A Neural Network Learning for Adaptively Extracting Cross-Correlation Features Between Two High-Dimensional Data Streams

Da-Zheng Feng, Member, IEEE, Xian-Da Zhang, Senior Member, IEEE, and Zheng Bao, Senior Member, IEEE

Abstract—This paper proposes a novel cross-correlation neural network (CNN) model for finding the principal singular subspace of a cross-correlation matrix between two high-dimensional data streams. We introduce a novel nonquadratic criterion (NQC) for searching the optimum weights of two linear neural networks (LNN). The NQC exhibits a single global minimum attained if and only if the weight matrices of the left and right neural networks span the left and right principal singular subspace of a cross-correlation matrix, respectively. The other stationary points of the NQC are (unstable) saddle points. We develop an adaptive algorithm based on the NQC for tracking the principal singular subspace of a cross-correlation matrix between two high-dimensional vector sequences. The NQC algorithm provides a fast online learning of the optimum weights for two LNN. The global asymptotic stability of the NQC algorithm is analyzed. The NQC algorithm has several key advantages such as faster convergence, which is illustrated through simulations.

Index Terms—Cross-correlation features, cross-correlation neural network (CNN), global asymptotic stability, learning rate, nonquadratic criterion (NQC), stationary point, SVD.

I. INTRODUCTION

SIGNAL PROCESSING approaches based on singular value decomposition (SVD) to data matrix or correlation matrix are usually robust [1]. Many signal-processing tasks can efficiently be achieved by SVD of a rectangular matrix or cross-correlation matrix. Due to the importance of SVD in signal processing, a variety of iterative methods have been proposed by researchers who are experts in matrix algebra [2]–[5]. These algorithms of updating SVD for tracking subspace can get the exact or approximate SVD of a cross-correlation matrix in a low complexity—$O(N^2)$ per update. Especially, the IV-PAST algorithm given in [32] has a complexity $3Nr + O(r^2)$ and is an efficient singular subspace tracker of the cross-correlation matrix between the output sequence and the instrumental variable. On the other hand, neural networks have provided effective and parallel processing methods for algebraic computations such as the principal component analysis [6]–[9], [15], [23], [34]. Neural networks can also provide an alternative approach to SVD of a rectangular matrix [10]–[14], [33]. By continuation of Oja’s algorithm, Yuile et al. [10] and Samardzija et al. [11] developed several recurrent neural networks, which extract the principal components of the auto-correlation matrix of a random data stream. These neural networks can also obtain SVD of a rectangular matrix, if and only if their weight matrix is taken as $A^T A$ or $AA^T$. However, if the data matrix $A$ is ill-conditioned, then the operation $A^T A$ or $AA^T$ usually is numerically unstable and should be avoided [1]. The gradient flow based on the least squares measure of differential equations for the SVD [13]–[19] is proved to be asymptotically convergent if all the singular values of $A$ are distinct. Diamantaras and Kung [20] proposed the cross-correlation neural network (CNN) models that can be directly used for extracting the cross-correlation features between two high-dimensional data streams. The networks can efficiently extract the principal cross-correlation features between two high-dimensional time series in real time, whereas their deterministic form can directly be used for performing SVD of a rectangular matrix. However, the cross-correlation neural network models are sometimes divergent for some initial states [21].

Both analytical and experimental studies [34] show that convergence of the above neural networks depends on the appropriate selection of the learning rate that is difficult to determine in advance, since the learning rate is directly related to the specific matrix. Although the above neural networks converge to the principal singular subspace (PSS), their convergence speed is slow due to the small learning rate and the slow convergence of simple gradient searching of quadratic criterion. Hence, it is important to find a neural network model with the larger fixed learning rate that can be chosen in advance.

In order to improve the convergence speed of the CNN models [20], we propose a novel CNN model in which the learning rate is independent of the singular value distribution of a cross-correlation matrix. The performance of the CNN is evaluated via computer simulations. The novel CNN has two advantages: 1) the learning rate can be fixed as a constant independent of the singular value distribution of the underlying matrix, which evidently increases the convergence speed of the CNN and 2) its state matrix maintains orthonormality if the initial state matrix is orthonormal. To track the PSS of a cross-correlation matrix, a novel learning algorithm is proposed, and its performance is shown by analysis and simulations.
II. PRELIMINARIES

A. Notations, Properties, Facts, and Acronyms

A capital boldfaced letter is used to denote a matrix. Moreover, a small boldfaced letter denotes a column vector, unless specified otherwise. Some notational symbols are listed below.

- $A^T$: Transpose of a real matrix $A$.
- $\text{tr}(A)$: Trace of $A$.
- $\text{rank}(A)$: Rank of $A$.
- $\ln(A)$: Natural logarithm of a positive definite matrix $A$.
- $||A||_F$: Frobenius norm of $A$.
- $||x||$: Euclidean norm of vector $x$.
- $\text{vec}(A)$: Vector formed by stacking columns of $A$ one beneath the other.
- $A \otimes B$: Kronecker product of two matrices.
- $A > (\geq) B$: Difference matrix $A - B$ is positive (non-negative) definite.
- $I$: Identity matrix.
- $0$: Null matrix or vector.
- $P_r$: $N \times r$ permutation matrix in which each column and each row have exactly one nonzero element that is equal to 1, where $N \geq r$.
- $\Lambda$: An $N \times N$ diagonal matrix given by $\text{diag} (\sigma_1, \cdots, \sigma_N)$.
- $j_i$: An integer number such that the permutation matrix $P_{j_i}$ has exactly one nonzero entry equal to 1 in row $i$ and column $j_i$.
- $\Lambda$: A diagonal matrix $\text{diag} (\sigma_{1j}, \cdots, \sigma_{nj})$ associated with $P_r$ and $\Lambda$.
- $Q_1, Q_2, Q_3, \cdots, r \times r$ orthogonal matrices, unless specified otherwise.
- $K_{pq}$: $pq \times pq$ permutation matrix such that $K_{pq} \text{vec}(A) = \text{vec}(A^T)$.
- $E$: Expectation operator.
- $d, \partial, \nabla, H$: Differential, partial differential, gradient, and Hessian operators (also see Appendix A), respectively.

We use some acronyms in this paper.

CNN: Cross-correlation neural network.
EVD: Eigenvalue decomposition.
LNN: Linear neural networks.
NQC: Nonquadratic criterion.
PSC: Principal singular components.
PSS: Principal singular subspace.
SVD: Singular value decomposition.

**Definition 2.1 [1]:** Given a $r \times r$ matrix $B$, then its EVD is represented as $B = \Phi \Psi \Phi^{-1}$, where $\Phi$ denotes a $r \times r$ matrix formed from all the eigenvectors, and $\Psi = \text{diag}(\lambda_1, \cdots, \lambda_r) > 0$ is a diagonal matrix constructed from all the eigenvalues.

**Property 2.1 [1]:** The determinant of a matrix has the following properties:

1. $\det(AB) = \det(A) \det(B)$;
2. $\det(B) = \prod_{i=1}^{r} \lambda_i$;
3. since $Q^T Q = Q Q^T = I$, $\det(Q) = \pm 1$, where $Q$ is an orthogonal matrix;
4. if $B = \Phi \Psi \Phi^{-1}$, then a matrix function for $B$ is given by $f(B) = \Phi \text{diag}(f(\lambda_1), \cdots, f(\lambda_r)) \Phi^{-1}$ (see Corollary 11.1.2 in [1]), and especially $\ln(B) = \Phi \text{diag}(\ln(\lambda_1), \cdots, \ln(\lambda_r)) \Phi^{-1}$ and $\det(\ln(B)) = \prod_{i=1}^{r} \ln(\lambda_i)$.

**Property 2.2 [1]:** The trace of a matrix $B$ can be computed by $\text{tr}(B) = \sum_{i=1}^{r} \lambda_i$.

**Property 2.3 [1]:** Let two different $\tilde{U}$ and $\tilde{V} \in R^{N \times r}$ for $r \leq N$ satisfy condition $\text{span} (\tilde{U}) = \text{span} (\tilde{V})$, then there always exists a $r \times r$ rank-full matrix $B$ such that $\tilde{U} = \tilde{V} B$.

**Property 2.4:** Given two different matrices $\tilde{U} = [\tilde{u}_1, \cdots, \tilde{u}_r]$ and $\tilde{V} = [\tilde{v}_1, \cdots, \tilde{v}_r] \in R^{N \times r}$, if $N > r$ and $\text{rank}(\tilde{U}^T \tilde{V}) = r$, then $\text{span}(\tilde{U}) = \text{span}(\tilde{V})$.

**Proof:** Obviously, $\text{rank}(\tilde{U}^T \tilde{V}) = r$ and $N > r$ imply that the numbers of the independent column vectors in $\tilde{U}$ and $\tilde{V}$ are equal to $r$. Let $V = [v_1, \cdots, v_r]$ be a column-orthonormal matrix and $\text{span}(V) = \text{span}(V)$, from Property 2.3 we have $\tilde{V} = VB$, where $B = \tilde{U}^T \tilde{V}$ is a full rank matrix.

By reduction to absurdity, we shall show that $\text{span}(\tilde{U}) = \text{span}(\tilde{V})$. Without loss of generality, assume that $\tilde{u}_1 \notin \text{span}(\tilde{V})$, i.e., $\tilde{u}_1 \notin \text{span}(V)$, then we have $\tilde{U} \tilde{V} = 0$. This shows that $\text{rank}(\tilde{U}^T \tilde{V}) \leq r - 1$ and $\text{rank}(\tilde{U}^T \tilde{V}) B \leq r - 1$. Which conflicts with condition $\text{rank}(\tilde{U}^T \tilde{V}) = r$. This completes the proof of Property 2.4.

