Improving Convergence Rate of Sign Algorithm using Natural Gradient Method

Taiyo Mineo
Graduate School of Informatics and Engineering
The University of Electro-Communications
Tokyo, JAPAN
m2040009@edu.cc.uec.ac.jp

Hayaru Shouno
Graduate School of Informatics and Engineering
The University of Electro-Communications
Tokyo, JAPAN
shouno@uec.ac.jp

***Abstract*—In lossless audio compression, it is essential for predictive residuals to remain sparse when applying entropy codings. Hence, developing an accurate predictive method is crucial. The sign algorithm (SA) is a conventional method for minimizing the magnitude of residuals; however, it exhibits poor convergence performance compared with the least mean square (LMS) algorithm. To overcome the convergence performance degradation, we proposed novel adaptive algorithms based on a natural gradient: the natural gradient sign algorithm (NGSA) and normalized NGSA (NNGSA). We also propose an efficient update method for the natural gradient based on the $\mathrm{AR}(p)$ model. It requires $\mathcal{O}(p)$ multiply-add operations at every adaptation step. Through experiments conducted using toy data and real music data, we showed that the proposed algorithms achieve better convergence performance than the SA does. The NNGSA suggested having good compression ability in lossless audio coding.**

***Index Terms*—Lossless audio coding, adaptive algorithm, sign algorithm, natural gradient method, autoregressive model**

I. Introduction

A greater storage capacity is required to enrich digital audio content further [1]. Therefore, lossless audio coding, which involves compressing audio data without information loss, is a vital technology for various applications, such as lossless music delivery, editing, and recording. Figure 1 depicts the general structure of a lossless audio codec [2]. First, a codec converts the audio signal to a residual via prediction using a mathematical model. Second, it compresses the residual through entropy coding. If a model obtains an accurate prediction, the residual signal is sparse, and thus, high compression performance can be achieved. Several codecs that follow the structure shown in Fig. 1 have been implemented since the introduction of the Shorten lossless codec [3]. For example, MPEG4-ALS [4] and FLAC [5] use linear predictive coding (LPC) as the predictive model, while WavPack [6] and TTA [7] use adaptive filters. In entropy coding, the Golomb-Rice code [8] is generally used. This code is optimal when the residual follows a Laplace distribution. However, the aforementioned predictive models are generally formulated based on the assumption that the residual follows a Gaussian distribution. To solve this problem, [9] improved compression rate by formulating an LPC under a Laplace distribution. The sign algorithm (SA) [10] is a practical choice for the adaptive algorithm when the residual follows a Laplace distribution; however, the SA converges considerably slower than the least mean square (LMS) algorithm does [11]. To overcome this performance gap, several variants of the SA have been proposed, such as the convex combination [12] and the logarithmic cost function [13]. However, these attempts did not show better convergence performance than that of the normalized LMS (NLMS) did. For the LMS, meanwhile, [14] outperforms NLMS by employing a natural gradient [15].

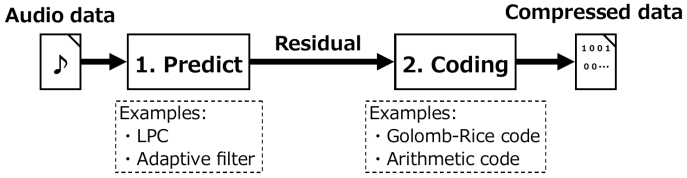

Fig. 1. Lossless audio codec

In this study, we improve the convergence performance of the SA using a natural gradient. The major contributions of this study are as follows:

1) We proposed novel adaptive algorithms: natural gradient sign algorithm (NGSA) and normalized NGSA (NNGSA).
2) These algorithms employ $\mathcal{O}(p)$ multiply-add operations for calculating the natural gradient at every step based on the p-th order autoregressive model assumption for input data.
3) The proposed algorithms achieve better convergence performance than the SA does.

The remainder of this paper is organized as follows: Section II provides an overview of the adaptive algorithm, SA, and autoregressive model. Section III presents the NGSA and NNGSA and describes the efficient calculation processes in these algorithms. Computer-based experiments conducted to demonstrate the performance of the proposed algorithms are discussed in Section IV. Finally, Sections V and VI present the discussion and the conclusion, respectively.

