# Two Variable Step-Size Adaptive Algorithms for

# Non-Gaussian Interference Environment Using

# Fractionally Lower-Order Moment Minimization

Yahong Rosa Zheng and Vítor H. Nascimento

## Abstract

Two variable step-size adaptive algorithms using fractionally lower-order moment minimization are proposed for system identification in non-Gaussian interference environment. The two algorithms automatically adjust their step-sizes and adapt the weight vector by minimizing the $p$-th moment of the *a posteriori* error, where $p$ is the order with $1 \le p \le 2$, thus they are named as variable step-size normalized least mean p-th norm (VSS-NLMP) algorithms. The proposed adaptive VSS-NLMP algorithms are applied to both real- and complex-valued systems using low-complexity time-averaging estimation of the lower-order moments. Simulation results show that the misalignment of the proposed VSS-NLMP algorithms with a smaller $p$ converges faster and achieves lower steady-state error in impulsive interference and/or colored input environment. The adaptive VSS-NLMP algorithms also perform better than the adaptive Fixed Step-Size (FSS) NLMP in both Gaussian and impulsive interference environment. A theoretical model for the steady-state excess mean-square error is also provided for convergence rate in both Gaussian and Bernoulli-Gaussian interference.

## Index Terms

Y. R. Zheng is with Department of Electrical & Computer Engineering, Missouri University of Science and Technology, Rolla, MO 65409, USA. Email: zhengyr@mst.edu.

V. Nascimento is with Department of Electronic Systems Engineering, Univ. of São Paulo, São Paulo, Brazil, Email: vitor@lps.usp.br

Non-Gaussian interference suppression, robust adaptive filter, variable step size, Least Mean P-Moment (LMP) algorithm, fractionally lower-order moment (FLOM) algorithm, Bernoulli-Gaussian distribution, compound K distribution.

## I. INTRODUCTION

Adaptive filters have been commonly used in various applications of system identification, such as channel estimation [1], noise cancelation [2], echo cancelation [3], image restoration [4], [5], and seismic system identification [6], [7]. The most popular adaptive filtering algorithms are the least mean square (LMS) algorithm and normalized LMS (NLMS) algorithm, which have the advantage of simplicity. However, their major drawbacks are slow convergence [8] and performance degradation in colored input or non-Gaussian interference [9], and a compromise set by the step-size between low steady-state error and convergence and tracking rates.

Over the past two decades, many variants of LMS have been proposed to overcome these problems [5], [9]–[19]. Most of these algorithms take one or a combination of four approaches: 1) using variable step size to achieve both fast convergence and low steady-state errors [10]–[16]; 2) using a small-order affine projection [16], [17] to reach a better compromise between fast convergence and low computational complexity for colored inputs; 3) using nonlinear filtering such as median filtering, least mean p-norm (LMP), least mean absolute deviation (LMAD), higher order moment algorithms to combat non-Gaussian interference [5], [9], [18]–[25]; and 4) using combinations of multiple adaptive algorithms with different properties [26], [27].

Impulsive, non-Gaussian interference often occurs in practical applications and the LMS algorithm, as a method optimized for Gaussian models, suffers performance degradation in non-Gaussian environments. Interfering signals with heavy-tailed distributions produce more outliers than those assumed by Gaussian models. The characterization of a non-Gaussian signal by its second order moment is no longer optimal and many studies have shown that higher or lower order statistics can lead to improved convergence or improved robustness against non-Gaussian interference. One solution is based on Huber robust statistics, which leads to filters based on a saturation non-linearity, which can be approximated by a least-mean $p$-norm filter with $1 \leq p < 2$ [28]–[30]. The approach using higher or lower order statistics yields several families of algorithms including the normalized sign algorithms (NSA) or least mean absolute deviation (LMAD) algorithms [5], [31], [32], fractional lower-order moment/statistic (FLOM or FLOS)

algorithms [9], [20], and least mean fourth-moment (LMF) algorithms [33]. These algorithms are based on the norms $L_1$, $L_p$ with $1 < p < 2$, and $L_4$, respectively, rather than the $L_2$ norm. We refer this class of algorithms as the least mean p-norm (LMP) algorithms, where $p = 2$ leads to the conventional LMS algorithm and $p = 1$ leads to the LMAD algorithm.

