Exploiting Sparsity in Feed-Forward Active Noise Control with Adaptive Douglas-Rachford Splitting

Masao Yamagishi* and Isao Yamada*
*Department of Communications and Computer Engineering, Tokyo Institute of Technology, Japan
E-mail: {myamagi,isao}@sp.ce.titech.ac.jp

***Abstract*—Observing that a typical primary path in Active Noise Control (ANC) system is sparse, i.e., having a few significant coefficients, we propose an adaptive learning which promotes the sparsity of the concatenation of the adaptive filter and the secondary path. More precisely, we propose to suppress a time-varying sum of the data-fidelity term and the weighted ℓ_1 norm of the concatenation by the adaptive Douglas-Rachford splitting scheme. Numerical examples demonstrate that the proposed algorithm shows excellent performance of the ANC by exploiting the sparsity and has robustness against a violation of the sparsity assumption.**

I. Introduction

Active noise control (ANC) [1], [2], [3], [4], [5] is a technique to cancel the unwanted noise based on the principle of superposition. The unwanted noise is filtered through the primary acoustic path after observed by the reference microphone. The anti-noise signal from the secondary speaker is generated by the reference noise filtered through an adaptive filter of which learning algorithm has been studied extensively, e.g. [6], [7], [8], [9], [10], [11], [12] because it affects directly the anti-noise performance.

Recently, in the adaptive learning, *sparsity* of the desired coefficients of the adaptive filter was utilized to improve performance of the ANC [13], where the sparsity implies that only a few coefficients are significant and other coefficients are zero (or near zero). In [13], the sparsity is presumed by employing a long adaptive filter length, and is exploited in the adaptive learning by adopting a convex combination of the update of the standard adaptive filter and the so-called proportionate-type update.

In this paper, motivated by observations in [14] and [15] that typical acoustic paths are sparse in practical situations, we propose an effective use of an inherent sparsity of the primary acoustic path to improve performance of the adaptive learning further. To exploit this sparsity, we adopt a time-varying sum of the weighted ℓ_1-norm of the concatenation of the adaptive filter and the secondary path, as a sparsity promoting term, and the data-fidelity term to measure consistency with observations. To suppress the time-varying sum in an online way, we derive an adaptive learning algorithm by applying the adaptive Douglas-Rachford splitting (ADRS) scheme [16].

This work was supported in part by JSPS Grants-in-Aid (B-21300091,24800022).

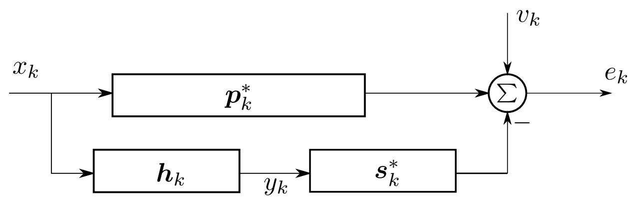

Fig. 1. A model of the feed-forward ANC.

Although the update of the ADRS scheme consists of two auxiliary convex minimization problems, we derive computationally efficient closed form solutions for the two auxiliary problems by reformulating the minimization of the time-varying sum into a higher dimensional minimization problem and by using special structure of the concatenation. Moreover, thanks to the flexibility of the ADRS scheme, we can also extend the proposed algorithm to exploit simultaneously the sparsity of both the primary path and the desired adaptive filter coefficients.

A numerical example for a sparse primary path demonstrates that the proposed algorithm achieves best performance of the ANC compared with popular conventional algorithms. In addition, a numerical example for a dense primary path shows that the proposed algorithm is robust against a violation of the sparsity assumption by achieving a comparable performance for a sparsity-unaware algorithm.

A preliminary version of this paper appeared as a technical report [17].

II. Preliminaries

A. Feed-Forward Active Noise Control

Let $\mathbb{R}$ and $\mathbb{N}$ denote the sets of all real numbers and nonnegative integers, respectively. Denote the set $\mathbb{N} \setminus \{0\}$ by $\mathbb{N}^*$ and transposition of a matrix or a vector by $(\cdot)^t$. Suppose that we observe the output $(e_k)_{k\in\mathbb{N}} \subset \mathbb{R}$ (i.e., $e_k \in \mathbb{R}, \forall k \in \mathbb{N}$)

$$e_k = (\boldsymbol{p}_k^*)^t \boldsymbol{x}_k - (\boldsymbol{s}_k^*)^t \boldsymbol{y}_k + v_k \tag{1}$$

at the error microphone (See Fig. 1), where $k \in \mathbb{N}$ denotes the time index, the reference noise $(x_k)_{k\in\mathbb{N}} \subset \mathbb{R}$ (with $\boldsymbol{x}_k := [x_k, x_{k-1}, \ldots, x_{k-N_1+1}]^t \in \mathbb{R}^{N_1}$) is filtered through the unknown primary acoustic path $\boldsymbol{p}_k^* \in \mathbb{R}^{N_1}$ (of tap length $N_1 \in \mathbb{N}^*$) between the reference noise source and the error microphone, the anti-noise signal $(y_k)_{k\in\mathbb{N}} \subset \mathbb{R}$

(with $\boldsymbol{y}_k := [y_k, y_{k-1}, \ldots, y_{k-N_2+1}]^t \in \mathbb{R}^{N_2}$) is filtered through the unknown secondary acoustic path $\boldsymbol{s}_k^* \in \mathbb{R}^{N_2}$ (of tap length $N_2 \in \mathbb{N}^*$), and $(v_k)_{k\in\mathbb{N}} \subset \mathbb{R}$ is the observation noise at the error microphone. In general, the anti-noise signal $(y_k)_{k\in\mathbb{N}} \subset \mathbb{R}$ is generated by the reference noise $(x_k)_{k\in\mathbb{N}}$ filtered by an adaptive filter $\boldsymbol{h}_k \in \mathbb{R}^{N_0}$:

$$y_k = \boldsymbol{h}_k^t \boldsymbol{x}_k.$$

A major goal of the active noise control is to create a zone of silence in the vicinity of the error microphone by designing adaptively $\boldsymbol{h}_k$ with the knowledge on $(x_i, e_i)_{i=0}^k$ and initial estimate $\boldsymbol{h}_0 \in \mathbb{R}^{N_0}$.

An observation in the case of time-invariant $\boldsymbol{p}_k^*$ and $\boldsymbol{s}_k^*$, i.e., $(\boldsymbol{p}_k^*, \boldsymbol{s}_k^*) = (\boldsymbol{p}_*, \boldsymbol{s}_*)$ for any $k \in \mathbb{N}^*$, leads to an underlying linear model for active noise control. Assume that there exists a desired filter $\boldsymbol{h}_*$, which minimizes $E[e_k^2]$, i.e.,

$$\boldsymbol{h}_* \in \underset{\boldsymbol{h}\in\mathbb{R}^{N_0}}{\arg\min}\ E[((\tilde{\boldsymbol{p}}_*^t - \boldsymbol{h}^t\tilde{\boldsymbol{S}}_*)\tilde{\boldsymbol{x}}_k + v_k)^2], \tag{2}$$

where[1] $\tilde{\boldsymbol{x}}_k$ and v_k are considered as random variables, and $\boldsymbol{h}^t\tilde{\boldsymbol{S}}_*$ implies the concatenation of the adaptive filter and the secondary path. Then by denoting the resulting error signal as n_k we obtain a standard linear model of $\boldsymbol{h}_*$:

$$\tilde{\boldsymbol{p}}_*^t\tilde{\boldsymbol{x}}_k = \boldsymbol{h}_*^t\tilde{\boldsymbol{S}}_*\tilde{\boldsymbol{x}}_k - v_k + n_k.$$