II. Theoretical overview

A. Adaptive filter

An overview of an adaptive filter is shown in Fig. 2. The input signal $x[n]$ and observation noise $v[n]$ are discrete-time

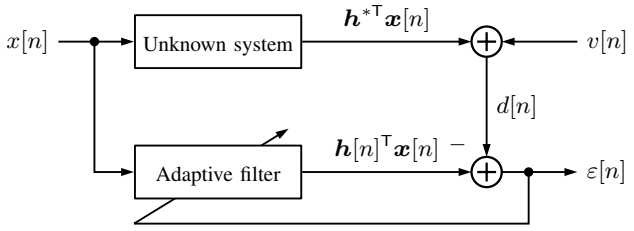

Fig. 2. Adaptive filter

signal sequences. In this study, the input signal $x[n]$ is assumed to have weak stationarity and to be an ergodic process. Let $\boldsymbol{h}[n] = [h_1[n], ..., h_N[n]]^{\mathsf{T}}$ be the adaptive filter coefficients, where T represents the matrix transposition. This study employs a finite impulse response filter. Hence, the filter output is denoted as $\boldsymbol{h}[n]^{\mathsf{T}}\boldsymbol{x}[n]$, where $\boldsymbol{x}[n] = [x[n-N+1], ..., x[n]]^{\mathsf{T}}$ represents the input vector. We denote the coefficient vector for an unknown system as $\boldsymbol{h}^*$. Filter adaptation is performed by updating the coefficients $\boldsymbol{h}[n]$ through the observed signal

$$d[n] := \boldsymbol{h}^{*\mathsf{T}}\boldsymbol{x}[n] + v[n], \tag{1}$$

and the residual

$$\varepsilon[n] := d[n] - \boldsymbol{h}[n]^{\mathsf{T}}\boldsymbol{x}[n]. \tag{2}$$

B. Sign algorithm (SA)

The SA is derived using the maximum likelihood method under the assumption that the residual $\varepsilon[n]$ follows a Laplace distribution. The probability density function of the Laplace distribution $p(\varepsilon[n] \mid \boldsymbol{h})$ is

$$p(\varepsilon[n] \mid \boldsymbol{h}) = \frac{1}{2\sigma}\exp\left[-\frac{|\varepsilon[n]|}{\sigma}\right], \tag{3}$$

where $\sigma > 0$ represents the deviation. The likelihood $L(\boldsymbol{h})$ and log-likelihood $\log L(\boldsymbol{h})$ functions for independent and identically distributed (i.i.d.) M samples are expressed as

$$L(\boldsymbol{h}) = \frac{1}{(2\sigma)^M}\prod_{k=1}^{M}\exp\left[-\frac{|\varepsilon[k]|}{\sigma}\right]. \tag{4}$$

$$\log L(\boldsymbol{h}) = -M\log(2\sigma) - \frac{1}{\sigma}\sum_{k=1}^{M}|\varepsilon[k]|. \tag{5}$$

We let $M = 1$ because the SA adapts at each step. To maximize the likelihood, we take the partial differentiation of $\log L(\boldsymbol{h})$ by $\boldsymbol{h}$

$$\frac{\partial \log L(\boldsymbol{h})}{\partial \boldsymbol{h}} = \frac{1}{\sigma}\operatorname{sgn}(\varepsilon[n])\boldsymbol{x}[n], \tag{6}$$

where $\operatorname{sgn}(\cdot)$ denotes the sign function, which is defined as

$$\operatorname{sgn}(x) = \begin{cases} 1 & (x > 0) \\ 0 & (x = 0) \\ -1 & (x < 0) \end{cases}. \tag{7}$$

The SA adaptation rule is shown in Eq. (6):

$$\boldsymbol{h}[n+1] = \boldsymbol{h}[n] + \mu\operatorname{sgn}(\varepsilon[n])\boldsymbol{x}[n], \tag{8}$$

where $\mu > 0$ denotes the step-size parameter.

