Study of Sparsity Emanating from NKPD and its Utilization to Enhance NKPD based Adaptive Algorithms

Sankha Subhra Bhattacharjee, Mads Græsbøll Christensen
Audio Analysis Lab, CREATE, Aalborg University,
Denmark, {ssbh; mgc}@create.aau.dk

Jacob Benesty
INRS-EMT, University of Quebec,
Montreal, Canada, {jacob.benesty}@inrs.ca

***Abstract*—Recently, the nearest Kronecker product (NKP) decomposition has become popular in several adaptive filtering (AF) applications owing to its fast convergence and tracking ability. In this paper, we study the nature of the smaller weight vectors resulting from NKP decomposition (NKPD) of a wide range of acoustic impulse responses (IRs). The study shows that the smaller weight vectors resulting from NKPD exhibit moderate to high degree of sparsity. To exploit this knowledge in AF problems, we propose a class of proportionate update based NKP normalized least-mean-square (NKP-NLMS) type algorithms: namely, the improved proportionate NKP-NLMS (NKP-IPNLMS) algorithm which uses the ℓ_1-norm of the smaller weight vectors and the NKP-IPNLMS-ℓ_0 which uses an approximation of the ℓ_0-norm. Further, we propose a new approximation of the ℓ_0-norm with reduced computational complexity, using which we also propose the NKP-IPNLMS-ℓ_0-2 algorithm. Next, we present a comparison of computational complexity of the proposed algorithms. Simulation results show the improved performance achieved by the proposed algorithms, showing the advantage of exploiting sparsity in the smaller weight vectors in NKPD based adaptive algorithms.**

***Index Terms*—Sparsity, nearest Kronecker product, Adaptive filter, Proportionate algorithms, System identification.**

I. Introduction

Today adaptive signal processing (ASP) has found applications in numerous scientific avenues. Several adaptive algorithms have been proposed for ASP, among which the most widely used due to their low computational complexity and simple implementation, are the least-mean-square (LMS) & normalized LMS (NLMS) algorithm and their variants [1]–[3]. One of the problems frequently studied in ASP is the adaptive modelling of an unknown system. The LMS & NLMS algorithm and their variants have been extensively studied for this problem. In practical applications such as network and acoustic echo cancellation, the underlying system can have long impulse responses (IRs). Moreover, the underlying system may be time variant, and hence the adaptive algorithm is expected to model such changes as quickly as possible. Modelling such long IRs using AFs pose challenges such as slow convergence and slow tracking performance [4], [5].

It is well known that the convergence speed of an adaptive algorithm is inversely proportional to the number of unknown coefficients it has to model, i.e., the length of the adaptive weight vector [6]. Recently the technique of nearest Kronecker product (NKP) decomposition was introduced into the framework of adaptive filters [4], [5], [7], [8]. The NKP decomposition (NKPD) breaks down the problem of modelling a system with a long IR, into smaller problems of identifying smaller weight vectors, thus reducing the number of adaptive parameters, and therefore improving the convergence and tracking speed [4], [8]. This property has inspired the study of NKPD based iterative Wiener filter [4], [9], RLS filter [5], [9], [10], Kalman filter [7], LMS and NLMS filters [8], including applications such as, feedback cancellation in hearing aids [11], modelling nonlinear systems [12], [13], active noise control [13], microphone beamforming [14].

In this paper, first we take a deeper look into the optimal smaller weight vectors following the NKPD of an IR (for example, IR of the unknown system to be modelled) and show that the smaller optimal weight vectors contain moderate to high degree of sparsity, especially one of the two smaller optimal weight vectors. It is known that exploiting the *a priori* knowledge of sparsity of the underlying system improves the convergence characteristics of traditional adaptive filters [15]–[17]. Secondly, to take advantage of the sparsity in the smaller optimal weight vectors following the NKPD and therefore improve the convergence and tracking performance, we incorporate the proportionate filtering framework into the NKPD based adaptive filters and propose the NKPD based IPNLMS (NKP-IPNLMS) algorithm, where the proportionate weighting is based on ℓ_1-norm of the smaller weight vectors. We also propose the NKP-IPNLMS-ℓ_0 and NKP-IPNLMS-ℓ_0-2 algorithms, where the proportionate weighting are based on approximations of the ℓ_0-norms of the smaller weight vectors. In the simulation studies, we first show that optimal smaller weight vectors resulting from the NKPD of a large set of measured acoustic IRs from several databases [18]–[21], have underlying sparse characteristics, which motivates the development of proportionate class of NKP-NLMS algorithms. Then the proposed proportionate NKPD based algorithms are evaluated for adaptive system identification (ASI).

