. Note that the outputs of the model (10) are estimations of the original segments after preprocessing using RT, in order to obtain the exact \(P\) original segments, it is necessary to process the outputs using the inverse regularization transformation (IRT), after solving the ambiguities of permutation and scale. The source signal \(s(t)\) is recovered frame-by-frame using the same method.

**IV. CONSTRUCTION OF NLF AND LINEAR MIXING MATRIX**

As mentioned above, for the proposed cryptosystem, the security of the encryption relies mainly on the NLF and the level of the decryption depends partly on the linear mixing matrix. As such, the construction of the NLF and the linear mixing matrix will be exploited in this Section.

**A. Construction of NLF**

In order to simplify the encryption and decryption, the assumption \(f_{it} = g_{kl}\), for \(\forall i, t, k, l\), is used.

The NLF matrix is given by

\[
G(U, S^n, A^n) = \tilde{\Phi}(U) - \bar{\Phi}(U, S^n, A^n) \quad (14)
\]

where \(\bar{\Phi}(U, S^n, A^n) = \exp(-\beta \ast (U + A^n S^n))\). The parameter \(\beta \neq 0\) is a amplitude modulation such that \(U\) and \(S^n\) are well masked mutually in \(G\). It is used repeatedly together with \(U\). The selection of \(\beta\) is adaptive, since \(U, A^n\) and \(S^n\) are known at the stage of encryption. Function \(\tilde{\Phi}\) is composed of the principal components (PC) of \(U\), i.e., letting \(B\) be the eigenvector matrix of the covariance matrix \(E[UU^T]\), where \(E[\cdot]\) denotes mathematical expectation, the function is

\[
\tilde{\Phi}(U) = B^T U
\]

Then the plaintext \(S^n\) is encrypted as

\[
X^n = \tilde{\Phi}(U) - \tilde{\Phi}^n(U, S^n, A^n) \quad (15)
\]

i.e.,

\[
X^n = B^T U - \exp(-\beta \ast (U + A^n S^n)) \quad (16)
\]
where $B$ is the eigenvector matrix of the covariance matrix $E[UU^T]$. The model (16) can be rewritten in a scalar form as

$$x_i^p(t) = g_{id}(u_i(t)) - \exp(-\beta(u_i(t) + y_i^p(t)))$$  \hspace{1cm} (17)$$

If $\beta \neq 0$ in the model (17), then the function $g_{id}$ is $p$ -conditionally invertible with respect to $y_i^p(t)$, and the corresponding inverse function is

$$y_i^p(t) = -\log(\hat{g}_{id}(u_i(t)) - x_i^p(t))/\beta - u_i(t).$$  \hspace{1cm} (18)$$

According to the models (16) and (17), the encryption of the proposed cryptosystem is simple. It is also convenient for decryption, when $x_i^p(t)$ and $u_i(t)$ are known at the stage of the decryption. Thus, $y_i^p(t)$ can be obtained from model (18), and the data matrix $Y^n = A^nS^n$ is attained. Then, $S^n$ (or $s_i^n(t)$) can be decrypted by the NMF algorithms. Using the same method, the source can be decrypted frame-by-frame.

Noting that $\hat{g}_{id}(u_i(t))$ is calculated using the same method in the encryption and the decryption, $A^n$ is used only for encryption, and $\beta$ can be inserted into the head data of the encrypted signal in a definite format for transmission.

### B. Construction of Linear Mixing Matrix

Since the accuracy of the NMF algorithm relies partly on the characteristics of the mixing matrix [26]–[28], it is necessary to select a proper mixing matrix $A^n$ for the $n$th frame. The following guidelines are given to select mixing matrix $A^n$:

i) nonnegative and full column rank;

ii) nontrivial, i.e., there are at least two nonzero entries in each row;

iii) as sparse as possible.

E.g., there are two nonzero elements in each row for simplicity:

$$A^n = \begin{bmatrix}
\gamma_{11}^n & \gamma_{12}^n & 0 & \cdots & 0 & 0 & 0 \\
0 & \gamma_{21}^n & \gamma_{22}^n & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \gamma_{(P-1)1}^n & \gamma_{(P-1)2}^n & 0 \\
0 & 0 & 0 & \cdots & 0 & \gamma_{P1}^n & 0 \\
\end{bmatrix}$$  \hspace{1cm} (19)$$

where $\gamma_{ij}^n, i \in [1,P], j \in \{1,2\}$ are random positive numbers less than one. Note that some other forms of $A^n$ can also be considered from the security point of view, e.g., the number of the nonzero elements in each row may be different regularly.