C. Autoregressive model

We denote $\mathrm{AR}(p)$ for the autoregressive model with order p, which satisfies the following equation for signal s:

$$s[n] = \sum_{i=1}^{p}\psi_i s[n-i] + \nu[n], \quad \psi_i \in \mathbb{R}\ (i = 1, ..., p), \tag{9}$$

where $\nu[n]$ is a sample from an independent standard normal distribution. The i-th row and j-th column element of the inverse autocovariance matrix for an $\mathrm{AR}(p)$ process $\boldsymbol{K}_p^{-1}$ is calculated explicitly as [16]

$$(\boldsymbol{K}_p^{-1})_{ij} = \begin{cases} \displaystyle\sum_{k=1}^{j}\psi_{i-k}\psi_{j-k} & 1 \le i \le p+1 \\ \displaystyle\sum_{k=1}^{L-i+1}\psi_{L-i+1-k}\psi_{L-j+1-k} & L-p \le j \le L \\ 0 & i \ge j+p+1 \\ \displaystyle\sum_{k=i-p}^{j}\psi_{i-k}\psi_{j-k} & \text{otherwise} \end{cases}, \tag{10}$$

where $i \ge j$, $\psi_0 = 1$, and L denotes a matrix size that satisfies $L > 2p$.

III. Proposed algorithms

A. Natural gradient sign algorithm (NGSA)

The natural gradient is derived from the multiplication of an inverse of a Fisher information matrix $\boldsymbol{F}^{-1}$ and a gradient of the cost function [15]. The matrix $\boldsymbol{F}$ is calculated using the covariance of the gradient for the log-likelihood function (Eq. (6)), as follows:

$$\boldsymbol{F} := \mathrm{E}\left[\left\{\frac{\partial \log L(\boldsymbol{h})}{\partial \boldsymbol{h}}\right\}\left\{\frac{\partial \log L(\boldsymbol{h})}{\partial \boldsymbol{h}}\right\}^{\mathsf{T}}\right] \tag{11}$$

$$= \mathrm{E}\left[\left\{\frac{\operatorname{sgn}(\varepsilon[n])}{\sigma}\right\}^2\boldsymbol{x}[n]\boldsymbol{x}[n]^{\mathsf{T}}\right] \tag{12}$$

$$= \frac{1}{\sigma^2}\mathrm{E}\left[\boldsymbol{x}[n]\boldsymbol{x}[n]^{\mathsf{T}}\right] \quad \text{(a.s.)} \tag{13}$$

$$= \frac{1}{\sigma^2}\boldsymbol{R}, \tag{14}$$

where $\boldsymbol{R}$ is the autocorrelation matrix of the input signal. Eq. (13) holds because $\{\operatorname{sgn}(x)\}^2 = 1$ is satisfied if $x \neq 0$. Using Eq. (14), we obtain the NGSA as follows:

$$\boldsymbol{h}[n+1] = \boldsymbol{h}[n] + \mu_{\mathrm{NGSA}}\operatorname{sgn}(\varepsilon[n])\boldsymbol{R}^{-1}\boldsymbol{x}[n], \tag{15}$$

where μ_{NGSA} denotes the step-size parameter, and $\boldsymbol{R}$ is assumed to be a regular matrix. In addition, the NGSA can be derived by replacing $\varepsilon[n]$ with $\operatorname{sgn}(\varepsilon[n])$ in the LMS/Newton algorithm [17], which is an approximation of the Newton method for the LMS algorithm.

The NGSA adaptation rule (Eq. (15)) satisfies the following inequality:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\mathrm{E}\left[|\varepsilon[k]|\right] \le \varepsilon_{\min} + \mu_{\mathrm{NGSA}}\frac{h}{\lambda_{\min}}, \tag{16}$$

where $\varepsilon_{\min} = \mathrm{E}\left[|v[n]|\right]$, $h = (1/2)\mathrm{E}\left[||\boldsymbol{x}[n]||_2^2\right]$, and $\lambda_{\min}$ denotes the minimum eigenvalue of $\boldsymbol{R}$. The proof of Eq. (16) follows that provided in [11].

B. Normalized natural gradient sign algorithm (NNGSA)

The NGSA experiences difficulty in determining the step-size parameter μ_{NGSA} because it has varying optimal settings based on the input signal. To overcome this difficulty, we introduce a variable step-size adaptation that minimizes the posterior residual criterion. This is identical to that in NLMS [18].