II. Background & Proposed Algorithms

A. Nearest Kronecker Product (NKP) Decomposition

We consider a linear system model, $y(n) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}(n) + v(n)$, where n is the sample index, $\boldsymbol{w} \in \mathbb{R}^{L\times 1}$ is a finite impulse response (FIR) filter and is the IR of the unknown system to be modelled, $\boldsymbol{x}(n) \in \mathbb{R}^{L\times 1}$ is a tap-delayed vector of the input signal $x(n)$, $v(n) \sim \mathcal{N}(0, \sigma_v^2)$ with variance σ_v^2 is the background noise and $[\cdot]^{\mathrm{T}}$ is the transpose operator. To adaptively model this system, we consider an adaptive FIR filter $\widehat{\boldsymbol{w}}(n) \in \mathbb{R}^{L\times 1}$, whose output is given by $\widehat{y}(n) = \widehat{\boldsymbol{w}}^{\mathrm{T}}(n)\boldsymbol{x}(n)$. The modelling error is given by $e(n) = y(n) - \widehat{y}(n)$. Lets consider the FIR filter $\boldsymbol{w} = [\boldsymbol{a}_1^{\mathrm{T}}\ \boldsymbol{a}_2^{\mathrm{T}}\ \ldots\ \boldsymbol{a}_{L_2}^{\mathrm{T}}]^{\mathrm{T}}$, where $\boldsymbol{a}_j \in \mathbb{R}^{L_1\times 1}$, $j \in \{1, 2, \ldots, L_2\}$ are sections of $\boldsymbol{w}$, such that $L = L_1 \times L_2$ & $L_1 \geq L_2$. Now, the sections of $\boldsymbol{w}$ can be arranged into an $L_1 \times L_2$ matrix as $\boldsymbol{W} = [\boldsymbol{a}_1\ \boldsymbol{a}_2\ \ldots\ \boldsymbol{a}_{L_2}]$. $\boldsymbol{W}$ may be decomposed using singular value decomposition (SVD) as $\boldsymbol{W} = \mathbf{Q}_1\boldsymbol{\Sigma}\mathbf{Q}_2^{\mathrm{T}} = \sum_{d=1}^{L_2}\sigma_d\mathbf{q}_{1,d}\mathbf{q}_{2,d}^{\mathrm{T}}$, where $\boldsymbol{\Sigma}$ is an $L_1 \times L_2$ rectangular diagonal matrix containing singular values of $\boldsymbol{W}$ in decreasing order $\sigma_1 \geq \ldots \geq \sigma_{L_2} \geq 0$ and $\mathbf{Q}_1$ ($\mathbf{Q}_2$) of size $L_1\times L_1$ ($L_2\times L_2$) contain the corresponding left (right) singular vectors of $\boldsymbol{W}$, i.e., $\mathbf{q}_{1,d}$ ($\mathbf{q}_{2,d}$), as its columns. Hence, the closest rank $D \leq L_2$ approximation of $\boldsymbol{W}$ in least squares (LS) sense can be obtained as $\overline{\boldsymbol{W}} = \sum_{d=1}^{D}\overline{\mathbf{w}}_{1,d}\overline{\mathbf{w}}_{2,d}^{\mathrm{T}}$, where $\overline{\mathbf{w}}_{1,d} = \sqrt{\sigma_d}\mathbf{q}_{1,d}$ and $\overline{\mathbf{w}}_{2,d} = \sqrt{\sigma_d}\mathbf{q}_{2,d}$, for $d = 1, 2, \ldots, D$ [4], [5]. Hence, the corresponding closest rank $D \leq L_2$ approximation of $\boldsymbol{w} = \mathrm{vec}(\boldsymbol{W})$ in the LS sense, equivalent to minimizing the misalignment $\mathcal{M} = ||\boldsymbol{w} - \overline{\boldsymbol{w}}||_2/||\boldsymbol{w}||_2$, can be obtained as

$$\overline{\boldsymbol{w}} = \mathrm{vec}(\overline{\boldsymbol{W}}) = \sum_{d=1}^{D}\mathrm{vec}(\overline{\mathbf{w}}_{1,d}\overline{\mathbf{w}}_{2,d}^{\mathrm{T}}) = \sum_{d=1}^{D}\overline{\mathbf{w}}_{2,d}\otimes\overline{\mathbf{w}}_{1,d} \quad (1)$$

where $\mathrm{vec}(\cdot)$ is the vectorization operation [8], $||\cdot||_p$ denotes the p-norm operator, $\otimes$ denotes the Kronecker product, $D = L_2$ gives back the exact $\boldsymbol{w}$. One may notice from (1) that the decomposition is not unique [4], i.e.,

$$\overline{\boldsymbol{w}} = \sum_{d=1}^{D}\overline{\mathbf{w}}_{2,d}\otimes\overline{\mathbf{w}}_{1,d} = \sum_{d=1}^{D}\frac{1}{\theta_d}\overline{\mathbf{w}}_{2,d}\otimes\theta_d\overline{\mathbf{w}}_{1,d} \quad (2)$$

i.e., the solution pairs at rank-d, $(\overline{\mathbf{w}}_{1,d}, \overline{\mathbf{w}}_{2,d})$ and $(\theta_d\overline{\mathbf{w}}_{1,d}, \frac{1}{\theta_d}\overline{\mathbf{w}}_{2,d})$, for a real-valued θ_d, are equivalent in terms of minimizing $\mathcal{M}$ and for any such solution pair $(\theta_d\overline{\mathbf{w}}_{1,d}, \frac{1}{\theta_d}\overline{\mathbf{w}}_{2,d})$, the obtained $\overline{\boldsymbol{w}}$ is the same. Following the NKPD in (1), the adaptive weight vector can be expressed as a combination of smaller weight vectors as

$$\widehat{\boldsymbol{w}}(n) = \sum_{d=1}^{D}\widehat{\mathbf{w}}_{2,d}(n)\otimes\widehat{\mathbf{w}}_{1,d}(n) \quad (3)$$