### C. Analysis of Security

**Theorem 3:** For the three typical attacks, the cryptosystem encrypted by the model (8) keeps secure even when the multitime padding method is used, if the function $x_i^p(t) = g_{id}(u_i(t), y_i^p(t))$ is not $p(p = 1)$ -conditionally invertible with respect to $u_i(t)$ when $y_i^p(t) = 0$, where $i \in [1,P], t \in [1,T]$.

The proof is given in Appendix II

**Remark:** Considering the proposed cryptosystem, when the nonlinear functions are selected properly, the security is equivalent to the BSS/ICA-based method in [11], although the multitime padding method is adopted.

For the NLF in the model (17), it is not $p(p = 1)$-conditionally invertible with respect to $u_i(t)$ when $y_i^p(t) = 0$, for that $g_{id}$ is a transcendental function. Also, it is quite difficult to obtain the cipher $U$ using numerical methods. In fact, under the conditions of $E[\hat{g}_{id}(u_i(t))] \neq 0$, $[A^n]_{ij} \in [0,1]$ and $s_i^n(t) \in [0,1]$, $x_i^p(t)$ is the mixture of $u_i(t), A^n$, and $S^n$ (including the amplitudes information of $U$ and $S^n$), which are mixed sufficiently by $g_{id}$. Then, to obtain the cipher $U$ numerically, the unknown and dynamic $A^n$ is one obstacle and $\hat{g}_{id}(u_i(t))$ is another one, even though $u_i(t)$ is fixed and a group of $S^n$ may be known (e.g., Attack 3). Function $\hat{g}_{id}(u_i(t))$ in the model (17) is not an invertible mapping with respect to $u_i(t)$, since $U$ and $WU$ have the same PCs for any orthogonal matrix $W \in \mathbb{R}^{P \times P}$.

Generally speaking, a secure NLF $G$ should be sensitive to $U$. Luckily, more complex NLF can be used to enhance the sensitivity in (17), especially for $\hat{G}$, e.g., $\hat{G}(U) = \exp(B^2U)$ or more complex such as hash or chaotic function. Moreover, it will fail for decryption using similar ciphers (see Section V.C). Note that the more complex NLF may be still a unilateral mapping to $U$, and the computational load does not increase remarkably for both the encryption and the decryption.

As mentioned above, the encryption scheme based on the model (17) is secure, according to the property of the NLF and Theorem 3.

### V. Simulations and Discussion

Three simulations are made to express the shortcomings of the cryptosystem proposed in [11] and the advantages of the cryptosystem proposed here. Signal-to-interference ratio (SIR) is used to evaluate the recovery of the original segments:

$$\text{SIR}_i = 10 \log \frac{E[s_i^n(t)]^2}{E[s_i^n(t) - \hat{s}_i^n(t)]^2}$$  \hspace{1cm} (20)$$

where $s_i^n(t), i = 1, \ldots, P$ are the original segments of the $n$th frame of the source signal, while $\hat{s}_i^n(t)$ is the recovered segment. $E[.]$ denotes mathematical expectation. If $\hat{s}_i^n(t)$ in formula (20) is replaced by the ciphertext $x_i^p(t)$, then it can also be used to evaluate the mask of the original segments. A smaller SIR is better for encryption while a larger SIR is better for decryption. In general, SIR levels below 8–12 dB indicate a failure in obtaining the desired segments [31], [43].