**Property 2.5:** Known $P_r$, $\Lambda$, and $\Psi$ given in the notation list, then there is $\Lambda P_r = P_r \Lambda$, where $\Lambda = \text{diag}(\sigma_{1j}, \cdots, \sigma_{nj})$.

Property 2.5 is easily verified.

**Fact 2.1 [28]:** From Properties 2.1 and 2.2, we have $\text{tr}(\ln(B)) = \ln(\det(B)) = \sum_{i=1}^{r} \ln(\lambda_i)$.

**Fact 2.2:** Let $P_r \in R^{N \times r}$ ($N \geq r$) be a permutation matrix and $\Psi = \text{diag}(\rho_1, \cdots, \rho_N)$ be a diagonal matrix, then $\Psi P_r (P_r^T \Psi P_r)^{-1} = P_r$.

**Proof:** Define $\Psi = \text{diag}(\rho_1, \cdots, \rho_N)$. By using $P_r $ $P_r$ $P_r$ $P_r$ $P_r$ $P_r$ $P_r$ $P_r$ $P_r$ $P_r$, and Property 2.5, we have $\Psi P_r (P_r^T \Psi P_r)^{-1} = P_r^T \Psi P_r (P_r^T)^{-1} = P_r$

**Fact 2.3:** Given two positive-definite matrices $A \in R^{r \times r}$ and $B \in R^{r \times r}$, if $A < B$ is positive-semidefinite, then we have $\text{tr}(\ln(A)) > \text{tr}(\ln(B))$.

**Proof:** See Appendix B.

B. Some Formulations Relative to PSC and PSS

Consider an $M$-dimensional sequence $x(t)$ and an $N$-dimensional sequence $y(t)$ with sampling number $k$ large enough. Without loss of generality, let $M \geq N$. If $x(t)$ and $y(t)$ are jointly stationary, their cross-correlation matrix [26] can be estimated by

$$C(k) = \frac{1}{k} \sum_{j=1}^{k} x(j)y^T(j) \in R^{M \times N}$$

(1)

and if $x(t)$ and $y(t)$ are jointly nonstationary and slowly time-varying, then their cross-correlation matrix can be estimated by

$$C(k) = \sum_{j=1}^{k} \alpha^{-j} x(j)y^T(j) \in R^{M \times N}$$

(2)
where $0 < \alpha \leq 1$ denotes the forgetting factor which makes the past data samples be less weighted than the recent ones. The exact value for $\alpha$ depends on specific applications. Generally speaking, for slow time-varying system, $\alpha$ is chosen near one, whereas for fast time-varying system, $\alpha$ should be chosen near zero [26].

Let $\sigma_1, I_1$ and $r_i$, $i = 1, \ldots, N$ denote the singular values, the corresponding left and right singular vectors of $C$, respectively. We shall arrange the orthonormal singular vectors $I_1, I_2, \ldots, I_M$ and $r_1, r_2, \ldots, r_N$ such that the corresponding singular values are in nonincreasing order: $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_N \geq 0$. Note that since these left singular vectors $I_{N+1}, I_{N+2}, \ldots, I_M$ are associated with the null subspace of $C$, we shall not consider them. Let $\Lambda = \text{diag}(\sigma_1, \ldots, \sigma_N)$, $L = [I_1, \ldots, I_N]$ and $R = [r_1, \ldots, r_N]$, then the SVD of $C$ is described by $C = \sum_{i=1}^{N} \sigma_i x_i^T = \Lambda R^T$. Usually, all $I_1, I_2, \ldots, I_r$ and $r_1, r_2, \ldots, r_r$ are called the principal singular components (PSC), and $L_0 = [I_1, \ldots, I_r]$ and $R_0 = [r_1, \ldots, r_r]$ are called the left and right singular vector matrix associated with signal [24], [25], respectively, while the corresponding principal singular values can construct a diagonal matrix $\Lambda = \text{diag}(\sigma_1, \ldots, \sigma_r)$, where $r$ denotes the number of the PSC.

An efficient estimation can be achieved by Akaike information criterion [37] based on the distribution of the singular values $\sigma_1, \sigma_2, \ldots, \sigma_N$. In some applications [24], [25], we require only to find a PSS spanned by $I_1, I_2, \ldots, I_r$ or $r_1, r_2, \ldots, r_r$ and given by $L_0 Q_1$ and $R_0 Q_2$.

Remark 2.1: In array processing [24], [25], noise is assumed to be spatially-temporally white, and the larger singular values are associated with signal and the other singular values are associated only with noise. Number $r$ can be chosen such that $\sigma_r > \sigma_{r+1}$ and $\sigma_{r+1} \approx \sigma_{r+2} \approx \cdots \approx \sigma_N$. Moreover, the signal components in array output $y(t)$ and instrumental variable $x(t)$ can be spanned by $L_0 Q_1$ and $R_0 Q_2$, respectively.

Consider the following two linear transformations:

$$
\begin{align}
\mathbf{u}(k) &= U^T x(k) \in R^{r \times 1} \\
\mathbf{v}(k) &= V^T y(k) \in R^{r \times 1}
\end{align}
$$

where $U \in R^{M \times r}$ and $V \in R^{N \times r}$ denote the optimal weight matrices whose columns span the same space as $L_0$ and $R_0$, respectively; $\mathbf{u}(k)$ and $\mathbf{v}(k)$ are the low-dimensional representation of $x(t)$ and $y(t)$, respectively. If $U = L_0$ and $V = R_0$, then $\mathbf{u}(k)$ and $\mathbf{v}(k)$ are principal components of $x(t)$ and $y(t)$.

Conventionally, the PSS is formulated in term of the following optimization framework [20], [21], [32]:

$$
\begin{align}
\max_{U,V} J_{\text{PSS}}(U, V) \\
J_{\text{PSS}}(U, V) &= \text{tr} \left[ E \{ \mathbf{u}(k) \mathbf{v}^T(k) \} \right] = \text{tr}[U^T C V] \\
\text{subject to } U^T U &= I_r \text{ and } V^T V = I_r.
\end{align}
$$

It is worth noting that $J_{\text{PSS}}(U, V)$ is an indefinite objective function with upper and lower bound. An alternative form of (4) is

$$
\begin{align}
\min_{U,V} \tilde{J}_{\text{PSS}}(U, V) \\
\tilde{J}_{\text{PSS}}(U, V) &= -\text{tr} \left[ E \{ \mathbf{u}(k) \mathbf{v}^T(k) \} \right] = -\text{tr}[U^T C V] \\
\text{subject to } U^T U &= I_r \text{ and } V^T V = I_r.
\end{align}
$$

It is easily shown that

$$
J_{\text{PSS}}(U, V) = -\tilde{J}_{\text{PSS}}(-U, V) = -\tilde{J}_{\text{PSS}}(U, -V).
$$

The landscape of $J_{\text{PSS}}(U, V)$ has been shown to have a global maximum at the PSS with all the other stationary points being saddles [21]. Based on this fact, some approximate gradient searching algorithms were derived [20], [21]. Although these algorithms are able to converge to the PSC, their convergence speed is slow due to the small learning rate and the slow convergence of approximate gradient searching of a quadratic criterion.

Remark 2.2: The relation between the PSC and the PSS is given by

$$
U = L_0 Q_1 \text{ and } V = R_0 Q_2.
$$

From the standpoint of vector space [1], (7) only rotates the coordinate system and changes the bases of left and right singular subspace. In fact, since $Q_1$ and $Q_2 \in R^{r \times r}$ are undetermined, $U = L_0 Q_1$ and $V = R_0 Q_2$ are not unique. Like the indeterminacy in blind source separation problem [35], [36], the PSS has the order-arranged indeterminacy associated with the arranged order of the columns in $U$ or $V$. This means that $UP$ spans the same subspace as that spanned by $U$, while $VP$ spans the same subspace as that spanned by $V$, where $P$ is a $r \times r$ permutation matrix. Moreover, the PSS includes the rotation indeterminacy associated with the rotation of $U$ or $V$. Therefore, $UQ$ spans the same subspace as that spanned by $U$, while $VQ$ also spans the same subspace as that spanned by $V$. The above results show that two projection matrices $UU^T$ and $VV^T$ are always unique. The indeterminacy in the PSS makes our analysis difficult. For convenience of analysis, the following two definitions are needed.

Definition 2.2 [35]: Two matrices $\hat{U}$ and $\hat{V} \in R^{N \times r}$ are said to be essentially equal if there always exists a $r \times r$ permutation matrix $P$ such that $\hat{U} = \hat{V}P$. This is denoted by $\hat{U} \hat{=} \hat{V}$. Moreover, two matrices $\hat{U}$ and $\hat{V}$ are said to be essentially unequal if there exists no $r \times r$ permutation matrix $P$ such that $\hat{U} = \hat{V}P$, and is denoted by $\hat{U} \not\hat{=} \hat{V}$.

Definition 2.3: Two column-orthonormal matrices $\hat{U}$ and $\hat{V} \in R^{N \times r}$ are said to be practically equal if there exists an $r \times r$ orthonormal $Q$ such that $\hat{U} = \hat{V}Q$. This is denoted by $\hat{U} \hat{=} \hat{V}$. Moreover, two column-orthonormal matrices $\hat{U}$ and $\hat{V} \in R^{N \times r}$ are said to be practically unequal if there exists no $r \times r$ orthonormal $Q$ such that $\hat{U} = \hat{V}Q$, and is denoted by $\hat{U} \not\hat{=} \hat{V}$.
It is easily shown that Definition 2.3 is an extension of Definition 2.2.

Fact 2.4: Obviously, \((\tilde{U}, \tilde{V}) \equiv (L_0, R_0)\) is a continuous point set \(\{(L_0Q_1, R_0Q_2)|Q_1^TQ_2 > 0\}\) that constructs a \(r\)-dimensional hypersphere, while \((\tilde{U}, \tilde{V}) \equiv (L_0, R_0)\) is a finite point set \(\{(L_\infty P, R_\infty P)\}\) in which a point lies in a corner of hypercube.