Let $\mu[n]$ be the adaptive step size and $\varepsilon^+[n]$ be the posterior residual at time n. Then, $\varepsilon^+[n]$ is calculated as

$$\varepsilon^+[n] := d[n] - \boldsymbol{h}[n+1]^{\mathsf{T}}\boldsymbol{x}[n] \tag{17}$$

$$= d[n] - \left\{\boldsymbol{h}[n] + \mu[n]\,\mathrm{sgn}(\varepsilon[n])\boldsymbol{R}^{-1}\boldsymbol{x}[n]\right\}^{\mathsf{T}}\boldsymbol{x}[n] \tag{18}$$

$$= \varepsilon[n] - \mu[n]\,\mathrm{sgn}(\varepsilon[n])\boldsymbol{x}[n]^{\mathsf{T}}\boldsymbol{R}^{-1}\boldsymbol{x}[n]. \tag{19}$$

We let $\varepsilon^+[n] = 0$; then, solving Eq. (19) according to $\mu[n]$, we obtain

$$\mu[n] = \frac{|\varepsilon[n]|}{\boldsymbol{x}[n]^{\mathsf{T}}\boldsymbol{R}^{-1}\boldsymbol{x}[n]}. \tag{20}$$

Substituting Eq. (20) into Eq. (15), we obtain the NNGSA as follows:

$$\boldsymbol{h}[n+1] = \boldsymbol{h}[n] + \mu_{\mathrm{NNGSA}}\frac{\varepsilon[n]}{\boldsymbol{x}[n]^{\mathsf{T}}\boldsymbol{R}^{-1}\boldsymbol{x}[n]}\boldsymbol{R}^{-1}\boldsymbol{x}[n], \tag{21}$$

where $\mu_{\mathrm{NNGSA}} > 0$ denotes the scale parameter. If $\mu_{\mathrm{NNGSA}} < 2$ holds and $\boldsymbol{h}[n]$ and $\boldsymbol{x}[n]$ are statistically independent, this adaptation rule achieves a first-order convergence rate. The proof of this proposition follows that of the NLMS provided in [18].

The NNGSA can be interpreted as a variable step-size modification of the LMS/Newton algorithm [19]. In [20], the authors state that [19] is a generalization of the recursive least squares (RLS). Furthermore, it is evident that Eq. (21) is identical to the NLMS if $\boldsymbol{R} = \boldsymbol{I}$, where $\boldsymbol{I}$ denotes the identity matrix.

C. Geometric interpretation of NNGSA

The adaptation rule in Eq. (21) is used to solve the following optimization problem:

$$\begin{aligned}&\underset{\boldsymbol{h}}{\mathrm{argmin}}\ (\boldsymbol{h} - \boldsymbol{h}[n])^{\mathsf{T}}\boldsymbol{R}(\boldsymbol{h} - \boldsymbol{h}[n]),\\ &\text{subject to: } d[n] = \boldsymbol{h}^{\mathsf{T}}\boldsymbol{x}[n].\end{aligned} \tag{22}$$

The Lagrange multiplier can be used to solve the aforementioned problem. Therefore, Eq. (21) projects $\boldsymbol{h}[n]$ onto hyperplane $W = \{\boldsymbol{h} \mid d[n] = \boldsymbol{h}^{\mathsf{T}}\boldsymbol{x}[n]\}$, where its metric is defined as $\boldsymbol{R}$ (see Fig. 3). Moreover, according to information geometry [21], the Kullback-Leibler divergence $\mathrm{KL}\left[\cdot||\cdot\right]$ for

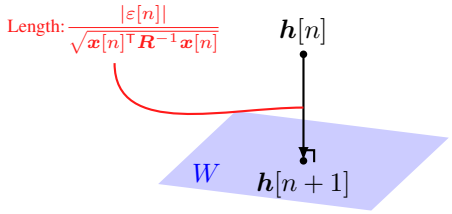

Fig. 3. Geometric interpretation of NNGSA

models that belong to the neighborhoods of parameter $\boldsymbol{h}[n]$ can be calculated as

$$\begin{aligned}&\mathrm{KL}\left[p(\varepsilon[n] \mid \boldsymbol{h}[n])||p(\varepsilon[n] \mid \boldsymbol{h})\right]\\ &\quad\approx \frac{1}{2}(\boldsymbol{h} - \boldsymbol{h}[n])^{\mathsf{T}}\boldsymbol{F}(\boldsymbol{h} - \boldsymbol{h}[n]) &(23)\\ &\quad= \frac{1}{2\sigma^2}(\boldsymbol{h} - \boldsymbol{h}[n])^{\mathsf{T}}\boldsymbol{R}(\boldsymbol{h} - \boldsymbol{h}[n]). &(24)\end{aligned}$$

Thus, Eq. (21) can be considered the m-projection from the model $p(\varepsilon[n] \mid \boldsymbol{h}[n])$ to the statistical manifold $M = \{p(\varepsilon[n] \mid \boldsymbol{h}) \mid d[n] = \boldsymbol{h}^{\mathsf{T}}\boldsymbol{x}[n]\}$ wherein elements have the minimum posterior residual.