#### A. Simulation 1

This simulation is used to verify the Theorem 2, i.e., the cryptosystem in [11] is not secure for the encryption of the sparse signal when the one-time padding method is extended into the multitime padding. Both of the source signal [see Fig. 2(a)] and the encryption scheme are the same as in [11] except the usage of ciphers. The source signal $s(t)$ is split into four frames and two segments for each frame after preprocessing [see Fig. 2(b)].
i.e., \( P = 2, T = 13750 \). The key signals \( \textbf{U} \) are generated uniformly with the numbers between -1 and 1, and they are used to encrypt the four frames above, where the four underdetermined mixing matrices are

\[
\begin{align*}
\textbf{A}_d^1 &= \begin{bmatrix} -0.7357 & 0.6773 & -1.4715 & 1.3545 \\
0.6773 & 0.7357 & 1.3545 & 1.4715 \end{bmatrix} & (21) \\
\textbf{A}_d^2 &= \begin{bmatrix} -0.8490 & -0.5284 & -3.3959 & -2.1137 \\
-0.5284 & 0.8490 & 3.3959 & 2.1137 \end{bmatrix} & (22) \\
\textbf{A}_d^3 &= \begin{bmatrix} -0.1302 & 0.9915 & 0.7814 & 0.9489 \\
0.9915 & -0.1302 & 0.9489 & 0.7814 \end{bmatrix} & (23) \\
\textbf{A}_d^4 &= \begin{bmatrix} -0.9046 & 0.4263 & 7.2366 & 2.4105 \\
0.4263 & 0.9046 & 2.4105 & 7.2366 \end{bmatrix}. & (24)
\end{align*}
\]

Fig. 2(c) and (d) shows the encrypted segments and uncovered segments under the Attack 1 (i.e., decrypted segments without the cipher), respectively. Note that the proof of Theorem 2 also supplies a method for decryption without the cipher. Table II shows the indexes SIRs of the encryption and the decryption. Since SIRs of the decrypted segments are all greater than 14 dB, the original segments are decrypted efficiently without the cipher [15], [31]. Theorem 2 is verified.

**B. Simulation 2**

In this simulation, the proposed scheme is tested for a speech encryption. The same source signal in [11] is used. After preprocessing, the source is split into two frames and each frame is split into four segments [see Fig. 3(a)], i.e., \( P = 4, T = 13750 \). The segments are mutual independent approximately. Fig. 3(b) shows the segments processed by RT in (4). Some indexes of each segment are shown in Table III, including kurt, \( s_{\text{max}} \) and \( s_{\text{min}} \). Fig. 3(c) shows the keys \( \textbf{U} \) which are generated uniformly between -1 and 1. Two frames of ciphertext signals which are encrypted with the same \( \textbf{U} \) are shown in Fig. 3(d). And the model (17) is used for encryption, where two linear mixing matrices are as following, respectively, the same NLF is used repeatedly and the parameter of the NLF is \( \beta = 0.5 \).

\[
\textbf{A}_1^1 = \begin{bmatrix} 0.8548 & 0.6344 & 0 & 0 \\
0 & 0.3656 & 0.7332 & 0 \end{bmatrix} \quad (25)
\]

\[
\textbf{A}_1^2 = \begin{bmatrix} 0.8908 & 0.7523 & 0 & 0 \\
0 & 0.2477 & 0.7661 & 0 \end{bmatrix} \quad (26)
\]

\[
\textbf{A}_2^1 = \begin{bmatrix} 0.1452 & 0 & 0 & 0.1190 \end{bmatrix}
\]

\[
\textbf{A}_2^2 = \begin{bmatrix} 0.1092 & 0 & 0 & 0.0973 \end{bmatrix}
\]
At the stage of the decryption, the seed $I_0$ received from a secure channel regenerates the key $U$. Combining $U$ with the ciphertexts, the nonlinear demixing is firstly operated based on the undisguised NLF, according to the model (18). Then, the linear demixing is operated using the NMF algorithm based on the output of the first step. At last, the waveform is reconstructed according to the information that is transmitted together with the ciphertext, including the operation of IRT. Fig. 3(e) shows the last decrypted segments, and the indexes SIRs of the encryption and the decryption are shown in Table IV. According to which, the source speech is well masked and efficiently decrypted. The performance of the proposed cryptosystem is comparable with the method in [11].