Fortunately, we can eliminate unknown $\tilde{\boldsymbol{p}}_*$ and v_k by (1):

$$e_k + \boldsymbol{s}_*^t\boldsymbol{y}_k = \boldsymbol{h}_*^t\tilde{\boldsymbol{S}}_*\tilde{\boldsymbol{x}}_k + n_k.$$

Since the complete knowledge of $\boldsymbol{s}_*$ is unavailable in general, an initial offline estimation (or an online modeling of $\boldsymbol{s}_*$) has been utilized (see e.g. [2], [3], [4], [18], [19]). Hence we can replace $\boldsymbol{s}_*$ and $\tilde{\boldsymbol{S}}_*$ by its estimate $\boldsymbol{s} \in \mathbb{R}^{N_2}$ and its corresponding Toeplitz matrix $\tilde{\boldsymbol{S}}$, which results a linear model:

$$e_k + \boldsymbol{s}^t\boldsymbol{y}_k = \boldsymbol{h}_*^t\tilde{\boldsymbol{S}}\tilde{\boldsymbol{x}}_k + n_k. \tag{3}$$

This suggests that we can apply standard adaptive filtering techniques to the linear model (3). In fact, a direct application of the least mean squares (LMS) algorithm [8] to the model (3) reproduces the modified filtered-x LMS (MFxLMS) algorithm [9].

In this paper, for simplicity, we assume that the primary and secondary paths are time-invariant, and also that an estimation $\boldsymbol{s}$ is obtained a priori, while the entire discussion can be extended to a time-varying case straightforwardly.

[1] $\tilde{\boldsymbol{x}}_k := [x_k, x_{k-1}, \ldots, x_{k-(N_0+N_2)}]^t \in \mathbb{R}^{\tilde{N}}$, $\tilde{N} := N_0 + N_2 - 1$, $\tilde{\boldsymbol{p}}_* \in \mathbb{R}^{\tilde{N}}$ and $\tilde{\boldsymbol{s}}_* \in \mathbb{R}^{\tilde{N}}$ are impulse response of which length are extended by zero padding, i.e.,

$$\tilde{\boldsymbol{p}}_* := [p_*^{(0)}, p_*^{(1)}, \ldots, p_*^{(N_1-1)}, 0, 0, \ldots, 0]^t,$$
$$\tilde{\boldsymbol{s}}_* := [s_*^{(0)}, s_*^{(1)}, \ldots, s_*^{(N_2-1)}, 0, 0, \ldots, 0]^t,$$

and $\tilde{\boldsymbol{S}}_* \in \mathbb{R}^{N_0\times\tilde{N}}$ is a Toeplitz matrix of which the first row is $\tilde{\boldsymbol{s}}_*$ and the first column is $(s^{(0)}, 0, \ldots, 0)^t \in \mathbb{R}^{N_0}$ in such a way that the multiplication $\tilde{\boldsymbol{S}}_*\tilde{\boldsymbol{x}}_k$ represents the convolution of $\tilde{\boldsymbol{s}}_*$ and $\tilde{\boldsymbol{x}}_k$.

B. Adaptive Douglas-Rachford Splitting Scheme

Define the inner product $\langle \boldsymbol{x}, \boldsymbol{y}\rangle := \boldsymbol{x}^t\boldsymbol{y}$ and its induced norm $\|\boldsymbol{x}\| := \sqrt{\langle \boldsymbol{x}, \boldsymbol{x}\rangle}$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$. We consider the situation where the time-varying cost function $\Theta_k \colon \mathbb{R}^N \to (-\infty, \infty]$, $k \in \mathbb{N}$, can be decomposed into the sum of two functions, i.e.,

$$\Theta_k(\boldsymbol{h}) := \varphi_k(\boldsymbol{h}) + \psi_k(\boldsymbol{h}), \tag{4}$$

where $\psi_k \colon \mathbb{R}^N \to (-\infty, \infty]$ and $\varphi_k \colon \mathbb{R}^N \to (-\infty, \infty]$ are proper lower semicontinuous convex functions (see e.g. [20]). To suppress the time-varying function Θ_k in an online way, the *adaptive Douglas-Rachford splitting* (ADRS) scheme [16] was proposed[2].

Scheme 1: **(Adaptive Douglas-Rachford Splitting Scheme)** For an arbitrary initial vector $\mathbf{g}_0 \in \mathbb{R}^N$ and any sequences $\gamma_k \in (0, \infty)$, $t_k \in (0, 2)$ $(k \in \mathbb{N})$, generate a sequence $\mathbf{h}_k \in \mathbb{R}^N$ $(k \in \mathbb{N})$ by

$$\mathbf{h}_k := \mathrm{prox}_{\gamma_k\psi_k}(\mathbf{g}_k) \tag{5}$$

with

$$\mathbf{g}_{k+1} := \mathbf{g}_k + t_k\left[\mathrm{prox}_{\gamma_k\varphi_k}(2\mathbf{h}_k - \mathbf{g}_k) - \mathbf{h}_k\right], \tag{6}$$

where, in general, a mapping $\mathrm{prox}_{\gamma f} \colon \mathbb{R}^N \to \mathbb{R}^N$ defined by

$$\mathrm{prox}_{\gamma f}(\boldsymbol{x}) := \underset{\boldsymbol{z}\in\mathbb{R}^N}{\arg\min}\left(\gamma f(\boldsymbol{z}) + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{z}\|^2\right), \forall \boldsymbol{x} \in \mathbb{R}^N$$

for $\gamma > 0$ and a proper lower semicontinuous convex function f, is called the proximity operator [21] of f of index $\gamma > 0$.

*Fact 1 (***Properties of Scheme 1):** Suppose that the functions φ_k and ψ_k satisfy the *qualification condition*[3] for every $k \in \mathbb{N}$.

Then the sequences $(\mathbf{h}_k)_{k\in\mathbb{N}}$ and $(\mathbf{g}_k)_{k\in\mathbb{N}}$ generated by Algorithm 1 satisfy the following

(i) $\begin{cases} \left\|\mathbf{h}_{k+1} - \mathrm{prox}_{\gamma_{k+1}\psi_{k+1}}(\mathbf{g}_{k+1}^*)\right\| \leq \|\mathbf{g}_{k+1} - \mathbf{g}_{k+1}^*\| \\ \|\mathbf{g}_{k+1} - \mathbf{g}_k^*\| \leq \|\mathbf{g}_k - \mathbf{g}_k^*\| \end{cases}$

for all $\mathbf{g}_{k+i}^* \in \left(\mathrm{prox}_{\gamma_{k+i}\psi_{k+i}}\right)^{-1}(\mathbf{\Omega}_{k+i}^*)$ with $\mathbf{\Omega}_{k+i}^* := \underset{\boldsymbol{h}\in\mathbb{R}^N}{\arg\min}\ \Theta_{k+i}(\boldsymbol{h})$ $(i = 0, 1)$.