where D is the rank of approximation. Following (2) and the discussion thereafter, we can state that at rank-d the optimal solution for the pair $(\widehat{\mathbf{w}}_{1,d}(n), \widehat{\mathbf{w}}_{2,d}(n))$ is $(\theta_d\overline{\mathbf{w}}_{1,d}, \frac{1}{\theta_d}\overline{\mathbf{w}}_{2,d})$, $d = 1, \ldots, D$, $\theta_d \in \mathbb{R}$. Using NKP-NLMS, the smaller weight vectors can be updated as [8]

$$\widehat{\mathbf{w}}_l(n+1) = \widehat{\mathbf{w}}_l(n) + [\mu e(n)/(\mathbf{x}_k^{\mathrm{T}}(n)\mathbf{x}_k(n) + \eta)]\mathbf{x}_k(n) \quad (4)$$

where μ is the step-size, η is the regularization parameter,

$$\widehat{\mathbf{w}}_l(n) = [\widehat{\mathbf{w}}_{l,1}^{\mathrm{T}}(n)\ \widehat{\mathbf{w}}_{l,2}^{\mathrm{T}}(n)\ \ldots\ \widehat{\mathbf{w}}_{l,D}^{\mathrm{T}}(n)]^{\mathrm{T}}, \quad (5)$$

$\widehat{\mathbf{x}}_{l,d}(n) = \widehat{\mathbf{W}}_{l,d}^{\mathrm{T}}(n)\boldsymbol{x}(n)$, $\mathbf{x}_l(n) = [\widehat{\mathbf{x}}_{l,1}^{\mathrm{T}}(n)\ \ldots\ \widehat{\mathbf{x}}_{l,D}^{\mathrm{T}}(n)]^{\mathrm{T}}$, $\widehat{\mathbf{W}}_{2,d}(n) = \widehat{\mathbf{w}}_{2,d}(n)\otimes\mathbf{I}_{L_1}$, $\widehat{\mathbf{W}}_{1,d}(n) = \mathbf{I}_{L_2}\otimes\widehat{\mathbf{w}}_{1,d}(n)$, with $\mathbf{I}_M$ being the identity matrix of size $M \times M$, $\widehat{y}(n) = \widehat{\mathbf{w}}_l^{\mathrm{T}}(n)\mathbf{x}_k(n)$, $l, k \in \{1, 2\}$, $l \neq k$, .

B. Proposed Proportionate NKPD based Algorithms

As we will see in Section III, the optimal solution pair for rank-d approximation contains good degree of sparsity. To solve for $\widehat{\mathbf{w}}_1(n)$ & $\widehat{\mathbf{w}}_2(n)$ and also exploit the sparsity arising from NKPD, we propose the following optimization problem

$$\min_{\widehat{\mathbf{w}}_l(n+1)}\frac{1}{2}||\widehat{\mathbf{w}}_l(n+1) - \widehat{\mathbf{w}}_l(n)||^2_{\boldsymbol{Z}_l^{-1}(n)} \quad \text{subject to } y(n) = \widehat{\mathbf{w}}_l^{\mathrm{T}}(n)\mathbf{x}_k(n) \quad (6)$$

where $l, k \in \{1, 2\}$, $l \neq k$, $\boldsymbol{Z}_l(n)$ is the diagonal gain matrix that adjusts the step-sizes of the individual coefficients of $\widehat{\mathbf{w}}_l(n)$, given by $\boldsymbol{Z}_l(n) = \mathrm{diag}\{z_{l,1}, z_{l,2}, \ldots, z_{l,DL_l}\}$, such that

$$z_{l,j} = \frac{1-\alpha}{2DL_l} + (1+\alpha)\frac{|\widehat{\mathrm{w}}_{l_j}(n)|}{2||\widehat{\mathbf{w}}_l(n)||_1 + \zeta}, \quad j \in \{1, 2, \ldots, DL_l\} \quad (7)$$

where $\widehat{\mathbf{w}}_l(n) \in \mathbb{R}^{DL_l\times 1}$ as in (5), $\widehat{\mathrm{w}}_{l_j}(n)$ is the j^{th} element of $\widehat{\mathbf{w}}_l(n)$ and $l \in \{1, 2\}$, $\alpha \in [-1, 1]$, ζ is a small positive number to avoid division by zero. Using the Lagrangian multiplier approach, the cost function for the optimization problem in (6) can be written as

$$J_l(n+1) = \frac{1}{2}||\widehat{\mathbf{w}}_l(n+1) - \widehat{\mathbf{w}}_l(n)||^2_{\boldsymbol{Z}_l^{-1}(n)} + \lambda_l[y(n) - \widehat{\mathbf{w}}_l^{\mathrm{T}}(n)\mathbf{x}_k(n)] \quad (8)$$

$l, k \in \{1, 2\}$, $l \neq k$, where λ_l is the Lagrange multiplier. Now, following the stochastic gradient descent approach [17], the update rules for $\widehat{\mathbf{w}}_1(n)$ and $\widehat{\mathbf{w}}_2(n)$ is derived as

$$\widehat{\mathbf{w}}_l(n+1) = \widehat{\mathbf{w}}_l(n) + \frac{\mu e(n)}{\mathbf{x}_k^{\mathrm{T}}(n)\boldsymbol{Z}_l(n)\mathbf{x}_k(n) + \eta}\boldsymbol{Z}_l(n)\mathbf{x}_k(n) \quad (9)$$