At the stage of preprocessing, the model (4) is used for RT, and the form of IRT is the same as RT. So, the sign information of the original segments will be used for the waveform reconstruction. Since the size of the matrix composed of the sign information is the same as the ciphertext signal matrix, the transmission of the data will increase remarkably (about 50%) if they are transmitted separately. Note that each entry of the sign matrix is 1 or $-1$, so it can be loaded to the ciphertext signal matrix.
at the corresponding position. It requires the latter to be transformed into a positive matrix. Thereby, the sign matrix can be extracted completely. Luckily, it is easy to implement by adding a sufficiently large positive constant \( C \) to the right side of the model (17) to satisfy that \( \forall x_i, t, x^2_i(t) > 0 \). The constant \( C \) can also be inserted into the head data of the encrypted signal in a definite format for transmission. Under this circumstance, receivers should extract the sign matrix from the ciphertext firstly; subtract the constant \( C \) from the rest ciphertext secondly; and decrypt the ciphertext at last.

### C. Simulation 3

In this simulation, an image signal is tested by our cryptosystem and the cryptosystem proposed in [11], respectively. At the stage of preprocessing, the source image is split into four segments \( s^1_1(t), s^2_1(t), s^3_1(t), s^4_1(t) \) as one frame \( s^2_1(t) \) [see Fig. 4(a)], and the segments are dependent. The matrix of the correlation coefficients of the segments is

\[
\text{Corr} = \begin{bmatrix}
1.0000 & 0.2324 & -0.1360 & 0.2769 \\
0.2324 & 1.0000 & 0.1155 & -0.0184 \\
-0.1360 & 0.1155 & 1.0000 & -0.1334 \\
0.2769 & -0.0184 & -0.1334 & 1.0000
\end{bmatrix}
\]  

(27)

Table V shows the indexes of these four segments, including \( \text{kur}, s_{\text{max}}, \) and \( s_{\text{min}} \). The same keys in the second simulation are used. Note that the source image is nonnegative. It needs not RT in the preprocessing of the proposed cryptosystem.

When the proposed cryptosystem is tested, Monte Carlo method with 95 times is used for decryption. The encrypted segments and the average of decrypted segments are shown in Figs. 4(b), and Fig. 5 shows the maximum and the minimum of SIRs of the recovered segments in every time. Since RT is not necessary, IRT is canceled. The corresponding indexes SIRs of the encryption and the decryption are shown in Table VI, where \( \text{Min} \) denotes the minimum SIR in the 95 experiments, \( \text{Max} \) denotes the maximum SIR, and \( \text{Mean} \) denotes the average SIR. At the stage of encryption, the linear mixing matrix is the same as the model (25) for simplicity, and the parameter of the NLF is also \( \beta = 0.5 \). The data in Table VI show that the source image is well masked and efficiently decrypted [see also Fig. 4(b)]. Under Attack 1, the recovered images are shown in Fig. 4(c), and the SIRs (dB) are \([-11.7005, -12.0832, -11.7852, -11.1920]\). It verifies

![Fig. 4. Encrypted segments and decrypted segments for image signal, using the proposed method and the BSS/ICA-based method, respectively. (a) Source image and four original segments \( s^1_1(t), s^3_1(t), s^4_1(t), \) and four decrypted segments \( \tilde{s}^1_1(t), \tilde{s}^3_1(t), \tilde{s}^4_1(t) \) using the proposed method. (b) Four encrypted segments \( s^1_1(t), s^3_1(t), s^4_1(t) \) and four decrypted segments using similar ciphers \( U \). (d) Four encrypted segments \( \frac{1}{2} \tilde{s}^1_1(t), \frac{1}{2} \tilde{s}^3_1(t), \frac{1}{2} \tilde{s}^4_1(t) \) using the BSS/ICA-based method.](image-url)
that the source images are secure under Attack 1. Also, the
sensitivity of the proposed scheme with respect to the ciphers
is tested. The correlation of the selected ciphers
sensitivity of the proposed scheme with respect to the ciphers
that the source images are secure under Attack 1. Also, the
SIRs (in decibels) of the decrypted images using the ciphers
$U_0$ are $[-6.7835 -6.1115 -6.5532 -6.2334]$, and the
images are shown in Fig. 4(c). Although $U_0$ and $U$ are quite
correlative, it fails to decrypt the plaintexts. Thus, the sensitivity
is verified.

When the cryptosystem proposed in [11] is tested, the re-
results of decryption using Comon’s Algorithm, the fast ICA algo-
rithm, and the Infomax algorithm are similar [16]. Thus, only the
fast ICA algorithm is utilized in this simulation. The encrypted
segments and the decrypted segments are shown in Fig. 4(d). Table
VII shows the indexes SIRs of the encryption and the de-
cription. The low SIRs show that the cryptosystem proposed in
[11] fails for the image signal. Also, Theorem 1 is verified.