Notice that we only consider those points \((U, V)(\equiv (L_0, R_0))\) that satisfy condition \(U^TCV > 0\).

III. NOVEL NONQUADRATIC CRITERION FORMULATION FOR PSS

To improve the cost surface for the PSS and the convergence of gradient searching, we present a novel NQC for the PSS.

Given \(U \in R^{M \times r}\) and \(V \in R^{N \times r}\) in the domain \(\{(U, V)|U^TCV > 0\}\), we propose the following framework for PSS:

\[
\begin{align}
\min_{U, V} J_{NQC}(U, V) & \quad (8a) \\
J_{NQC}(U, V) & = -\text{tr} [\ln(U^TCV)] + \frac{1}{2} \text{tr}[U^TU + V^TV], \quad (8b)
\end{align}
\]

It is worth mentioning that we have given the same criterion as (8) for finding the single principal singular component [33]. By using deflation transformation [20], the multiple principal singular components have been obtained in [33]. Since the single component case of criterion (8) is simple, better ideas and theory were not developed in [33]. Other criterions given in [7], [8]–[12], [15], [22], [23], [28], and [29] were used for tracking the principal components or the principal subspace of autocorrelation matrix of a high-dimensional data stream.

The cost functions defined in [16] and [17] are different from the above criterion, and although the relative algorithms in [13], [14], and [16] and [17] can be applied to perform the SVD of a rectangular matrix, their convergence is slow [33] and depends on the selection of the leaning rate. Moreover, the cost functions in [20] and [21] are also the single component case and can be used to track the multiple component by using deflation transformation, and the corresponding algorithms are of LMS-type [26] and have also the slow convergence.

Fact 3.1: \(J_{NQC}(U, V)\) with lower bound is unbounded as \([U^TU + V^TV] \rightarrow \infty\) and \((U^TCV) \rightarrow 0^t\).

Fact 3.2: If \(U\) and \(V\) are expanded by the left and right singular vector bases into

\[
U = L^T \tilde{U} \quad \text{and} \quad V = R^T \tilde{V}
\]

respectively, then we can find the NQC for the expanded coefficient matrices

\[
\begin{align}
\min_{\tilde{U}, \tilde{V}} J_{NQC}(\tilde{U}, \tilde{V}) & \quad (10a) \\
J_{NQC}(\tilde{U}, \tilde{V}) & = -\text{tr} [\ln(\tilde{U}^T\tilde{V})] + \frac{1}{2} \text{tr}[\tilde{U}^T\tilde{U} + \tilde{V}^T\tilde{V}] \quad (10b)
\end{align}
\]

where \(\tilde{U}\) and \(\tilde{V}\) \(\in R^{N \times r}\) are two expanded coefficient matrices. Clearly, (10) represents an equivalent form of (8).

The landscape of NQC is depicted by the following two theorems. Since the matrix differential method will be used extensively in deriving derivatives, we present some useful facts of the method in Appendix A. Interested readers may refer to [27] for details.

Theorem 3.1: \((U, V)\) is a stationary point of \(J_{NQC}(U, V)\) in the domain \(\{(U, V)|U^TCV > 0\}\) if and only if \(\tilde{U} \equiv L_e\) and \(\tilde{V} \equiv R_e\), where \(L_e \in R^{M \times r}\) and \(R_e \in R^{N \times r}\) consist of the \(r\) left and right singular vectors of \(C\), respectively.

Note that \((U, V) \equiv (L_e, R_e)\) shows a stationary set of \(J_{NQC}(U, V)\).

Considering Fact 3.2, we know that Theorem 3.1 is equivalent to the following Corollary 3.1.

Corollary 3.1: \((U, V)\) is a stationary point of \(J_{NQC}(U, V)\) in the domain \(\{(U, V)|U^T\Lambda \tilde{V} > 0\}\) if and only if \(\tilde{U} \equiv P_r\) and \(\tilde{V} \equiv P_r\), where \(P_r \in R^{N \times r}\) is a permutation matrix and consists of \(r\) eigenvectors of \(\Lambda\), respectively.

Proof: See Appendix B.

Clearly, \((\tilde{U}, \tilde{V}) \equiv (P_r, P_r)\) shows a stationary set of \(J_{NQC}(U, V)\).

Theorem 3.1 establishes the property of all the stationary points of \(J_{NQC}(U, V)\). The next theorem further distinguishes the global minimum point set attained by \((U, V)\) spanning the principal singular subspace from the other stationary points that are saddle (unstable) points.

Theorem 3.2: In the domain \(\{(U, V)|U^TCV > 0\}\), \(J_{NQC}(U, V)\) has a global minimum that is attained when and only when \(U \equiv L_0\) and \(V \equiv R_0\). At a global minimum, \(J_{NQC}(U, V) = \sum_{i=1}^{r} \ln \sigma_i - r\). All the other stationary points \((U, V) \neq (L_0, R_0)\) are saddle (unstable) points of \(J_{NQC}(U, V)\).

In fact, \((U, V) \equiv (L_\infty, R_\infty)\) shows a global minimum point set of \(J_{NQC}(U, V)\).

By Fact 3.2, again, we know that Theorem 3.2 is equivalent to the following Corollary 3.2.

Corollary 3.2: In the domain \(\{(\tilde{U}, \tilde{V})|\tilde{U}^T\Lambda \tilde{V} > 0\}\), \(J_{NQC}(\tilde{U}, \tilde{V})\) has a global minimum that is attained when and only when \(\tilde{U} \equiv P_r\) and \(\tilde{V} \equiv P_r\), where \(P_r \neq (I_r^0), I_r\) is a \(r \times r\) identity matrix. At a global minimum, \(J_{NQC}(\tilde{U}, \tilde{V}) = \sum_{i=1}^{r} \ln \sigma_i - r\). All the other stationary points \((\tilde{U}, \tilde{V}) \neq (P_r, P_r)\) are saddle (unstable) points of \(J_{NQC}(\tilde{U}, \tilde{V})\).

Proof: See Appendix C.

It is easily shown that \((\tilde{U}, \tilde{V}) \equiv (P_r, P_r)\) denotes a global minimum point set of \(J_{NQC}(\tilde{U}, \tilde{V})\).

Some following remarks are given in order.

Remark 3.1: From theorems 3.1 and 3.2, it is seen that the minimization of \(J_{NQC}(U, V)\) orthonormalizes automatically the columns of \(U\) and \(V\). Therefore, we need not impose any explicit constraints on \(U\) and \(V\).

Remark 3.2: At a minimum point of \(J_{NQC}(U, V)\), \(U\) and \(V\) only produce an arbitrary orthonormal base of the principal singular subspace. However, \(UU^T = L_0L_0^T\) and \(VV^T = R_0R_0^T\) are two orthogonal projection onto the principal singular subspace and can be uniquely determined.
Remark 3.3: Similar to the NIC landscape depicted in [23], the NQC has a global minimum set and no local ones. Thus, the iterative algorithms like the gradient descent search algorithm are guaranteed to globally converge to the desired principal subspace for the proper initializations of $\mathbf{U}$ and $\mathbf{V}$ in the domain $\{ \{ \mathbf{U}, \mathbf{V} \} | \mathbf{U}^T \mathbf{C} \mathbf{V} > 0 \}$ (also see Section V). The presence of saddle points does not cause any problem because they can be avoided through random perturbations of $\mathbf{U}$ and $\mathbf{V}$ in practice. Furthermore, it should be noted that the different landscapes of the NQC and PSS determine different convergence speed of their corresponding gradient searching algorithms. Especially, the gradient-searching algorithm of the NQC has a fixed learning rate independent of the specific problem, as will be seen in Section VI.

IV. NQC LEARNING ALGORITHM

At time instant $k$, we have $\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(k)$ and $\mathbf{y}(1), \mathbf{y}(2), \ldots, \mathbf{y}(k)$ available and are interested in estimating $\hat{\mathbf{U}}(k)$ and $\hat{\mathbf{V}}(k)$. Our objective is to establish a fast adaptive algorithm for calculating recursively an estimate of the principal singular subspace at time instant $k$ from the known estimate at $k - 1$ and the newly observed samples $\mathbf{x}(k)$ and $\mathbf{y}(k)$. We will apply the gradient descent searching to the unconstrained minimum of $J_{\text{NQC}}(\mathbf{U}, \mathbf{V})$.

A. Batch Implementation

Given the gradient of $J_{\text{NQC}}(\mathbf{U}, \mathbf{V})$ with respect to $\mathbf{U}$ and $\mathbf{V}$, we have the following gradient descent rule for updating $\mathbf{U}(k)$ and $\mathbf{V}(k)$:

\[
\mathbf{U}(k) = (1 - \eta) \mathbf{U}(k - 1) + \eta \mathbf{C}(k) \mathbf{V}(k) \times (\mathbf{U}^T(k) \mathbf{C}(k) \mathbf{V}(k))^{-1}
\]

\[
\mathbf{V}(k) = (1 - \eta) \mathbf{V}(k - 1) + \eta \mathbf{C}(k)^T \mathbf{U}(k) \times (\mathbf{V}^T(k) \mathbf{C}(k)^T \mathbf{U}(k))^{-1}
\]

\[
\mathbf{C}(k) = \alpha \mathbf{C}(k - 1) + \mathbf{x}(k) \mathbf{y}^T(k)
\]

(11a)

(11b)

(11c)

where $0 < \eta \leq 1$ denotes the learning rate, $0 < \alpha \leq 1$ is the forgetting factor. Equation (11c) is a batch implementation of (2).