$l, k \in \{1, 2\}$, $l \neq k$. We call (9) and (7) the NKPD based improved proportionate NLMS (NKP-IPNLMS) algorithm, which uses ℓ_1-norm of the weight vectors $\widehat{\mathbf{w}}_1(n)$ & $\widehat{\mathbf{w}}_2(n)$, as in (7), to induce sparsity into the corresponding adaptive weight vector solutions. For $\alpha = -1$, NKP-IPNLMS becomes equivalent to the NKP-NLMS algorithm and for α close to 1, NKP-IPNLMS behaves like NKPD based proportionate NLMS (NKP-PNLMS) algorithm, with a proportionate weighting similar to [22]. The ideal candidate to induce sparsity is known to be the ℓ_0-norm of the weight vectors, i.e., $||\widehat{\mathbf{w}}_1(n)||_0$ & $||\widehat{\mathbf{w}}_2(n)||_0$. However, since solving the ℓ_0-norm is non-polynomial (NP)-hard, an approximation of the ℓ_0-norm is generally used [16]. The NKP-IPNLMS algorithm uses ℓ_1-norm which is a poor approximation of the ℓ_0-norm [23]. A better & more natural approximation of ℓ_0-norm is given by [16], $||\widehat{\mathbf{w}}_l(n)||_0 \approx \sum_{j=1}^{DL_l}[1 - \exp\{-\beta|\widehat{\mathrm{w}}_{l_j}(n)|\}]$, $l \in \{1, 2\}$, which can be incorporated into the proposed NKPD based proportionate framework as

$$z_{l,j} = \frac{1-\alpha}{2DL_l} + (1+\alpha)\frac{1 - \exp\{-\beta|\widehat{\mathrm{w}}_{l_j}(n)|\}}{2\sum_{j=1}^{DL_l}[1 - \exp\{-\beta|\widehat{\mathrm{w}}_{l_j}(n)|\}] + \zeta} \quad (10)$$

for $j = 1, \ldots, DL_1$ & $j = 1, \ldots, DL_l$, $l \in \{1, 2\}$, respectively & $\beta > 0$. We call (9) and (10) the NKP-IPNLMS-ℓ_0 algorithm. As we can see from (10) and later in subsection II-C, the NKP-IPNLMS-ℓ_0 algorithm requires $DL_1 + DL_2$ exponential

TABLE I
COMPUTATIONAL COMPLEXITY COMPARISON

Algorithm	$\times$	$+$	$\div$	$\exp\{\cdot\}$
NLMS [1]	$3L+1$	$3L$	1	-
IPNLMS [17]	$4L+2$	$5L$	$L+1$	-
IPNLMS-ℓ_0 [23]	$5L+2$	$6L$	$L+1$	L
NKP-NLMS [8]	$2DL+2DL_1$ $+3DL_2+2$	$2DL+DL_1$ $+2DL_2$	2	-
NKP-IPNLMS	$2DL+3DL_1$ $+4DL_2+4$	$2DL+3DL_1$ $+4DL_2$	DL_1+DL_2 $+2$	-
NKP-IPNLMS-ℓ_0	$2DL+4DL_1$ $+5DL_2+4$	$2DL+4DL_1$ $+5DL_2$	DL_1+DL_2 $+2$	DL_1 $+DL_2$
NKP-IPNLMS-ℓ_0-2	$2DL+4DL_1$ $+5DL_2+4$	$2DL+4DL_1$ $+5DL_2$	DL_1+DL_2 $+2$	-

($\exp\{\cdot\}$) operations. Computing $\exp\{\cdot\}$ is known to require high computational complexity and it was shown in [24] that for implementation in the double logarithmic arithmetic method [25], implementing one $\exp\{\cdot\}$ requires one $\log(\cdot)$, one $+$ and two antilog$(\cdot)$ operations. In the proposed NKP-IPNLMS-ℓ_0 algorithm, $\exp\{\cdot\}$ operations are required to approximate the ℓ_0 norm, which are essentially decaying exponentials. A similar behaviour can be achieved using a $2^{\{\cdot\}}$ function in place of the $\exp\{\cdot\}$ function [26]. Moreover, the $2^{\{\cdot\}}$ function requires only one antilog$(\cdot)$ operation in architecture [24] and hence requires less computational complexity compared to the $\exp\{\cdot\}$ function. Hence, we propose a new reduced complexity approximation of the ℓ_0-norm as

$$||\widehat{\mathbf{w}}_l(n)||_0 \approx \sum_{j=1}^{DL_l}\left[1-2^{\{-\beta|\widehat{\mathrm{w}}_{l_j}(n)|\}}\right] \tag{11}$$

$l \in \{1,2\}$, where $\approx$ becomes $=$ for $\beta \to \infty$. Using (11), we propose NKP-IPNLMS-ℓ_0-2 algorithm, as (9), with

$$z_{l,j} = \frac{1-\alpha}{2DL_l} + (1+\alpha)\frac{1-2^{\{-\beta|\widehat{\mathrm{w}}_{l_j}(n)|\}}}{2\sum_{j=1}^{DL_l}[1-2^{\{-\beta|\widehat{\mathrm{w}}_{l_j}(n)|\}}]+\zeta} \tag{12}$$

$l \in \{1,2\}$.

