IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 26, 2017, accepted August 21, 2017, date of publication September 15, 2017, date of current version September 27, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2748144

Sparse Channel Estimation for Interference Limited OFDM Systems and Its Convergence Analysis

ABHIJEET BISHNU AND VIMAL BHATIA, (Senior Member, IEEE)
Signal and Software Group, Discipline of Electrical Engineering, IIT Indore, Indore 453552, India

Corresponding author: Abhijeet Bishnu (phd1501102014@iiti.ac.in).

This work was supported by the Ministry of Electronics and Information Technology, Government of India, under Grant PhD MLA/4(05)/2014 and Grant 14(3)/2014-CC&BT.

ABSTRACT Wireless communication channels are highly prone to interference in addition to the presence of additive white Gaussian noise (AWGN). Stochastic gradient (SG)-based non-parametric maximum likelihood (NPML) estimator, gives better channel estimates in the presence of Gaussian mixture (AWGN plus interference) noise processes, for subsequent use by the channel equalizer. However, for sparse channels, the SG-NPML-based channel estimator requires large iterations to converge. In this paper, we propose a natural gradient (NG)-based channel estimator for sparse channel estimation in the presence of high interference. We propose a generalized *p*th order warping transformation on channel coefficients space and then calculate the Riemannian metric tensor, thereby resulting in faster convergence in interference limited channels. The proposed algorithm is applied for IEEE 802.22 (based on orthogonal frequency division multiplexing) channel estimation in the presence of interference. Extensive simulations and experimental results show that the proposed NG-based algorithm converges faster than SG-NPML for the same mean squared error (MSE) floor with similar computational complexity per iteration as an SG-NPML algorithm. We also present convergence analysis of proposed NG-NPML algorithm in the presence of Gaussian mixture noise and derive an analytical expression for the steady-state MSE.

INDEX TERMS Orthogonal frequency division multiplexing non-parametric maximum likelihood, Riemannian geometry, Gaussian mixture noise, natural gradient, IEEE 802.22, convergence analysis, sparse channel estimation.

I. INTRODUCTION

Wireless communication channels are highly affected by interference from both the co-channel interference (CCI) [1] and adjacent channel interference due to the extensive growth of wireless services and applications, and due to hardware imperfections [2]. These interferences along with additive white Gaussian noise (AWGN) can be jointly modeled as Gaussian mixture noise which is non-Gaussian in nature [3]. In addition to interference in the radio channel, the example of non-Gaussian noise sources include symmetric alpha-stable noise [4], [5], double talk in echo cancellation [6], biological noise [7] in the underwater acoustic channels and variety of natural and man-made sources [8]. Stochastic gradient (SG) based non-parametric maximum likelihood (SG-NPML) adaptive algorithm gives better channel estimates in the presence of Gaussian mixture noise [3]. In the SG-NPML based channel estimator, first, the error signal (which is a Gaussian mixture) is estimated, then the probability density function (PDF) of this error signal is estimated with the help of kernel density estimators [8] and finally, the gradient of the cost function [2], [10], [11] is applied iteratively to approach maximum likelihood estimate. This resulting channel estimate has lower mean squared error (MSE) than least squares based channel estimator in the interference limited channels. However, in practice, there are various communication channels that are sparse in nature. Some examples include digital TV transmission channel [12], broadband system [13], [14], multi-input multi-output channel [15], underwater acoustic channel [16], and wireless multipath channel in cellular communication [17]. For sparse channels, SG-NPML based adaptive channel estimation requires a large number of iterations for convergence, and hence increases computational cost of the system.

In literature, there are various sparse channel estimators including least absolute shrinkage and selection operator (LASSO) [18], matching pursuit (MP) [19], and

orthogonal matching pursuit (OMP) [20], which have low computational complexity as compared to SG-NPML. The LASSO, MP, and OMP have good MSE performance in the presence of Gaussian noise, if *a priori* knowledge of an exact number of active taps is provided. However, MSE floor increases for LASSO, MP, and OMP for an interference limited sparse channel, thereby rendering them unsuitable for channel estimation.

One solution to circumvent the slow convergence of SG-NPML, and MSE degradation for LASSO, MP, and OMP, is to use a natural gradient (NG) based adaptive algorithms. NG learning provides an efficient algorithm which replaces stochastic gradient with a natural gradient. NG is the steepest descent direction in Riemannian space [21]. In NG adaptation, the weights are updated according to the non-Euclidean nature of the parameter space. The adaptation is based on a ''non-straight line'' distance metric defined by the Riemannian metric structure [22]. A Riemannian metric structure describes the parameter space for NG adaptation [23]. The optimum value of filter parameters is expected to be close to the coordinate axis for a sparse channel. Hence, the parameter space close to the axis should be warped in the sense that any direction which is orthogonal to the axes should be larger than the Euclidean distance [18]. Thus, once the channel coefficient vector comes close enough to coordinate axis, they swing to the same position, thereby leading to a sparse solution [22]. In sparse channel estimation, the improved-proportionate normalized least-mean square (IPNLMS) [24] and improved-proportionate affine projection algorithm (IPAPA) [25] are well known NG algorithm which outperforms their corresponding classical SG normalized least-mean square (SG-NLMS) and affine projection algorithm (APA) respectively. However, in interference limited communication channels (modeled by Gaussian mixture distribution) these algorithms also do not perform well and have high MSE floor. Additionally, for sparse channel estimation under symmetric alpha-stable distribution, natural gradient-based M-estimate affine projection algorithm (NGMAPA) and natural gradient-based p-norm affine projection algorithm (NGpNAPA) have been proposed in [4], which also have high MSE floor in the presence of interference.

Thus, to circumvent the limitations of existing algorithms, in this paper, we propose NG based channel estimator for channel estimation in interference limited sparse channels. The proposed algorithm uses p^{th} order transformation on the channel coefficients space which transforms the channel coefficients space from Euclidean space to Riemannian space. In [22] and [26], a linear transformation is used to calculate the channel update in the presence of Gaussian noise. However, in [22] NG is applied on NLMS, which has poor performance in the presence of high interference. Additionally, the algorithm in [22] requires a large (thousands) number of iterations to converge. In both [22] and [26], the generalization of this transformation and its analysis is not done. In this paper, we generalized this transformation, thereby resulting in faster convergence. The effect of this generalization on convergence rate and MSE is also highlighted to aid in selection transformation order for given application. In Riemannian space, the Euclidean distance based gradient does not give the steepest descent/ascent direction of the cost function [27]. Hence, once the channel coefficients space is warped, we apply NG adaptation on this space and calculate the Riemannian metric tensor (gain matrix) which is a positive-definite matrix that describes the local curvature of coefficient space for the proposed algorithm. The gain matrix is proportional to the updated channel coefficients; hence the effective NG's adaptation step-size is assumed to be large for active taps (the taps having a significant magnitudes of channel coefficients) and small for inactive taps (the taps having zero or nearly zero magnitudes of channel coefficients), leading to faster convergence.

To validate effectiveness of the proposed algorithm, we apply this to a standardized IEEE 802.22 based receiver for sparse channel estimation. Both the simulation and experimental results show that the proposed NG-NPML outperforms the classical SG-NPML and other algorithms. For both simulation and experimental setup we have chosen IEEE 802.22 standard (based on OFDM) as it is an evolving standard for TV white space (TVWS) communication and its transmission parameters such as bandwidth, sampling rate, and operating frequency band is same as digital terrestrial TV (DTTV) whose transmission channel is sparse [28]. Hence, the transmission channel for IEEE 802.22 standard is also sparse. In many countries, large bandwidth is unutilized (or underutilized) in terrestrial TV bands [29]. Moreover, these vacant bands are likely to be made unlicensed like in the UK [30] and Africa, or made lightly licensed for data transmission in rest of the world.

Additionally, in this paper, we also present convergence analysis of the proposed NG-NPML algorithm in mean weight error, and derive an approximate analytical expression for steady-state MSE based on Taylor series expansion. The analysis is based on 'transform' domain approach as used in [31]. Simulation results verify the theoretical calculation well. As an early work, the authors introduced preliminary quadratic warping transformation for sparse channel estimation in the presence of interference [32]. In summary, the main contributions of this paper are as follows:

- We propose a generalized robust (to various interferences) NG-NPML for sparse channel estimation for OFDM systems by transforming the coefficients space into Riemannian space. The complexity per iteration of proposed algorithm is similar to SG-NPML. However, by calculating the gain matrix, which provides large step-size for active taps and small step-size for non-active taps, a much faster convergence is achieved for the same MSE as SG-NPML, and superior to LASSO and OMP.
- We also present convergence analysis of the proposed generalized NG-NPML algorithm in mean weight error, and derive an approximate analytical expression for

steady-state MSE based on Taylor series expansion for the first time in literature. Simulation results validate the theoretical expressions for MSE and excess MSE (EMSE). Additionally, the convergence analysis of NG-NPML algorithm can be used to analyze the other NG based algorithms like NGMAPA and NGpNAPA in the presence of Gaussian mixture noise.

For validation of robustness of the proposed algorithm, we have performed simulations and experiments under various conditions:

- We have chosen IEEE 802.22 standard (as it is an evolving standard for TVWS communication and is based on OFDM) for simulation and experimental purpose to verify superiority of the proposed algorithm for Typical Urban COST-207 channel model. For simulations, we used another IEEE 802.22 transmitter as CCI.
- For experimental validation (on National Instruments Universal Software Radio Peripheral (NI-USRP) 2952R), we used two types of CCI; an IEEE 802.22 transmitter and DTTV transmitter. Both the simulation and experimental results show robustness of the proposed algorithm.

Rest of the paper is organized as follows: Section II describes the system model and provides an overview of NPML technique and the NG adaptation algorithm. Section III describes the proposed algorithm by using warping transformation with detailed derivation of the quadratic warping transformation. Section IV provides the stability condition of the proposed algorithm with respect to mean weight error and derive the analytical expression for steady-state MSE which is based on transform domain model of NG-NPML. Section V shows the computer simulation results and experimental results analysis with discussions and complexity analysis of the conventional and proposed algorithm. Finally, the conclusions are drawn in Section VI.

Notation: In this paper, the superscripts $(.)^T$, $(.)^*$ and $(.)^H$ denote the transposition, complex conjugate and Hermitian of (.) respectively. Bold capital letters denote matrix and bold small letters denote vector. |.| and ||.|| represent absolute value and Euclidean norm of (.) respectively. The $\mathbb{E}[.]$ is the statistical expectation operator, $Tr(.)$ is the trace of (.), and $\mathbf{I}$ is an identity matrix of dimension $L \times L$.