In particular, the adaptive FLOM algorithms (or LMP with $1 \leq p < 2$) have been developed for impulsive models. They were first proposed for systems with alpha-stable distributed inputs [9], and later for continuous-time filters in the context of Huber robust statistics [28]. More recently, they were reformulated to the normalized NLMP algorithm [19]. It has been shown [4], [5], [34] that the adaptive NLMP algorithms perform better than the NLMS in heavy-tailed non-Gaussian interference, but slightly worse in Gaussian interference environment. When the order $p$ is smaller, the adaptive NLMP algorithm achieves faster convergence but higher steady-state errors than the NLMS algorithm. The conflicting goals of fast convergence and low steady-state error are caused by the inherent limitation of the fixed step size in both the NLMS and NLMP algorithms. A large step size results in fast convergence but large steady-state errors; whilst a small step size achieves small steady-state errors but with slow convergence.

Several variable step-size NLMS algorithms have been proposed in the literature [10]–[16], [35]. The basic idea is to use a time-varying step size to achieve a compromise between fast convergence and small steady-state error. The step size is automatically adjusted according to a criterion. For example, the step size is selected in [10] based on the correlation between the *a priori* error and the *a posteriori* error, or in [16], [35] by minimizing the mean-square deviation. In [14] and references therein, the proportionate NLMS algorithms control the step size at each filter tap individually based on the difference between the current value of the coefficient and the averaged, past values. In [11], the step size is chosen by matching the *a posteriori* mean-square error (MSE) to the power of the background white noise rather than simply minimizing the MSE. This matching of powers leads to a quadratic function of the step size and an approximate solution to the quadratic function results in a nonparametric VSS-NLMS algorithm. In [12] and [13], a mixed-norm and a switched-norm algorithm are proposed, respectively, combining the NLMS with the NSA according to the error dynamics. In [26], [27] two different algorithms are run in parallel, and their outputs combined in a convex manner. While the resulting filter has excellent convergence and tracking properties, its complexity is double that of a conventional filter.

This paper proposes two variable step-size (VSS) adaptive algorithms for NLMP filters, thus combining the benefits of variable step sizes with the robustness of the lower order statistics algorithms against finite-variance impulsive interference. The weight vector is adapted by minimizing the $p$-th moment of the *a posteriori* error where $1 \leq p \leq 2$. The step-size is automatically controlled by approximating the power of the *a posteriori* error to that of the background white noise as in [11]. Although the The proposed adaptive VSS-NLMP algorithms extend the VSS-NLMS method [11] in two aspects: first, we extend the nonparametric variable step size approach to the lower order moment algorithms where the derivation using a lower-order moment $1 \leq p < 2$ is non-trivial; second, a different approximation is also derived to solve the quadratic equation of the step size along with the approximation used in [11], thus leading to two VSS-NLMP algorithms. In contrast, the exact solution to the quadratic equation does not lead to a good adaptive VSS algorithm.

The steady-state performance of excess mean square error (EMSE) of the proposed algorithms is analyzed using the energy conservation method [36, Chapter 15], for both Gaussian and Bernoulli-Gaussian noise models. The analysis results also apply to the VSS-NLMS method of [11], for which, to the best of our knowledge, no theoretical method currently exists. The proposed adaptive VSS-NLMP algorithms are also evaluated extensively by computer simulations under different interference and input signals, and for real- and complex-coefficient systems. The results indicate that the proposed VSS-NLMP algorithms achieve faster or comparable convergence rate and smaller steady-state error than the fixed step-size NLMP (FSS-NLMP) algorithms in all signal scenarios. The VSS-NLMP algorithms with $p = 1$ achieve the best performance among all scenarios and for both real- and complex-coefficient systems. The tracking performance and stability of the proposed algorithms are also investigated.