D. Efficient natural gradient updation

It is vital to calculate the natural gradient $\boldsymbol{R}^{-1}\boldsymbol{x}[n]$ at every step. Although the Sherman-Morrison formula is used to reduce the complexity of the RLS, this algorithm involves $\mathcal{O}(N^2)$ operations, which entails high costs in practical applications. Hence, we propose an efficient method to solve this problem.

Herein, we assume that input signals follow the $\mathrm{AR}(p)$ process. The natural gradient at time n $\boldsymbol{m}[n] = [m_1[n], ..., m_N[n]]^{\mathsf{T}} := \boldsymbol{K}_p^{-1}\boldsymbol{x}[n]$ can be updated as

$$\begin{aligned}\boldsymbol{K}_p^{-1}\boldsymbol{x}[n+1] = &\begin{bmatrix} m_2[n] \\ m_3[n] \\ \vdots \\ m_N[n] \\ 0 \end{bmatrix} + m_1[n]\begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_p \\ \boldsymbol{0}_{N-p} \end{bmatrix}\\ &- m_N[n+1]\begin{bmatrix} \boldsymbol{0}_{N-p-1} \\ \psi_p \\ \vdots \\ \psi_1 \\ -1 \end{bmatrix},\\ m_N[n+1] = x[n+1] - &\sum_{i=1}^{p}\psi_i x[n+1-i],\end{aligned} \tag{25}$$

where $\boldsymbol{0}_N$ is an $N \times 1$ zero vector. Eq. (25) is followed by a direct calculation employing Eq. (10). Furthermore, the Mahalanobis norm $\boldsymbol{x}[n]^{\mathsf{T}}\boldsymbol{K}_p^{-1}\boldsymbol{x}[n]$ can be updated as follows:

$$\begin{aligned}&\boldsymbol{x}[n+1]^{\mathsf{T}}\boldsymbol{K}_p^{-1}\boldsymbol{x}[n+1]\\ &\quad= \boldsymbol{x}[n]^{\mathsf{T}}\boldsymbol{K}_p^{-1}\boldsymbol{x}[n] - m_1[n]^2 + m_N[n+1]^2.\end{aligned} \tag{26}$$

Eq. (25) requires $3p$ multiply-add (subtract) calculations, and Eq. (26) requires 2. Hence, we can update the natural gradient

Algorithm 1 NNGSA with AR(p) assumption

Require: Desired output $d[n]$, Input $x[n]$, Number of inputs M, Filter order N, Scale size μ, Autoregressive order p, Small constant $c > 0$
Ensure: Residual $\varepsilon[n]$
Calculate ψ_i $(i = 1, ..., p)$ by Levinson-Durbin algorithm [24]
$\boldsymbol{m} \leftarrow \boldsymbol{0}$, $\boldsymbol{h} \leftarrow \boldsymbol{0}$, $w \leftarrow 0$
for $i = 1, \cdots, M$ **do**
  $\boldsymbol{x}[i] \leftarrow [x[i-N+1], ..., x[i]]^\mathsf{T}$ $x[j] = 0$ s.t. $j \leq 0$
  $\varepsilon[i] \leftarrow d[i] - \boldsymbol{h}^\mathsf{T}\boldsymbol{x}[i]$
  $w \leftarrow w - (\boldsymbol{m})_1^2$
  for $j = 1, \cdots, p$ **do**
    $(\boldsymbol{m})_{j+1} \leftarrow (\boldsymbol{m})_{j+1} + \psi_j(\boldsymbol{m})_1$
  end for
  for $j = 1, \cdots, N-1$ **do**
    $(\boldsymbol{m})_j \leftarrow (\boldsymbol{m})_{j+1}$
  end for
  $(\boldsymbol{m})_N \leftarrow x[i] - \sum_{j=1}^{p} \psi_j x[i-j]$
  for $j = 1, \cdots, p$ **do**
    $(\boldsymbol{m})_{N-j} \leftarrow (\boldsymbol{m})_{N-j} - \psi_j(\boldsymbol{m})_N$
  end for
  $w \leftarrow w + (\boldsymbol{m})_N^2$
  $\boldsymbol{h} \leftarrow \boldsymbol{h} + (\mu\varepsilon[i]/\max\{w, c\})\boldsymbol{m}$
end for