II. SYSTEM MODEL AND PRELIMINARIES

In this section, we briefly explain the system model for real valued channel estimation as shown in Fig. 1 and overview of SG-NPML and NG adaptive algorithm. Let $x(n)$, $\mathbf{x}(n) = [x(n), x(n-1), \ldots, x(n-L+1)]^T$, $\mathbf{h} = [h(0), h(1), \ldots, h(L-1)]^T$, $w(n)$, $i(n)$, $\hat{\mathbf{h}}_k = [\hat{h}_k(0), \hat{h}_k(1), \ldots, \hat{h}_k(L-1)]^T$, $\epsilon(n)$ and $e(n) = w(n) + i(n)$ represent the source signal, vector of source signal of dimension $L \times 1$, true channel impulse response vector, AWGN, CCI signal, adaptive (estimated) filter coefficient vector, residual error and Gaussian mixture noise respectively. Here L, k, and n represent the channel length, iteration index, and time index respectively. The joint PDF of AWGN and CCI follow the Gaussian mixture. Let the received signal be given by:

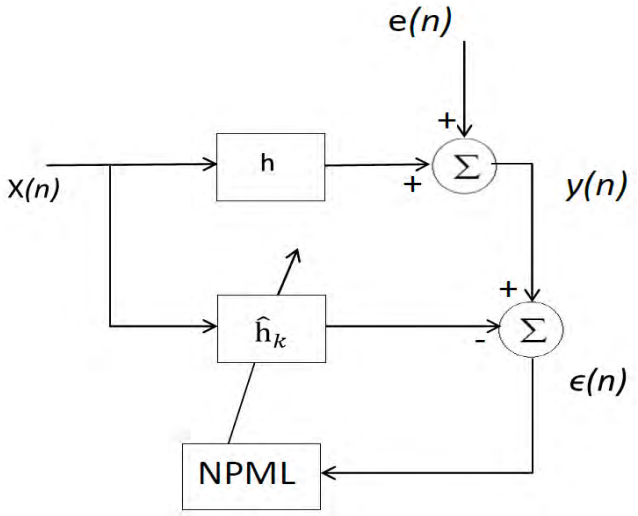

FIGURE 1. System model for channel estimator.

$$y(n) = \sum_{l=0}^{L-1} h(l)x(n-l) + e(n) \tag{1}$$

In SG-NPML technique, an iterative gradient ascent method is used for channel estimation [3]. In SG-NPML estimator, the cost function is given by [3]:

$$\mathcal{J}(\hat{\mathbf{h}}_k) = \mathcal{L}(\hat{\mathbf{h}}_k|\mathbf{y}) = \log f(\mathbf{y}|\hat{\mathbf{h}}_k) = \sum_{m=1}^{M} \log f(\epsilon(m)) \tag{2}$$

where $\mathcal{L}(.)$ represents the log likelihood function and $f(\epsilon(m))$ is the PDF of Gaussian mixture noise which is estimated by kernel density estimation with M measured data samples. $\epsilon(m)$ can be calculated as:

$$\epsilon(m) = y(m) - \sum_{l=0}^{L-1} \hat{h}_k(l)x(m-l) \tag{3}$$

PDF of the Gaussian mixture noise is estimated by (4):

$$\hat{f}(\epsilon(i)) = \frac{1}{M}\sum_{j=1}^{M} K(\epsilon(i) - \epsilon(j)) \quad i = 1, 2, \ldots, M \tag{4}$$

where $K(.)$ is assumed to be Gaussian kernel [9] and defined as:

$$K(t) = \frac{1}{\sqrt{(2\pi\sigma^2)}} \exp\left(\frac{-t^2}{2\sigma^2}\right) \tag{5}$$

where σ is the kernel width [9]. The channel update equation at $(k+1)^{th}$ iteration is given by

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k + \mu \nabla_{\hat{\mathbf{h}}_k} \mathcal{J}(\hat{\mathbf{h}}_k) \tag{6}$$

where μ is the step-size and for SG-NPML, the gradient of cost function which minimize (2) is given by [3]:

$$\nabla_{\hat{\mathbf{h}}_k}\mathcal{J}(\hat{\mathbf{h}}_k) = \frac{1}{\sigma^2}\sum_{i=1}^{M}\frac{\sum_{j=1}^{M}(\epsilon(i)-\epsilon(j))(\mathbf{x}(i)-\mathbf{x}(j))K(\epsilon(i)-\epsilon(j))}{\sum_{k=1}^{M}K(\epsilon(i)-\epsilon(k))} \tag{7}$$

After substituting (7) into (6), the channel update equation is given by:

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k + \frac{\mu}{\sigma^2}\sum_{i=1}^{M}\left(\frac{\sum_{j=1}^{M}\epsilon_k\mathbf{q}K(\epsilon_k)}{\sum_{j=1}^{M}K(\epsilon_k)}\right) \tag{8}$$

where $\epsilon_k = \epsilon(i)-\epsilon(j)$, $\mathbf{q} = \mathbf{x}(i)-\mathbf{x}(j)$ of dimension $L\times 1$. After iterations, the above equation tends to convergence and the ML estimate of the channel vector in additive Gaussian mixture noise is obtained [2]. However, for a large channel, the convergence is slow and is comparatively more complex.

On the other hand, if the channel is sparse i.e. most of the channel coefficients are nearly zero; an adaptation algorithm that updates the channel coefficients by taking advantage of sparseness of parameter solution space is needed. It is shown in [33], that the NG based adaptation provides a better solution for sparse channels. This is because in NG adaptation, Riemannian metric tensor provides large step-size for active tap coefficients, and small step-size for inactive tap coefficients and hence gives faster convergence. The estimator coefficients are updated based on a ''non-straight line'' distance metric defined by the Riemannian metric structure. In Riemannian space, distance is not measured according to the Euclidean norm. Once the parameter space is warped, then the NG update of $\mathbf{h}$ at $(k+1)^{th}$ iteration is given by [22]:

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k + \mu\mathbf{G}^{-1}_{\hat{\mathbf{h}}_k}\nabla_{\hat{\mathbf{h}}_k}\mathcal{J}(\hat{\mathbf{h}}_k) \tag{9}$$

where $\mathbf{G}_{\hat{\mathbf{h}}_k}$ is the Riemannian metric tensor (gain matrix) which is a positive-definite matrix that describes the local curvature of the parameter space at $\hat{\mathbf{h}}_k$, and $\mathbf{G}^{-1}_{\hat{\mathbf{h}}_k}$ is the inverse of $\mathbf{G}_{\hat{\mathbf{h}}_k}$.

To derive the natural gradient, we choose a distance metric (as given in [23]), which is not Euclidean, but Riemannian:

$$\begin{aligned} D &= |d_{\hat{h}_k(l)}(\hat{h}_k(l), \hat{h}_k(l)+r_{k+1}(l))|^2 \\ &= \sum_{l=0}^{L-1}|F(\hat{h}_k(l)+r_{k+1}(l)) - F(\hat{h}_k(l))|^2 \end{aligned} \tag{10}$$

where the $\mathbf{r}_{k+1}$ $(r_{k+1}(0),\ r_{k+1}(1),\ldots,\ r_{k+1}(L-1)^T)$ is a column vector of $L\times 1$ dimension having elements of small real value data and $F(.)$ is a warping transformation on (.). By using Taylor series approximations in the above equation, the distance metric can be written as [23]:

$$D = \mathbf{r}^T_{k+1}\mathbf{G}_{\hat{\mathbf{h}}_k}\mathbf{r}_{k+1} \tag{11}$$

$\mathbf{G}_{\hat{\mathbf{h}}_k}$ for the proposed NG-NPML algorithm is calculated using (10) and (11).

III. PROPOSED NG-NPML ALGORITHM

In this section, we derive the proposed iterative algorithm by transforming the parameter space and calculate the gain matrix. The gain matrix $\mathbf{G}_{\hat{\mathbf{h}}_k}$ is updated at each iteration, hence it provides large step-size for active taps and small step-size for inactive taps, thereby leading to faster convergence. The gain matrix $\mathbf{G}_{\hat{\mathbf{h}}_k}$ depends on the estimated channel coefficients at each iterations. As shown in (10) and (11), the gain matrix $\mathbf{G}_{\hat{\mathbf{h}}_k}$ also depends on the transformation. The transformation can include linear, quadratic, and higher orders. The transformation can be generalized to any order as given below:

$$F(\hat{h}_k(l)) = \sqrt{\alpha_k\left(\left|\,|\hat{h}_k(l)| - |\hat{h}_k(l)|^2 \ldots - |\hat{h}_k(l)|^{p-1} - |\hat{h}_k(l)|^p\right| + \beta\right)} \tag{12}$$

where p is the order of transformation, β is a regularization parameter and α_k is a normalization term which is given as:

$$\alpha_k = \frac{1}{L}\sum_{l=0}^{L-1}\left|\,|\hat{h}_k(l)| - |\hat{h}_k(l)|^2 \ldots - |\hat{h}_k(l)|^{p-1} - |\hat{h}_k(l)|^p\right| + \beta \tag{13}$$

The convergence of NG-NPML can be made faster by using higher order transformation. However, the computational complexity also increases as the order of the transformation increases. Hence, to balance computational complexity and faster convergence rate, we consider quadratic warping transformation on the channel coefficient space which is given by:

$$F(\hat{h}_k(l)) = \sqrt{\alpha_k\left(\left|\,|\hat{h}_k(l)| - |\hat{h}_k(l)|^2\right| + \beta\right)} \tag{14}$$

After substituting (14) in (10), the distance metric is given by (15), as shown at the bottom of the next page.

Here, we assume that $|r_{k+1}(l)| \ll 1$ and $||\hat{\mathbf{h}}_k|| \gg ||\mathbf{r}_{k+1}||$. In the above equation, we consider two cases, $\hat{h}_k(l)r_{k+1}(l) > 0$ and $\hat{h}_k(l)r_{k+1}(l) < 0$. For both the cases, the final channel update equation at $(k+1)^{th}$ iteration is given as:

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k + \mu\frac{\left(\left|\,|\hat{\mathbf{H}}_k| - |\hat{\mathbf{H}}_k|^2\right| + \beta\mathbf{I}\right)}{\alpha_k\mathbf{\Gamma}_2}\nabla_{\hat{\mathbf{h}}_k}\mathcal{J}(\hat{\mathbf{h}}_k) \tag{16}$$

where $\mathbf{\Gamma}_2$ is a matrix of dimension $L\times L$, $|\hat{\mathbf{H}}_k|$ and $|\hat{\mathbf{H}}_k|^2$ are $L\times L$ diagonal matrix. The proof of the above equation is given in Appendix A. Similarly, the channel update equation at $(k+1)^{th}$ iteration for p^{th} order transformation is given by:

$$\hat{\mathbf{h}}_{k+1} \approx \hat{\mathbf{h}}_k + \mu\frac{\left(\left|\,|\hat{\mathbf{H}}_k| - \ldots - |\hat{\mathbf{H}}_k|^p\right| + \beta\mathbf{I}\right)}{\alpha_k\mathbf{\Gamma}_p}\nabla_{\hat{\mathbf{h}}_k}\mathcal{J}(\hat{\mathbf{h}}_k) \tag{17}$$

where $\mathbf{\Gamma}_p = (\mathbf{I}+2|\hat{\mathbf{H}}_k| + \ldots + p|\hat{\mathbf{H}}_k|^{p-1})^2$. The proof of (17) is given in Appendix B. After substituting (7) into (16),

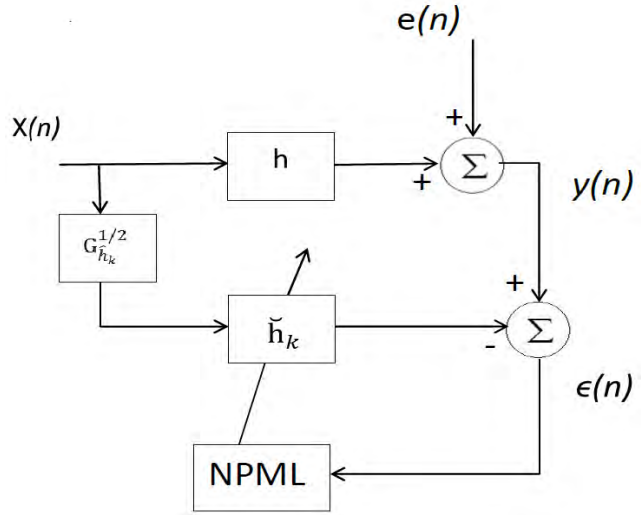

FIGURE 2. Transform domain model of NG-NPML estimator.

the channel update equation at $(k+1)^{th}$ iteration for quadratic transformation can be written as:

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k + \frac{\tilde{\mu}}{\sigma^2}\sum_{i=1}^{M}\frac{\sum_{j=1}^{M}\epsilon_k \mathbf{G}_{\hat{\mathbf{h}}_k}\mathbf{q}K(\epsilon_k)}{\sum_{j=1}^{M}K(\epsilon_k)} \tag{18}$$

where $\mathbf{G}_{\hat{\mathbf{h}}_k} = \left(\left||\hat{\mathbf{H}}_k| - |\hat{\mathbf{H}}_k|^2\right| + \beta\mathbf{I}\right)/(\alpha_k\mathbf{\Gamma}_2)$ is a diagonal gain matrix with $g_k(l) = \mathbf{G}_k(l,l)$ and satisfying $\sum_{l=0}^{L-1} g_k(l) = 1\ (0 < g_k(l) < 1)$.