## II. THE PROPOSED VSS-NLMP ALGORITHM

Consider a system identification problem where the output signal from an unknown system described by a complex coefficient vector $\mathbf{w}_o$ is

$$y(k) = \mathbf{w}_o^H \mathbf{x}(k) + v(k) \tag{1}$$

where $\mathbf{x}(k)$ is the input signal vector of length $L$ and $v(k)$ is the background noise plus interference signal. The superscript $()^H$ denotes conjugate transpose (Hermitian). Let $\hat{\mathbf{w}}(k)$ be an estimate for $\mathbf{w}_o$ at iteration $k$ and define

the *a priori* and *a posteriori* errors as

$$e_k = y(k) - \hat{\mathbf{w}}^H(k-1)\mathbf{x}(k) = [\mathbf{w}_o - \hat{\mathbf{w}}(k-1)]^H \mathbf{x}(k) + v(k), \tag{2}$$

$$\varepsilon_k = y(k) - \hat{\mathbf{w}}^H(k)\mathbf{x}(k) = [\mathbf{w}_o - \hat{\mathbf{w}}(k)]^H \mathbf{x}(k) + v(k), \tag{3}$$

respectively. The adaptive filter must minimize the cost function selected as the $p$-th order moment of $\varepsilon_k$

$$J(\hat{\mathbf{w}}(k)) = E\left\{|y(k) - \hat{\mathbf{w}}^H(k)\mathbf{x}(k)|^p\right\}, \tag{4}$$

where $E\{\cdot\}$ is the expectation operator.

The adaptive FLOM algorithm [9] provides an approach similar to the LMS algorithm for updating the filter coefficients along the steepest descent of the cost function (4). Using the stochastic gradient approach, the filter coefficients are solved iteratively by [9]

$$\hat{\mathbf{w}}(k) = \hat{\mathbf{w}}(k-1) + \mu(k)g(e_k)\mathbf{x}(k) \tag{5}$$

where $\mu(k)$ is the step size, and $g(e_k) = e_k^*|e_k|^{p-2}$ is a function with odd symmetry, the superscript $^*$ denotes the conjugate and $|\cdot|$ the absolute value.

In the absence of noise and interference, a reasonable method for selecting a variable step size is to set $\varepsilon_k$ equal to 0. However, in the presence of noise and interference, a better criterion [11] is to set $[\mathbf{w}_o - \hat{\mathbf{w}}(k)]^H \mathbf{x}(k)$ equal to 0 for all $k$. This implies, based on (3), that the variable step size is selected to satisfy

$$\mathcal{S}_\varepsilon(k) = E\{\varepsilon_k \varepsilon_k^*\} \approx E\{v(k)v^*(k)\} = \mathcal{S}_v. \tag{6}$$

where $\mathcal{S}_v$ is the noise-plus-interference power which is often estimated during the absence of the input signal (for example, during periods of silence in speech). Subtracting (2) from (3) and substituting (5), the *a posteriori* error in terms of $\mu(k)$ can be expressed as,

$$\varepsilon_k = e_k + [\hat{\mathbf{w}}(k-1) - \hat{\mathbf{w}}(k)]^H \mathbf{x}(k),$$

$$= e_k - \mu(k)g^*(e_k)\mathbf{x}^H(k)\mathbf{x}(k). \tag{7}$$

Substituting (7) in (6) yields

$$\mathcal{S}_\varepsilon(k) = E\left[|e_k|^2\right] - 2\mu(k)E\left[|e_k|^p \mathbf{x}^H(k)\mathbf{x}(k)\right] + \mu^2(k)E\left[|e_k|^{2p-2}|\mathbf{x}^H(k)\mathbf{x}(k)|^2\right] \tag{8}$$

Equating (8) to $S_v$ gives a quadratic equation in $\mu(k)$:

$$1 - 2\frac{b(k)}{\mathcal{S}_e(k)}\mu(k) + \frac{a(k)}{\mathcal{S}_e(k)}\mu^2(k) \approx \frac{S_v}{\mathcal{S}_e(k)}. \tag{9}$$

where $S_e(k) = E\left[|e_k|^2\right]$ and

$$b(k) \quad := \quad E\left[|e_k|^p \mathbf{x}^H(k)\mathbf{x}(k)\right], \tag{10}$$

$$a(k) \quad := \quad E\left[|e_k|^{2p-2}(\mathbf{x}^H(k)\mathbf{x}(k))^2\right]. \tag{11}$$