C. Computational Complexity

Table I shows the comparison of computational complexity of the proposed algorithms with standard NLMS [1], IPNLMS [17] and IPNLMS-ℓ_0 [23], in terms of number of multiplications ($\times$), additions ($+$), divisions ($\div$) and $\exp\{\cdot\}$ operations required per sample. From Table I, we can see that NKP-IPNLMS requires slightly higher computations compared to NKP-NLMS. NKP-IPNLMS-ℓ_0 requires calculation of extra $\exp\{\cdot\}$ operations compared to NKP-IPNLMS, increasing its complexity. However, the complexity due to $\exp\{\cdot\}$ operation is eliminated in NKP-IPNLMS-ℓ_0-2. In general, proportionate NKP-NLMS algorithms have a higher complexity compared to the standard NKP-NLMS algorithm, but the motivation & advantages of using the proportionate method in NKP-NLMS framework will be apparent from studies in Section III.

III. SIMULATION STUDY

A. Investigation of Sparsity in Smaller Optimal Weight Vectors

To measure the sparseness of a vector $\boldsymbol{b} \in \mathbb{R}^{M\times 1}$, we consider the following sparseness measure based on the ℓ_1 and ℓ_2 norms of $\boldsymbol{b}$ [5], [27], given by

$$\xi_{12} = \frac{M}{M-\sqrt{M}}\left(1-\frac{||\boldsymbol{b}||_1}{\sqrt{M}||\boldsymbol{b}||_2}\right) \tag{13}$$

For a certain rank of approximation D of an IR $\boldsymbol{w}$, as in (1), the optimal weight vectors for $\widehat{\mathbf{w}}_1(n)$ & $\widehat{\mathbf{w}}_2(n)$ can be obtained from the SVD of the matrix $\boldsymbol{W}$ constructed from $\boldsymbol{w}$, as $\overline{\mathbf{w}}_1 = [\theta_1\overline{\mathbf{w}}_{1,1}^{\mathrm{T}} \ \dots \ \theta_D\overline{\mathbf{w}}_{1,D}^{\mathrm{T}}]^{\mathrm{T}}$ & $\overline{\mathbf{w}}_2 = [\frac{1}{\theta_1}\overline{\mathbf{w}}_{2,1}^{\mathrm{T}} \ \dots \ \frac{1}{\theta_D}\overline{\mathbf{w}}_{2,D}^{\mathrm{T}}]^{\mathrm{T}}$, & $\overline{\mathbf{w}}_{i,d} = \sqrt{\sigma_d}\mathbf{q}_{i,d}$, $i \in \{1,2\}$, $\theta_d \in \mathbb{R}$, $d = 1,2,\cdots,D$. L_1 and L_2 are chosen such that $L_1L_2 = L$, $L_1 \geq L_2$ and L_1+L_2 is minimum, so that NKPD based algorithms require minimum number of modelling parameters (weight coefficients) [4], [8].

We study the sparsity degree ξ_{12} of $\overline{\mathbf{w}}_1$ & $\overline{\mathbf{w}}_2$ for a wide range of acoustic IRs: (i) 154 measured IRs for active noise control [18]; (ii) 1008 measured room IRs (SMARD database) [19]; (iii) 10704 measured head related IRs (HRIRs) [20]; (iv) 34560 room IRs (RIRs) [21]. Fig. 1 (i)-(iv) respectively, show the sparsity degree ξ_{12} of $\overline{\mathbf{w}}_1$ & $\overline{\mathbf{w}}_2$, resulting from each acoustic IRs in the corresponding databases for $D=5$ & $\theta_d=1$, and Fig. 1 (v)-(viii) respectively, for $D=7$ & $\theta_d=1$. From Fig. 1, we can see that for different acoustic IRs, $\overline{\mathbf{w}}_2$ has moderate to a high degree of sparsity. On the other hand, for different acoustic IRs, $\overline{\mathbf{w}}_1$ has moderate degree of sparsity. Similar observations can be made from ξ_{12} for other random values of θ_d and other approximation ranks $1 \leq D \leq L_2$, but are not presented here due to page constraints. This observation motivates the use of sparsity aware techniques in NKPD based adaptive algorithms.

B. Comparison of Convergence Performance

We consider an ASI problem with an AR(1) process as input signal, obtained by filtering a white Gaussian noise $\sim \mathcal{N}(0,1)$ via first order system $1/(1-0.9z^{-1})$ and a signal to noise ratio (σ_y^2/σ_v^2) of 20 dB at the output of the system [4], [8]. The IR of the unknown system is considered to be one of the IRs of length $L=500$ shown in Fig. 2 (a) & (b), which are echo paths from the G168 Recommendation [4], [28]. Here we choose $L_1=25$ and $L_2=20$, such that $L_1L_2=L$, $L_1 \geq L_2$ and L_1+L_2 is minimum [4], [8]. The decaying nature of the singular values of the matrix $\boldsymbol{W}$ for the IRs in Fig. 2 (a) & (b) have been studied in detail in [4], and hence has not been presented here. It was shown in [4] that the IRs in Fig. 2(a) & (b) represent low rank systems and low rank approximation with $D=3$ & $D=5$, respectively, can give the exact IRs. Fig. 2 (a.1) & (a.2) shows $\overline{\mathbf{w}}_1$ & $\overline{\mathbf{w}}_2$, for the IR in Fig. 2 (a) with $D=3$, $\theta_d=1$ and Fig. 2 (b.1) & (b.2) shows $\overline{\mathbf{w}}_1$ & $\overline{\mathbf{w}}_2$, for the IR in Fig. 2 (b) with $D=5$, $\theta_d=1$. The normalized misalignment (in dB) is considered as the performance measure, given by $\mathcal{M}(\widehat{\mathbf{w}}_1,\widehat{\mathbf{w}}_2) = 20\log_{10}[(||\boldsymbol{w}-\widehat{\boldsymbol{w}}(n)||_2)/||\boldsymbol{w}||_2]$, with $\widehat{\boldsymbol{w}}(n)$ calculated as in (3) and the results shown are averaged over 20 independent trials.