IV. CONVERGENCE ANALYSIS OF PROPOSED NG-NPML ALGORITHM

In this section, convergence analysis of NG-NPML algorithm is described. The convergence analysis of NG-NPML algorithm is not straightforward as used in the least mean square algorithm because of the presence of $\mathbf{G}_{\hat{\mathbf{h}}_k}$ and $K(\epsilon_k)$. In order to analyze convergence of the proposed algorithm in the presence of $\mathbf{G}_{\hat{\mathbf{h}}_k}$, we use 'transform' domain model of proportionate-type normalized least mean square algorithm [31], [34], [35]. Further, since $K(\epsilon_k)$ consists of exponential term, it is approximated by its Taylor series expansion for mathematical tractability. Let $\mathbf{R}_{s,k} = \mathbb{E}[\tilde{\mathbf{q}}_k\tilde{\mathbf{q}}_k^T]$, $\mathrm{R} = \mathbb{E}[\mathbf{q}\mathbf{q}^T]$ and $\tilde{\mathbf{R}}_k = \mathbb{E}[\tilde{\mathbf{h}}_k\tilde{\mathbf{h}}_k^T]$ is the correlation matrix (diagonal) of transformed input $\tilde{\mathbf{q}}_k$, $\mathbf{q}$ and weight error vector $\tilde{\mathbf{h}}_k$ respectively of dimension $L \times L$.

A. TRANSFORM DOMAIN MODEL OF THE NG-NPML ALGORITHM

The transform domain model of the NG-NPML algorithm is shown in Fig. 2, where $\mathbf{G}^{1/2}_{\hat{\mathbf{h}}_k}$ is a transform domain diagonal matrix of dimension $L \times L$, $\mathbf{s}(n)$ and $\breve{\mathbf{h}}_k = [\breve{h}_k(0), \breve{h}_k(1), \ldots, \breve{h}_k(L-1)]^T$ are the transformed input and filter coefficient vector respectively with $\mathbf{s}(n) = \mathbf{G}^{1/2}_{\hat{\mathbf{h}}_k}\mathbf{x}(n)$ and $\breve{\mathbf{h}}_k = \mathbf{G}^{-1/2}_{\hat{\mathbf{h}}_k}\hat{\mathbf{h}}_k$, such that $\breve{\mathbf{h}}_k^T\mathbf{s}(n) = \hat{\mathbf{h}}_k^T\mathbf{x}(n)$. The update equation of $\breve{\mathbf{h}}$ at $(k+1)^{th}$ iteration is given according to the SG-NPML:

$$\breve{\mathbf{h}}'_{k+1} = \breve{\mathbf{h}}_k + \frac{\mu}{\sigma^2}\sum_{i=1}^{M}\frac{\sum_{j=1}^{M}\epsilon_k\tilde{\mathbf{q}}_k K(\epsilon_k)}{\sum_{j=1}^{M}K(\epsilon_k)} \tag{19}$$

where, $\tilde{\mathbf{q}}_k = \mathbf{s}(i) - \mathbf{s}(j)$ of dimension $L \times 1$. From (18), the $\hat{\mathbf{h}}_{k+1}$ can be computed as $\hat{\mathbf{h}}_{k+1} = \mathbf{G}^{1/2}_{\hat{\mathbf{h}}_k}\breve{\mathbf{h}}'_{k+1}$. Now, $\breve{\mathbf{h}}_{k+1} = \mathbf{G}^{-1/2}_{\hat{\mathbf{h}}_{k+1}}\hat{\mathbf{h}}_{k+1}$ which can be written as $\breve{\mathbf{h}}_{k+1} = \mathbf{G}^{1/2}_{\hat{\mathbf{h}}_{k+1}}\mathbf{G}^{1/2}_{\hat{\mathbf{h}}_k}\breve{\mathbf{h}}'_{k+1}$. It is reasonable to assume $\mathbf{G}^{1/2}_{\hat{\mathbf{h}}_{k+1}} \approx \mathbf{G}^{1/2}_{\hat{\mathbf{h}}_k}$ near convergence and/or large order channel [35] and this implies $\breve{\mathbf{h}}'_{k+1} = \breve{\mathbf{h}}_{k+1}$. Hence, (19) can be written as:

$$\breve{\mathbf{h}}_{k+1} = \breve{\mathbf{h}}_k + \frac{\mu}{\sigma^2}\sum_{i=1}^{M}\frac{\sum_{j=1}^{M}\epsilon_k\tilde{\mathbf{q}}_k K(\epsilon_k)}{\sum_{j=1}^{M}K(\epsilon_k)} \tag{20}$$

It is also assumed that the $\tilde{\mathbf{q}}_k$ is zero mean Gaussian random variable for large order channel by the central limit theorem [36]. The above equation is used for the convergence analysis of the proposed algorithm in terms of mean weight error.

B. MEAN WEIGHT ERROR CONVERGENCE ANALYSIS

With a zero-mean Gaussian transformed input $\tilde{\mathbf{q}}_k$ and its correlation matrix $\mathbf{R}_{s,k}$, the proposed algorithm produces stable performance with respect to mean weight error if the step-size μ satisfies the below criterion:

$$0 < \mu < \frac{2\sigma^2}{M\sum_{l=1}^{L}\lambda_{s,k}(l)} \tag{21}$$

where $\lambda_{s,k}(l)$ is the eigenvalue of $\mathbf{R}_{s,k}$. The proof of the above equation is given in Appendix C.

C. STEADY-STATE MEAN SQUARE ERROR PERFORMANCE ANALYSIS

In this subsection, we derive the analytical expression for the steady-state MSE based on Taylor series expansion and show that to achieve lower MSE floor, step-size must be small. With a zero-mean Gaussian transformed input $\tilde{\mathbf{q}}_k$, correlation matrix $\mathbf{R}_{s,k}$ and using the Taylor series expansion of the exponential function, the steady-state MSE is given by:

$$\xi_\infty = \xi_0 + \lim_{k\to\infty}\frac{\mu MTr(\mathbf{R}_{s,k})\mathbb{E}[\exp\left(\frac{-\tilde{e}^2}{\sigma^2}\right)\tilde{e}^2]\mathbb{E}[\exp\left(\frac{-\tilde{e}^2}{2\sigma^2}\right)]}{2\sigma^2\mathbb{E}[\exp\left(\frac{-\tilde{e}^2}{2\sigma^2}\right)\left(1-\frac{\tilde{e}^2}{\sigma^2}\right)]\mathbb{E}[\exp\left(\frac{-\tilde{e}^2}{\sigma^2}\right)]} \tag{22}$$

$$D = \sum_{l=0}^{L-1}\left|\sqrt{\alpha_k\left(\left||\hat{h}_k(l) + r_{k+1}(l)| - |\hat{h}_k(l) + r_{k+1}(l)|^2\right| + \beta\right)} - \sqrt{\alpha_k\left(\left||\hat{h}_k(l)| - |\hat{h}_k(l)|^2\right| + \beta\right)}\right|^2 \tag{15}$$

where, $\xi_0 = \mathbb{E}[\tilde{e}^2]$ is the minimum MSE (MMSE), $\tilde{e} = e(i) - e(j)$, $\xi_\infty = \mathbb{E}[\epsilon_k^2]$ at steady-state. The proof of the above equation is given in Appendix D. With the help of above equation, observations can be derived as corollary

Corollary 1: The steady-state MSE of the proposed NG-NPML algorithm is independent of the gain matrix which is given as:

$$\xi_\infty = \xi_0 + \frac{\mu M \sigma_{\mathbf{q}}^2 \mathbb{E}[\exp\left(\frac{-\tilde{e}^2}{\sigma^2}\right)\tilde{e}^2]\mathbb{E}[\exp\left(\frac{-\tilde{e}^2}{2\sigma^2}\right)]}{2\sigma^2 \mathbb{E}[\exp\left(\frac{-\tilde{e}^2}{2\sigma^2}\right)\left(1-\frac{\tilde{e}^2}{\sigma^2}\right)]\mathbb{E}[\exp\left(\frac{-\tilde{e}^2}{\sigma^2}\right)]} \tag{23}$$

where $\sigma_{\mathbf{q}}^2$ is the variance of $\mathbf{q}$.

Proof: Since, $\tilde{\mathbf{q}}_k = \mathbf{G}_{\hat{\mathbf{h}}_k}^{1/2}\mathbf{q}$, $\mathbf{R}_{s,k} = \mathbb{E}[\tilde{\mathbf{q}}_k \tilde{\mathbf{q}}_k^T]$ which is equal to $\mathbb{E}[\mathbf{G}_{\hat{\mathbf{h}}_k}^{1/2}\mathbf{q}\mathbf{q}^T\mathbf{G}_{\hat{\mathbf{h}}_k}^{1/2}]$, the gain matrix $\mathbf{G}_{\hat{\mathbf{h}}_k}$ satisfy $\sum_{l=0}^{L-1} g_k(l) = 1$ and at steady-state the variation of gain matrix is very small. Hence, at steady-state $Tr(\mathbf{R}_{s,k}) = Tr(\mathbb{E}[\mathbf{G}_{\hat{\mathbf{h}}_k}^{1/2}\mathbf{R}\mathbf{G}_{\hat{\mathbf{h}}_k}^{1/2}]) = \sigma_{\mathbf{q}}^2$ and by using above condition in (22) we arrive at (23). ■

Remark 1: The steady-state EMSE which is the second term on right hand side of (23) does not depend on the gain matrix. Hence both MSE and EMSE at steady-state does not depend on the transformation $F(.)$.

V. RESULTS AND DISCUSSIONS

In this section, effectiveness of the proposed algorithm and its convergence analysis is validated by numerical simulation and experimental results. In subsection *A*, numerical simulation is done followed by the experimental result in subsection *B*, and finally, the computational complexity of the proposed and the conventional algorithms is given in subsection *C*.

A. SIMULATION RESULTS

In this sub-section, we have compared the performance of conventional SG-NPML algorithm, and the proposed NG-NPML algorithm on IEEE 802.22 based transceiver as its transmission channel may be considered as sparse, and it is an evolving standard for TVWS [37], [38] for rural broadband. The IEEE 8022.22 is based on OFDM with fast Fourier transform (FFT) of size 2048, cyclic prefix of length 1/4, 1/18, 1/16 and 1/32. Data is modulated by quadrature phase shift keying, 16-quadrature amplitude modulation (16-QAM) and 64-QAM as per the specification [39]. The Gaussian mixture noise considered in this paper, is due to the presence of unknown (at receiver) CCI and AWGN. For simulation, we have assumed only one strong co-channel interferer. We have considered $M = 2048$ (large) samples for better PDF estimation of Gaussian mixture noise and Typical Urban COST-207 channel model [40] which gives sparse channel of length $L = 51$ with 6 active (non-zero) tap coefficient at sampling rate of 6 MHz [41]. The channel coefficients are Rayleigh faded. For both the algorithms, the adaptation step-size, μ, was taken to be 5×10^{-6}, $\beta = 0.01$, and all the simulation results were obtained by taking an ensemble

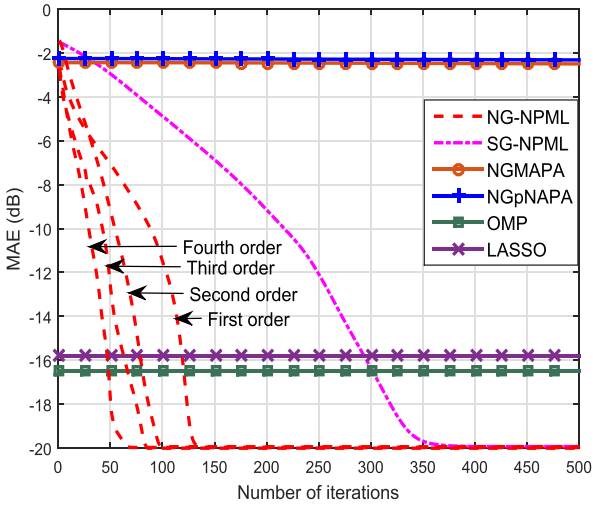

FIGURE 3. MAE of sparse channel estimation at SIR 5dB and SNR 30dB.

of 250 runs. The channel is estimated with the help of preamble which consists of long training sequence (LTS) and it is binary phase shift keying (BPSK) modulated as per the standard. We assume knowledge of the position of non-zero coefficient. However, there are various techniques to identify the position of active taps coefficients [42], [43].