We also noticed that in some situations the following approximations lead to better filter performance:

$$b(k) \quad \approx \quad \mathcal{L}_p(e_k)\mathcal{S}_{xx}(k), \tag{12}$$

$$a(k) \quad \approx \quad \mathcal{L}_{2p-2}(e_k)\mathcal{P}_{xx}(k)). \tag{13}$$

where $\mathcal{L}_p(e_k) := E[|e_k|^p]$, $\mathcal{L}_{2p-2}(e_k) := E[|e_k|^{2p-2}]$, $\mathcal{S}_{xx}(k) := E[\mathbf{x}^H(k)\mathbf{x}(k)]$, and $\mathcal{P}_{xx}(k) := E[|\mathbf{x}^H(k)\mathbf{x}(k)|^2]$.

These approximations are equivalent to assuming that $e_k$ and $\mathbf{x}(k)$ are independent. This is approximately true only after the filter converged; however, as our simulations will show, the alternative filters also show good performance in the transient.

The quadratic function (9) can be easily solved using time-averaged estimates of $a(k)$, $b(k)$ and $S_e(k)$. However, the exact roots of the quadratic equation are often complex or out of the stability range of $\mu(k)$. Even if one or both roots of (9) are real and valid at some instant $k$, using those exact roots for $\mu(k)$ can cause large jumps in the step size and does not lead to a good variable step size algorithm. We show in the Appendix that the ratio $b^2(k)/[a(k)\mathcal{S}_e(k)]$ is very close to one for $p = 2$ and large filter length $L$, thus the approximation $b(k)/\mathcal{S}_e(k) \approx \sqrt{a(k)/\mathcal{S}_e(k)}$ yields a good VSS adaptive algorithm, as shown in [11]. This approximation, as can be seen from the discussion below, does not suffer from these problems listed above. For small $L$ and $p \ll 2$, however, the ratio $b^2(k)/[a(k)\mathcal{S}_e(k)]$ is significantly smaller than one. In this case, we use two approximate solutions to (9) to derive two variable step-size algorithms that have low computational complexity and stable $\mu(k)$. We name them the adaptive VSS-NLMPa and VSS-NLMPb algorithms, as will be presented next.

First, replacing $b(k)/\mathcal{S}_e(k)$ by $\sqrt{a(k)/\mathcal{S}_e(k)}$ in (9) yields the adaptive VSS-NLMPa algorithm

$$\left[1 - \sqrt{\frac{a(k)}{\mathcal{S}_e(k)}}\mu(k)\right]^2 \approx \frac{S_v}{\mathcal{S}_e(k)}. \tag{14}$$

The variable step size is then formulated as

$$\mu(k) = \sqrt{\frac{\mathcal{S}_e(k)}{a(k)}} \left[ 1 - \sqrt{\frac{\mathcal{S}_v}{\mathcal{S}_e(k)}} \right], \qquad \text{if} \quad \mathcal{S}_v \le \mathcal{S}_e(k) \tag{15}$$

where $\mathcal{S}_e(k)$ and $a(k)$ can be estimated by time averaging using a forgetting factor $\lambda$

$$\hat{\mathcal{S}}_e(k) = \lambda \hat{\mathcal{S}}_e(k-1) + (1-\lambda)|e_k|^2, \tag{16}$$

$$\hat{a}(k) = \lambda \hat{a}(k-1) + (1-\lambda)|e_k|^{2p-2} \left( \mathbf{x}^H(k)\mathbf{x}(k) \right)^2. \tag{17}$$

Similarly, $\mathcal{S}_v$ must be estimated off-line or during periods of absence of the input signal. We denote the estimated

noise-plus-interference power as $\hat{\mathcal{S}}_v$.