D. Discussions

Considering the traditional cryptosystems, the sources can be
decrypted analytically and the ciphers are used with the mul-
time padding method [1], [2]. However, the security relies on
the apparently intractability of the computational problems. It is
still not ensured theoretically. The widely used cryptosystems
have been faced with more and more challenges, since the colli-
sion of MD5 is proclaimed by Wang [32].

Note that the sources are composed of partly distorted sam-
ple in the speech and image communication, and they need
only be decrypted approximately. So, BSS technique may be a
useful tool for decryption [33]–[35]. Recently, a fast and attrac-
tive scheme based on BSS has been applied to speech encryption
[11]. The security of which is based on the difficulty of solving
the underdetermined BSS problem, and it can be ensured theo-
retically. However, the implicit method using one-time padding
costs too much in practice because of the frequent transfor-
mation, and it cannot be extended in to the multime padding
method according to Theorem 2 (see also Simulation 1). The
data in Table II show that the source signal is decrypted effi-
ciently without the cipher when the multime padding method
is used, though it is masked quite well. It can also be verified
in Fig. 2(c) and (d). Furthermore, the plaintext signals are as-
sumed mutually independent in [11]. That is to say, their cryp-
tosystem may fail for dependent signals which exist widely in
reality. Fig. 4(d) and the data in Table VII show that the accu-
racny of the decrypted image is not high, and it is even unaccept-
able for the first segment of the source image according to the
measure in [31], [43]. Also, the correlation levels of these four
segments are still low (see the model (27)), and it seems that
the effect of the decryption may be even worse if the segments are
more correlative.

From the implementation point of view about the proposed
cryptosystem, at the stages of preprocessing and encryption, the
operations are very fast, and the computational load is quite
low. Because of the NMF algorithm, the decryption may be
time-consuming. For alternate least-squares approach for NMF
TABLE VII
SIRs (DB) of Encryption and Decryption for Image Signal Using BSS/ICA-Based Method

<table>
<thead>
<tr>
<th></th>
<th>$x_1(t)$</th>
<th>$x_2(t)$</th>
<th>$x_3(t)$</th>
<th>$x_4(t)$</th>
<th>FastICA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_1(t)$</td>
<td>-89.7577</td>
<td>-83.9823</td>
<td>-85.3231</td>
<td>-55.2292</td>
<td>8.9675</td>
</tr>
<tr>
<td>$s_2(t)$</td>
<td>-83.7709</td>
<td>-77.9985</td>
<td>-79.3412</td>
<td>-49.2766</td>
<td>20.5568</td>
</tr>
<tr>
<td>$s_3(t)$</td>
<td>-81.4183</td>
<td>-75.6339</td>
<td>-76.9857</td>
<td>-46.9439</td>
<td>19.1056</td>
</tr>
<tr>
<td>$s_4(t)$</td>
<td>-82.7279</td>
<td>-76.9526</td>
<td>-78.2918</td>
<td>-48.2395</td>
<td>14.8015</td>
</tr>
</tbody>
</table>

Fig. 6. Number of the nonzero entries in four nested areas according to the optimal seeking method.

utilized in this proposed cryptosystem, we found that the computational complexity is $O(\mathcal{P} \times \mathcal{P} \times \mathcal{T})$ approximately [40], [41]. Therefore, roughly, the computational complexity of the proposed scheme is also $O(\mathcal{P} \times \mathcal{P} \times \mathcal{T})$.

Compared with the cryptosystem proposed in [11], there are mainly two advantages of our cryptosystem: 1) the method of multitime padding makes it be more convenient; 2) the effect of decryption is much better for dependent segments and similar for independent segments. It also denotes that the decrypted signal is the efficient estimation of the source signal, according to the Listening Test in [11]. It is a prior cryptosystem to the traditional DES method, since 1) the length of cipher is dynamic, the key space is nearly infinite, and the security can be ensured theoretically; 2) the processes of encryption and decryption are more simplified.