Fig. 3 shows the convergence curve for SG-NPML and NG-NPML (with different order of transformation) which shows the mean absolute error (MAE) of sparse channel estimation ($10\log_{10}(\mathbb{E}[\frac{1}{L_a}\sum_{l=0}^{L-1}|h(l) - \hat{h}(l)|])$, where L_a is the number of active taps) against number of iterations at 5 dB signal to interference ratio (SIR) and 30 dB signal to noise ratio (SNR). Fig. 3 shows that the proposed NG-NPML convergences much faster than SG-NPML for the same MSE floor and requires about $1/4^{th}$ iterations as compared to SG-NPML for convergence, thereby resulting in huge savings in both computations and time. Fig. 3 also shows that the proposed NG-NPML algorithm converges much faster as we increase the order of transformation, however the complexity is also increased with the increase in the order of transformation. In Fig. 3, we also compare the proposed algorithm with other popular algorithms namely NGMAPA, NGpNAPA, LASSO, and OMP which shows that the proposed algorithm outperforms all existing techniques for sparse channel estimation. Fig. 3 also shows that the performance of NGMAPA, NGpNAPA, LASSO, and OMP degrade under additive CCI and AWGN (Gaussian mixture noise). Fig. 4 shows good agreement between the theoretical expression of steady-state EMSE ($\mathbb{E}[\epsilon^2(n)]_{e,\infty}$) calculated in (80) and Monte Carlo simulation of the proposed algorithm, for comparison, we also show the Monte Carlo simulation of the SG-NPML algorithm. In Fig. 4, the legend 'SS Theoretical' represents the theoretical steady-state of EMSE. A small deviation is attributed to the approximate expression of (80). The steady-state EMSE for both the algorithm is same because the proposed algorithm is independent of the gain matrix at

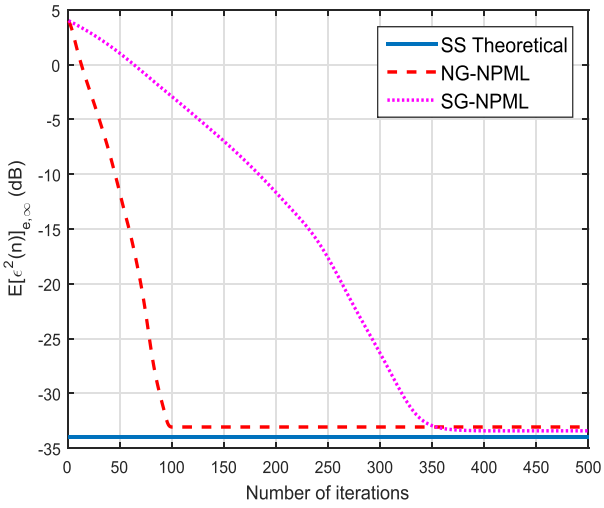

FIGURE 4. Theoretical and Monte Carlo simulation of EMSE of residual error for SG-NPML and NG-NPML at SIR 5dB and SNR 30dB.

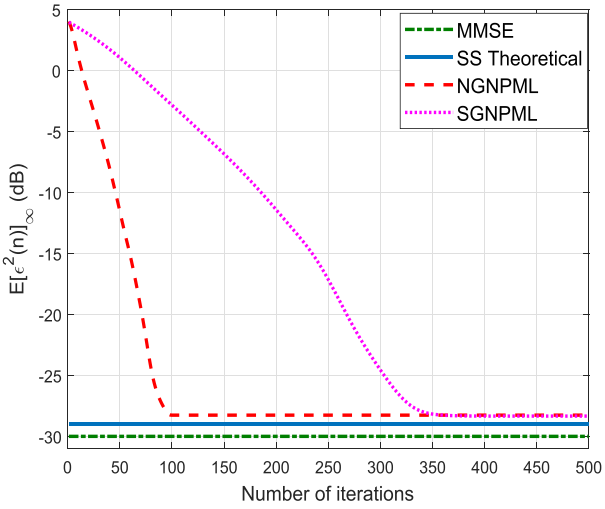

FIGURE 5. Theoretical and Monte Carlo simulation of MSE of residual error for SG-NPML and NG-NPML at SIR 5dB and SNR 30dB.

steady-state as given in *Corollary 1*. Fig. 5 shows good agreement between the theoretical expression of steady-state MSE ($\mathbb{E}[\epsilon^2(n)]_\infty$) calculated in (82) and Monte Carlo simulation of proposed algorithm. Again for comparison, we also show the Monte Carlo simulation of the SG-NPML algorithm. In Fig. 5, the legend 'SS Theoretical' represents the theoretical steady-state of MSE. We have chosen MMSE (ξ_0) = -30 dB (which is equivalent to the variance of AWGN) since the NPML based algorithms mitigate the effect of additive interference.

B. EXPERIMENTAL RESULTS

In this subsection, we validate robustness and applicability of the proposed NG-NPML algorithm in the presence of system nonlinearities, fixed point implementation and other impairments introduced by the hardware. The performance of proposed algorithm is also tested practically for IEEE 802.22 based transceiver using NI-USRP 2952R. The USRP is a software defined radio (SDR) based radio frequency (RF) hardware designed to test digital communication systems [44]. With support from Laboratory Virtual Instrument Engineering Workbench (LabVIEW) which is a graphical programming software developed by National Instruments, real time processing is made to and from USRP via Ethernet connection [45]. Detailed specification of USRP 2925R is given in [46]. In the case of interference from another IEEE 802.22 transmitter, the experimental setup using USRP boards is shown in Fig. 6.

The FFT size of OFDM and other specification is as per IEEE 802.22 standard. The frame synchronization and channel estimation are done with the help of preamble which consists of the short training sequence (STS) and LTS of frame of length 2560 symbols including 512 symbols for cyclic prefix. The training sequence is BPSK modulated, while the data sequence was QPSK modulated and then transmitted at a symbol rate of 312.5K symbols/sec onto a 450MHz carrier frequency. Long distance transmission is not possible due to the limited power of USRP; hence we used a Typical Urban COST-207 sparse channel to emulate sparsity before transmitting the signal over the air. In this paper, two types of CCI are considered as follows:

1) ANOTHER IEEE 802.22 AS AN INTERFERER

This interference is created by using another USRP 2952R with same IEEE 802.22 specification at 450MHz. In Fig. 6, the left hand side and right hand side USRP act as transceiver and interferer respectively. The distance between transmit and receive antenna is close hence the effect of the real channel is less as compared to emulated channel. The distance between receiver and interferer is 1.2 meter. The MSE performance of both SG-NPML and NG-NPML algorithm is shown in Fig. 7 for 5dB SIR and 30dB SNR. This experimental result also validates the faster convergence rate of proposed NG-NPML algorithm as shown in the simulation results in Fig. 4 and 5.

2) DTTV AS AN INTERFERER

In this part, we have considered two cases. In the first case, real time DTTV signal acts as an interference and in the second case, the recorded Advanced Television Systems Committee (ATSC) DTTV signal [47] is used as interference. In the case of real time DTTV signal, the desired signal is transmitted at 514 MHz and 519 MHz with 6 MHz bandwidth, since the DTTV signal in Indore, India is present at 514 MHz with bandwidth of 8 MHz [48]. The DTTV transmitter transmits at 6.4 KW power and 6 Km far from the receiver. At 514 MHz, the whole desired signal experiences interference by DTTV and hence results in low SIR (5 dB). However, at 519 MHz, only a small part of the desired signal is affected and leading to high SIR (20 dB). Fig. 8 shows the MSE of SG-NPML and NG-NPML at 30 dB SNR, and at 514 MHz and 519 MHz. We observe that the proposed NG-NPML converges much faster than SG-NPML.

FIGURE 6. Experimental setup for IEEE 802.22 acts as an interferer.

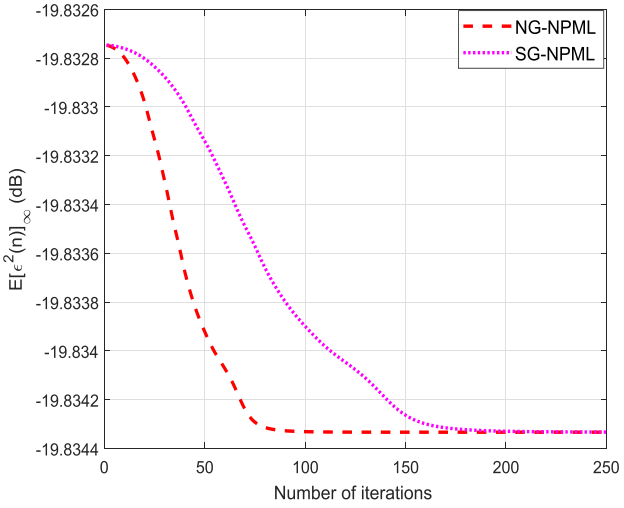

FIGURE 7. Experimental result of MSE of residual error for SG-NPML and NG-NPML at SIR 5dB and SNR 30dB.

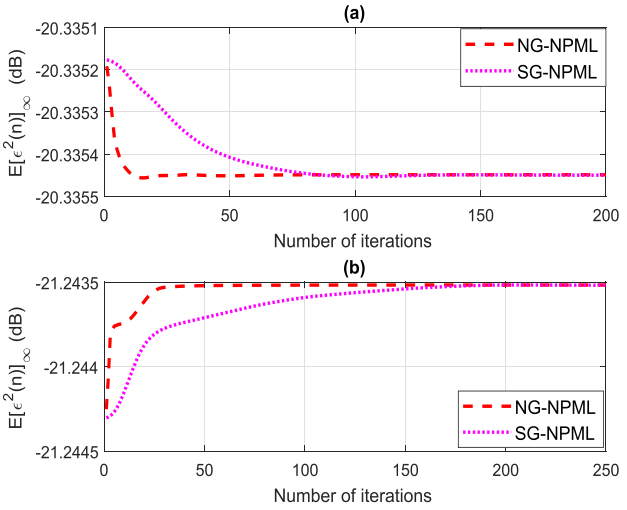

FIGURE 8. Experimental result of MSE of residual error for SG-NPML and NG-NPML at SNR 30dB (a) 514 MHz with 5 dB SIR (b) 519 MHz with 20 dB SIR.

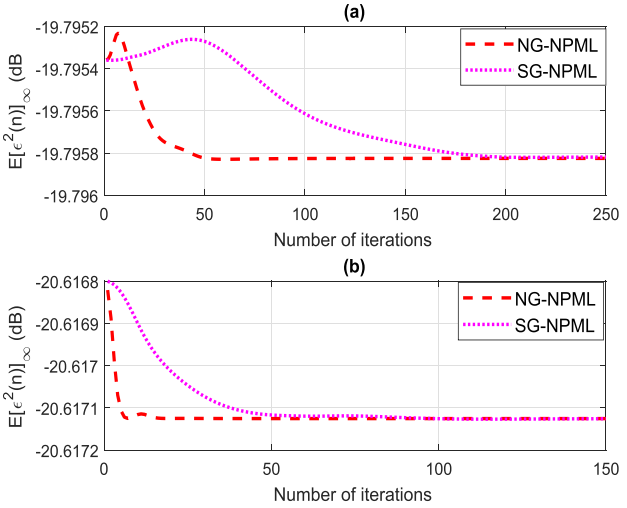

FIGURE 9. Experimental result of MSE of residual error for SG-NPML and NG-NPML at SNR 30dB (a) 514 MHz with 5 dB SIR (b) 519 MHz with 20 dB SIR without knowing the active tap positions.

Fig. 9 shows the same MSE plot as shown in Fig. 8, however without knowing the active tap positions and hence, the MSE floor of Fig. 9 is higher than Fig. 8. Fig. 9 also shows that even without knowing the active tap positions the proposed NG-NPML converges much faster than the SG-NPML. Fig. 10 shows the constellation diagram of the received signal before and after NG-NPML based channel estimation and equalization at 514 MHz and 519 MHz. It is observed from Fig. 10 (b) and (d) that the received signal after equalization is less affected by interference at 519 MHz as compared to 514 MHz, since at 514 MHz SIR is high as compared to 519 MHz.