A variant of VSS-NLMPa uses (13) instead of (11) and the variable step size becomes

$$\mu(k) = \sqrt{\frac{\mathcal{S}_e(k)}{\mathcal{L}_{2p-2}(e_k)\mathcal{P}_{xx}(k)}} \left[ 1 - \sqrt{\frac{\mathcal{S}_v}{\mathcal{S}_e(k)}} \right], \qquad \text{if} \quad \mathcal{S}_v \le \mathcal{S}_e(k) \tag{18}$$

using time-averaging estimates $\mathcal{L}_{2p-2}(e_k)$ and $\mathcal{P}_{xx}(k)$ instead of $\hat{a}(k)$.

$$\hat{\mathcal{P}}_{xx}(k) = \lambda \hat{\mathcal{P}}_{xx}(k-1) + (1-\lambda)|\mathbf{x}^H(k)\mathbf{x}(k)|^2, \tag{19}$$

$$\hat{\mathcal{L}}_{2p-2}(e_k) = \lambda \hat{\mathcal{L}}_{2p-2}(e_{k-1}) + (1-\lambda)|e_k|^{2p-2}. \tag{20}$$

Next, an alternative approximation to (14) is to replace $a(k)/\mathcal{S}_e(k)$ by $[b(k)/\mathcal{S}_e(k)]^2$, then we have

$$\left[ 1 - \frac{b(k)}{\mathcal{S}_e(k)}\mu(k) \right]^2 \approx \frac{\mathcal{S}_v}{\mathcal{S}_e(k)} \tag{21}$$

This results in the adaptive VSS-NLMPb algorithm with variable step size as

$$\mu(k) = \frac{\mathcal{S}_e(k)}{b(k)} \left[ 1 - \sqrt{\frac{\mathcal{S}_v}{\mathcal{S}_e(k)}} \right] \qquad \text{if} \quad \mathcal{S}_v \le \mathcal{S}_e(k) \tag{22}$$

Again, a variant of the VSS-NLMPb using (12) instead of (10) yields

$$\mu(k) = \frac{\mathcal{S}_e(k)}{\mathcal{L}_p(e_k)\mathcal{S}_{xx}(k)} \left[ 1 - \sqrt{\frac{\mathcal{S}_v}{\mathcal{S}_e(k)}} \right], \qquad \text{if} \quad \mathcal{S}_v \le \mathcal{S}_e(k) \tag{23}$$

In practical implementations, the expectation $\mathcal{S}_{xx}(k) = E[\mathbf{x}^H(k)\mathbf{x}(k)]$ can be replaced by the instantaneous signal

energy $\mathbf{x}^H(k)\mathbf{x}(k)$. This is the approach taken by [11] for VSS-NLMS. It is interesting to notice that the step-size

$\mu(k)$ in (23) is inversely proportional to $E\left[\mathbf{x}^H(k)\mathbf{x}(k)\right]$, so that the filter includes normalization. All other variants

of VSS-NLMP, (15), (18), (22) also include some form of normalization, in the sense that the denominator for

$\mu(k)$ includes a term that is related to $\|\mathbf{x}(k)\|^2$.