(ii) Suppose that there exists a $N' \in \mathbb{N}$ such that $\mathbf{\Omega}_* := \cap_{i\geq N'}\mathbf{\Omega}_i^* \neq \varnothing$, $\psi_i = \psi$, and $\gamma_i = \gamma$ for all $i \geq N'$. Then we have

$$\left\|\mathbf{h}_{k+1} - \mathrm{prox}_{\gamma\psi}(\mathbf{g}_*)\right\| \leq \|\mathbf{g}_{k+1} - \mathbf{g}_*\| \leq \|\mathbf{g}_k - \mathbf{g}_*\|$$

for all $k \geq N'$ and all $\mathbf{g}_* \in \left(\mathrm{prox}_{\gamma\psi}\right)^{-1}(\mathbf{\Omega}_*)$.

[2] Note that the original ADRS scheme has an adaptively defined matrix $\boldsymbol{Q}_k$ which improves convergence performance significantly. However, for simplicity, we consider the case of $\boldsymbol{Q}_k$ as the identity matrix in this paper.

[3] Qualification condition [22]: The set

$$\bigcup_{\lambda>0}\{\lambda\mathbf{x} \mid \mathbf{x} \in \mathrm{dom}(\varphi_k) - \mathrm{dom}(\psi_k)\}$$

is a subspace of $\mathbb{R}^N$, where

$$\mathrm{dom}(\varphi_k) - \mathrm{dom}(\psi_k) := \{\mathbf{x}_1 - \mathbf{x}_2 \in \mathbb{R}^N \mid \forall(\mathbf{x}_1, \mathbf{x}_2) \in \mathrm{dom}(\varphi_k) \times \mathrm{dom}(\psi_k)\}.$$

(iii) Suppose that $\varphi_k = \varphi$, $\psi_k = \psi$ (i.e., $\boldsymbol{\Omega}_k^* = \boldsymbol{\Omega}_*$) and $\gamma_k = \gamma$ for all $k \in \mathbb{N}$. Then by using $(t_k)_{k\in\mathbb{N}}$ satisfying $\sum_{k\in\mathbb{N}} t_k(2-t_k) = \infty$, we have

$$\left\|\mathbf{h}_k - \mathrm{prox}_{\gamma\psi}(\mathbf{g}_*)\right\| \le \|\mathbf{g}_k - \mathbf{g}_*\| \overset{k\to\infty}{\longrightarrow} 0$$

for some $\mathbf{g}_* \in \left(\mathrm{prox}_{\gamma\psi}\right)^{-1}(\boldsymbol{\Omega}_*)$.

Note that Fact 1(ii) implies a monotone decrease of a sequence of upper bounds $(\|\mathbf{g}_k - \mathbf{g}_*\|)_{k\in\mathbb{N}}$ of the distance[4] $d(\mathbf{h}_k, \boldsymbol{\Omega}_*)$ without assuming $\varphi_i = \varphi$ for any $i \ge N'$. This property is useful for adaptive filtering applications.

III. Proposed Method

We propose a sparsity-aware adaptive learning algorithm of $\boldsymbol{h}_*$ in the frame of the ADRS scheme, based on the fact that the concatenation $\boldsymbol{h}^t\tilde{\boldsymbol{S}}$ is desired to approximate the primary acoustic path (see (2)) and to be sparse. More precisely, we suppress the time-varying sum of the data-fidelity term and a sparsity promoting term of the concatenation

$$\min_{\boldsymbol{h}\in\mathbb{R}^{N_0}} f_k(\langle\tilde{\boldsymbol{S}}\tilde{\boldsymbol{x}}_k, \boldsymbol{h}\rangle) + \lambda_k\|\tilde{\boldsymbol{S}}^t\boldsymbol{h}\|_1^{\boldsymbol{w}_k} \tag{7}$$

with a vector $\boldsymbol{w}_k := (w_1^{(k)}, w_2^{(k)}, \ldots, w_{\tilde{N}}^{(k)})^t \in \mathbb{R}^{\tilde{N}}$ of nonnegative coefficients for weighting the concatenation,

$$\|\cdot\|_1^{\boldsymbol{w}_k} : \mathbb{R}^{\tilde{N}} \to \mathbb{R}, \tilde{\boldsymbol{z}} = (\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_{\tilde{N}}) \mapsto \sum_{i=1}^{\tilde{N}} w_i^{(k)}|\tilde{z}_i|,$$
$$f_k : \mathbb{R} \to \mathbb{R}, r \mapsto (e_k + \boldsymbol{s}^t\boldsymbol{y}_k - r)^2,$$

and a regularization parameter $\lambda_k \ge 0$. Since the proximity operator of the latter term in (7) is hard to compute directly, we introduce an equivalent problem of (7) through the following two steps: (i) embedding the Toeplitz matrix $\tilde{\boldsymbol{S}}^t$ into a circulant matrix $\mathcal{C}(\tilde{\boldsymbol{S}}^t)$ (of which leading submatrix is $\tilde{\boldsymbol{S}}^t$ [23]), i.e.,

$$\tilde{\boldsymbol{S}}^t\boldsymbol{h} = \boldsymbol{\Pi}_{\tilde{N}}\mathcal{C}(\tilde{\boldsymbol{S}}^t)\hat{\boldsymbol{h}}$$

for any $(\boldsymbol{h}, \hat{\boldsymbol{h}}) \in \mathbb{R}^{N_0} \times C_0$ such that

$$\boldsymbol{h} = \boldsymbol{\Pi}\hat{\boldsymbol{h}}, \tag{8}$$

where $\boldsymbol{\Pi} := [\boldsymbol{I}_{N_0}\ \boldsymbol{O}] \in \mathbb{R}^{N_0\times\hat{N}}$, $\boldsymbol{\Pi}_{\tilde{N}} := [\boldsymbol{I}_{\tilde{N}}\ \boldsymbol{O}] \in \mathbb{R}^{\tilde{N}\times\hat{N}}$ with[5] $\hat{N} := \tilde{N} + N_0 - 1$, and $C_0 := \{\hat{\boldsymbol{h}} \in \mathbb{R}^{\hat{N}} \mid \hat{h}_{N_0+i} = 0,\ \forall i \in \{1, 2, \ldots, \hat{N} - N_0\}\}$, and (ii) introducing an auxiliary variable $\hat{\boldsymbol{\xi}} = \mathcal{C}(\tilde{\boldsymbol{S}}^t)\hat{\boldsymbol{h}} \in \mathbb{R}^{\hat{N}}$. That is, we obtain an equivalent problem of (7)

$$\min_{(\hat{\boldsymbol{h}},\hat{\boldsymbol{\xi}})\in\mathbb{R}^{\hat{N}}\times\mathbb{R}^{\hat{N}}} f_k(\langle\boldsymbol{\Pi}^t\tilde{\boldsymbol{S}}\tilde{\boldsymbol{x}}_k, \hat{\boldsymbol{h}}\rangle) + \lambda_k\|\boldsymbol{\Pi}_{\tilde{N}}\hat{\boldsymbol{\xi}}\|_1^{\boldsymbol{w}_k} + \iota_{C_0}(\hat{\boldsymbol{h}}) + \iota_{C_1}(\hat{\boldsymbol{h}}, \hat{\boldsymbol{\xi}}) \tag{9}$$

with[6] a closed convex set $C_1 := \{(\hat{\boldsymbol{h}}, \hat{\boldsymbol{\xi}}) \in \mathbb{R}^{\hat{N}} \times \mathbb{R}^{\hat{N}} \mid \hat{\boldsymbol{\xi}} = \mathcal{C}(\tilde{\boldsymbol{S}}^t)\hat{\boldsymbol{h}}\}$, of which any solution $\hat{\boldsymbol{h}}_*$ recovers a solution $\boldsymbol{h}_*$ of problem (7) by equation (8). By applying Scheme 1 to problem (9) with

$$\varphi_k(\hat{\boldsymbol{h}}, \hat{\boldsymbol{\xi}}) = f_k(\langle\boldsymbol{\Pi}^t\tilde{\boldsymbol{S}}\tilde{\boldsymbol{x}}_k, \hat{\boldsymbol{h}}\rangle) + \lambda_k\|\boldsymbol{\Pi}_{\tilde{N}}\hat{\boldsymbol{\xi}}\|_1^{\boldsymbol{w}_k} + \iota_{C_0}(\hat{\boldsymbol{h}}),$$
$$\psi_k(\hat{\boldsymbol{h}}, \hat{\boldsymbol{\xi}}) = \iota_{C_1}(\hat{\boldsymbol{h}}, \hat{\boldsymbol{\xi}}),$$

we propose an adaptive learning of $\boldsymbol{h}_*$ (see the following remark for computation of the proximity operators and Algorithm 1 for the resulting algorithm).