The convergence of the algorithm depends greatly on its two adaptive weighting matrices $(\mathbf{U}^T(k) \mathbf{C}(k) \mathbf{V}(k))^{-1}$ and $(\mathbf{V}^T(k) \mathbf{C}(k)^T \mathbf{U}(k))^{-1}$, which, respectively, make the convergence speed of $\mathbf{U}^T(k) \mathbf{U}(k)$ and $\mathbf{V}^T(k) \mathbf{V}(k)$ independent of the cross-correlation matrix $\mathbf{C}(k)$. The batch NQC algorithm (11) has a computational complexity of $2r(M \times N) + O(q^2 M) + O(r^2 N)$ flops per update. The operations involved in (11) are simple matrix addition, multiplication, and inversion, which are easy for the systolic array implementation [26]. This batch implementation, however, is mainly suitable for the adaptive singular subspace estimation and tracking, where the cross-correlation matrix $\mathbf{C}(k)$ is explicitly involved in computations. For the online learning of neural networks, it is expected that the network should learn the principal singular subspace directly from the input data sequences $\mathbf{x}(k)$ and $\mathbf{y}(k)$. Such a learning algorithm is derived below.

B. Recursive Implementation

Note that $\mathbf{U}^T(k - 1) \mathbf{x}(i)$ and $\mathbf{V}^T(k - 1) \mathbf{y}(i)$ are the projection of $\mathbf{x}(i)$ and $\mathbf{y}(i)$ onto the columns of $\hat{\mathbf{U}}(k - 1)$ and $\hat{\mathbf{V}}(k - 1)$, respectively. It is assumed that $\mathbf{U}^T(k - 1) \mathbf{x}(i)$ and $\mathbf{V}^T(k - 1) \mathbf{y}(i)$ can be approximated by $\mathbf{g}(i) = \mathbf{U}^T(i - 1) \mathbf{x}(i)$ and $\mathbf{h}(i) = \mathbf{V}^T(i - 1) \mathbf{y}(i)$ for all $1 \leq i \leq k$. In this way, $\mathbf{g}(i)$ and $\mathbf{h}(i)$ become immediately available at any time instant $i$ given the weight estimates $\hat{\mathbf{U}}(k - 1)$ and $\hat{\mathbf{V}}(k - 1)$ for $i = 1, \ldots, k$. Without this approximation, $\mathbf{U}^T(k - 1) \mathbf{x}(i)$ and $\mathbf{V}^T(k - 1) \mathbf{y}(i)$ would only be available for $i = k$ when $\hat{\mathbf{U}}(k - 1)$ and $\hat{\mathbf{V}}(k - 1)$ have been obtained. Substituting (2) into (11a) and (11b), we have

\[
\mathbf{U}(k) = (1 - \eta) \mathbf{U}(k - 1) + \eta \left( \sum_{i=1}^{k} \alpha^{k-i} \mathbf{x}(i) \mathbf{h}^T(i) \right)^{-1} \times \left( \sum_{i=1}^{k} \alpha^{k-i} \mathbf{g}(i) \mathbf{h}^T(i) \right)
\]

\[
\mathbf{V}(k) = (1 - \eta) \mathbf{V}(k - 1) + \eta \left( \sum_{i=1}^{k} \alpha^{k-i} \mathbf{y}(i) \mathbf{g}^T(i) \right)^{-1} \times \left( \sum_{i=1}^{k} \alpha^{k-i} \mathbf{g}(i) \mathbf{h}^T(i) \right)^T
\]

(12a)

(12b)

Note that we generally expect that the difference between $\mathbf{U}^T(k - 1) \mathbf{x}(i)$ and $\mathbf{U}^T(i - 1) \mathbf{x}(i)$ and difference between $\mathbf{V}^T(k - 1) \mathbf{y}(i)$ and $\mathbf{V}^T(i - 1) \mathbf{y}(i)$ are small if $k \eta$ is small. The similar consideration was used in [15], [29], [32]. The above approximation is called a projection approximation, and the similar projection approximation has been utilized in [23], [29], [30], and [32].

Now, we define $\mathbf{P}(k) = \left( \sum_{i=1}^{k} \alpha^{k-i} \mathbf{g}(i) \mathbf{h}^T(i) \right)^{-1}$, $\hat{\mathbf{U}}(k) = \left( \sum_{i=1}^{k} \alpha^{k-i} \mathbf{x}(i) \mathbf{h}^T(i) \right) \mathbf{P}(k)$, and $\hat{\mathbf{V}}(k) = \left( \sum_{i=1}^{k} \alpha^{k-i} \mathbf{y}(i) \mathbf{g}^T(i) \right) \mathbf{P}(k)$. Applying the matrix inversion lemma [1] to $\mathbf{P}(k)$, we get

\[
\mathbf{P}(k) = \left( \sum_{i=1}^{k} \alpha^{k-i} \mathbf{g}(i) \mathbf{h}^T(i) \right)^{-1}
\]

\[
= \frac{\alpha \mathbf{P}^{-1}(k - 1) + \mathbf{g}(k) \mathbf{h}^T(k)}{\alpha^{-1} \mathbf{P}(k - 1) - \gamma(k) \mathbf{g}(k) \mathbf{h}^T(k)}
\]

(13a)

\[
\hat{\mathbf{g}}(k) = \mathbf{P}(k - 1) \mathbf{g}(k)
\]

(13b)

\[
\hat{\mathbf{h}}(k) = \mathbf{P}^T(k - 1) \mathbf{h}(k)
\]

(13c)

\[
\gamma(k) = \frac{1}{\alpha + \mathbf{h}^T(k) \hat{\mathbf{g}}(k)}
\]

(13d)

With (13), we have

\[
\hat{\mathbf{U}}(k) = \left( \alpha \sum_{i=1}^{k-1} \alpha^{k-1-i} \mathbf{x}(i) \mathbf{h}^T(i) + \mathbf{x}(k) \mathbf{h}^T(k) \right) \times \alpha^{-1} \left[ \mathbf{P}(k - 1) - \gamma(k) \hat{\mathbf{g}}(k) \mathbf{h}^T(k) \right]
\]

\[
= \hat{\mathbf{U}}(k - 1) - \gamma(k) \hat{\mathbf{U}}(k - 1) \mathbf{g}(k) \mathbf{h}^T(k)
\]

\[
+ \alpha^{-1} \mathbf{x}(k) \mathbf{h}^T(k) - \alpha^{-1} \gamma(k) \mathbf{h}^T(k) \mathbf{g}(k) \mathbf{x}(k) \mathbf{h}^T(k)
\]

\[
= \hat{\mathbf{U}}(k - 1) - \gamma(k) \hat{\mathbf{U}}(k - 1) \mathbf{g}(k) \mathbf{h}^T(k) + \gamma(k) \mathbf{x}(k) \mathbf{h}^T(k)
\]

\[
= \hat{\mathbf{U}}(k - 1) + \gamma(k) \left[ \mathbf{x}(k) - \hat{\mathbf{U}}(k - 1) \mathbf{g}(k) \right] \mathbf{h}^T(k)
\]

(14)
Similarly, we have also

\[ \dot{V}(k) = \dot{V}(k-1) + \gamma(k) \left[ y(k) - \dot{V}(k-1)h(k) \right] \hat{g}^T(k), \]

(15)

Summarizing the above derivations, we can describe our recursive implementation of the NQC algorithm as follows.

Initialization:
- \( P(0) = \varepsilon I_r \) (\( \varepsilon \) is a very large positive number)
- \( U(0) = 0 \) and \( \dot{V}(0) = 0 \)
- \( U(0) \) is a random \( M \times r \) matrix with very small Frobenius norm
- \( V(0) \) is a random \( N \times r \) matrix with very small Frobenius norm.

Update equations:

\[ g(k) = U^T(k-1)x(k) \]  
(16a)
\[ h(k) = V^T(k-1)y(k) \]  
(16b)
\[ \dot{g}(k) = P(k-1)g(k) \]  
(16c)
\[ \dot{h}(k) = P(k-1)h(k) \]  
(16d)
\[ \gamma(k) = \frac{1}{\alpha + h^T(k)g(k)} \]  
(16e)
\[ P(k) = \alpha^{-1} \left[ P(k-1) - \gamma(k)g(k) h^T(k) \right] \]  
(16f)
\[ \dot{U}(k) = \dot{U}(k-1) + \gamma(k) \left[ x(k) - \dot{U}(k-1)g(k) \right] \dot{h}^T(k) \]  
(16g)
\[ \dot{V}(k) = \dot{V}(k-1) + \gamma(k) \left[ y(k) - \dot{V}(k-1)h(k) \right] \dot{g}^T(k) \]  
(16h)
\[ U(k) = (1 - \eta)U(k-1) + \eta \dot{U}(k) \]  
(16i)
\[ V(k) = (1 - \eta)V(k-1) + \eta \dot{V}(k) \]  
(16j)

The update equation (16) yields an online learning algorithm for two linear neural networks shown in Fig. 1. \( U^T(k) \) and \( V^T(k) \) denote the weight matrices of these linear neural networks.

The initialization of this data-driven NQC algorithm is similar to that of the standard recursive algorithm [26, p. 809]. \( U(0) = 0 \) and \( \dot{V}(0) = 0 \) are their natural choice. In order to ensure that \( \dot{P}(k) \) is positive–definite, good choices of \( P(0) \) are typically given by \( P(0) = \varepsilon I_r \) (\( \varepsilon \) is a very large positive number), \( U(0) = 0 \) is a very small random \( M \times r \) matrix, and \( V(0) = 0 \) is a very small random \( N \times r \) matrix.