TABLE I

THE VSS-NLMP ALGORITHMS

| Parameters | $\lambda = 1 - 1/(\eta L)$, forgetting factor, with $2 \leq \eta \leq 10$ | |
|---|---|---|
| | $\hat{\mathcal{S}}_v$ = estimated variance of background noise and interference | |
| | $\delta = CL\sigma_x^2$, $C$ = constant, $\sigma_x^2$ = power of input signal | |
| | $\epsilon > 0$, small constant to avoid division by zero; $\mu_0$ = step size multiplier | |
| Initialization | $\hat{\mathbf{w}}(0) = \mathbf{0}$, $\hat{\mathcal{S}}_e(0) = 0$ | |
| Init (original) | original $\hat{a}(0) = 0$ | $\hat{b}(0) = 0$ |
| Init. (variant) | $\hat{\mathcal{L}}_{2p-2}(e_0) = 0$, $\hat{\mathcal{P}}_{xx}(0) = 0$ | $\hat{\mathcal{L}}_p(e_0) = 0$, $\hat{\mathcal{S}}_{xx}(0) = 0$ |
| Error | $e_k = y(k) - \hat{\mathbf{w}}^H(k-1)\mathbf{x}(k)$ | |
| Estimates | $\hat{\mathcal{S}}_e(k) = \lambda\hat{\mathcal{S}}_e(k-1) + (1-\lambda)|e_k|^2$, $\gamma(k) = \left[1 - \sqrt{\frac{\hat{\mathcal{S}}_v}{\hat{\mathcal{S}}_e(k)+\epsilon}}\right]$, $\mathcal{L}_x(k) = \mathbf{x}^H(k)\mathbf{x}(k)$ | |
| Est. (original) | $\hat{a}(k) = \lambda\hat{a}(k-1) + (1-\lambda)|e_k|^{2p-2}\mathcal{L}_x^2(k)$ | $\hat{b}(k) = \lambda\hat{b}(k-1) + (1-\lambda)|e_k|^p\mathcal{L}_x(k)$ |
| Est. (variant) | $\hat{\mathcal{P}}_{xx}(k) = \lambda\hat{\mathcal{P}}_{xx}(k-1) + (1-\lambda)\mathcal{L}_x^2(k)$ | $\hat{\mathcal{S}}_{xx}(k) = \lambda\hat{\mathcal{S}}_{xx}(k-1) + (1-\lambda)\mathcal{L}_x(k)$ |
| | $\hat{\mathcal{L}}_{2p-2}(e_k) = \lambda\hat{\mathcal{L}}_{2p-2}(e_{k-1}) + (1-\lambda)|e_k|^{2p-2}$ | $\hat{\mathcal{L}}_p(e_k) = \lambda\hat{\mathcal{L}}_p(e_{k-1}) + (1-\lambda)|e_k|^p$ |
| Update (original) | $\alpha(k) = \sqrt{\frac{\hat{\mathcal{S}}_e(k)}{\hat{a}(k)+\delta^2}}$ | $\beta(k) = \frac{\hat{\mathcal{S}}_e(k)}{\hat{b}(k)+\delta}$ |
| Update (variant) | $\alpha(k) = \sqrt{\frac{\hat{\mathcal{S}}_e(k)}{\hat{\mathcal{L}}_{2p-2}(e_k)\hat{\mathcal{P}}_{xx}(k)+\delta^2}}$ | $\beta(k) = \frac{\hat{\mathcal{S}}_e(k)}{\hat{\mathcal{L}}_p(e_k)\hat{\mathcal{S}}_{xx}(k)+\delta}$ |
| Step-size | $\mu(k) = \begin{cases} \mu_0\alpha(k)\gamma(k), & \text{if } \hat{\mathcal{S}}_v \leq \hat{\mathcal{S}}_e(k) \\ 0, & \text{otherwise} \end{cases}$ | $\mu(k) = \begin{cases} \mu_0\beta(k)\gamma(k), & \text{if } \hat{\mathcal{S}}_v \leq \hat{\mathcal{S}}_e(k) \\ 0, & \text{otherwise} \end{cases}$ |
| Weight vector | $\hat{\mathbf{w}}(k) = \hat{\mathbf{w}}(k-1) + \mu(k)\mathbf{x}(k)e^{<p-1>}(k)$ | |

The proposed two adaptive VSS-NLMP algorithms and their variants are summarized in Table I, where $\delta$ and $\epsilon$ are small positive constants to avoid division by zero, and $\mu_0$ is a step size multiplier often selected as $\mu_0 \leq 1$. The left column gives the VSS-NLMPa algorithms and the right column shows the VSS-NLMPb.

## III.  PERFORMANCE EVALUATION VIA SIMULATION

The proposed VSS-NLMP algorithms were evaluated through computer simulations for two system identification applications. One was a real-coefficient system with real-valued signals and coefficients often used in speech or image processing applications, such as acoustic echo cancelation and watermark detection. Another was a complex-coefficient system, often found in communications and radar signal processing, such as channel estimation and adaptive beamforming [1]. In both cases, the adaptive filter had a length of $L = 128$ taps and the input was chosen to be a white or colored Gaussian process. The colored input, denoted as AR(1) signal, was generated by filtering a

white Gaussian noise through a first order system with a pole at $0.8$ for the real-coefficient system and a pole at $0.5$ for the complex-coefficient system. An independent white Gaussian noise was added to the system background and the signal-to-noise ratio (SNR) was $30$ dB. In addition, a strong interference signal was also added to the system output $y(k)$ with an interference-to-noise ratio (INR) of $20$ dB.