[Remark 1]**(Efficient Computation of Proximity Operator of φ_k and ψ_k)** (a) For the function φ_k, we have

$$\mathrm{prox}_{\gamma_k\varphi_k}(\hat{\boldsymbol{h}}, \hat{\boldsymbol{\xi}}) = \left(\boldsymbol{\Pi}^t\mathrm{prox}_{\gamma_k f_k(\langle\tilde{\boldsymbol{S}}\tilde{\boldsymbol{x}}_k,\cdot\rangle)}(\boldsymbol{\Pi}\hat{\boldsymbol{h}}), \mathrm{prox}_{\gamma_k\lambda_k\|\boldsymbol{\Pi}_{\tilde{N}}\cdot\|_1^{\boldsymbol{w}_k}}(\hat{\boldsymbol{\xi}})\right) \tag{10}$$

with[7]

$$\mathrm{prox}_{\gamma_k f_k(\langle\tilde{\boldsymbol{S}}\tilde{\boldsymbol{x}}_k,\cdot\rangle)}(\boldsymbol{h}) = \boldsymbol{R}_k\left(\boldsymbol{h} + 2\gamma_k(e_k + \boldsymbol{s}^t\boldsymbol{y}_k)\tilde{\boldsymbol{S}}\tilde{\boldsymbol{x}}_k\right),$$
$$\boldsymbol{R}_k := \boldsymbol{I}_{N_0} - \left((2\gamma_k)^{-1} + \|\tilde{\boldsymbol{S}}\tilde{\boldsymbol{x}}_k\|^2\right)^{-1}\tilde{\boldsymbol{S}}\tilde{\boldsymbol{x}}_k(\tilde{\boldsymbol{S}}\tilde{\boldsymbol{x}}_k)^t \in \mathbb{R}^{N_0\times N_0},$$

$$\mathrm{prox}_{\gamma_k\lambda_k\|\boldsymbol{\Pi}_{\tilde{N}}\cdot\|_1^{\boldsymbol{w}_k}}(\hat{\boldsymbol{\xi}}) = \sum_{i=1}^{\tilde{N}}\mathrm{sgn}(\hat{\xi}_i)\max\left\{|\hat{\xi}_i| - \gamma_k\lambda_k w_i^{(k)},\ 0\right\}\boldsymbol{e}_i + \sum_{i=\tilde{N}+1}^{\hat{N}}\hat{\xi}_i\boldsymbol{e}_i \tag{11}$$

for any $\hat{\boldsymbol{h}}, \hat{\boldsymbol{\xi}} \in \mathbb{R}^{\hat{N}}$ and any $\boldsymbol{h} \in \mathbb{R}^{N_0}$, where (11) is *adaptively weighted soft-thresholding* proposed in [24] (as an extension of [25]).

(b) For the function ψ_k, we utilize the following three facts: (i) the circulant matrix $\mathcal{C}(\tilde{\boldsymbol{S}}^t)$ can be diagonalized by the Discrete Fourier Transform (DFT)[8] $\boldsymbol{U}^t \in \mathbb{C}^{\hat{N}\times\hat{N}}$ [23], i.e.,

$$\mathcal{C}(\tilde{\boldsymbol{S}}^t) = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^t$$

with[9] $\boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{U}^t\hat{\boldsymbol{s}})$ and $\hat{\boldsymbol{s}} := (\boldsymbol{s}^t, 0, \ldots, 0)^t \in \mathbb{R}^{\hat{N}}$;
(ii) by introducing $\phi_k := \psi_k \circ \boldsymbol{Z}$ with

$$\boldsymbol{Z} := \begin{pmatrix}\boldsymbol{U} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{U}\end{pmatrix} \in \mathbb{C}^{2\hat{N}\times2\hat{N}},$$

we have[10] $\mathrm{prox}_{\gamma\psi_k} \circ \boldsymbol{Z} = \boldsymbol{Z} \circ \mathrm{prox}^{\mathbb{C}}_{\gamma\phi_k}$;

[4]The distance between an arbitrary point $\mathbf{x} \in \mathbb{R}^N$ and a closed convex set $C \subset \mathbb{R}^N$ is defined by $d(\mathbf{x}, C) := \min_{\mathbf{y}\in C}\|\mathbf{x} - \mathbf{y}\|$.

[5]For any $N \in \mathbb{N}$, $\boldsymbol{I}_N \in \mathbb{R}^{N\times N}$ implies the identity matrix.

[6]For a given nonempty closed convex set $C \subset \mathbb{R}^N$, the indicator function $\iota_C : \mathbb{R}^N \to (-\infty, \infty]$ is defined by $\iota_C(\boldsymbol{x}) := 0$ if $\boldsymbol{x} \in C$, $\iota_C(\boldsymbol{x}) := \infty$ otherwise.

[7]The signum function $\mathrm{sgn}(\cdot)$ is defined by $\mathrm{sgn}(x) := x/|x|$ if $x \ne 0$, $\mathrm{sgn}(x) := 0$ otherwise, for all $x \in \mathbb{R}$, $\{\boldsymbol{e}_i\}_{i=1}^{\hat{N}}$ is the standard orthonormal basis of $\mathbb{R}^{\hat{N}}$ (i.e., $\boldsymbol{e}_i := [0, \ldots, 0, 1, 0, \ldots, 0]^t$ and $i \in \{1, 2, \ldots, \hat{N}\}$, with the value 1 assigned to its ith position).

[8]$\mathbb{C}^{\hat{N}}$ implies the $\hat{N}$-dimensional complex space with inner product $\langle\cdot,\cdot\rangle_{\mathbb{C}^{\hat{N}}} : \mathbb{C}^{\hat{N}} \times \mathbb{C}^{\hat{N}} \to \mathbb{C}, (\boldsymbol{x}, \boldsymbol{y}) \mapsto \bar{\boldsymbol{x}}^t\boldsymbol{y}$ and its induced norm $\|\cdot\|_{\mathbb{C}^{\hat{N}}} := \sqrt{\langle\cdot,\cdot\rangle_{\mathbb{C}^{\hat{N}}}}$, where $\bar{\boldsymbol{x}}$ implies the complex conjugate of $\boldsymbol{x}$. For a given vector $\boldsymbol{x} \in \mathbb{C}^{\hat{N}}$, $\boldsymbol{U}^t\boldsymbol{x}$ represents the DFT of $\boldsymbol{x}$ and $\boldsymbol{U}\boldsymbol{x}$ the inverse of DFT (IDFT).