in $\mathcal{O}(p)$ operations. Eq. (25) is essentially the same as that in [22], which applies a lattice filter (with partial autocorrelation coefficients) to update the gradient. Meanwhile, our method is applicable for updating norms.

We can apply this procedure to the LMS/Newton algorithm:

$$\boldsymbol{h}[n+1] = \boldsymbol{h}[n] + \mu_{\mathrm{LMSN}}\boldsymbol{R}_p^{-1}\boldsymbol{x}[n], \ \ \boldsymbol{R}_p^{-1} := \sigma_p^{-1}\boldsymbol{K}_p^{-1}, \tag{27}$$

where $\mu_{\mathrm{LMSN}} > 0$ denotes the step-size parameter, and σ_p is a constant that depends on p. For $p = 1$, Eq. (27) achieves a first-order convergence if $\mu_{\mathrm{LMSN}} < 2(1-\psi_1)/(N(1+\psi_1))$. The proof of this proposition follows that of the LMS provided in [17] and the employs range of eigenvalues of $\boldsymbol{R}_1$ [23].

E. Algorithm

Algorithm 1 describes the NNGSA coding procedure under the AR(p) assumption.

IV. Experiments

A. Toy-data experiments

We observed the convergence performance under the artificial setting as follows: The elements of the unknown parameter $\boldsymbol{h}^*$ were randomly chosen with a uniform distribution of $[-1, 1]$, filter order N was set to 5, and the observation noise $v[n]$ was white Gaussian noise with a variance of -40 [dB]. These settings were referenced from [13]. We calculated the mean square deviation (MSD) criteria $||\boldsymbol{h}^* - \boldsymbol{h}||_2$ from 200 independent trials. We set $p = 1$ and step sizes for the proposed algorithms as $\mu_{\mathrm{NGSA}} = 0.01$, $\mu_{\mathrm{NNGSA}} = 0.1$, and $\mu_{\mathrm{LMSN}} = 0.01$. We implemented algorithms by the Python 3.8.1 and simulated on the Intel(R) Core-i7 2.8 [GHz] Dual Core CPU with 16 [GB] RAM.

First, we tracked the MSD learning curves for $x[n]$ with a variance of 0 [dB]. Fig. 4 shows a comparison between the proposed algorithms and the SA, NLMS, and RLS. We set various step sizes for the SA and NLMS and employed forgetting factors λ for the RLS. Fig. 4 shows that the NGSA and NNGSA achieved almost the same performance as the SA and NLMS did, respectively. This is because $\boldsymbol{R}_1^{-1} \approx \boldsymbol{I}$ holds for i.i.d. noise input.