Fig. 11 (a) and (b) show the MSE performance of SG-NPML and NG-NPML at 30 dB SNR and 5 dB SIR for recorded ATSC DTTV signal in New York city (NYC/200/44/01) and Washington DC (WAS-082/35/01) respectively as an interference. Fig. 11 again validates the faster convergence of proposed algorithm. It is observed from the Fig. 7-9, and Fig. 11 that the MSE of residual error starts at a very low value due to close proximity of transmit and receive antennas as shown in Fig. 6. Since the transmit and receive antennas are close to each other, the transmit power is reduced. In order to reduce transmit power in the experimental setup, we can either reduce the gain of the USRP or scale down the signal before transmission. In this paper, we scale down the signal before transmission, since the lowest gain of the USRP leads the received signal in nonlinear range. Hence, the scaling down of the signal before transmission leads to a lower absolute value of MSE even at initial iterations as observed in Fig. 7-9, and Fig. 11.

There is some difference in convergence rate, and MSE of the residual error between simulation and experimental results due to the difference in the real and simulated channel. This also results in different channel coefficients for

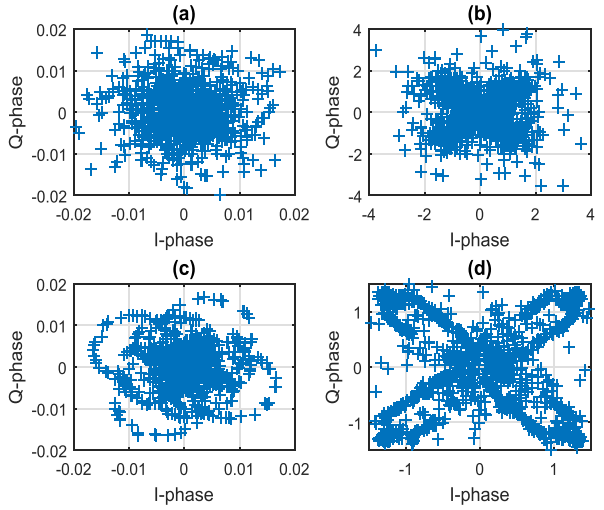

FIGURE 10. Constellation diagram of received signal at (a) 514 MHz before equalization (b) 514 MHz after equalization (c) 519 MHz before equalization (d) 519 MHz after equalization.

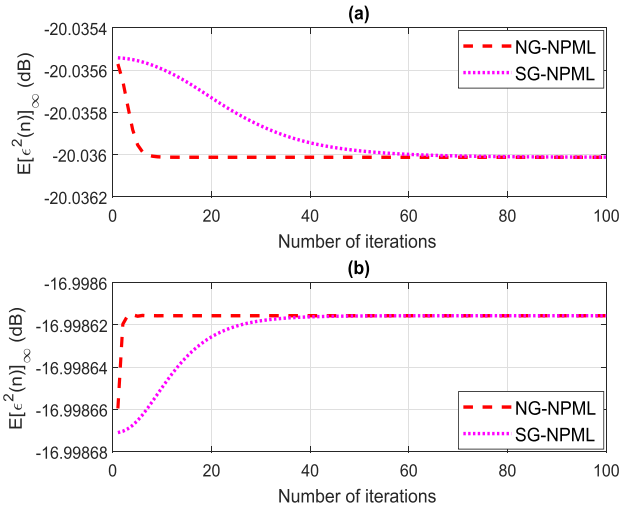

FIGURE 11. Experimental result of MSE of residual error for SG-NPML and NG-NPML at SIR 5dB and SNR 30dB for recorded ATSC DTTV signal acts as an interferer.

simulation and experimental results, less ensemble for the experimental result (ensemble for different value of channel gains), artifacts (such as synchronization, frequency offset etc.), limitations of the hardware, non-synchronized interference, and dynamicity of noise and interference. Hence, we have plotted separate graphs for simulation and experimental results. However, both the simulation and experimental results validate faster convergence rate of the proposed NG-NPML algorithm over the conventional SG-NPML. It is to be noted that the proposed algorithm is not limited to IEEE 802.22 standard, it can be applied to any OFDM based system like Long-Term Evolution in unlicensed spectrum for sparse channel estimation.

C. COMPLEXITY ANALYSIS

We have calculated the per iteration computational cost of SG-NPML and the proposed NG-NPML algorithm for real data in terms of the number of additions, multiplications, and divisions. While calculating the computational cost, constant term multiplication is ignored (using a look up table implementation) and the value of σ^2 is assumed to be known fixed value. Table 1 describes the estimated computational cost of SG-NPML based on (8). In Table 1, as $e_{ij} = -e_{ji}$ in term 3^{rd}, hence e_{ji} is not calculated. From the Table, the 6^{th} term, the exponential term is calculated using explicitly lookup table. In term 9^{th}, X_{ij} is a vector of dimension $L \times 1$ and due to the constant multiplication in 5^{th} and 7^{th} term, hence the computational cost of operation is ignored. Thus, the approximated cost per iteration for SG-NPML is $O(M^2L)$ multiplications, $O(M^2L)$ additions and $O(M)$ division.

TABLE 1. Estimated computational cost of SG-NPML per iteration.

S. NO.	TERM	x	+	/
1.	$\hat{\mathbf{h}}_i\mathbf{x}(i)$	ML	$M(L-1)$	-
2.	$e(i) = y(i) - \hat{\mathbf{h}}_i\mathbf{x}(i)$	-	M	-
3.	$e_{ij(i\neq j)} = e(i) - e(j)$	-	$\frac{M^2-M}{2}$	-
4.	e_{ij}^2	$\frac{M^2-M}{2}$	-	-
5.	$\frac{e_{ij}^2}{2\sigma^2}$	-	-	-
6.	$p_{ij} = \exp\left(-\frac{e_{ij}^2}{2\sigma^2}\right)$	-	-	-
7.	$k_{ij} = \frac{p_{ij}}{\sqrt{2\pi\sigma^2}}$	-	-	-
8.	$q_i = \sum_{j=1}^{M} k_{ij}$	-	$M^2 - M$	-
9.	$\mathbf{q}_{ij(i\neq j)} = [\mathbf{x}(i) - \mathbf{x}(j)]$	-	$L\left(\frac{M^2-M}{2}\right)$	-
10.	$\mathbf{r}_{i(i\neq j)} = \sum_{j=1}^{M} e_{ij}\mathbf{q}_{ij}k_{ij}$	$2LM^2$ $-2LM$	$M^2 - 2M$	-
11.	$\mathbf{t} = \sum_{i=1}^{M} \frac{\mathbf{r}_i}{q_i}$	-	$M-1$	M
12.	$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k + \frac{\mu\mathbf{t}}{\sigma^2}$	-	L	-
	TOTAL	$2M^2L + \frac{M^2}{2}$ $-ML - \frac{M}{2}$	$\frac{5M^2}{2} + \frac{LM^2}{2}$ $+\frac{ML}{2} + L$ $-\frac{5M}{2} - 1$	M

Table 2 describes the additional terms computational cost of NG-NPML based on (18). In the 1^{st} term, the sign of $\hat{\mathbf{h}}_k$ is multiplied by itself to obtain $|\hat{\mathbf{h}}_k|$, hence it require L multiplications, and then creates a diagonal matrix of dimension $L \times L$ with $|\hat{h}_k(l)|$ term as diagonal element. In the 3^{rd} term, the computational complexity of α_k is calculated using (13). In the 6^{th} term, $\mathbf{t}$ is a vector of dimension $1 \times L$ and $\mathbf{W}$ is diagonal matrix of dimension $L \times L$. Hence, the total computational cost is the sum of Table 1 and 2 for NG-NPML. The computational cost of the gain matrix of proposed algorithm is $O(L)$, since $M \gg L$. Hence, the total estimated cost of NG-NPML per iteration is $O(M^2L)$ multiplications, $O(M^2L)$ additions and $O(M)$ division. Thus, we can conclude that the proposed NG-NPML has same computational

TABLE 2. Estimated computational cost of additional terms in NG-NPML per iteration for quadratic warping.

S. NO.	TERM	x	+	/
1.	$\lvert\hat{\mathbf{h}}_k\rvert$	L	-	-
2.	$\lvert\hat{\mathbf{h}}_k\rvert^2$	L	-	-
3.	α_k (as given in (13)	L	$2L$	-
4.	$\mathbf{\Gamma}_2$ (as given in (35))	L	L	-
5.	$\mathbf{W}=\left(\left\lvert\lvert\hat{\mathbf{H}}_k\rvert-\lvert\hat{\mathbf{H}}_k\rvert^2\right\rvert+\beta\mathbf{I}\right)/(\alpha_k\mathbf{\Gamma}_2)$	L	L	L
6.	$\mathbf{tW}$	L	-	-
	TOTAL	$6L$	$4L$	L

complexity as SG-NPML per iteration. However, as the order of transformation increases (order of $p > 2$), additional terms of NG-NPML requires additional L multiplications and L additions for each increase of the order of transformation. It is observed from Fig. 8, Fig. 9, and Fig. 11 that in terms of total computational complexity NG-NPML is 8-10 times faster than SG-NPML. The computational complexity of NG-NPML is slightly high as compared to OMP ($O(ML)$) and LASSO ($O(L^3)$), and high as compared to NGMAPA ($O(M)$) and NGpNAPA ($O(L)$) per iteration. However, the improvement in MSE/bit error rate is substantial by using proposed algorithm. Additionally, the complexity of PDF estimation in NG-NPML can be reduced by using fast or reduced density estimation techniques.

VI. CONCLUSION

In this paper, we proposed NG-NPML algorithm for sparse channel estimation in the interference limited environments. NG is applied on coefficient space which is transformed from Euclidean distance space to Riemannian space by using warping transformation function. In this paper, we used quadratic warping transformation to balance the computational complexity of transformation and faster convergence rate. The proposed algorithm is found to be robust to CCI as compared with other algorithms. The proposed algorithm is applied on IEEE 802.22 based transceiver and the simulation results, as well as experimental results, show that the proposed NG-NPML algorithm has much faster convergence compared to conventional SG-NPML in high interference.

We also derive the stability condition of proposed algorithm in terms of mean weight error and the approximated analytical expression of steady state MSE and EMSE. The simulation results show good agreement with the derived expressions.