[9]For a vector $\boldsymbol{x}$, $\mathrm{diag}(\boldsymbol{x})$ denotes the diagonal matrix whose entry is $\boldsymbol{x}$.

[10]By adopting an inner product $\langle(\boldsymbol{x}_1, \boldsymbol{x}_2), (\boldsymbol{y}_1, \boldsymbol{y}_2)\rangle_{\mathbb{C}^{\hat{N}}\times\mathbb{C}^{\hat{N}}} := \langle\boldsymbol{x}_1, \boldsymbol{y}_1\rangle_{\mathbb{C}^{\hat{N}}} + \langle\boldsymbol{x}_2, \boldsymbol{y}_2\rangle_{\mathbb{C}^{\hat{N}}}$ $(\forall(\boldsymbol{x}_1, \boldsymbol{x}_2), (\boldsymbol{y}_1, \boldsymbol{y}_2) \in \mathbb{C}^{\hat{N}} \times \mathbb{C}^{\hat{N}})$ and its induced norm $\|\cdot\|_{\mathbb{C}^{\hat{N}}\times\mathbb{C}^{\hat{N}}}$, for a proper lower semicontinuous convex function $f : \mathbb{C}^{\hat{N}} \times \mathbb{C}^{\hat{N}} \to (-\infty, \infty]$, define $\mathrm{prox}^{\mathbb{C}}_f : \mathbb{C}^{\hat{N}} \times \mathbb{C}^{\hat{N}} \to \mathbb{C}^{\hat{N}} \times \mathbb{C}^{\hat{N}}$,

$$\mathrm{prox}^{\mathbb{C}}_f(\boldsymbol{w}) := \mathop{\arg\min}_{\boldsymbol{z}\in\mathbb{C}^{\hat{N}}\times\mathbb{C}^{\hat{N}}}\left(f(\boldsymbol{z}) + \frac{1}{2}\|\boldsymbol{z} - \boldsymbol{w}\|^2_{\mathbb{C}^{\hat{N}}\times\mathbb{C}^{\hat{N}}}\right)$$

(see for an extended proximity operator in the complex space [26]).

Algorithm 1 Proposed Sparsity-Aware Adaptive Learning

Require: $(\hat{\boldsymbol{g}}_0, \hat{\boldsymbol{\zeta}}_0) \in \mathbb{R}^{\hat{N}} \times \mathbb{R}^{\hat{N}}$, $(t_k)_{k\in\mathbb{N}} \subset (0, 2)$, $(\gamma_k)_{k\in\mathbb{N}} \in (0, \infty)$, $k = 0$, $\boldsymbol{\sigma} := \mathrm{DFT}(\hat{\boldsymbol{s}})$,

$$\boldsymbol{\tau} := \mathrm{diag}\left((|\sigma_1|^2 + 1)^{-1}, \ldots, (|\sigma_{\hat{N}}|^2 + 1)^{-1}\right).$$

Repeat the following step:

(Compute $\mathrm{prox}_{\gamma_k \psi_k}$ by (12))

$$\begin{aligned}\boldsymbol{Q}_{\hat{\boldsymbol{g}}_c, \hat{\boldsymbol{\zeta}}_c} &= \boldsymbol{\tau}\mathrm{diag}\left(\mathrm{DFT}(\hat{\boldsymbol{\zeta}}_k) - \mathrm{diag}(\boldsymbol{\sigma})\mathrm{DFT}(\hat{\boldsymbol{g}}_k)\right)\\ \hat{\boldsymbol{h}}_k &= \hat{\boldsymbol{g}}_k + \mathrm{IDFT}(\boldsymbol{Q}_{\hat{\boldsymbol{g}}_c, \hat{\boldsymbol{\zeta}}_c}\bar{\boldsymbol{\sigma}})\\ \hat{\boldsymbol{\xi}}_k &= \hat{\boldsymbol{\zeta}}_k - \mathrm{IDFT}(\boldsymbol{Q}_{\hat{\boldsymbol{g}}_c, \hat{\boldsymbol{\zeta}}_c}\boldsymbol{1})\end{aligned}$$

(Compute $\mathrm{prox}_{\gamma_k \varphi_k}$ by (10))

$$\begin{aligned}\boldsymbol{R}_k &= \boldsymbol{I}_{N_0} - \left((2\gamma_k)^{-1} + \|\tilde{\boldsymbol{S}}\tilde{\boldsymbol{x}}_k\|^2\right)^{-1} \tilde{\boldsymbol{S}}\tilde{\boldsymbol{x}}_k(\tilde{\boldsymbol{S}}\tilde{\boldsymbol{x}}_k)^t\\ \hat{\boldsymbol{\mu}}_k &= \boldsymbol{\Pi}^t\boldsymbol{R}_k\left(\boldsymbol{\Pi}(2\hat{\boldsymbol{h}}_k - \hat{\boldsymbol{g}}_k) + 2\gamma_k(e_k + \boldsymbol{s}^t\boldsymbol{y}_k)\tilde{\boldsymbol{S}}\tilde{\boldsymbol{x}}_k\right)\end{aligned} \tag{13}$$

$$\begin{aligned}\hat{\boldsymbol{\chi}}_k &= 2\hat{\boldsymbol{\xi}}_k - \hat{\boldsymbol{\zeta}}_k\\ \hat{\boldsymbol{\nu}}_k &= \sum_{i=1}^{\tilde{N}} \mathrm{sgn}\big(\hat{\chi}_i^{(k)}\big)\max\left\{|\hat{\chi}_i^{(k)}| - \gamma_k\lambda_k w_i^{(k)},\ 0\right\}\boldsymbol{e}_i\\ &\quad + \sum_{i=\tilde{N}+1}^{\hat{N}} \hat{\chi}_i^{(k)}\boldsymbol{e}_i\end{aligned} \tag{14}$$

(Update $(\hat{\boldsymbol{g}}_{k+1}, \hat{\boldsymbol{\zeta}}_{k+1})$ by (6))

$$\begin{aligned}\hat{\boldsymbol{g}}_{k+1} &= \hat{\boldsymbol{g}}_k + t_k(\hat{\boldsymbol{\mu}}_k - \hat{\boldsymbol{h}}_k)\\ \hat{\boldsymbol{\zeta}}_{k+1} &= \hat{\boldsymbol{\zeta}}_k + t_k(\hat{\boldsymbol{\nu}}_k - \hat{\boldsymbol{\xi}}_k)\end{aligned}$$