In the real system, two types of distributions, a Gaussian and an impulsive Bernoulli-Gaussian [18], were considered for the interference signals. The Bernoulli-Gaussian interference was generated as the product of a Bernoulli process and a Gaussian process

$$v(k) = \omega(k)N(k) \tag{24}$$

where $N(k)$ was a white Gaussian random sequence with zero mean and variance $\sigma_N^2$, and $\omega(k)$ was a Bernoulli process with the probability mass function given as

$$P(\omega) = \begin{cases} 1 - P_r, & \omega = 0 \\ P_r, & \omega = 1 \end{cases} \tag{25}$$

The average power of the BG process was $P_r \cdot \sigma_N^2$. In general, a BG process is spikier when $P_r$ is smaller and it reduces to a Gaussian process when $P_r = 1$. BG interference is often seen in seismic system identification [6], [7] and can also model double talk in network echo cancellation [12], [37], [38].

In the complex system, the input, noise, and interference signals were all assumed complex. The real and imaginary parts of the input and noise were generated independently as white or filtered Gaussian and the complex interference signal was either complex Gaussian or compound K distributed [39]. The compound K distributed signal was generated as the product of two random processes: $Z_c(k) = \sqrt{G(k)} \cdot N(k)$, where $G(k)$ was a Gamma-distributed texture and $N(k)$ was a complex Gaussian speckle. The envelope of $Z_c(k)$, denoted as $r(k)$, exhibited compound K distribution as [39]

$$f_R(r) = \frac{4}{\sqrt{\rho}\Gamma(\nu)} \left( \frac{r}{\sqrt{\rho}} \right)^\nu K_{\nu-1}\left( \frac{2r}{\sqrt{\rho}} \right), \quad r \geq 0 \tag{26}$$

where $K_\nu(\cdot)$ is the modified Bessel function of the second kind and with order $\nu$, and $\nu$ and $\rho$ are the shape and scale parameters, respectively. The average power of the compound K process is $(\rho \cdot \nu)$. With a fixed average power, the compound K distribution with a smaller $\nu$ has a higher envelope tail. When $\nu \to \infty$, it reduces to the Rayleigh

distribution, which is the envelope distribution of a complex Gaussian random variable. Compound K distribution is often found in radar array applications and underwater acoustic communications [39]–[41].

For all simulation studies in both the real and complex systems, the VSS-NLMP algorithms used a forgetting factor $\lambda = 1 - 1/(\eta L)$, where $\eta = 2$ for white input signal, and $\eta = 6$ for AR(1) input signal, unless specified otherwise. The small constants were selected as $\delta = 0.04L\sigma_x^2$ and $\epsilon = 10^{-3}$. The step size multiplier for VSS-NLMP was $\mu_0 = 1$ and the step size for FSS-NLMP was $\mu = 0.1$. The multipliers were chosen so that the initial rate of convergence of all filters is the same.

The convergence performance was evaluated by the normalized misalignment $\mathcal{M}(k)$ defined as [1]

$$\mathcal{M}(k) = 20\log_{10}\frac{||\hat{\mathbf{w}}(k) - \mathbf{w}_o||}{||\mathbf{w}_o||} \tag{27}$$

An ensemble average of 100 trials was used for the evaluation of $\mathcal{M}(k)$.

*A. Comparison of VSS-NLMP algorithms and variants*

The performance of the proposed two VSS-NLMP algorithms and their variants were compared with the FSS-NLMP algorithms using the example of $L = 128$ real-coefficient system. With white input and Gaussian interference, all VSS-NLMP algorithms performed better than the FSS-NLMP, as shown in Fig. 1(a). The two VSS-NLMP algorithms had convergence rates very similar to their variants for all values of $p$. Their steady-state errors were slightly different, but the differences were less than 2 dB. With AR(1) input, BG interference and $p = 2$, both VSS-NLMP algorithms and their variants performed similarly, as shown in Fig. 1(b). Their convergence rate was slightly inferior to the FSS-NLMP, but the steady-state errors were slightly lower than that of the FSS-NLMP. For $p = 1$, the VSS-NLMPa performed better than the VSS-NLMPb with more than 5 dB lower steady-state error. Both VSS-NLMP algorithms performed better than the FSS-NLMP. The variant of the VSS-NLMPb algorithm exhibited 2 dB lower steady-state error than the original VSS-NLMPb algorithm; while the original VSS-NLMPa converged slightly faster than its variant but achieved the same steady-state error.