CASE I: Fig. 3 shows the normalized misalignment variation for the competing algorithms for the IR in Fig. 2(a). From Fig. 3, we can see that the proportionate based NKP algorithms, i.e., NKP-IPNLMS, NKP-IPNLMS-ℓ_0 and NKP-IPNLMS-ℓ_0-2 reach lower steady state misalignment compared to the

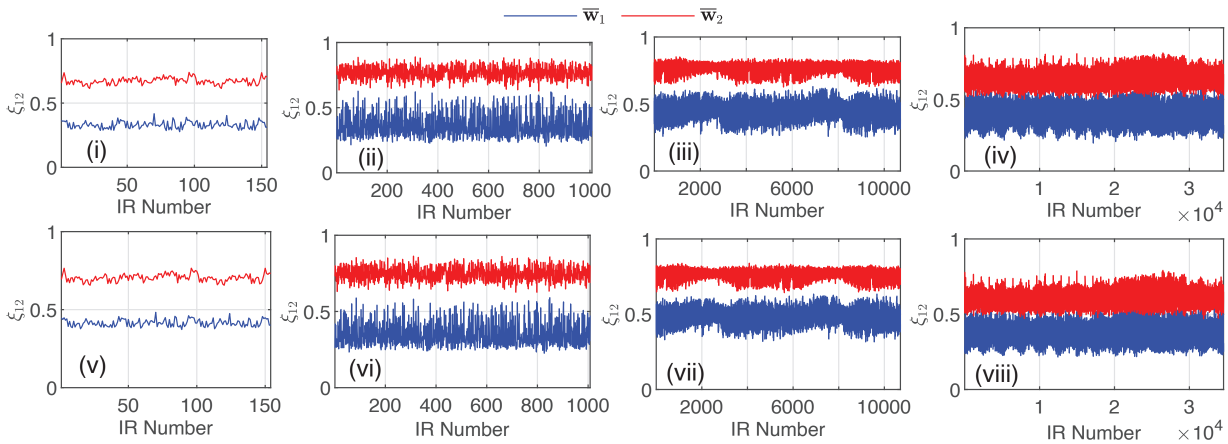

Fig. 1. Sparseness Degree of $\overline{\mathbf{w}}_1$ & $\overline{\mathbf{w}}_2$ for IRs from: (i) & (v) [18], with $L_1 = 25, L_2 = 20$, (ii) & (vi) [19], with $L_1 = 60, L_2 = 50$; (iii) & (vii) [20], with $L_1 = 16, L_2 = 16$, (iv) & (viii) [21], with $L_1 = 32, L_2 = 32$; (i) - (iv) $D = 5$, (v) - (viii) $D = 7$; $\theta_d = 1$.

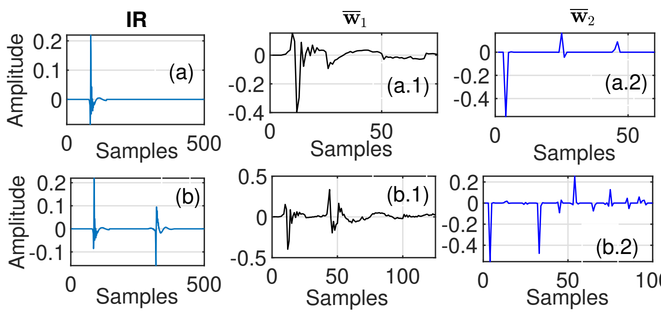

Fig. 2. (a),(b): Impulse Responses (IRs); (a.1): $\overline{\mathbf{w}}_1$ for (a) with $\xi_{12} = 0.5398$; (a.2): $\overline{\mathbf{w}}_2$ for (a) with $\xi_{12} = 0.9253$, $D = 3$; (b.1): $\overline{\mathbf{w}}_1$ for (b) with $\xi_{12} = 0.4970$; (b.2): $\overline{\mathbf{w}}_2$ for (b) with $\xi_{12} = 0.8243$, $D = 5$. $\theta_d = 1$.

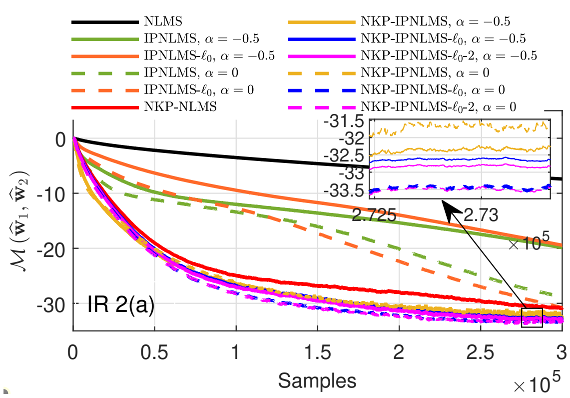

Fig. 3. CASE I: Misalignment (dB) for IR in Fig. 2(a), $\mu = 0.01$, $\beta = 100$; $D = 3$ for NKPD based algorithms.