$k \to k + 1$

(iii) $\mathrm{prox}^{\mathbb{C}}_{\gamma\phi_k}\colon \mathbb{C}^{\hat{N}} \times \mathbb{C}^{\hat{N}} \to \mathbb{C}^{\hat{N}} \times \mathbb{C}^{\hat{N}}$,

$$\mathrm{prox}^{\mathbb{C}}_{\gamma\phi_k}(\hat{\boldsymbol{h}}_c, \hat{\boldsymbol{\xi}}_c) = \left(\hat{\boldsymbol{h}}_c + \boldsymbol{Q}(\hat{\boldsymbol{h}}_c, \hat{\boldsymbol{\xi}}_c)\bar{\boldsymbol{\sigma}}, \hat{\boldsymbol{\xi}}_c - \boldsymbol{Q}(\hat{\boldsymbol{h}}_c, \hat{\boldsymbol{\xi}}_c)\boldsymbol{1}\right),$$

where $\boldsymbol{\sigma} := \boldsymbol{U}^t\hat{\boldsymbol{s}} \in \mathbb{C}^{\hat{N}}$, $\bar{\boldsymbol{\sigma}}$ is the complex conjugate of $\boldsymbol{\sigma}$,

$$\begin{aligned}&\boldsymbol{1} := (1, 1, \ldots, 1)^t \in \mathbb{R}^{\hat{N}},\\ &\boldsymbol{\tau} := \mathrm{diag}\left((|\sigma_1|^2 + 1)^{-1}, \ldots, (|\sigma_{\hat{N}}|^2 + 1)^{-1}\right) \in \mathbb{R}^{\hat{N}\times\hat{N}},\\ &\boldsymbol{Q}\colon\ \mathbb{C}^{\hat{N}} \times \mathbb{C}^{\hat{N}} \to \mathbb{C}^{\hat{N}\times\hat{N}}, (\hat{\boldsymbol{h}}_c, \hat{\boldsymbol{\xi}}_c) \mapsto \boldsymbol{\tau}\mathrm{diag}\left(\hat{\boldsymbol{\xi}}_c - \boldsymbol{\Sigma}\hat{\boldsymbol{h}}_c\right).\end{aligned}$$

These facts lead to a closed form expression of $\mathrm{prox}_{\gamma_k\psi_k}$:

$$\begin{aligned}\mathrm{prox}_{\gamma_k\psi_k}(\hat{\boldsymbol{h}}, \hat{\boldsymbol{\xi}}) = \big(&\hat{\boldsymbol{h}} + \boldsymbol{U}\boldsymbol{Q}(\boldsymbol{U}^t\hat{\boldsymbol{h}}, \boldsymbol{U}^t\hat{\boldsymbol{\xi}})\bar{\boldsymbol{\sigma}},\\ &\hat{\boldsymbol{\xi}} - \boldsymbol{U}\boldsymbol{Q}(\boldsymbol{U}^t\hat{\boldsymbol{h}}, \boldsymbol{U}^t\hat{\boldsymbol{\xi}})\boldsymbol{1}\big).\end{aligned} \tag{12}$$

[Remark 2]**(Extension to a simultaneous use[11] of the sparsity of $\boldsymbol{p}_*$ and $\boldsymbol{h}_*$)** By adding a weighted ℓ_1-norm of the filter $\boldsymbol{h}$ with its weight vector $\boldsymbol{w}'_k \in \mathbb{R}^{N_0}$ to the cost function in problem (7), we have

$$\min_{\boldsymbol{h}\in\mathbb{R}^{N_0}} f_k(\langle\tilde{\boldsymbol{S}}\tilde{\boldsymbol{x}}_k, \boldsymbol{h}\rangle) + \lambda_k\|\tilde{\boldsymbol{S}}^t\boldsymbol{h}\|_1^{\boldsymbol{w}_k} + \lambda'_k\|\boldsymbol{h}\|_1^{\boldsymbol{w}'_k},$$

and its equivalent form

$$\begin{aligned}\min_{(\hat{\boldsymbol{h}}, \hat{\boldsymbol{\xi}})\in\mathbb{R}^{\hat{N}}\times\mathbb{R}^{\hat{N}}} &f_k(\langle\boldsymbol{\Pi}^t\tilde{\boldsymbol{S}}\tilde{\boldsymbol{x}}_k, \hat{\boldsymbol{h}}\rangle) + \lambda_k\|\boldsymbol{\Pi}_{\tilde{N}}\hat{\boldsymbol{\xi}}\|_1^{\boldsymbol{w}_k}\\ &+ \iota_{C_0}(\hat{\boldsymbol{h}}) + \iota_{C_1}(\hat{\boldsymbol{h}}, \hat{\boldsymbol{\xi}}) + \lambda'_k\|\boldsymbol{\Pi}\hat{\boldsymbol{h}}\|_1^{\boldsymbol{w}'_k}.\end{aligned} \tag{15}$$

Then a direct application of Scheme 1 to problem (15) produces an algorithm same as the proposed algorithm except the update (13) of $\hat{\boldsymbol{h}}_k$, which is replaced by

$$\begin{aligned}\hat{\boldsymbol{\mu}}_k = \mathrm{prox}_{\gamma_k\lambda'_k\|\boldsymbol{\Pi}\cdot\|_1^{\boldsymbol{w}'_k}}\Big[&\boldsymbol{\Pi}^t\boldsymbol{R}_k\Big(\boldsymbol{\Pi}(2\hat{\boldsymbol{h}}_k - \hat{\boldsymbol{g}}_k)\\ &+ 2\gamma(e_k + \boldsymbol{s}^t\boldsymbol{y}_k)\tilde{\boldsymbol{S}}\tilde{\boldsymbol{x}}_k\Big)\Big].\end{aligned}$$

[11] Note that the sparsity of the concatenation does not imply that of $\boldsymbol{h}_*$. In fact, we can theoretically generate a pair of a sparse concatenation and a dense $\boldsymbol{h}_*$. Empirically, the concatenation is sparse if $\boldsymbol{p}_*$ is sparse, and $\boldsymbol{h}_*$ is sparse if $\boldsymbol{p}_*$ is similar to $\boldsymbol{s}_*$.

TABLE I

PARAMETER SETTINGS.

Size of $\boldsymbol{p}^*$	N_1	100
Size of $\boldsymbol{s}^*$	N_2	20
Size of $\boldsymbol{h}^*$	N_0	100
Reference noise	$(x_k)_{k\geq 0}$	(16)
Observation noise	$(v_k)_{k\geq 0}$	(17)
Stepsize of FxLMS		4.0×10^{-5}
Stepsize of MFxLMS		2.5×10^{-2}
Stepsize of Sun's Algorithm		5.0×10^{-5}
Stepsize of Akhtar's Algorithm		5.0×10^{-5}
Parameters of Proposed 1.	$(\lambda_k, t_k, \gamma_k)$	$(0.3, 1.0, 5.0 \times 10^{-5})$
Parameters of Proposed 2.	$(\lambda_k, t_k, \gamma_k)$	$(0.0, 1.0, 5.0 \times 10^{-5})$

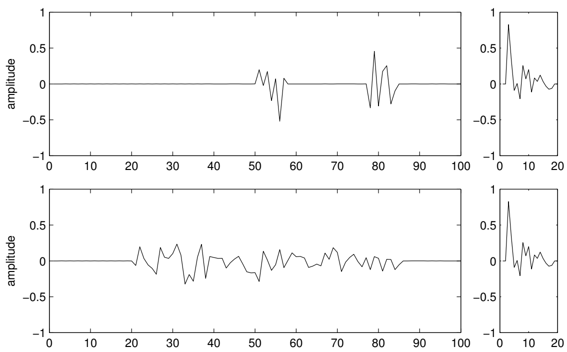

Fig. 2. Primary path of sparse (top left)/dense (bottom left) and secondary path (right). The secondary path is same in the both cases. These paths are generated artificially.