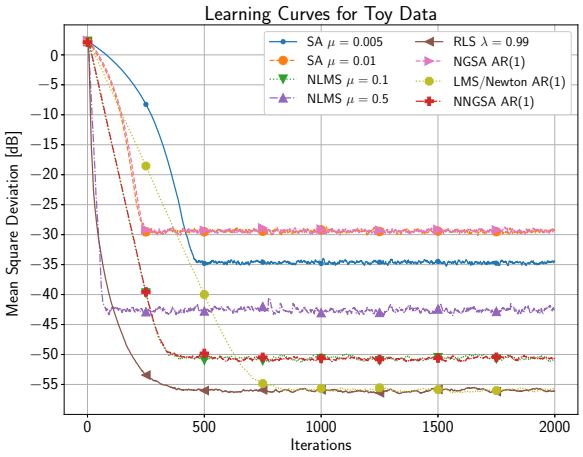

Fig. 4. Learning curves for the white Gaussian noise input

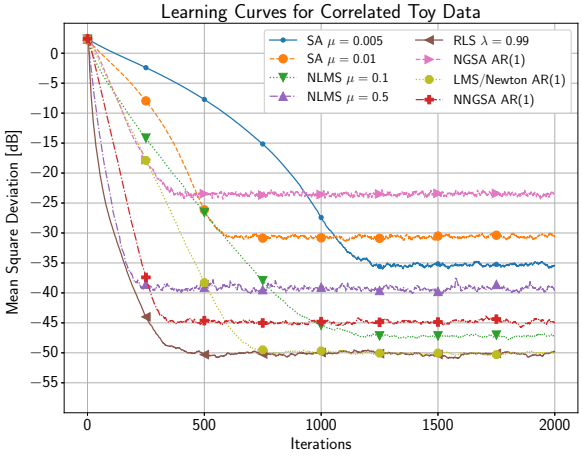

Fig. 5. Learning curves for correlated noise input

Second, we observed the case in which the Gaussian noise is correlated with $x[n] \leftarrow x[n] + x[n-1] \times 0.8$. Fig. 5 shows the results for the correlation. The convergence performance of the SA and NLMS is worse than that of the non-correlated noise input (Fig. 4). Moreover, the steady-state errors for the proposed algorithms also deteriorated. This is because $\boldsymbol{R}$ is close to being ill-conditioned, and the right-hand side of Eq. (16) is large.

B. Real-data experiments

We observed the absolute error (AE) for filter prediction using real music data [25]. In this experiment, we assumed that the input data is only an audio data signal and that the

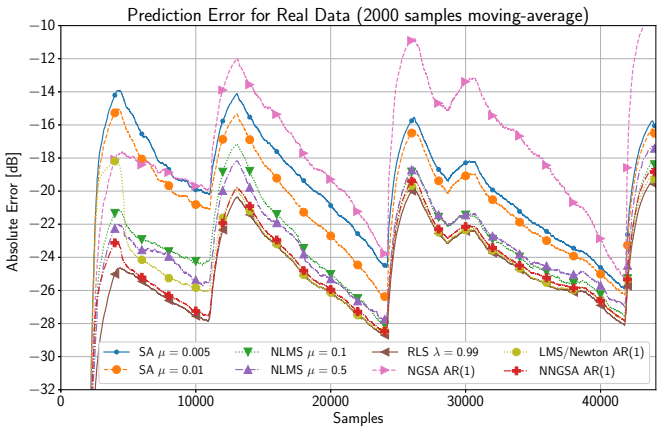

Fig. 6. AE comparison for real music data

reference output and observation noise are zero (silence). We set the same configurations for the proposed algorithms as in the toy-data experiments. Fig. 6 shows the AE curves in the first one second (at a 44100 [Hz] sampling rate) for the left channel of the "When the Saints Go Marching in". Fig. 6 indicates that the NNGSA and LMS/Newton performs better than the NLMS does and approximately the same as the RLS does. However, the NGSA with AR(1) exhibits considerably worse performance. We assume that this is because of a worse steady-state error of the NGSA, which was arisen from a long-term ($\approx$ 10000 samples) signal stationary.

V. Discussion

The proposed algorithms evidently achieved better convergence performance than the SA and NLMS did for correlated signal inputs. Furthermore, the NNGSA and LMS/Newton algorithms exhibited similar performance to that of the RLS, as indicated in [20]. We concluded that the NNGSA suggested being a more accurate predictive algorithm than the SA and practical for lossless audio codec. However, the proposed algorithms suffer from two major problems with regard to practical applications. First, the matrix $\boldsymbol{R}$ needs to be a singular matrix that depends on input signals, such as a constant wave. To address this, we consider introducing regularization, that is, we calculate the inverse matrix for $\boldsymbol{R} + \gamma \boldsymbol{I}$ $(\gamma > 0)$ instead of $\boldsymbol{R}$. Second, the AR coefficients ψ_i $(i = 1, ..., p)$ need to be calculated before the adaptation process. This can lead to some difficulties in streaming data processing.

VI. Conclusion

We proposed two novel adaptive algorithms that introduce a natural gradient into the SA. The adaptive step-size algorithm (NNGSA) exhibited certain similarities with well-known algorithms such as the NLMS and RLS. Furthermore, we demonstrated the superior performance of the proposed algorithms compared with that of the SA via toy-data and real-music-data experiments. In a future study, we will introduce an iterative method for the estimation of AR coefficients and expansion methods for affine projection algorithms [18], and integrate them into lossless audio compression to evaluate the proposed algorithms.