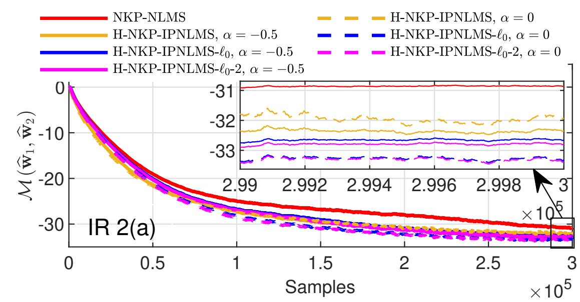

Fig. 4. CASE II: Misalignment (dB) with hybrid algorithms for IR in Fig. 2(a), $\mu = 0.01$, $\beta = 100$; $D = 3$ for NKPD based algorithms.

NKP-NLMS algorithm. For NKP-IPNLMS, $\alpha = -0.5$ gives a lower steady state misalignment compared to $\alpha = 0$, whereas for NKP-IPNLMS-ℓ_0 & NKP-IPNLMS-ℓ_0-2, $\alpha = 0$ reaches lower steady state misalignment compared to $\alpha = -0.5$. Also, NKP-IPNLMS-ℓ_0 and NKP-IPNLMS-ℓ_0-2 algorithms have similar convergence, however NKP-IPNLMS-ℓ_0-2 has lower complexity compared to NKP-IPNLMS-ℓ_0 (Table I), and both reach a lower steady state misalignment compared to NKP-IPNLMS algorithm. Also, NKPD based algorithms provide a faster convergence compared to standard NLMS and IPNLMS.

CASE II: As observed previously, $\overline{\mathbf{w}}_1$ generally contains less sparsity than $\overline{\mathbf{w}}_2$. Hence, we also study the case where $\widehat{\mathbf{w}}_1$ is updated using NKP-NLMS and $\widehat{\mathbf{w}}_2$ is updated by NKP-IPNLMS, NKP-IPNLMS-ℓ_0 or NKP-IPNLMS-ℓ_0-2, which we call the hybrid NKP-IPNLMS (H-NKP-NLMS), hybrid NKP-IPNLMS-ℓ_0 (H-NKP-IPNLMS-ℓ_0) and hybrid NKP-IPNLMS-ℓ_0-2 (H-NKP-IPNLMS-ℓ_0-2), respectively. Fig. 4 shows the variation of normalized misalignment for this case. From Fig. 4, similar observations can be made as in Case I. Moreover, the convergence characteristics and steady state misalignment of the NKP-IPNLMS, NKP-IPNLMS-ℓ_0 & NKP-IPNLMS-ℓ_0-2 in Fig. 3 are similar to the H-NKP-IPNLMS, H-NKP-IPNLMS-ℓ_0 & H-NKP-IPNLMS-ℓ_0-2 in Fig. 4, respectively, and all of them provide better convergence characteristics compared to NKP-NLMS. Hence, using a proportionate update appears to be more beneficial for $\widehat{\mathbf{w}}_2(n)$ and has minimal effect for $\widehat{\mathbf{w}}_1(n)$, which can be related to the previous observation that $\overline{\mathbf{w}}_2$ has high sparsity & $\overline{\mathbf{w}}_1$ has moderate sparsity.

CASE III: We also consider a path change ASI problem to study the tracking performance of the algorithms, where the IR of the unknown system changes midway in the simulation, from the IR in Fig. 2(a) in the first half of the simulation

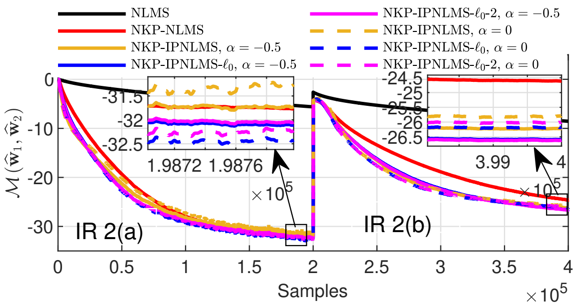

Fig. 5. CASE III: Misalignment (dB) with path change from Fig. 2(a) to (b). $\mu = 0.01$, $\beta = 10$; $D = 5$ for NKPD based algorithms.

to Fig. 2(b) in the later half. Fig. 5 shows the convergence characteristics for this case. From Fig. 5, we can observe that NKP-IPNLMS, NKP-IPNLMS-ℓ_0 & NKP-IPNLMS-ℓ_0-2 provide faster tracking performance when there is a sudden change in the modelled system, compared to the NLMS & NKP-NLMS algorithms. Similar to Case I and II, we observe that NKP-IPNLMS-ℓ_0-2 provides similar performance as that of NKP-IPNLMS-ℓ_0 but at a lower computational complexity. Also, depending on the value of α, there is slight variation in performance of the NKP-IPNLMS, NKP-IPNLMS-ℓ_0 & NKP-IPNLMS-ℓ_0-2 algorithms.