IV. NUMERICAL EXAMPLES

We examine the performance of the proposed sparsity-aware adaptive learning algorithm. To clarify effect of the sparsity, we adopt a primary path of sparse or dense (see Fig. 2), as well as employ the exact secondary path as its estimate, i.e., $\boldsymbol{s} = \boldsymbol{s}_*$. Since the reference noise signal $(x_k)_{k\geq 0}$ and the observation noise $(v_k)_{k\geq 0}$ are impulsive in practical situations, these noises are generated from random variables x_k^I and v_k^I

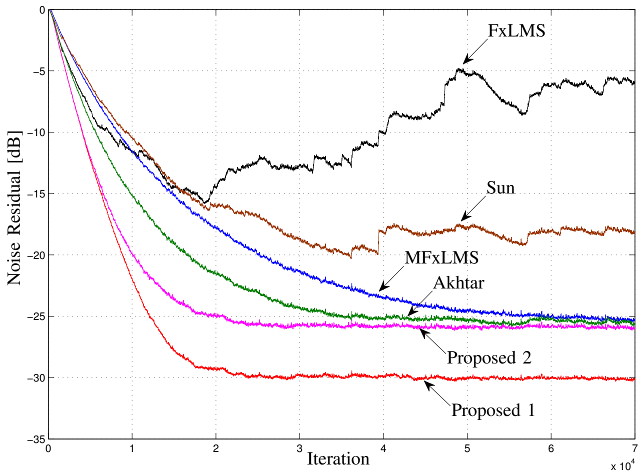

Fig. 3. A comparison of noise residual averaged over 300 trials for the sparse primary path. Although FxLMS and Sun's algorithm are unstable against the impulsive reference noise, the others are stable. Proposed 1 achieves the best performance in all the algorithms. Since Proposed 1 differs from Proposed 2 only the parameter λ_k of the sparsity promoting term in (7), this behavior shows that our use of the sparsity improves the performance of the ANC.

with α-stable distribution[12] [27] of $\alpha = 1.63$:

$$x_k = x_k^G + x_k^I, \tag{16}$$

$$v_k = v_k^G + 0.1 \times v_k^I, \tag{17}$$

where x_k^G and v_k^G are drawn from a zero mean white Gaussian distribution with variance 1. We apply the filtered-x least mean squares (FxLMS) [6], the modified filtered-x least mean squares (MFxLMS) [9], Sun's algorithm [11], Akhtar's Algorithm [12][13], Algorithm 1 (Proposed 1), and Algorithm 1 of $\lambda_k = 0$ (Proposed 2) which is unaware of the sparsity. Table I shows parameter settings of the model and the algorithms. The stepsizes of conventional algorithms are chosen in a way that

[12]The α-stable distribution was utilized to generate impulsive noise sequences, e.g., in [11], [12], because a small α ($\in (0, 2]$) implies a heavy tailed distribution.

[13]Sun's algorithm

$$\boldsymbol{h}_{k+1} = \boldsymbol{h}_k + \mu e_k \tilde{\boldsymbol{S}} \tilde{\boldsymbol{x}}'_k,$$

utilizes a modified reference signal $\tilde{\boldsymbol{x}}'_k$

$$x'_k = \begin{cases} 0 & x_k < c_1 \\ x_k & x_k \in [c_1, c_2] \\ 0 & x_k > c_2 \end{cases}$$

with user-defined parameters $c_1, c_2 \in \mathbb{R}\colon c_1 < c_2$, i.e., significant values of $\tilde{\boldsymbol{x}}_k$ are removed. Akhtar's algorithm

$$\boldsymbol{h}_{k+1} = \boldsymbol{h}_k + \mu e''_k \tilde{\boldsymbol{S}} \tilde{\boldsymbol{x}}''_k,$$

also modifies significant reference signals and error signals as $x''_k = P_{[c_1,c_2]}(x_k)$ and $e''_k = P_{[c_1,c_2]}(e_k)$ with $P_{[c_1,c_2]}\colon \mathbb{R} \to \mathbb{R}$,

$$P_{[c_1,c_2]}(r) = \begin{cases} c_1 & r < c_1 \\ r & r \in [c_1, c_2] \\ c_2 & r > c_2. \end{cases}$$

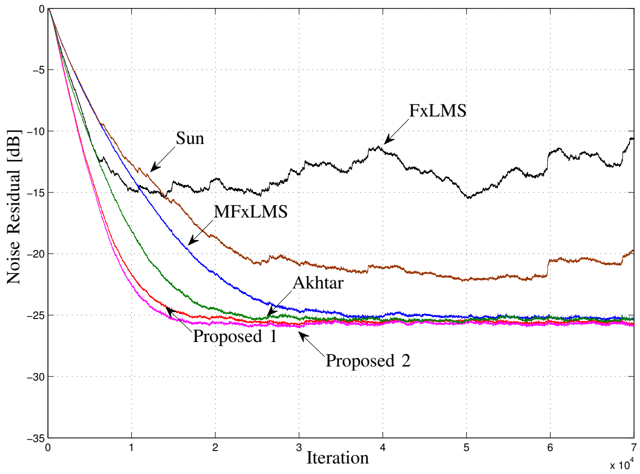

Fig. 4. A comparison of noise residual averaged over 300 trials for the dense primary path. Proposed 1 achieves a comparable performance for Proposed 2, which shows a robustness of Proposed 1 against a violation of the sparsity assumption.

the suppression speed are same in early iterations. In Sun's algorithm and Akhtar's algorithm, the paramters c_1, c_2 are selected as 0.01 and 99.99 percentile of the reference noise.

We adopt as a criterion the noise residual defined by a power ratio of the uncontrolled noise $d_k := \boldsymbol{p}_*^t \boldsymbol{x}_k$ and the controlled noise at the error microphone, i.e., $e'_k = d_k - \boldsymbol{s}_*^t \boldsymbol{y}_k$:

$$(\mathrm{NR}) = 10 \log_{10} \left(\frac{A_{e'_k}}{A_{d_k}} \right),$$

where a low-pass filter is utilized to clarify its behavior, i.e.,

$$A_{e'_k} := \eta A_{e'_{k-1}} + (1 - \eta)(e'_k)^2,$$
$$A_{d_k} := \eta A_{d_{k-1}} + (1 - \eta) d_k^2,$$

with $\eta = 0.99$.

As an example, we employ the design of $\boldsymbol{w}_k$ introduced in [24] to assign the small threshold $\gamma_k \lambda_k w_i^{(k)}$ in (14) for significant coefficients (see for other designs e.g. [28]):

$$w_i^{(k)} := \nu(\hat{\chi}_i^{(k)}),$$

$$\nu\colon \mathbb{R} \to (0, \infty), \nu(x) := \begin{cases} \delta, & \text{if } |x| > \tau, \\ 1, & \text{otherwise}, \end{cases}$$

for any $i \in \{1, 2, \ldots, \tilde{N}\}$, where $\delta := 10^{-9}$ and $\tau = 0.02$.

Fig. 3 shows a comparison of the noise residual for the sparse primary path. Although FxLMS and Sun's algorithm are unstable against the impulsive reference noise, other algorithms succeed to control the noise. Proposed 1 achieves the best performance in all the algorithms and improves approximately 4dB compared to Proposed 2, which shows that the proposed sparsity promoting term of the concatenation improves the performance.

Fig. 4 illustrates that Proposed 1 is robust against a violation of the sparsity assumption on the primary path because it achieves a comparable performance for Proposed 2.

V. Concluding Remark

This paper has proposed an efficient use of the sparsity of the primary path for the active noise control (ANC) in the frame of the adaptive Douglas-Rachford splitting (ADRS) scheme. Although our discussion in this paper assumes for a simplicity a time-invariant primary/secondary path, we can extend straightforwardly for a time-varying case by using online secondary path modeling techniques (see e.g. [18], [19]).