V. GLOBAL CONVERGENCE ANALYSIS

We now study a convergence property of the NQC algorithm by considering the gradient rule (11). If \( x(k) \) and \( y(k) \) are the two jointly stationary processes and the learning rate \( \eta \) is small enough, then the discrete-time difference equation (11) approximates the continuous-time ordinary differential equation (ODE) [6, 7, 15, 23, 28, 29]

\[ \frac{dU(t)}{dt} = CV(t) \left[ U^T(t)CV(t) \right]^{-1} - U(t) \]  
(17a)
\[ \frac{dV(t)}{dt} = C^T U(t) \left[ V^T(t)C^T U(t) \right]^{-1} - V(t). \]  
(17b)

By analyzing the global convergence of (17), we can establish the condition for the global convergence of (11) and that of (16).

By the Lyapunov function approach [31], we need to answer the following questions.

1) Is the dynamical system described by (17) able to globally converge to the principal singular subspace solution?
2) What is the domain of attraction around the equilibrium point attained at the principal singular subspace, or equivalently, what is the initial condition to ensure the global convergence?

It can be shown that \( J_{\text{NQC}}(U, V) \) is a Lyapunov function for the ODE (17). To show this, let us define a region \( \Omega = \{ (U, V) \mid J_{\text{NQC}}(U, V) < \infty \} = \{ (U, V) \mid 0 < (U^T CV) < \infty, \| U \|_F < \infty \text{ and } \| V \|_F < \infty \} \). Within this region, \( J_{\text{NQC}}(U, V) \) is continuous and has a continuous first-order derivative.

**Theorem 5.1:** Given the ODE (17) and \( (U(0), V(0)) \in \Omega \), then \((U(t), V(t))\) converges to a point in set \((U, V) = (L_o, R_s)\) with probability 1, as \( t \rightarrow \infty \).

**Proof:** See Appendix E.

Let us establish an essential result, which can partly illustrate that (17) is fast convergent.

**Lemma 5.1:** Let \( U(t) \) and \( V(t) \) be the solution of the ODE (17) on \( \Omega \). Then for all \( t \in [0, \infty) \), we have

\[ U^T(t)U(t) = I_r + [U^T(0)U(0) - I_r] \exp(-2t), \]  
(18a)
\[ V^T(t)V(t) = I_r + [V^T(0)V(0) - I_r] \exp(-2t). \]  
(18b)

**Proof:** From (17), \( U^T(t)U(t) \) and \( V^T(t)V(t) \) satisfy the following ODEs:

\[ \frac{dU^T(t)U(t)}{dt} = 2I_r - 2U^T(t)U(t) \]  
(19a)
\[ \frac{dV^T(t)V(t)}{dt} = 2I_r - 2V^T(t)V(t). \]  
(19b)

Since they are linear with respect to \( U^T(t)U(t) \) and \( V^T(t)V(t) \), they admit the solution (18).

**Remark 5.1:** From (19), the iteration algorithm for computing \( U^T(k)U(k) \) and \( V^T(k)V(k) \) can be written as

\[ U^T(k)U(k) = (1 - 2\eta)U^T(k-1)U(k-1) + 2\eta I_r; \]  
(20a)
\[ V^T(k)V(k) = (1 - 2\eta)V^T(k-1)V(k-1) + 2\eta I_r. \]  
(20b)
where an appropriate learning rate is $0 < \eta < 0.5$. Since (20) is linear, the learning rate can be taken as a fixed constant near 0.5, for example $\eta = 0.49$.

From Lemma 5.1, we directly deduce the following corollary.

**Corollary 5.1 (Boundedness):** For any bounded initial values $U(0)$ and $V(0)$, then $\|U(t)\|_F$ and $\|V(t)\|_F$ in nonlinear system (17) are bounded.

VI. SIMULATIONS

Three simulations are presented, where $\eta = 0.49$. In Simulation 1, two high-dimensional sequences with an ill-conditioned cross-correlation matrix are used to evaluate the efficiency of this parallel neural network. In Simulation 2, the solution of a harmonic retrieval problem is obtained by neural-network-based principal singular subspace tracking for the cross-correlation matrix. Simulation 3 gives a singular subspace-based direction-of-arrival (DOA) estimation technique. Fifty independent runs in each simulation are performed. In each run, we use the EIV-PAST algorithm in [32], the batch and recursive NQC ones in this paper.

To quantify performance, a measure of deviation from orthonormality is given by

$$\text{dist}(k) = \|U^T(k)U(k) - I\|_F^2.$$  

(21)

**Simulation 1:** Generate randomly the nine-dimensional Gaussian white sequence $y(k)$, whereas $x(k) = Ay(k)$, where $A$ is an ill-conditioned matrix

$$A = [u_0, \ldots, u_8]$$

$$\times \text{diag}(10, 10, 10, 10^{-1}, 10^{-1}, 10^{-3}, 10^{-3}, 10^{-3})$$

$$\times [v_0, \ldots, v_8]^T$$

(22)

where $u_i$ and $v_i$ ($i = 0, \ldots, 8$) are the $i$th components of 11- and 9-dimensional orthogonal discrete cosine basis functions. The above matrix has nine nonzero singular values in which the three distinct singular values $10, 10^{-1}$, and $10^{-3}$ have multiplicity 3, respectively. In theory, the cross-correlation matrix of $y(k)$ and $x(k)$ is ill-conditioned and has the condition number $10^4$ [1]. The principal singular subspace spanned by the first 6 singular components will be tracked.

Data size is 200, i.e., $x(1), \ldots, x(200)$ and $y(1), \ldots, y(200)$. In order to ensure stationarity we have repeated the sequence, periodically, so that the convergence performance of the algorithm can be studied. Each period of the data is called a sweep and contains 200 data samples. Fig. 2 shows the evolution curves of all the measure dist($k$) against iteration number $k$.

**Simulation 2:** In the high-resolution harmonic retrieval, it is expected to estimate accurately closely spaced frequencies using as small a sample size as possible. Consider the more realistic case where the signal $x(k)$ is embedded in an additive stationary white noise $n(k)$ with variance $\sigma^2$, namely

$$x(k) = \sum_{i=1}^{m} a_i \exp(jk\omega_i + \phi_i) + n(k).$$

(23)

Let $K$ be the total number of samples and denote

$$x(k) = [x(k), x(k+1), \ldots, x(k+M-1)]^T$$

$$y(k) = [x(k), x(k+1), \ldots, x(k+N-1)]^T$$

$$a = [a_1 \exp(j\phi_1), \ldots, a_m \exp(j\phi_m)]^T$$

$$S = \begin{bmatrix}
1 & \cdots & 1 \\
\exp(j\omega_1) & \cdots & \exp(j\omega_m) \\
\vdots & \ddots & \vdots \\
\exp(j(M-1)\omega_1) & \cdots & \exp(j(M-1)\omega_m)
\end{bmatrix}$$

$$\hat{S} = \begin{bmatrix}
1 & \cdots & 1 \\
\exp(j\omega_1) & \cdots & \exp(j\omega_m) \\
\vdots & \ddots & \vdots \\
\exp(j(N-1)\omega_1) & \cdots & \exp(j(N-1)\omega_m)
\end{bmatrix}$$

$$D = \text{diag}[\exp(j\omega_1), \ldots, \exp(j\omega_m)].$$

In the noise-free case we have

$$x(k) = SD^ka$$

(25a)

$$y(k) = \hat{S}D^ka$$

(25b)

$S_1$ and $S_2$ are composed of the first $(N-1)$ and the last $(N-1)$ rows of $\hat{S}$, respectively. It is easily shown that

$$S_2 = S_1D.$$

(26)

Let the estimated covariance matrix

$$\hat{C} = \frac{1}{K-N+1} \sum_{k=0}^{K-M} x(k)y^H(k)$$

(27)

where superscript “$H$” denotes complex conjugate transposition. If the principal right singular component matrix of $\hat{C}$ is

![Plot of the deviation from orthonormality for tracking the singular subspace spanned by the first six left singular vectors.](image-url)
The eigenvalues of \( \mathbf{L}_s \) which is gotten by the proposed algorithm, then we have \( \text{span}(\mathbf{S}) = \text{span}(\mathbf{L}_s) \), which implies

\[
\mathbf{S} = \mathbf{L}_s \mathbf{F}
\]

where \( \mathbf{F} \) is an \( m \times m \) transformation matrix. \( \mathbf{U}_{s,1} \) and \( \mathbf{U}_{s,2} \) are composed of the first \( (N - 1) \) and the last \( (N - 1) \) rows of \( \mathbf{L}_s \), we immediately derive

\[
\mathbf{S}_2 = \mathbf{S}_1 \mathbf{D} = \mathbf{U}_{s,2} \mathbf{F} = \mathbf{U}_{s,1} \mathbf{F} \mathbf{D}.
\]

We performed simulation using two sinusoidal signals of very closely spaced frequencies and additive white noise

\[
x(k) = \exp(j(k + 0.513)) + \exp(j(1.1k + 0.714)) + n(k),
\]

where

\[
S_2 = S_1 D = U_{s,2} F = U_{s,1} F D.
\]

\[
(31)
\]

\[
(32)
\]

\[
(29)
\]

\[
(28)
\]

\[
\mathbf{F} \mathbf{D} \mathbf{F}^{-1} = \mathbf{G} = (\mathbf{U}_{s,1}^{T} \mathbf{U}_{s,1})^{-1} \mathbf{U}_{s,1}^{T} \mathbf{U}_{s,2}.
\]

The least squares solution of the previous equation is given by

\[
\mathbf{S}_2 = \mathbf{S}_1 \mathbf{D} = \mathbf{U}_{s,2} \mathbf{F} = \mathbf{U}_{s,1} \mathbf{F} \mathbf{D}.
\]