APPENDIX A
DERIVATION OF (16)

Case 1 ($\hat{h}_k(l)r_{k+1}(l) > 0$): For the first case, after expanding (15) yields

$$D=\sum_{l=0}^{L-1}\left|\sqrt{A+\alpha_k|r_{k+1}(l)|-2B}-\sqrt{A}\right|^2 \tag{24}$$

where

$$A=\alpha_k\left(\left||\hat{h}_k(l)|-|h_k(l)|^2\right|+\beta\right) \tag{25}$$

$$B=\alpha_k|\hat{h}_k(l)||r_{k+1}(l)| \tag{26}$$

and $\alpha_k|r_{k+1}^2(l)|$ is neglected. The (24) can be written as:

$$D=\sum_{l=0}^{L-1}\left|\sqrt{A}\left(1+\frac{\alpha_k|r_{k+1}(l)|\left(1-2|\hat{h}_k(l)|\right)}{A}\right)^{1/2}-\sqrt{A}\right|^2 \tag{27}$$

By using Taylor series expansion and keeping the significant terms, the above equation can be approximated as:

$$D\approx\sum_{l=0}^{L-1}\left|\sqrt{A}\left(1+\frac{\alpha_k|r_{k+1}(l)|\left(1-2|\hat{h}_k(l)|\right)}{2A}\right)-\sqrt{A}\right|^2 \tag{28}$$

The above equation can be further simplified as:

$$D\approx\sum_{l=0}^{L-1}\frac{\alpha_k^2 r_{k+1}(l)r_{k+1}(l)\left(1-2|\hat{h}_k(l)|\right)^2}{4A} \tag{29}$$

The above equation can be written as:

$$D\approx\frac{\mathbf{r}_{k+1}^T\alpha_k\left(\mathbf{I}-2|\hat{\mathbf{H}}_k|\right)^2\mathbf{r}_{k+1}}{4\left(\left||\hat{\mathbf{H}}_k|-|\hat{\mathbf{H}}_k|^2\right|+\beta\mathbf{I}\right)} \tag{30}$$

$|\hat{\mathbf{H}}_k|$ is given as:

$$|\hat{\mathbf{H}}_k|=\begin{bmatrix}|\hat{h}_k(0)| & 0 & . & . & 0\\ 0 & |\hat{h}_k(1)| & . & . & 0\\ . & & . & & .\\ . & & & . & .\\ 0 & 0 & . & . & |\hat{h}_k(L-1)|\end{bmatrix} \tag{31}$$

By comparing (11) with (30), we get the Riemannian metric tensor as:

$$\mathbf{G}_{\hat{\mathbf{h}}_k}=\frac{\alpha_k\left(\mathbf{I}-2|\hat{\mathbf{H}}_k|\right)^2}{4\left(\left||\hat{\mathbf{H}}_k|-|\hat{\mathbf{H}}_k|^2\right|+\beta\mathbf{I}\right)} \tag{32}$$

Hence,

$$\mathbf{G}_{\hat{\mathbf{h}}_k}^{-1}=\frac{4\left(\left||\hat{\mathbf{H}}_k|-|\hat{\mathbf{H}}_k|^2\right|+\beta\mathbf{I}\right)}{\alpha_k\left(\mathbf{I}-2|\hat{\mathbf{H}}_k|\right)^2} \tag{33}$$

The β should be small so that it does not dominate the Riemannian metric tensor at any stage of iteration. After substituting (33) in (9) and dropping the constant "4", the NG update of channel at $(k+1)^{th}$ iteration is given by:

$$\hat{\mathbf{h}}_{k+1}=\hat{\mathbf{h}}_k+\mu\frac{\left(\left||\hat{\mathbf{H}}_k|-|\hat{\mathbf{H}}_k|^2\right|+\beta\mathbf{I}\right)}{\alpha_k\left(\mathbf{I}-2|\hat{\mathbf{H}}_k|\right)^2}\nabla_{\hat{\mathbf{h}}_k}\mathcal{J}(\hat{\mathbf{h}}_k) \tag{34}$$

Assume that at any iteration $(\mathbf{I} - 2|\hat{\mathbf{H}}_k|)$=0, then the above equation goes to infinity and there is no solution. Hence, to overcome this problem we modified (34) as:

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k + \mu \frac{\left(\left||\hat{\mathbf{H}}_k| - |\hat{\mathbf{H}}_k|^2\right| + \beta\mathbf{I}\right)}{\alpha_k \mathbf{\Gamma}_2} \nabla_{\hat{\mathbf{h}}_k} \mathcal{J}(\hat{\mathbf{h}}_k) \quad (35)$$

where $\mathbf{\Gamma}_2 = (\mathbf{I} + 2|\hat{\mathbf{H}}_k|)^2$.

Case 2 ($\hat{h}_k(l) r_{k+1}(l) < 0$): For the second case, after expanding (15) yields

$$D = \sum_{l=0}^{L-1} \left|\sqrt{A - \alpha_k |r_{k+1}(l)| + 2B} - \sqrt{A}\right|^2 \quad (36)$$

where A and B are given by (25) and (26) respectively. After solving the above equation as in *Case 1*, the channel update equation for the *Case 2* is given by:

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k + \mu \frac{\left(\left||\hat{\mathbf{H}}_k| - |\hat{\mathbf{H}}_k|^2\right| + \beta\mathbf{I}\right)}{\alpha_k \left(\mathbf{I} + 2|\hat{\mathbf{H}}_k|\right)^2} \nabla_{\hat{\mathbf{h}}_k} \mathcal{J}(\hat{\mathbf{h}}_k) \quad (37)$$

Hence, for both the cases, the final channel update equation is given by (37).

APPENDIX B
DERIVATION OF (17)

After substituting (12) into (10), the distance metric for $p = 3$ is given by (38), as shown at the bottom of this page. For $\hat{h}_k^2(l) r_{k+1}(l) < 0$, after expanding (38) yield (39)

$$D = \sum_{l=0}^{L-1} \left|\sqrt{A + B} - \sqrt{A}\right|^2 \quad (39)$$

where

$$A = \alpha_k \left(\left||\hat{h}_k(l)| - |h_k(l)|^2 - |h_k(l)|^3\right| + \beta\right) \quad (40)$$

$$B = \alpha_k |r_{k+1}(l)| + 2\alpha_k |\hat{h}_k(l)||r_{k+1}(l)| + 3\alpha_k |\hat{h}_k(l)|^2 |r_{k+1}(l)| \quad (41)$$

and $\alpha_k |r_{k+1}^2(l)|$, $\alpha_k |r_{k+1}^3(l)|$, $\alpha_k |\hat{h}_k(l)||r_{k+1}^2(l)|$ are neglected because of $|r_{k+1}(l)| \ll 1$. The (39) can be written as (42), as shown at the bottom of this page.

By using Taylor series expansion and keeping the first two terms only, the (42) can be approximated as (43), as shown at the bottom of this page. The (43) can be further simplified as:

$$D \approx \sum_{l=0}^{L-1} \frac{\alpha_k^2 r_{k+1}(l) r_{k+1}(l) \left(1 + 2|\hat{h}_k(l)| + 3|\hat{h}_k(l)|^2\right)^2}{4A} \quad (44)$$

The above equation can be written as:

$$D \approx \frac{\mathbf{r}_{k+1}^T \alpha_k \left(\mathbf{I} + 2|\hat{\mathbf{H}}_k| + 3|\hat{\mathbf{H}}_k|^2\right)^2 \mathbf{r}_{k+1}}{4\left(\left||\hat{\mathbf{H}}_k| - |\hat{\mathbf{H}}_k|^2 - |\hat{\mathbf{H}}_k|^3\right| + \beta\mathbf{I}\right)} \quad (45)$$

By comparing (11) with (45), we can get inverse of Riemannian metric tensor as:

$$\mathbf{G}_{\hat{\mathbf{h}}_k}^{-1} = \frac{4\left(\left||\hat{\mathbf{H}}_k| - |\hat{\mathbf{H}}_k|^2 - |\hat{\mathbf{H}}_k|^3\right| + \beta\mathbf{I}\right)}{\alpha_k \left(\mathbf{I} + 2|\hat{\mathbf{H}}_k| + 3|\hat{\mathbf{H}}_k|^2\right)^2} \quad (46)$$

Similarly for p^{th} order transformation, the above equation can be written as:

$$\mathbf{G}_{\hat{\mathbf{h}}_k}^{-1} = \frac{\left(\left||\hat{\mathbf{H}}_k| - \ldots - |\hat{\mathbf{H}}_k|^p\right| + \beta\mathbf{I}\right)}{\alpha_k \left(\mathbf{I} + 2|\hat{\mathbf{H}}_k| + \ldots + p|\hat{\mathbf{H}}_k|^{p-1}\right)^2} \quad (47)$$

After substituting (47) into (9), the channel update equation at $(k+1)^{th}$ iteration for p^{th} order transformation is given by:

$$\hat{\mathbf{h}}_{k+1} \approx \hat{\mathbf{h}}_k + \mu \frac{\left(\left||\hat{\mathbf{H}}_k| - \ldots - |\hat{\mathbf{H}}_k|^p\right| + \beta\mathbf{I}\right)}{\alpha_k \mathbf{\Gamma}_p} \nabla_{\hat{\mathbf{h}}_k} \mathcal{J}(\hat{\mathbf{h}}_k) \quad (48)$$

APPENDIX C
DERIVATION OF (21)

Let $\tilde{\mathbf{h}}_k$ is the weight error vector which is defined as:

$$\tilde{\mathbf{h}}_k = \mathbf{h} - \hat{\mathbf{h}}_k \quad (49)$$

then the transformed domain weight error vector can be represented as:

$$\acute{\mathbf{h}}_k = \mathbf{G}_{\hat{\mathbf{h}}_k}^{-1/2} \tilde{\mathbf{h}}_k = \mathbf{G}_{\hat{\mathbf{h}}_k}^{-1/2} \mathbf{h} - \breve{\mathbf{h}}_k \quad (50)$$

From Fig. 2, $\epsilon(n) = \mathbf{h}^T \mathbf{x}(n) + e(n) - \hat{\mathbf{h}}_k^T \mathbf{s}(n)$, which can be further written as $\epsilon(n) = \mathbf{h}^T \mathbf{G}_{\hat{\mathbf{h}}_k}^{-1/2} \mathbf{s}(n) + e(n) - \breve{\mathbf{h}}_k^T \mathbf{s}(n)$ and finally

$$\epsilon(n) = \acute{\mathbf{h}}_k^T \mathbf{s}(n) + e(n) \quad (51)$$

$$D = \sum_{l=0}^{L-1} \left|\sqrt{\alpha_k \left(\left||\hat{h}_k(l) + r_{k+1}(l)| - |\hat{h}_k(l) + r_{k+1}(l)|^2 - |\hat{h}_k(l) + r_{k+1}(l)|^3\right| + \beta\right)} - \sqrt{\alpha_k \left(\left||\hat{h}_k(l)| - |\hat{h}_k(l)|^2 - |\hat{h}_k(l)|^3\right| + \beta\right)}\right|^2 \quad (38)$$

$$D = \sum_{l=0}^{L-1} \left|\sqrt{A}\left(1 + \frac{\alpha_k r_{k+1}(l)\left(1 + 2|\hat{h}_k(l)| + 3|\hat{h}_k(l)|^2\right)}{A}\right)^{1/2} - \sqrt{A}\right|^2 \quad (42)$$

$$D \approx \sum_{l=0}^{L-1} \left|\sqrt{A}\left(1 + \frac{\alpha_k r_{k+1}(l)\left(1 + 2|\hat{h}_k(l)| + 3|\hat{h}_k(l)|^2\right)}{2A}\right) - \sqrt{A}\right|^2 \quad (43)$$

or

$$\epsilon_k = \acute{\mathbf{h}}_k^T \tilde{\mathbf{q}}_k + \tilde{e} = \tilde{\mathbf{q}}_k^T \acute{\mathbf{h}}_k + \tilde{e} \tag{52}$$

Now (20) can be written in terms of transformed weight error vector as:

$$\mathbf{G}_{\hat{\mathbf{h}}_{k+1}}^{-1/2}\mathbf{h} - \acute{\mathbf{h}}_{k+1} = \mathbf{G}_{\hat{\mathbf{h}}_k}^{-1/2}\mathbf{h} - \acute{\mathbf{h}}_k + \frac{\mu}{\sigma^2}\sum_{i=1}^{M}\frac{\sum_{j=1}^{M}\epsilon_k\tilde{\mathbf{q}}_k K(\epsilon_k)}{\sum_{j=1}^{M}K(\epsilon_k)} \tag{53}$$

We can assume $\mathbf{G}_{\hat{\mathbf{h}}_{k+1}}^{-1/2}\mathbf{h} \approx \mathbf{G}_{\hat{\mathbf{h}}_k}^{-1/2}\mathbf{h}$ because the gain matrix do not change significantly from k^{th} to $(k+1)^{th}$ index (near convergence), so the above equation can be written as:

$$\acute{\mathbf{h}}_{k+1} = \acute{\mathbf{h}}_k - \frac{\mu}{\sigma^2}\sum_{i=1}^{M}\frac{\sum_{j=1}^{M}\tilde{\mathbf{q}}_k\epsilon_k K(\epsilon_k)}{\sum_{j=1}^{M}K(\epsilon_k)} \tag{54}$$

After substituting (52) into (54), the transformed weight error vector at $(k+1)^{th}$ iteration is given as:

$$\acute{\mathbf{h}}_{k+1} = \acute{\mathbf{h}}_k - \frac{\mu}{\sigma^2}\sum_{i=1}^{M}\frac{\sum_{j=1}^{M}\tilde{\mathbf{q}}_k(\tilde{\mathbf{q}}_k^T\acute{\mathbf{h}}_k + \tilde{e})K(\epsilon_k)}{\sum_{j=1}^{M}K(\epsilon_k)} \tag{55}$$