Future work includes (i) applications, (ii) complexity reductions, and (iii) performance improvements. (i) Exploiting the sparsity of the concatenation can be applicable in various noise control situations, e.g. multichannel active noise control and narrow band active noise control (see for recent developments [5], [29]). (ii) Though our update of the proposed algorithm is efficient, further complexity reduction is necessary because the discrete Fourier transform has been required at each iteration, which demands computational cost compared with the FxLMS-type algorithms. To reduce the computational cost of the proposed algorithm, the so-called homotopy algorithm for the generalized LARS [30] will be effective if the concatenation is sparse. (iii) For further performance improvements of the proposed algorithm, we can utilize the so-called variable metric technique (e.g., in the original ADRS scheme [16]) as well as can adopt multiple data-fidelity terms defined using observations at previous time (see e.g. [24], [31], [32]). In such a situation, the adaptive primal-dual splitting scheme [33], which utilizes the gradient descent update for the multiple smooth data-fidelity terms, will be useful to avoid use of proximity operators requiring huge computational cost in the ADRS. To robustify the performance of the ANC against unknown noise n_k in the underlying linear model (3), we can adopt in the ADRS a special design of the data-fidelity term introduced in [34]. Such extensions will be discussed elsewhere.

References

[1] S. J. Elliot and P. A. Nelson, "Active noise control," *IEEE Signal Processing Magazine*, vol. 10, no. 4, pp. 12–35, 1993.

[2] P. A. Nelson and S. J. Elliot, *Active control of sound*, Academic Press, Inc., 1993.

[3] S. M. Kuo and D. R. Morgan, *Active Noise Control Systems: Algorithms and DSP Implementation*, Wiley, New York, 1996.

[4] S. M. Kuo and D. R. Morgan, "Active noise control: a tutorial review," *Proceedings of the IEEE*, vol. 87, no. 6, pp. 943–973, 1999.

[5] Y. Kajikawa, W.-S. Gan, and S. M. Kuo, "Recent advances on active noise control: open issues and innovative applications," *APSIPA Transactions on Signal and Information Processing*, vol. 1, no. 1, 2012.

[6] D.R. Morgan, "An analysis of multiple correlation cancellation loops with a filter in the auxiliary path," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 28, no. 4, pp. 454–467, 1980.

[7] B. Widrow, D. Shur, and S. Shaffer, "On adaptive inverse control," in *Proc. 15th Asilomar Conf.*, 1981, pp. 185–189.

[8] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, 1985.

[9] E. Bjarnason, "Active noise cancellation using a modified form of the filtered-x LMS algorithm," in *Proc. EUSIPCO*, 1992, vol. 2, pp. 1053–1056.

[10] F. Jiang, H. Tsuji, H. Ohmori, and A. Sano, "Adaptation for active noise control," *IEEE Control Systems*, vol. 17, no. 6, pp. 36–47, 1997.

[11] X. Sun, S. M. Kuo, and G. Meng, "Adaptive algorithm for active control of impulsive noise," *Journal of Sound and Vibration*, vol. 291, no. 1, pp. 516–522, 2006.

[12] M. T. Akhtar and W. Mitsuhashi, "Improving performance of FxLMS algorithm for active noise control of impulsive noise," *Journal of Sound and Vibration*, vol. 327, no. 3, pp. 647–656, 2009.

[13] J. Arenas-García, M. de Diego, L. A. Azpicueta-Ruiz, M. Ferrer, and A. Gonzalez, "Combinations of proportionate adaptive filteres in acoustics: an application to active noise control," in *Proc. EUSIPCO*, 2011, pp. 1269–1273.

[14] A. Asaei, M. J. Taghizadeh, H. Bourlard, and V. Cevher, "Multi-party speech recovery exploiting structured sparsity models," in *Proc. INTERSPEECH*, 2011, pp. 185–188.

[15] C. Paleologu, J. Benesty, and S. Ciochina, *Sparse Adaptive Filters for Echo Cancellation*, Morgan and Claypool Publishers, 2010.

[16] I. Yamada, S. Gandy, and M. Yamagishi, "Sparsity-aware adaptive filtering based on a Douglas-Rachford splitting," in *Proc. EUSIPCO*, 2011, pp. 1929–1933.

[17] M. Yamagishi and I. Yamada, "Sparsity-aware feed-forward active noise control with the adaptive Douglas-Rachford splitting," *IEICE Technical Report*, vol. 113, no. 27, pp. 151–156, 2013.

[18] L. J. Eriksson and M. C. Allie, "Use of random noise for on-line transducer modeling in an adaptive active attenuation system," *The Journal of the Acoustical Society of America*, vol. 85, no. 2, pp. 797–802, 1989.

[19] L. J. Eriksson, *Active Acoustic Attenuation System with Overall Modeling*, Patent US4987598, Nelson Industries, Inc., 1991.

[20] J. B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms*, vol. 1–2, Springer-Verlag, 1993.

[21] J. J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," *C. R. Acad. Sci. Paris Sér*, vol. 255, pp. 2897–2899, 1962.

[22] P. L. Combettes and J.-C. Pesquet, "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE J. Sel. Top. Signal Process.*, vol. 1, pp. 564–574, 2007.

[23] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 3 edition, 1996.

[24] Y. Murakami, M. Yamagishi, M. Yukawa, and I. Yamada, "A sparse adaptive filtering using time-varying soft-thresholding techniques," in *Proc. IEEE ICASSP*, 2010, pp. 3734–3737.

[25] D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[26] L. Chaâri, A. Benazza-Benyahia, J.-C. Pesquet, and P. Ciuciu, "Wavelet-based parallel MRI regularization using bivariate sparsity promoting priors," in *Proc. IEEE ICIP*, 2009, pp. 1725–1728.

[27] J. M. Chambers, C. L. Mallows, and B. W. Stuck, "A method for simulating stable random variables," *Journal of the American Statistical Association*, vol. 71, no. 354, pp. 340–344, 1976.

[28] M. Yukawa, Y. Tawara, M. Yamagishi, and I. Yamada, "Sparsity-aware adaptive filters based on lp-norm inspired soft-thresholding technique," in *IEEE ISCAS*, May 2012, pp. 2749–2752.

[29] B. Huang, Y. Xiao, J. Sun, and G. Wei, "A variable step-size fxlms algorithm for narrowband active noise control," *IEEE Trans. audio, speech, and language processing*, vol. 21, no. 2, pp. 301–312, 2013.

[30] J.R. Tibshirani, *The solution path of the generalized lasso*, Stanford University, 2011.

[31] M. Yamagishi, M. Yukawa, and I. Yamada, "Sparse system identification by exponentially weighted adaptive parallel projection and generalized soft-thresholding," in *Proc. APSIPA ASC*, 2010, pp. 367–370.

[32] M. Yamagishi, M. Yukawa, and I. Yamada, "Acceleration of adaptive proximal forward-backward splitting method and its application to sparse system identification," in *Proc. IEEE ICASSP*, 2011, pp. 4296–4299.

[33] S. Ono, M. Yamagishi, and I. Yamada, "A sparse system identification by using adaptively-weighted total variation via a primal-dual splitting approach," in *Proc. IEEE ICASSP*, 2013, pp. 6029–6033.

[34] T. Yamamoto, M. Yamagishi, and I. Yamada, "Adaptive proximal forward-backward splitting for sparse system identification under impulsive noise," in *Proc. EUSIPCO*, 2012, pp. 2620–2624.