IV. Conclusion & Future Directions

In this paper, we analyse the smaller optimal weight vectors resulting from NKPD of numerous acoustic IRs and show that the smaller weight vectors exhibit moderate to highly sparse characteristics, where one of them is generally seen to have a high sparsity while the other having moderate sparsity. To take advantage of this observation, we propose a set of proportionate update NKPD based NLMS class of algorithms, which can take advantage of the sparse nature of the smaller weight vectors and improve the convergence and tracking characteristics. This would potentially benefit other Kronecker product decomposition based algorithms, including the ones involving bilinear and trilinear forms. The observation also opens the potential for using zero attraction penalty to exploit the sparsity in the decomposed smaller weight vectors. Since, performance of the proposed algorithms depend on parameter $\alpha \in [-1, 1]$, making it adaptive in the framework of NKPD based proportionate algorithms can also be studied.

References

[1] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*. New York, NY, USA: Springer, 2013.
[2] A. H. Sayed, *Adaptive filters*. New York, NY, USA: Wiley, 2008.
[3] K.-A. Lee, W.-S. Gan, and S. M. Kuo, *Subband adaptive filtering: theory and implementation*. John Wiley & Sons, 2009.
[4] C. Paleologu, J. Benesty, and S. Ciochina, "Linear system identification based on a Kronecker product decomposition," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 26, no. 10, pp. 1793–1808, 2018.
[5] C. Elisei-Iliescu, C. Paleologu, J. Benesty, C. Stanciu, C. Anghel, and S. Ciochina, "Recursive least-squares algorithms for the identification of low-rank systems," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 27, no. 5, pp. 903–918, 2019.
[6] M. Rupp and S. Schwarz, "A tensor LMS algorithm," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 3347–3351.
[7] L.-M. Dogariu, C. Paleologu, J. Benesty, and S. Ciochină, "An efficient Kalman filter for the identification of low-rank systems," *Signal Processing*, vol. 166, p. 107239, 2020.
[8] S. S. Bhattacharjee and N. V. George, "Nearest Kronecker product decomposition based normalized least mean square algorithm," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 476–480.
[9] L.-M. Dogariu, J. Benesty, C. Paleologu, and S. Ciochină, "Identification of room acoustic impulse responses via Kronecker product decompositions," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 2828–2841, 2022.
[10] C. Elisei-Iliescu, C. Paleologu, J. Benesty, and S. Ciochina, "A recursive least-squares algorithm based on the nearest Kronecker product decomposition," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 4843–4847.
[11] S. S. Bhattacharjee and N. V. George, "Fast and efficient acoustic feedback cancellation based on low rank approximation," *Signal Processing*, vol. 182, p. 107984, 2021.
[12] ——, "Nearest Kronecker product decomposition based linear-in-the-parameters nonlinear filters," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 2111–2122, 2021.
[13] S. S. Bhattacharjee, V. Patel, and N. V. George, "Nonlinear spline adaptive filters based on a low rank approximation," *Signal Processing*, vol. 201, p. 108726, 2022.
[14] J. Benesty, I. Cohen, and J. Chen, *Array Processing*. Springer, 2019.
[15] Y. Chen, Y. Gu, and A. O. Hero, "Sparse LMS for System Identification," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009, pp. 3125–3128.
[16] Y. Gu, J. Jin, and S. Mei, "ℓ_0 Norm Constraint LMS Algorithm for Sparse System Identification," *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 774–777, 2009.
[17] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *2002 IEEE International conference on Acoustics, Speech, and Signal Processing*, vol. 2. IEEE, 2002, pp. II–1881.
[18] S. Liebich, J. Fabry, P. Jax, and P. Vary, *Acoustic path database for ANC in-ear headphone development*. Universitätsbibliothek der RWTH Aachen, 2019.
[19] J. K. Nielsen, J. R. Jensen, S. H. Jensen, and M. G. Christensen, "The single-and multichannel audio recordings database (SMARD)," in *2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE, 2014, pp. 40–44.
[20] C. Oreinos and J. M. Buchholz, "Measurement of a full 3D set of HRTFs for in-ear and hearing aid microphones on a head and torso simulator (HATS)," *Acta Acustica united with Acustica*, vol. 99, no. 5, pp. 836–844, 2013.
[21] "Room impulse responses of the GTAC listening room at the UPV," https://gtac.webs.upv.es/room.php, accessed: 2023-02-01.
[22] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 5, pp. 508–518, 2000.
[23] C. Paleologu, J. Benesty, and S. Ciochină, "An improved proportionate NLMS algorithm based on the ℓ_0 norm," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2010, pp. 309–312.
[24] T. Deb, D. Ray, and N. V. George, "Design of nonlinear filters using affine projection algorithm based exact and approximate adaptive exponential functional link networks," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 11, pp. 2757–2761, 2019.
[25] D. M. Ellaithy, M. A. El-Moursy, G. H. Ibrahim, A. Zaki, and A. Zekry, "Double logarithmic arithmetic technique for low-power 3-D graphics applications," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 25, no. 7, pp. 2144–2152, 2017.
[26] S. S. Bhattacharjee and N. V. George, "Nonlinear system identification using exact and approximate improved adaptive exponential functional link networks," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 12, pp. 3542–3546, 2020.
[27] C. Paleologu, J. Benesty, and S. Ciochina, "Sparse adaptive filters for echo cancellation," *Synthesis Lectures on Speech and Audio Processing*, vol. 6, no. 1, pp. 1–124, 2010.
[28] *Digital Network Echo Cancellers,* ITU-T Recommendations G.168, 2002.