VI. CONCLUSION

A novel nonlinear cryptosystem using multitime padding is proposed in this paper. The encryption relies on the construction of the nonlinear function that can ensure the security theoretically, and the decryption is equivalent to the NMF that is blind or semiblind. Since the NMF does not rely on the statistical characteristics such as Gaussianity and dependence, a large of sources for communication can be processed efficiently by our cryptosystem. Speech and image sources are tested in Simulation 2 and Simulation 3, respectively. It is very fast, and the computational load is quite low.

Although the local extremum is a well-known trap for the NMF, it can be overcome efficiently by the GPCG algorithm in [28]. And the sparse transformation to the sources may improve the accuracy of the decryption [26]–[28]. Furthermore, the special construction of the linear mixing matrix and the waveform information ($s_{\text{max}}$ and $s_{\text{min}}$) of the original segments can guide the selection of the initial values in the NMF algorithm. However, since the NMF is an iterated algorithm, the speed of the decryption is still a problem to be exploited. Also, considering the cryptosystems based on the blind or semiblind method, the secure is enhanced at the cost of the reasonable decrease of the decryption level, and they may be not useful when the sources are required to be decrypted completely or analytically.

APPENDIX A

PROOF OF THEOREM 2

Since the cipher can be used repeatedly when multitime padding is performed, let three frames $s_i^{m}(t), s_i^{n2}(t), s_i^{n3}(t)$ be encrypted by the same cipher $u(t)$ using the method in [11]. According to the corresponding encryption scheme, the following three mixing segments can be obtained from the ciphertext without the cipher using ICA algorithms [14], [36]:

\[
\begin{align*}
\begin{cases}
y_i^{m}(t) = s_i^{n1}(t) + \beta_{n1} \cdot u(t) \\
y_i^{n2}(t) = s_i^{n2}(t) + \beta_{n2} \cdot u(t) \\
y_i^{n3}(t) = s_i^{n3}(t) + \beta_{n3} \cdot u(t)
\end{cases}
\end{align*}
\]  

or equivalently in a scalar form as

\[
\begin{align*}
\begin{cases}
y_i^{m}(t) = s_i^{n1}(t) + \beta_{n1} u_i(t) \\
y_i^{n2}(t) = s_i^{n2}(t) + \beta_{n2} u_i(t) \\
y_i^{n3}(t) = s_i^{n3}(t) + \beta_{n3} u_i(t)
\end{cases}
\end{align*}
\]  

where $i = 1, \ldots, \mathcal{P}, \beta_{n1}, \beta_{n2}, \beta_{n3} \neq 0$. On the other hand, as $s_i^{n1}(t)$ and $s_i^{n2}(t)$ are sparse and $u_i(t)$ is often not sparse (otherwise, the plaintext can not be well masked), the extremum of the following objective function exists

\[
\min : f(\mu_1) = \|y_i^{m}(t) - \mu_1 y_i^{n2}(t)\|_0
\]  

where $\|0\|_0$-norm $\| \cdot \|_0$ denotes the number of nonzero entries ($\text{nnz}$). In fact, the minimum of $f(\mu_1)$ appears at $\mu_1 = \beta_{n1}/\beta_{n2}$. Note that the model (31) represents a single peak function with
single variable, and \( \mu_1 \) can be obtained using the optimal seeking method [37]. The values \( f(\mu_1) \) (i.e., nonne) in four nested search areas are shown in Fig. 6. Then we get the linear mixing of \( s^{(P_1)}(t) \) and \( s^{(P_2)}(t) \)

\[
w_1(t) = s^{(P_1)}(t) - \mu_1 s^{(P_2)}(t)
\]

where \( \mu_1 = \beta_{P_1}/\beta_{P_2} \), similarly,

\[
w_2(t) = s^{(P_1)}(t) - \mu_2 s^{(P_2)}(t)
\]

where \( \mu_2 = \beta_{P_1}/\beta_{P_2} \). The models (32), (33) can be rewritten as

\[
\begin{bmatrix}
w_1(t) \\
w_2(t)
\end{bmatrix} =
\begin{bmatrix}
1 & -\mu_1 \\
1 & 0
\end{bmatrix}
\begin{bmatrix}
s^{(P_1)}(t) \\
s^{(P_2)}(t)
\end{bmatrix}
\]

where \( i = 1, \ldots, P \).