Let $\Delta_k = \mathbb{E}[\acute{\mathbf{h}}_k]$, then the above equation can be written as:

$$\Delta_{k+1} \approx \Delta_k - \frac{\mu M}{\sigma^2}\frac{\left(\mathbb{E}[\tilde{\mathbf{q}}_k\tilde{\mathbf{q}}_k^T\acute{\mathbf{h}}_k K(\epsilon_k)] + \mathbb{E}[\tilde{\mathbf{q}}_k\tilde{e}K(\epsilon_k)]\right)}{\mathbb{E}[K(\epsilon_k)]} \tag{56}$$

Without loss of generality, $\acute{\mathbf{h}}_k$ is independent with $\tilde{\mathbf{q}}_k$ and $K(\epsilon_k)$, and therefore $\acute{\mathbf{h}}_k$ with $\tilde{\mathbf{q}}_k$ and $K(\epsilon_k)$ [31], [49]. We also assume that $\tilde{\mathbf{q}}_k$ is asymptotically uncorrelated with $K(\epsilon_k)$ [36], [50], [51]. Hence, by using above assumptions, $\mathbb{E}[\tilde{\mathbf{q}}_k\tilde{\mathbf{q}}_k^T\acute{\mathbf{h}}_k K(\epsilon_k)]$ can be written as $\mathbb{E}[\tilde{\mathbf{q}}_k\tilde{\mathbf{q}}_k^T]\Delta_k\mathbb{E}[K(\epsilon_k)]$. Further, $\mathbb{E}[\tilde{\mathbf{q}}_k\tilde{e}K(\epsilon_k)] = 0$ as $\tilde{\mathbf{q}}_k$ is a zero mean and statistically independent of $\tilde{e}$.

The above equation can then be written as:

$$\Delta_{k+1} \approx \Delta_k - \frac{\mu M}{\sigma^2}\mathbf{R}_{s,k}\Delta_k = \left(\mathbf{I} - \frac{\mu M}{\sigma^2}\mathbf{R}_{s,k}\right)\Delta_k \tag{57}$$

For l^{th} tap, we have

$$\Delta_{k+1}(l) \approx \left(1 - \frac{\mu M}{\sigma^2}\lambda_{s,k}(l)\right)\Delta_k(l) \tag{58}$$

For stability or convergence of the proposed algorithm, the following condition satisfy

$$-1 < 1 - \frac{\mu M}{\sigma^2}\lambda_{s,k}(l) < 1 \quad \forall l \tag{59}$$

For the stability of the proposed NG-NPML algorithm, the step-size μ satisfying

$$0 < \mu < \frac{2\sigma^2}{M\sum_{l=1}^{L}\lambda_{s,k}(l)} \tag{60}$$

It is observed from the above equation that the upper bound (on right hand side of (60)) on μ varies at each iteration. However, this upper bound does not change significantly near convergence since $Tr(\mathbf{R}_{s,k}) = Tr(\mathbb{E}[\mathbf{G}_{\hat{\mathbf{h}}_k}^{1/2}\mathbf{R}\mathbf{G}_{\hat{\mathbf{h}}_k}^{1/2}])$, $\mathbf{R}$ is constant (as $\mathbf{q}$ is constant), and it is assumed that near convergence $\mathbf{G}_{\hat{\mathbf{h}}_k}$ does not change significantly, and hence from (60), the upper bound is constant near convergence.

APPENDIX D
DERIVATION OF (22)

By using (51), the MSE is given by:

$$\xi_k = \xi_0 + \xi_{e,k} \tag{61}$$

where $\xi_k = \mathbb{E}[\epsilon_k^2]$, $\xi_{e,k} = \mathbb{E}[\acute{\mathbf{h}}_k^T\tilde{\mathbf{q}}_k\tilde{\mathbf{q}}_k^T\acute{\mathbf{h}}_k]$ is the EMSE and cross term is zero. To evaluate the steady-state MSE, firstly we calculate the steady-state EMSE. Let us define

$$f(\epsilon_k) = \epsilon_k K(\epsilon_k) = \epsilon_k \exp\left(\frac{-\epsilon_k^2}{2\sigma^2}\right) \tag{62}$$

$$\tilde{f}(\epsilon_k) = K(\epsilon_k) = \exp\left(\frac{-\epsilon_k^2}{2\sigma^2}\right) \tag{63}$$

and

$$\zeta_k = \acute{\mathbf{h}}_k^T\tilde{\mathbf{q}}_k \tag{64}$$

Hence, the steady-state EMSE is given by:

$$\xi_{e,\infty} = \lim_{k\to\infty}\mathbb{E}[\zeta_k^2] \tag{65}$$

By using (59) and some mathematical manipulation and approximation, the following relation holds for energy conservation [36], [50], [51]:

$$\mathbb{E}[||\acute{\mathbf{h}}_{k+1}||^2] \approx \mathbb{E}[||\acute{\mathbf{h}}_k||^2] - \frac{2\mu M\mathbb{E}[\zeta_k f(\epsilon_k)]}{\sigma^2\mathbb{E}[\tilde{f}(\epsilon_k)]} + \frac{\mu^2M^2\mathbb{E}[||\tilde{\mathbf{q}}_k||^2 f^2(\epsilon_k)]}{\sigma^4\mathbb{E}[\tilde{f}^2(\epsilon_k)]} \tag{66}$$

where $||\tilde{\mathbf{q}}_k||^2 = Tr(\mathbf{R}_{s,k})$. Assume that the estimator is in steady-state such that

$$\lim_{k\to\infty}\mathbb{E}[||\acute{\mathbf{h}}_{k+1}||^2] = \lim_{k\to\infty}\mathbb{E}[||\acute{\mathbf{h}}_k||^2] \tag{67}$$

Hence, in the steady-state (67) becomes

$$2\lim_{k\to\infty}\frac{\mathbb{E}[\zeta_k f(\epsilon_k)]}{\mathbb{E}[\tilde{f}(\epsilon_k)]} = \frac{\mu M}{\sigma^2}\lim_{k\to\infty}\frac{\mathbb{E}[||\tilde{\mathbf{q}}_k||^2 f^2(\epsilon_k)]}{\mathbb{E}[\tilde{f}^2(\epsilon_k)]} \tag{68}$$

We do Taylor series expansion of the function $f(.)$ [51] for the derivation of steady-state EMSE. Taking the Taylor series expansion of $f(\epsilon_k)$ and $\tilde{f}(\epsilon_k)$ with respect to ζ_k around $\tilde{e}$ yields

$$f(\epsilon_k) = f(\zeta_k + \tilde{e}) = f(\tilde{e}) + f'(\tilde{e})\zeta_k + \frac{1}{2}f''(\tilde{e})\zeta_k^2 + o(\zeta_k^2) \tag{69}$$

and

$$\tilde{f}(\epsilon_k) = \tilde{f}(\zeta_k + \tilde{e}) = \tilde{f}(\tilde{e}) + \tilde{f}'(\tilde{e})\zeta_k + \frac{1}{2}\tilde{f}''(\tilde{e})\zeta_k^2 + o(\zeta_k^2) \tag{70}$$

$$2\frac{\mathbb{E}[f'(\tilde{e})]\xi_{e,k}}{\mathbb{E}[\tilde{f}(\tilde{e})]+\frac{1}{2}\mathbb{E}[\tilde{f}''(\tilde{e})]\xi_{e,k}}=\frac{\mu MTr(\mathbf{R}_{s,k})\Big(\mathbb{E}[f^2(\tilde{e})]+\mathbb{E}[f(\tilde{e})f''(\tilde{e})+|f'(\tilde{e})|^2]\xi_{e,k}\Big)}{\sigma^2\Big(\mathbb{E}[\tilde{f}^2(\tilde{e})]+\mathbb{E}[\tilde{f}(\tilde{e})\tilde{f}''(\tilde{e})+|\tilde{f}'(\tilde{e})|^2]\xi_{e,k}\Big)} \tag{75}$$

$$\xi_{e,k}\approx\frac{\mu MTr(\mathbf{R}_{s,k})\mathbb{E}[f^2(\tilde{e})]\mathbb{E}[\tilde{f}(\tilde{e})]}{2\sigma^2\mathbb{E}[f'(\tilde{e})]\mathbb{E}[\tilde{f}^2(\tilde{e})]-\mu MTr(\mathbf{R}_{s,k})\Big(\mathbb{E}[\tilde{f}(\tilde{e})]\mathbb{E}[f(\tilde{e})f''(\tilde{e})+|f'(\tilde{e})|^2]+\mathbb{E}[f^2(\tilde{e})]\frac{1}{2}\mathbb{E}[\tilde{f}''(\tilde{e})]\Big)} \tag{76}$$

where $o(\zeta_k^2)$ is the third and higher-order terms. We assume $\mathbb{E}[o(\zeta_k^2)]$ is very small and $||\tilde{\mathbf{q}}_k||^2$ is asymptotically uncorrelated with $f^2(\epsilon_k)$ [50], [52], then we derive

$$\mathbb{E}[\zeta_k f(\epsilon_k)]\approx\mathbb{E}[\zeta_k f(\tilde{e})+f'(\tilde{e})\zeta_k^2]\approx\mathbb{E}[f'(\tilde{e})]\xi_k \tag{71}$$

$$\mathbb{E}[f^2(\epsilon_k)]\approx\mathbb{E}[f^2(\tilde{e})]+\mathbb{E}[f(\tilde{e})f''(\tilde{e})+|f'(\tilde{e})|^2]\xi_k \tag{72}$$

$$\mathbb{E}[\tilde{f}(\epsilon_k)]\approx\mathbb{E}[\tilde{f}(\tilde{e})]+\frac{1}{2}\mathbb{E}[\tilde{f}''(\tilde{e})]\xi_k \tag{73}$$

and

$$\mathbb{E}[\tilde{f}^2(\epsilon_k)]\approx\mathbb{E}[\tilde{f}^2(\tilde{e})]+\mathbb{E}[\tilde{f}(\tilde{e})\tilde{f}''(\tilde{e})+|\tilde{f}'(\tilde{e})|^2]\xi_k \tag{74}$$

After substituting (71)-(74) into (68) yields (75), as shown at the top of this page. After some simple manipulation and assumption that $\xi_{e,k}^2$ is very small at steady-state, (75) can be approximated as (76), as shown at the top of this page. For small value of μ, (76) can be simplified as:

$$\xi_{e,k}\approx\frac{\mu MTr(\mathbf{R}_{s,k})\mathbb{E}[f^2(\tilde{e})]\mathbb{E}[\tilde{f}(\tilde{e})]}{2\sigma^2\mathbb{E}[f'(\tilde{e})]\mathbb{E}[\tilde{f}^2(\tilde{e})]} \tag{77}$$

Here,

$$f'(\tilde{e})=\exp\left(\frac{-\tilde{e}^2}{2\sigma^2}\right)\left(1-\frac{\tilde{e}^2}{\sigma^2}\right) \tag{78}$$

After substituting (62), (63) and (78) into (77), the steady state EMSE can be approximated as:

$$\lim_{k\to\infty}\xi_{e,k}=\lim_{k\to\infty}\frac{\mu MTr(\mathbf{R}_{s,k})\mathbb{E}[\exp\left(\frac{-\tilde{e}^2}{\sigma^2}\right)\tilde{e}^2]\mathbb{E}[\exp\left(\frac{-\tilde{e}^2}{2\sigma^2}\right)]}{2\sigma^2\mathbb{E}[\exp\left(\frac{-\tilde{e}^2}{2\sigma^2}\right)\left(1-\frac{\tilde{e}^2}{\sigma^2}\right)]\mathbb{E}[\exp\left(\frac{-\tilde{e}^2}{\sigma^2}\right)]} \tag{79}$$

At steady-state, the mean of residual error, $\mathbb{E}[\epsilon(n)]$, is zero and hence $\mathbb{E}[\epsilon_k]$ is also zero. Therefore, the EMSE of residual error at steady-state is given as:

$$\mathbb{E}[\epsilon^2(n)]_{e,\infty}=\lim_{k\to\infty}\xi_{e,k}/2 \tag{80}$$

After substituting (79) into (61), the steady-state MSE is given by:

$$\xi_\infty=\xi_0+\lim_{k\to\infty}\frac{\mu MTr(\mathbf{R}_{s,k})\mathbb{E}[\exp\left(\frac{-\tilde{e}^2}{\sigma^2}\right)\tilde{e}^2]\mathbb{E}[\exp\left(\frac{-\tilde{e}^2}{2\sigma^2}\right)]}{2\sigma^2\mathbb{E}[\exp\left(\frac{-\tilde{e}^2}{2\sigma^2}\right)\left(1-\frac{\tilde{e}^2}{\sigma^2}\right)]\mathbb{E}[\exp\left(\frac{-\tilde{e}^2}{\sigma^2}\right)]} \tag{81}$$

Similarly, the MSE of residual error at steady-state is given as:

$$\mathbb{E}[\epsilon^2(n)]_\infty=\xi_\infty/2 \tag{82}$$

ACKNOWLEDGEMENT

The authors would like to thank IIT Indore for all the support.