References

[1] K. Konstantinides, "An introduction to super audio cd and dvd-audio," *IEEE signal processing magazine*, vol. 20, no. 4, pp. 71–82, 2003.

[2] M. Hans and R. W. Schafer, "Lossless compression of digital audio," *IEEE Signal processing magazine*, vol. 18, no. 4, pp. 21–32, 2001.

[3] T. Robinson, "Shorten: Simple lossless and near-lossless waveform compression," *Technical Report, Cambridge Univ., Eng. Dept.*, 1994.

[4] T. Liebchen, "Mpeg-4 als-the standard for lossless audio coding," *the Journal of the Acoustical Society of Korea*, vol. 28, no. 7, pp. 618–629, 2009.

[5] "Flac - free lossless audio codec," https://xiph.org/flac/.

[6] "Wavpack audio compression," http://www.wavpack.com.

[7] "Tta lossless audio codec - true audio compressor algorithms," http://tausoft.org/wiki/True_Audio_Codec_Overview.

[8] R. F. Rice, "Practical universal noiseless coding," in *Applications of Digital Image Processing III*, vol. 207. International Society for Optics and Photonics, 1979, pp. 247–267.

[9] H. Kameoka, Y. Kamamoto, N. Harada, and T. Moriya, "A linear predictive coding algorithm minimizing the golomb-rice code length of the residual signal," *The Transactions of the Institute of Electronics,Information and Communication Engineers. A*, vol. 91, pp. 1017–1025, 2008.

[10] P. S. Diniz, *Adaptive filtering*. Springer, 1997.

[11] A. Gersho, "Adaptive filtering with binary reinforcement," *IEEE Transactions on Information Theory*, vol. 30, no. 2, pp. 191–199, 1984.

[12] L. Lu, H. Zhao, K. Li, and B. Chen, "A novel normalized sign algorithm for system identification under impulsive noise interference," *Circuits, Systems, and Signal Processing*, vol. 35, no. 9, pp. 3244–3265, 2016.

[13] M. O. Sayin, N. D. Vanli, and S. S. Kozat, "A novel family of adaptive filtering algorithms based on the logarithmic cost," *IEEE Transactions on signal processing*, vol. 62, no. 17, pp. 4411–4424, 2014.

[14] S. L. Gay and S. C. Douglas, "Normalized natural gradient adaptive filtering for sparse and non-sparse systems," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2. IEEE, 2002, pp. II–1405.

[15] S.-I. Amari, "Natural gradient works efficiently in learning," *Neural computation*, vol. 10, no. 2, pp. 251–276, 1998.

[16] M. Siddiqui, "On the inversion of the sample covariance matrix in a stationary autoregressive process," *The Annals of Mathematical Statistics*, vol. 29, no. 2, pp. 585–588, 1958.

[17] W. Bernard and D. S. Samuel, *Adaptive signal processing*. Englewood Cliffs, NJ: Prentice Hall, 1985.

[18] S. S. Haykin, *Adaptive filter theory*. Pearson Education India, 2005.

[19] P. S. Diniz and L. W. Biscainho, "Optimal variable step size for the lms/newton algorithm with application to subband adaptive filtering," *IEEE Transactions on signal processing*, vol. 40, no. 11, pp. 2825–2829, 1992.

[20] P. S. Diniz, M. L. de Campos, and A. Antoniou, "Analysis of lms-newton adaptive filtering algorithms with variable convergence factor," *IEEE Transactions on Signal Processing*, vol. 43, no. 3, pp. 617–627, 1995.

[21] S.-I. Amari, *Differential-geometrical methods in statistics*. Springer Science & Business Media, 2012, vol. 28.

[22] B. Farhang-Boroujeny, "Fast lms/newton algorithms based on autoregressive modeling and their application to acoustic echo cancellation," *IEEE Transactions on Signal Processing*, vol. 45, no. 8, pp. 1987–2000, 1997.

[23] U. Grenander and G. Szegö, *Toeplitz forms and their applications*. Univ of California Press, 1958.

[24] J. E. Markel and A. H. Gray, *Linear Prediction of Speech*. Berlin, Heidelberg: Springer-Verlag, 1982.

[25] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "Rwc music database: Popular, classical and jazz music databases." in *Ismir*, vol. 2, 2002, pp. 287–288.