As sparse signals \( s^{(P_1)}(t), s^{(P_2)}(t), \) and \( s^{(P_3)}(t) \) can be recovered approximately from the model (34), using the methods proposed in [20]–[22], then the plaintext can be decrypted efficiently by segment by segment (or frame by frame) from the ciphertext without the cipher. Note that if \( s^{(P_1)}(t), s^{(P_2)}(t), \) and \( s^{(P_3)}(t) \) are not sparse, according to the model (30), the cipher \( u(t) \) can be obtained directly by Attack 2, then the ambiguity of permutation and scale which does not affect the decryption when the BSS technology is used. By this reason, the BSS/ICA-based cryptosystem using the multtime padding is not secure. Theorem 2 is proved.

**APPENDIX B**

**PROOF OF THEOREM 3**

The scalar form of the model (8) is \( x_i(t) = g_{dt}(y(t), g_{dt}(t), i), i = 1, \ldots, mt = 1, \ldots, T \). And \( x_i(t) \) is a binary function of \( u(t), y(t) \), i.e., \( r = 2 \) in Definition 2. The security of the proposed cryptosystem is proved under the Attack 1–3.

i) Under Attack 1, since the ciphertext \( x^{(t)}(t) \) or \( x^{(P)}(t) \) is the known information only, for \( \forall t, p = 0 < r - 1 \) in the \( P \)-conditional function \( g_{dt} \), the function \( g_{dt} \) is not \( P \)-conditionally invertible. Thus, \( y(t) \) and \( u(t) \) cannot be calculated. Also, \( s^{(t)}(t) \) cannot be recovered directly using nonlinear BSS [38], [39]. As such, the plaintext \( s^{(t)}(t) \) is secure.

ii) Under Attack 2, since the linear mixing matrix \( A^{(t)} \) is nonnegative and can be constructed, it is easy to select the proper matrix \( A^{(t)} \) such that its inverse matrix is not nonnegative. At this time, the NMF is unilateral, i.e., one cannot get \( y^{(t)}(t) \) from \( s^{(t)}(t) \) using the NMF. As a result, the ciphertext \( x^{(t)}(t) \) or \( x^{(P)}(t) \) is still the known useful information only, and the cipher \( u(t) \) is secure.

iii) Under Attack 3, if one wants to obtain \( u(t) \) from \( x^{(t)}(t) \), it is necessary that \( y^{(t)}(t) \) can be calculated through the proper selection of the plaintext \( s^{(t)}(t) \). Since the exact mixing matrix \( A^{(t)} \) is unknown and dynamic in different encryption (though its form is fixed and public), there exists only one possibility to get \( y^{(t)}(t) \) which does not rely on \( A^{(t)} \). It is, the selected \( s^{(t)}(t) \) satisfies \( y^{(t)}(t) = A^{(t)} s^{(t)}(t) = 0 \), according to the public form of \( A^{(t)} \). Because the function \( g_{dt}(u(t), g_{dt}(t)) \) is not \( p(1)- \)conditionally invertible with respect to \( u(t) \) when \( g_{dt}(t) = 0 \), the security of the cipher \( u(t) \) is ensured, even multtime padding way is used. Theorem 3 is proved.


**Shengli Xie** (M’01–SM’02) was born in Hubei Province, China, in 1958. He received the M.S. degree in mathematics from Central China Normal University, Wuhan, China, in 1992, and the Ph.D. degree in control theory and applications from South China University of Technology, Guangzhou, China, in 1997.

He is presently a Full Professor with the South China University of Technology and a vice head of the Institute of Automation and Radio Engineering. His research interests include automatic control and blind signal processing. He is the author or coauthor of two books and more than 70 scientific papers in journals and conference proceedings.

**Zuyuan Yang** was born in Hubei Province, China, in 1982. He is currently working towards the Ph.D. degree with the Intelligent Information and Signal Processing Group (led by Prof. Shengli Xie), South China University of Technology, Guangzhou, China. His research interests include information security, blind signal processing and machine learning.

**Yuli Fu** received the Ph.D. degree from Huazhong University of Science and Technology, Wuhan, China, in 2000. From 2000 to 2002, he was a Postdoctoral Researcher with School of Electronics and Information Engineering, South China University of Technology, Guangzhou China, where he is currently an Full Professor. His research interests include adaptive signal processing and nonlinear dynamical systems.