The least squares solution of the previous equation is given by

\[
\mathbf{F} \mathbf{D} \mathbf{F}^{-1} = \mathbf{G} = (\mathbf{U}_{s,1}^{T} \mathbf{U}_{s,1})^{-1} \mathbf{U}_{s,1}^{T} \mathbf{U}_{s,2}.
\]

\[
(31)
\]

\[
(32)
\]

\[
(28)
\]

\[
(29)
\]

\[
\mathbf{F} \mathbf{D} \mathbf{F}^{-1} = \mathbf{G} = (\mathbf{U}_{s,1}^{T} \mathbf{U}_{s,1})^{-1} \mathbf{U}_{s,1}^{T} \mathbf{U}_{s,2}.
\]

The least squares solution of the previous equation is given by

\[
\mathbf{S}_2 = \mathbf{S}_1 \mathbf{D} = \mathbf{U}_{s,2} \mathbf{F} = \mathbf{U}_{s,1} \mathbf{F} \mathbf{D}.
\]

The least squares solution of the previous equation is given by

\[
\mathbf{F} \mathbf{D} \mathbf{F}^{-1} = \mathbf{G} = (\mathbf{U}_{s,1}^{T} \mathbf{U}_{s,1})^{-1} \mathbf{U}_{s,1}^{T} \mathbf{U}_{s,2}.
\]

The least squares solution of the previous equation is given by

\[
\mathbf{F} \mathbf{D} \mathbf{F}^{-1} = \mathbf{G} = (\mathbf{U}_{s,1}^{T} \mathbf{U}_{s,1})^{-1} \mathbf{U}_{s,1}^{T} \mathbf{U}_{s,2}.
\]

\[
(31)
\]

\[
(32)
\]

\[
(28)
\]

\[
(29)
\]

\[
\mathbf{F} \mathbf{D} \mathbf{F}^{-1} = \mathbf{G} = (\mathbf{U}_{s,1}^{T} \mathbf{U}_{s,1})^{-1} \mathbf{U}_{s,1}^{T} \mathbf{U}_{s,2}.
\]

The least squares solution of the previous equation is given by

\[
\mathbf{F} \mathbf{D} \mathbf{F}^{-1} = \mathbf{G} = (\mathbf{U}_{s,1}^{T} \mathbf{U}_{s,1})^{-1} \mathbf{U}_{s,1}^{T} \mathbf{U}_{s,2}.
\]

The least squares solution of the previous equation is given by

\[
\mathbf{F} \mathbf{D} \mathbf{F}^{-1} = \mathbf{G} = (\mathbf{U}_{s,1}^{T} \mathbf{U}_{s,1})^{-1} \mathbf{U}_{s,1}^{T} \mathbf{U}_{s,2}.
\]
Fig. 4. The recursive-based NQC, the batch NQC and EIV-PAST used as tracking the DOA of signals. (a) Evolution curves of the deviation from orthonormality. (b) Evolution curves for estimating the three DOA. (c) Error evolution curves for estimating the three DOA.

[25] is used to get the DOA estimates. Fig. 4 shows the DOA learning curves of the three algorithms.

From Figs. 2–4, we can see that the proposed NQC algorithm may be emerged as an efficient neural-network-based learning technique for the signal subspace tracking, adaptive high-resolution harmonic retrieval and DOA estimation.

VII. CONCLUSION

This paper proposes a new neural network learning algorithm for adaptive principal singular subspace tracking of a cross-correlation matrix, which can be used for adaptive high-resolution harmonic retrieval and DOA estimation. The dynamical performance of this nonlinear system is studied. The state matrices premultiplied by their transpose are governed by the stable ordinary differential equation independent of the data matrix and globally exponentially converge to the identity matrix in the fixed decay rate equal to 2. Simulation results show that the time-step length or leaning rate in the iterative algorithm associated with this neural network is independent of the data matrix.

The NQC algorithm has also emerged as a useful neural-network-based technique when the data matrix is close to a singular matrix. The global analysis shows that the stable equilibrium points associated with the principal singular subspace have the large domain of attraction. The NQC algorithm may be useful in real-time signal processing applications.

APPENDIX A

SOME FACTS OF THE MATRIX DIFFERENTIAL METHOD

We briefly explain the procedure of using the matrix differential method to compute the derivative of a function of matrix. Since the computation of differentials is relatively easy, the computation of derivatives can be performed simply based on the following lemma [27].

Lemma A.1: Let $\Phi(X)$ is a twice-differentiable real-valued function of an $\eta \times \eta$ matrix $X$. Then, the following relationships hold

$$d\Phi(X) = \text{tr}(X^T dX) \Leftrightarrow \nabla \Phi(X) = A$$  \hfill (A.1)
\[ d^2 \Phi(X) = \text{tr} (B (dX)^T C dX) = H \Phi(X) \]
\[ = \frac{1}{2} (B^T \otimes C + B \otimes C^T) \quad (A.2) \]
\[ d^2 \Phi(X) = \text{tr} (B (dX)^T C dX) = H \Phi(X) \]
\[ = \frac{1}{2} K_{tr} (B^T \otimes C + C^T \otimes B) \quad (A.3) \]

where \( d \) denotes the differential, \( A, B, \) and \( C \) are matrices, each of which may be a function of \( X \). The gradient of \( \Phi(X) \) with respect to \( X \) and the Hessian matrix of \( \Phi(X) \) are defined, respectively, as
\[ \nabla \Phi(X) = \frac{\partial \Phi(X)}{\partial X} \quad (A.4) \]
and
\[ H \Phi(X) = \frac{\partial}{\partial (\text{vec}X)^T} \left( \frac{\partial \Phi(X)}{\partial (\text{vec}X)^T} \right)^T \quad (A.5) \]

Based on this lemma, we are ready to derive the gradient and Hessian matrix of \( J_{\text{NQC}}(U, V) \). From (7) and after some calculations, we have
\[ d J_{\text{NQC}}(U, V) = \text{tr} \left\{ \left[ (U^T CV)^{-1} U^T C dV + (V^T C U)^{-1} V^T C dU \right] + \frac{1}{2} (U^T dU + V^T dV) \right\} \quad (A.6) \]

Define a matrix \( X = [U^T \ V^T]^T \), then (A.6) can be rewritten as
\[ d J_{\text{NQC}}(U, V) = \text{tr} \left\{ \left[ [(U^T CV)^{-1} U^T C dV + (V^T C U)^{-1} V^T C dU \times (U^T C V)^{-1} U^T C + U^T V^T] \right] [dU^T \ dV^T]^T \right\} \quad (A.7) \]

Applying (A.1) to (A.7), the gradient of \( J_{\text{NQC}}(U, V) \) with respect to \( U \) and \( V \) is given by
\[ \nabla_U J_{\text{NQC}}(U, V) = -CV(U^T CV)^{-1} + U \quad (A.8a) \]
\[ \nabla_V J_{\text{NQC}}(U, V) = -CU(V^T CV)^{-1} + V \quad (A.8b) \]

From (A.6), we calculate the second-order differential
\[ d^2 J_{\text{NQC}}(U, V) = \text{tr} \left\{ \left[ \frac{U}{2} (U^T CV)^{-1} U^T C dV + \frac{V}{2} (V^T C U)^{-1} V^T C dU \times (U^T C V)^{-1} U^T C + U^T V^T \right] dU^T dV^T \right\} \]
\[ + \text{tr} \left\{ \left[ \frac{U}{2} (U^T CV)^{-1} (dU^T CV + U^T C dV) \times (U^T C V)^{-1} U^T C + U^T V^T \right] dU^T dV^T \right\} \]
\[ \times (U^T C V)^{-1} U^T C + U^T V^T \}
\[ \times \left[ \frac{V}{2} (V^T C U)^{-1} (dV^T C U + V^T C dU) \times (V^T C U)^{-1} V^T C + V^T U^T \right] dU^T dV^T \] \quad (A.9)

The Hessian matrix follows directly from the superposition principle on (A.2), (A.3) and (A.9) and is a high-dimensional complicated matrix that is hard analyzed.

**APPENDIX B**

**PROOF OF FACT 2.3**

From Fact 2.1 we have
\[ \text{tr}(\ln A) - \text{tr}(\ln B) = \ln \left( \frac{\text{det}(A)}{\text{det}(B)} \right) \quad (B.1) \]

We require only proving \( \text{det}(A)/\text{det}(B) > 1 \). It is easily shown from Property 2.1 that
\[ \frac{\text{det}(A)}{\text{det}(B)} = \text{det}(AB^{-1}) \quad (B.2) \]

Let \( \Phi = A - B \), then we have
\[ AB^{-1} = I + \Phi B^{-1} \quad (B.3) \]

Since \( \Phi \) is positive–semidefinite, \( \Phi B^{-1} \) is also positive–semidefinite and its EVD is
\[ \Phi B^{-1} = \Psi(I + \Theta)\Psi^{-1} \quad (B.4) \]

where \( \Theta \) is a positive–semidefinite diagonal matrix. Substituting (B.4) into (B.3) yields
\[ AB^{-1} = \Psi(I + \Theta)\Psi^{-1} \quad (B.5) \]

Since all the elements of the diagonal matrix \( I + \Theta \) are larger than or equal to 1, we have
\[ \text{det}(AB^{-1}) = \text{det}(I + \Theta) > 1 \quad (B.6) \]

This completes the proof of Fact 2.3.