REFERENCES

[1] A. C. Cirik, O. Taghizadeh, L. Lampe, R. Mathar, and Y. Hua, "Linear transceiver design for full-duplex multi-cell MIMO systems," *IEEE Access*, vol. 4, pp. 4678–4689, Aug. 2016.

[2] P. K. Singya, N. Kumar, and V. Bhatia, "Mitigating NLD for wireless networks," *IEEE Microw. Mag.*, vol. 18, no. 5, pp. 73–90, Jun. 2017.

[3] V. Bhatia and B. Mulgrew, "Non-parametric likelihood based channel estimator for Gaussian mixture noise," *Elsevier Signal Process.*, vol. 87, no. 11, pp. 2569–2586, Nov. 2007.

[4] K. Pelekanakis and M. Chitre, "Adaptive sparse channel estimation under symmetric alpha-stable noise," *IEEE Trans. Wireless Commun.*, vol. 13, no. 6, pp. 3183–3195, Jun. 2014.

[5] Y. Chen, "Suboptimum detectors for AF relaying with Gaussian noise and SαS interference," *IEEE Trans. Vehi. Tech.*, vol. 64, no. 10, pp. 4833–4839, Oct. 2015.

[6] T. Zhang, H. Q. Jiao, and Z. C. Lei, "Individual-activation-factor memory proportionate affine projection algorithm with evolving regularization," *IEEE Access*, vol. 5, pp. 4939–4946, Mar. 2017.

[7] M. A. Chitre, J. R. Potter, and S. H. Ong, "Optimal and near-optimal signal detection in snapping shrimp dominated ambient noise," *IEEE J. Ocean. Eng.*, vol. 31, no. 2, pp. 497–503, Apr. 2006.

[8] V. G. Chavali and C. R. C. M. da Silva, "Classification of digital amplitude-phase modulated signals in time-correlated non-Gaussian channels," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2408–2419, Jun. 2013.

[9] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Boca Raton, FL, USA: CRC Press, 1998.

[10] V. Bhatia, B. Mulgrew, and D. D. Falconer, "Non-parametric maximum likelihood channel estimator for OFDM in presence of interference," *IET Commun.*, vol. 1, no. 4, pp. 647–654, Aug. 2007.

[11] V. Bhatia and B. Mulgrew, "A minimum error entropy based channel estimator in presence of co-channel interference with lower bounds," in *Proc. IEEE SPCOM*, Dec. 2004, pp. 41–45.

[12] W. Schreiber, "Advanced television systems for terrestrial broadcasting: Some problems and some proposed solutions," *Proc. IEEE*, vol. 83, no. 6, pp. 958–981, Jun. 1995.

[13] G. Dziwoki and J. Izydorczyk, "Iterative identification of sparse mobile channels for TDS-OFDM systems," *IEEE Trans. Broadcast.*, vol. 62, no. 2, pp. 384–397, Jun. 2016.

[14] X. Ma *et al.*, "Novel approach to design time-domain training sequence for accurate sparse channel estimation," *IEEE Trans. Broadcast.*, vol. 62, no. 3, pp. 512–520, Sep. 2016.

[15] F. Wan, W. P. Zhu, and M. N. S. Swamy, "Semiblind sparse channel estimation for MIMO-OFDM systems," *IEEE Trans. Vehi. Tech.*, vol. 60, no. 6, pp. 2569–2582, Jul. 2011.

[16] X. Zhang, K. Song, C. Li, and A. L. Yang, "Parameter estimation for multi-scale multi-lag underwater acoustic channels based on modified particle swarm optimization algorithm," *IEEE Access*, vol. 5, pp. 4808–4820, Mar. 2017.

[17] W. U. Bajwa, J. Haupt, G. Raz, and R. Nowak, "Compressed channel sensing," in *Proc. IEEE CISS*, Mar. 2008, pp. 5–10.

[18] A. Liu, V. Lau, and W. Dai, "Joint burst LASSO for sparse channel estimation in multi-user massive MIMO," in *Proc. IEEE ICC*, May 2016, pp. 1–6.

[19] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374–377, Mar. 2002.

[20] G. Z. Karabulut and A. Yongacoglu, "Sparse channel estimation using orthogonal matching pursuit algorithm," in *Proc. IEEE 60th Veh. Technol. Conf. (VTC-Fall)*, vol. 6. Sep. 2004, pp. 3880–3884.
[21] S. I. Amari, "Natural gradient works efficiently in learning," *Neural Comput.*, vol. 10, no. 2, pp. 251–276, 1998.
[22] S. L. Gay and S. C. Douglas, "Normalized natural gradient adaptive filtering for sparse and non-sparse systems," in *Proc. IEEE ICASSP*, vol. 2, May 2002, pp. 1405–1408.
[23] W. Wan, "Implementing online natural gradient learning: Problems and solutions," *IEEE Trans. Neural Netw.*, vol. 17, no. 2, pp. 317–329, Mar. 2006.
[24] K. Pelekanakis and M. Chitre, "New sparse adaptive algorithms based on the natural gradient and the L_0-norm," *IEEE J. Ocean. Eng.*, vol. 38, no. 2, pp. 323–332, Apr. 2013.
[25] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proc. IEEE ICASSP*, vol. 2, May 2002, pp. 1881–1884.
[26] K. Pelekanakis and M. Chitre, "Natural gradient-based adaptive algorithms for sparse underwater acoustic channel identification," in *Proc. ICUAM*, 2011, pp. 1403–1410.
[27] S. Amari and S. C. Douglas, "Why natural gradient?" in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, vol. 2. May 1998, pp. 1213–1216.
[28] C. Pladdy *et al.*, "Semiblind BLUE channel estimation with applications to digital television," *IEEE Trans. Vehi. Tech.*, vol. 55, no. 6, pp. 526–532, Nov. 2006.
[29] G. Naik, S. Singhal, A. Kumar, and A. Karandikar, "Quantitative assessment of TV white space in India," in *Proc. NCC*, 2014, pp. 1–6.
[30] *Regulatory Requirements for White Space Device in the UHF TV Band*, OFCOM, Warrington, U.K., Jul. 2012.
[31] R. L. Das and M. Chakraborty, "On convergence of proportionate-type normalized least square algorithms," *IEEE Trans. Circuits, Syst., Exp. Briefs*, vol. 62, no. 5, pp. 491–495, May 2015.
[32] A. Bishnu and V. Bhatia, "Natural gradient non-parametric maximum likelihood algorithm for sparse channel estimation in non-Gaussian noise," in *Proc. IEEE SPCOM*, Jun. 2016, pp. 1–5.
[33] S. Haykin, *Unsupervised Adaptive Filtering, Blind Deconvolution*, vol. 2, 1st ed. Hoboken, NJ, USA: Wiley, 2000.
[34] Z. Yang, Y. R. Zeng, and S. L. Grant, "Proportionate affine projection sign algorithms for network echo cancellation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 8, pp. 2273–2284, Nov. 2011.
[35] M. Yukawa and I. Yamada, "A unified view of adaptive variable-metric projection algorithms," *EURASIP J. Adv. Signal Process.*, vol. 2009, pp. 1–13, Dec. 2009.
[36] T. Y. Al-Naffouri and A. H. Sayed, "Adaptive filters with error nonlinearities: Mean-square analysis and optimum design," *EURASIP J. Appl. Signal Process.*, vol. 4, no. 1, pp. 192–205, 2001.
[37] C. Wang, M. Ma, and L. Zhang, "An efficient EAP-based pre-authentication for inter-WRAN handover in TV white space," *IEEE Access*, vol. 5, pp. 9785–9796, May 2017.
[38] K. Ishizu *et al.*, "Field experiment of long-distance broadband communications in TV white space using IEEE 802.22 and IEEE 802.11af," in *Proc. Int. Symp. Wireless Pers. Multimedia Commun. (WPMC)*, Sep. 2014, pp. 468–473.
[39] C. C. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. J. Shellhammer, and W. Caldwell, "IEEE 802.22: The first cognitive radio wireless regional area network standard," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 130–138, Jan. 2009.
[40] A. F. Molisch, *Wireless Communications*, 2nd ed. Hoboken, NJ, USA: Wiley, 2011.
[41] A. Bishnu and V. Bhatia, "On performance analysis of IEEE 802.22 (PHY) for COST-207 channel models," in *Proc. IEEE CSCN*, Oct. 2015, pp. 67–72.
[42] F. Wan, W. P. Zhu, and M. N. S. Swamy, "Semi-blind most significant tap detection for sparse channel estimation of OFDM systems," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 3, pp. 703–713, Mar. 2010.
[43] A. Bishnu and V. Bhatia, "Iterative time-domain-based sparse channel estimation for IEEE 802.22," *IEEE Wireless Commun. Lett.*, vol. 6, no. 3, pp. 290–293, Mar. 2017.
[44] M. E. Hajjar, Q. A. Nguyen, R. G. Maunder, and S. X. Ng, "Demonstrating the practical challenges of wireless communications using USRP," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 194–201, May 2014.
[45] T. B. Welch and S. Shearman, "Teaching software defined radio using the USRP and LabVIEW," in *Proc. IEEE ICASSP*, Mar. 2012, pp. 1792–2789.
[46] Accessed: Aug. 22, 2016. [Online]. Available: http://sine.ni.com/ds/app/doc/p/id/ds-538/lang/en
[47] V. Tawil. (2006). *51 Captured DTV Signal*. [online]. Available: http://grouper.ieee.org/groups/802/22/Meetingdocuments/
[48] [Online]. Available: http://www.ddpunjabi.in/dvbfaq.pdf
[49] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.
[50] B. Chen, L. Xing, J. Liang, N. Zheng, and J. C. Principe, "Steady-state mean-square error analysis for adaptive filtering under the maximum correntropy criterion," *IEEE Signal Process. Lett.*, vol. 21, no. 7, pp. 880–884, Jul. 2014.
[51] N. R. Yousef and A. H. Sayed, "A unified approach to the steady-state and tracking analyses of adaptive filters," *IEEE Trans. Signal Process.*, vol. 49, no. 2, pp. 314–324, Feb. 2001.
[52] B. Lin, R. He, X. Wang, and B. Wang, "The steady-state mean-square error analysis for least mean p-order algorithm," *IEEE Signal Process. Lett.*, vol. 16, no. 3, pp. 176–179, Mar. 2009.

ABHIJEET BISHNU received the B.E. degree in electronics and communication engineering from the Technocrat Institute of Technology, Bhopal, India, in 2010, the M.E. degree in electronics and telecommunication engineering from S.G.S.I.T.S. Indore, India, in 2013. He is currently pursuing the Ph.D. degree with IIT Indore, India. His research interests include OFDM system and signal processing for wireless communications.

VIMAL BHATIA received the Ph.D. degree from the Institute for Digital Communications, The University of Edinburgh, U.K., in 2005. He is currently an Associate Professor with IIT Indore. His research interests are in the algorithms and solutions for future communication and optical systems.