**APPENDIX C**

**PROOF OF COROLLARY 3.1**

Since \( \tilde{U}^T \tilde{A} \tilde{V} \) is positive–definite, it is invertible. The gradient of \( J_{\text{NQC}}(\tilde{U}, \tilde{V}) \) with respect to \( \tilde{U} \) and \( \tilde{V} \) exists and is given by (see Appendix A)
\[ \nabla_{\tilde{U}} J_{\text{NQC}}(\tilde{U}, \tilde{V}) = -\tilde{A} \tilde{V} (\tilde{U}^T \tilde{A} \tilde{V})^{-1} + \tilde{U} \quad (C.1a) \]
\[ \nabla_{\tilde{V}} J_{\text{NQC}}(\tilde{U}, \tilde{V}) = -\tilde{A} \tilde{V} (\tilde{U}^T \tilde{A} \tilde{V})^{-1} + \tilde{V} \quad (C.1b) \]

Given a point \((P_r, Q_1, P_r, Q_2)\) in \((\tilde{U}, \tilde{V}) \equiv (P_r, P_r)\) and considering Fact 2.2, then we have
\[ \nabla_{\tilde{U}} J_{\text{NQC}}(P_r, Q_1, P_r, Q_2) = -\Lambda P_r Q_2 (P_r^T \Lambda P_r)^{-1} Q_1 + P_r Q_1 \]
\[ = -\Lambda P_r Q_2 Q_2^T (P_r^T \Lambda P_r)^{-1} Q_1 + P_r Q_1 \]
\[ = -\Lambda P_r (P_r^T \Lambda P_r)^{-1} Q_1 + P_r Q_1 = 0 \quad (C.2) \]
Similarly, we can get the following equation:

\[ \nabla_{\hat{\mathbf{V}}} J_{\text{NQC}}(\hat{\mathbf{P}}_r, \mathbf{Q}_2, \mathbf{P}_r \mathbf{Q}_2) = 0. \]  
(C.3)

Conversely, \( \hat{J}_{\text{NQC}}(\mathbf{U}, \mathbf{V}) \) at a stationary point should satisfy
\[ \nabla_{\mathbf{U}} \hat{J}_{\text{NQC}}(\mathbf{U}, \mathbf{V}) = 0 \text{ and } \nabla_{\mathbf{V}} \hat{J}_{\text{NQC}}(\mathbf{U}, \mathbf{V}) = 0, \]
which yields

\[ \mathbf{\Lambda} \hat{\mathbf{V}} = \hat{\mathbf{U}} (\hat{\mathbf{U}}^T \mathbf{\Lambda} \hat{\mathbf{V}}) \]  
(C.4a)

\[ \mathbf{\Lambda} \hat{\mathbf{U}} = \hat{\mathbf{V}} (\hat{\mathbf{V}}^T \mathbf{\Lambda} \hat{\mathbf{U}}). \]  
(C.4b)

Premultiplying both sizes of (C.4a) and (C.4b) by \( \hat{\mathbf{U}}^T \) and \( \hat{\mathbf{V}}^T \), respectively, we obtain

\[ \hat{\mathbf{U}}^T \hat{\mathbf{U}} = \mathbf{I}_r \]  
(C.5a)

\[ \hat{\mathbf{V}}^T \hat{\mathbf{V}} = \mathbf{I}_r \]  
(C.5b)

which implies that the columns of \( \hat{\mathbf{U}} \) and \( \hat{\mathbf{V}} \in \mathbb{R}^{N \times r} \) are column-orthonormal at a stationary point of \( \hat{J}_{\text{NQC}}(\hat{\mathbf{U}}, \hat{\mathbf{V}}) \). Obviously, from (C.5a) we have

\[ \text{rank}(\hat{\mathbf{U}}^T \mathbf{\Lambda} \hat{\mathbf{U}}) = r. \]  
(C.6)

Moreover, premultiplying both sizes of (C.4b) by \( \hat{\mathbf{U}}^T \), we can get the following relation

\[ \text{rank} \left( \hat{\mathbf{U}}^T \hat{\mathbf{V}} (\hat{\mathbf{V}}^T \mathbf{\Lambda} \hat{\mathbf{U}}) \right) = \text{rank}(\hat{\mathbf{U}}^T \mathbf{\Lambda} \hat{\mathbf{U}}) = r \]  
(C.7)

which shows

\[ \text{rank}(\hat{\mathbf{U}}^T \hat{\mathbf{V}}) = \text{rank}(\hat{\mathbf{V}}^T \mathbf{\Lambda} \hat{\mathbf{U}}) = r. \]  
(C.8)

From Property 2.4 and (C.8), we deduce

\[ \text{span}(\hat{\mathbf{U}}) = \text{span}(\hat{\mathbf{V}}) = \text{span}(\mathbf{\Lambda} \hat{\mathbf{U}}) = \text{span}(\mathbf{\Lambda} \hat{\mathbf{V}}) \]  
(C.9)

From Property 2.3, we can obtain that

\[ \hat{\mathbf{U}} = \hat{\mathbf{V}} \mathbf{Q}, \text{ i.e. } \hat{\mathbf{U}} \equiv \hat{\mathbf{V}}. \]  
(C.10)

Substituting (C.10) into (C.4a) and noticing \( \hat{\mathbf{U}} \hat{\mathbf{U}}^T = \hat{\mathbf{V}} \hat{\mathbf{V}}^T \), it can be obtained that

\[ \mathbf{\Lambda} \hat{\mathbf{V}} = \hat{\mathbf{V}} \hat{\mathbf{V}}^T \mathbf{\Lambda} \hat{\mathbf{V}}. \]  
(C.11)

Let \( \hat{\mathbf{V}} = [\hat{\psi}_1^T, \ldots, \hat{\psi}_N^T]^T \), where \( \hat{\psi}_i \) \((i = 1, \ldots, N)\) is a row vector, and take \( \mathbf{B} = \hat{\mathbf{V}} \hat{\mathbf{V}}^T \mathbf{\Lambda} \mathbf{V} \) that is a \( r \times r \) symmetric positive–definite matrix, then the alternative form of (C.11) is

\[ \sigma_i \hat{\psi}_i = \hat{\psi}_i \mathbf{B} \quad (i = 1, \ldots, N), \]  
(C.12)

Obviously, (C.12) shows the EVD for \( \mathbf{B} \). Since \( \mathbf{B} \) is a \( r \times r \) symmetric positive–definite matrix, it has only the \( r \) nonzero orthonormal left eigenvectors, which shows that \( \hat{\mathbf{V}} \) has only the \( r \) nonzero orthonormal row vectors. Moreover, all the \( r \) nonzero row vectors in \( \hat{\mathbf{V}} \) form an orthonormal matrix, which shows that \( \hat{\mathbf{V}} \) can always be represented as

\[ \hat{\mathbf{V}} = \hat{\mathbf{P}}_r \mathbf{Q}_2 \]  
(C.13a)

i.e.,

\[ \hat{\mathbf{V}} \equiv \hat{\mathbf{P}}_r. \]  
(C.13b)

Similarly, we can get

\[ \hat{\mathbf{U}} \equiv \hat{\mathbf{P}}_r. \]  
(C.14)

Since in the domain \( \{ (\hat{\mathbf{U}}, \hat{\mathbf{V}}) | \hat{\mathbf{U}}^T \mathbf{\Lambda} \hat{\mathbf{V}} > 0 \} \),

\[ \text{rank}(\hat{\mathbf{U}}^T \mathbf{\Lambda} \hat{\mathbf{V}}) = \text{rank} \left( \mathbf{Q}_r^T \hat{\mathbf{P}}_r^T \hat{\mathbf{P}}_r \mathbf{Q}_2 \right) = r, \]  
(C.15)

from Property 2.5 we have

\[ \text{rank} \left( \hat{\mathbf{P}}_r^T \hat{\mathbf{P}}_r \right) = \text{rank} \left( \hat{\mathbf{P}}_r^T \hat{\mathbf{P}}_r \mathbf{\Lambda} \right) = r \]  
(C.16)

where \( r \times r \) diagonal matrix \( \mathbf{\Lambda} \) is similar to \( \hat{\mathbf{\Lambda}} \) in Property 2.5. This shows that

\[ \text{rank} \left( \hat{\mathbf{P}}_r^T \hat{\mathbf{P}}_r \right) = r \]  
(C.17)

or equivalently

\[ \hat{\mathbf{P}}_r \equiv \hat{\mathbf{P}}_r. \]  
(C.18)

Thus, (C.13b) and (C.14) can always be rewritten as

\[ (\hat{\mathbf{U}}, \hat{\mathbf{V}}) = (\mathbf{P}_r, \mathbf{P}_r) \]  
(C.19)

This completes the proof of Corollary 3.1.
a global minimum of $\tilde{J}_{\text{NQC}}(\tilde{U}, \tilde{V})$ is attained when and only when

$$\tilde{U}, \tilde{V} \in \{ (PQ_1, PQ_2) | Q_2^T \Lambda Q_2 > 0 \}.$$  \hfill (D.1)

From $Q_2^T \Lambda Q_2 > 0$, we directly deduce

$$\det(Q_1) = \det(Q_2) = \pm 1.$$  \hfill (D.2)

By substituting (D.1) into (10b) and performing some algebra operations, we can get the global minimum of $\tilde{J}_{\text{NQC}}(U, V)$ as follows

$$\tilde{J}_{\text{NQC}}(U, V) = - \sum_{i=1}^{r} \ln \sigma_i + r.$$  \hfill (D.3)

Note that if $Q_1^T \Lambda Q_2 \leq 0$, $\tilde{J}_{\text{NQC}}(PQ_1, PQ_2)$ is complex or infinite. Thus, we do not consider the case of $Q_1^T \Lambda Q_2 \leq 0$. Moreover, we can determine a stationary point of $\tilde{J}_{\text{NQC}}(U, V)$ being saddle (unstable) in such a way that within an infinitesimal neighborhood near the stationary point, there is a point $(U', V')$ such that its value $\tilde{J}_{\text{NQC}}(U', V')$ is less than $\tilde{J}_{\text{NQC}}(U, V)$.

Let $P_r \neq \bar{P}$, then there, at least, exists a nonzero element in the row vectors from $r + 1$ to $N$ for $P_r$. From Property 2.5 it is easily known that

$$P_r^T \Lambda P_r - P_r^T \Lambda P_r = \bar{\Lambda} - \bar{\Lambda}$$  \hfill (D.4)

where diagonal matrix $(\bar{\Lambda} - \bar{\Lambda})$, at least, has a positive diagonal element for $P_r \neq \bar{P}$. Then from Property 2.4 we have

$$\ln \left( \det(P_r^T \Lambda P_r) \right) > \ln \left( \det(P_r^T \Lambda P_r) \right).$$  \hfill (D.5)

Since

$$\tilde{J}_{\text{NQC}}(P, Q_1, P, Q_2) = - \ln \left( \det(Q_1^T P_r^T \Lambda P_r Q_2) \right) + r = - \ln \left( \det(P_r^T \Lambda P_r) \right) + r$$  \hfill (D.6)

$$\tilde{J}_{\text{NQC}}(\bar{P}, Q_1, \bar{P}, Q_2) = - \ln \left( \det(Q_1^T \bar{P}_r^T \Lambda Q_2) \right) + r = - \ln \left( \det(P_r^T \Lambda P_r) \right) + r$$  \hfill (D.7)

we get

$$\tilde{J}_{\text{NQC}}(P, Q_1, P, Q_2) > \tilde{J}_{\text{NQC}}(\bar{P}, Q_1, \bar{P}, Q_2)$$  \hfill (D.8)

which shows that set \(\{ (P, Q_1, P, Q_2) | Q_2^T \Lambda Q_2 > 0, \) and $P_r \neq \bar{P}$\} is not a global minimum point set.

Since $P_r \neq \bar{P}$, we can always select a column $p_i (1 \leq i \leq r)$ from $P = [p_1, \cdots, p_r]$ such that

$$p_i^T P_r = 0.$$  \hfill (D.9)

otherwise $p_r \neq \bar{P}$. Moreover, we can always select a column $p_{r,j} (1 \leq i \leq r)$ from $P_r = [p_{r,1}, \cdots, p_{r,r}]$ such that

$$p_{r,j}^T P_r \neq 0.$$  \hfill (D.10)

otherwise $P_r \neq \bar{P}$. Let $p_i$ have only the nonzero element in row $j_i$ and $p_{r,j}$ have only the nonzero element in row $j_r$. Obviously, $j_i < j_r$ and $\sigma_{j_i} < \sigma_{j_r}$, otherwise $P_r \neq \bar{P}$. Define an orthonormal matrix as $B = [p_{r,1}, \cdots, (p_{r,i} + \varepsilon p_{r,i})/\sqrt{1 + \varepsilon^2}, \cdots, p_{r,r-r}]$, where $\varepsilon$ is a positive infinitesimal, from Property 2.5 we know that

$$\Lambda B = \left[ \begin{array}{c} \sigma_{j_1} p_{r,1} + \varepsilon \sigma_{j_1} p_{r,r} \\ \vdots \\ \sigma_{j_r} p_{r,1} + \varepsilon \sigma_{j_r} p_{r,r} \end{array} \right].$$  \hfill (D.11)

Combining (D.9), (D.10), and (D.11), we have

$$B^T \Lambda B = \text{diag} \left[ \sigma_{j_1}, \cdots, \frac{(\sigma_{j_i} + \varepsilon \sigma_{j_i})}{\sqrt{1 + \varepsilon^2}}, \cdots, \sigma_{j_r} \right].$$  \hfill (D.12)

On the other hand, from Property 2.5, we can deduce

$$P_r^T \Lambda P_r = \text{diag} \left[ \sigma_{j_1}, \cdots, \sigma_{j_i}, \cdots, \sigma_{j_r} \right].$$  \hfill (D.13)

Let

$$B^T \Lambda B - P_r^T \Lambda P_r$$

$$= \text{diag} \left[ \sigma_{j_1}, \cdots, \frac{(\sigma_{j_i} + \varepsilon \sigma_{j_i})}{\sqrt{1 + \varepsilon^2}}, \cdots, \sigma_{j_r} \right]$$

$$- \text{diag} \left[ \sigma_{j_1}, \cdots, \sigma_{j_i}, \cdots, \sigma_{j_r} \right]$$

$$= \text{diag} \left[ 0, \cdots, 0, \frac{-\sigma_{j_i}(\sqrt{1 + \varepsilon^2} - 1) + \varepsilon \sigma_{j_i}}{\sqrt{1 + \varepsilon^2}}, 0, \cdots, 0 \right].$$  \hfill (D.14)

Since $(\sqrt{1 + \varepsilon^2} - 1)$ is the second-order infinitesimal of $\varepsilon$ and $\sigma_{j_i} < \sigma_{j_r}$, $B^T \Lambda B - P_r^T \Lambda P_r$ is a nonzero positive–semidefinite matrix. Thus, we deduce

$$\tilde{J}_{\text{NQC}}(BQ_1, BQ_2) = - \ln \left( \det(Q_1^T B^T \Lambda B Q_2) \right) + r$$

$$< \tilde{J}_{\text{NQC}}(P, Q_1, P, Q_2)$$

$$= - \ln \left( \det(Q_1^T P_r^T \Lambda P_r Q_2) \right) + r.$$  \hfill (D.15)

This shows that $\{ (P, Q_1, P, Q_2) | Q_2^T \Lambda Q_2 > 0 \) and $P_r \neq \bar{P}$\} is a saddle (unstable) point set. This completes the proof of Corollary 3.2.
PROOF OF THEOREM 5.1

By using the chain rule, we have
\[
\frac{dJ_{\text{NQC}}(U, V)}{dt} = \text{tr} \left[ \nabla_U J_{\text{NQC}}(U, V) \cdot \frac{dU^T(t)}{dt} + \nabla_V J_{\text{NQC}}(U, V) \cdot \frac{dV^T(t)}{dt} \right].
\]  

(E.1)

Substituting (C.1) and (17) into (E.1) yields
\[
\frac{dJ_{\text{NQC}}(U, V)}{dt} = -\text{tr} \left[ \frac{dU(t)}{dt} \cdot \frac{dU^T(t)}{dt} + \frac{dV(t)}{dt} \cdot \frac{dV^T(t)}{dt} \right]
= \begin{cases} 
0 & \text{for } \frac{dU(t)}{dt} \not= 0 \text{ or } \frac{dV(t)}{dt} \not= 0 \\
0 & \text{for } \frac{dU(t)}{dt} = 0 \text{ and } \frac{dV(t)}{dt} = 0.
\end{cases}
\]

(E.2)

This shows that \( J_{\text{NQC}}(U, V) \) is a Lyapunov function, and \( (dU(t)/dt) \to 0 \) and \( (dV(t)/dt) \to 0 \), i.e., \( \nabla_U J_{\text{NQC}}(U(t), V(t)) \to 0 \) and \( \nabla_V J_{\text{NQC}}(U(t), V(t)) \to 0 \) as \( t \to \infty \). Thus, \( (U(t), V(t)) \) globally converges to a stationary point of \( J_{\text{NQC}}(U, V) \). Moreover, \( J_{\text{NQC}}(U, V) \) attains a uniquely global minimum in set \( \{ (Q_1, R_1, Q_2) | Q_1^T \Lambda Q_2 > 0 \} \) (see Theorem 3.2). Since all the other stationary points of \( J_{\text{NQC}}(U, V) \) are saddle and, thus, are unstable, we conclude that \( (U(t), V(t)) \) converges to a point in \( \{ (Q_1, R_1, Q_2) | Q_1^T \Lambda Q_2 > 0 \} \) with probability 1, as \( t \to \infty \). This completes the proof of Theorem 5.1.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their constructive comments and suggestions for improving the manuscript.

REFERENCES

Da-Zheng Feng (M’02) was born in Shaxxi, China, in 1959. He received the B.S. degree in water-electrical engineering from Xi’an Science and Technology University, Xi’an, China, in 1982, the M.S. degree in electrical engineering from Xi’an Jiaotong University, Xi’an, China, in 1986, and the Ph.D. degree in electronic Engineering from Xidian University, Xi’an, China, in 1996.

From 1986 to 1992, he was with Northwest Polytechnology University, X’an, China, where he was a Lecturer. From 1996 to 1998, he was a Postdoctoral Researcher and Associate Professor at the Department of Mechatronic Engineering, Xi’an Jiaotong University. He is currently a Professor at Xidian University. He has published more than 50 journal papers. His current research interests are signal processing, intelligent signal processing, artificial brain, and InSAR.

Xian-Da Zhang (SM’93) was born in Jiangxi, China, in 1946. He received the B.S. degree in radar engineering from Xidian University, Xi’an, China, in 1969, the M.S. degree in instrument engineering from Harbin Institute of Technology, China, in 1982, and the Ph.D. degree in electrical engineering from Tohoku University, Sendai, Japan, in 1987.

From 1990 to 1991, he was a Postdoctoral Researcher with the Department of Electrical and Computer Engineering, University of California at San Diego. In 1992, he joined Tsinghua University, Beijing, China, where he was a Professor. From 1999 to 2002, he was with the Key Laboratory for Radar Signal Processing, Xidian University, Xi’an, China, as a Specially Appointed Professor awarded by the Minister of Education of China and the Cheung Kong Scholars Program. He is currently a Professor with Department of Automation, Tsinghua University, Beijing. His current research interests are signal processing and intelligent signal processing with applications in radar and communications. He has published over 20 papers in several IEEE Transactions, and is the author of five books (all in Chinese). He holds four patents for invention.

Zheng Bao (M’80–SM’90) received the B.S. degree in radar engineering from Xidian University, Xi’an, China, in 1953.

He is currently a Professor at Xidian University. He has been working in a number of areas, including radar system, signal processing, neural networks, and automatic target recognition. He has published more than 200 journal papers and is the author of ten books. His current research interests include radar signal processing, array signal processing, and SAR/ISAR imaging.

Prof. Bao is a member of the Chinese Academy of Sciences.