The final misalignments after convergence are compared in Table II for Gaussian and Bernoulli-Gaussian interference cases. It is clear that the proposed algorithms with $p = 1.5$ provide good tradeoff between Gaussian and non-Gaussian interference cases, while $p = 2$ only performs well in Gaussian interference. It is interesting

(a) White input, Gaussian interference
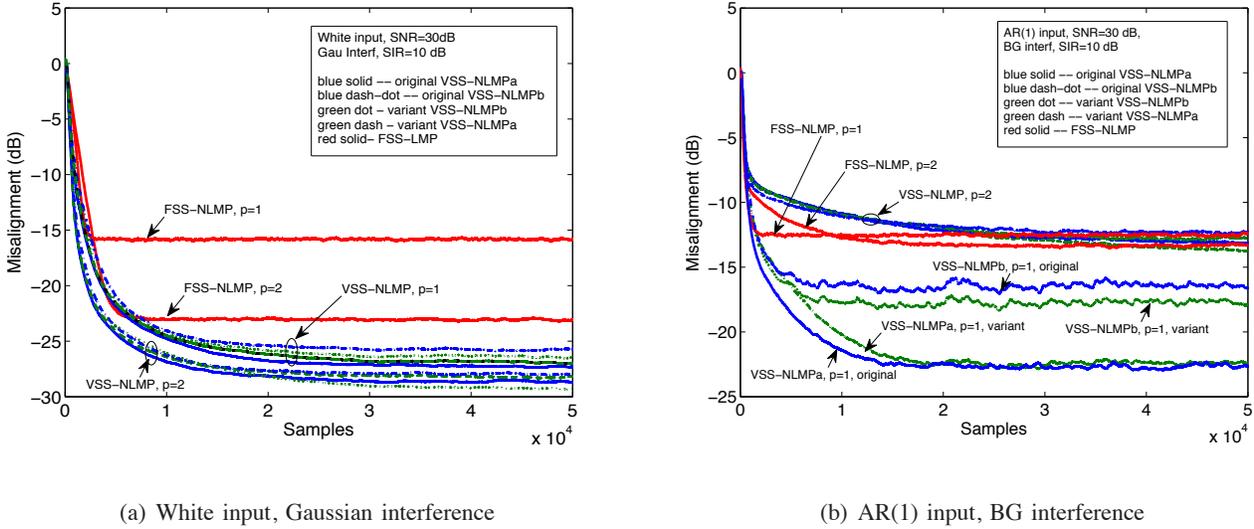


(b) AR(1) input, BG interference

Fig. 1.    Misalignment of VSS-NLMP algorithms in a real-coefficient system. SNR=30 dB. SIR=10 dB. $\delta = 0.04 L \sigma_x^2, \eta = 4$. The two VSS-NLMP algorithms had very similar performance as their variants for all orders of $p$. The results were the average of 100 simulations trials.

TABLE II

MISALIGNMENTS OF VSS-NLMP ALGORITHMS IN GAUSSIAN AND BG INTERFERENCE (DB)

|  | Gaussian interference | | | BG interference | | |
|---|---|---|---|---|---|---|
|  | $p = 1$ | $p = 1.5$ | $p = 2$ | $p = 1$ | $p = 1.5$ | $p = 2$ |
| VSS-NLMPa | $-27.2$ | $-28.0$ | $-28.7$ | $-22.5$ | $-19.8$ | $-15.0$ |
| VSS-NLMPa variant | $-26.4$ | $-26.4$ | $-29.3$ | $-22.5$ | $-19.3$ | $-14.9$ |
| VSS-NLMPb | $-25.7$ | $-26.6$ | $-27.9$ | $-16.2$ | $-18.1$ | $-12.7$ |
| VSS-NLMPb variant | $-27.0$ | $-27.9$ | $-28.4$ | $-18.3$ | $-18.7$ | $-13.0$ |

<span style="color:red">to note that, although the usual application of the NLMP algorithms is the case where the second order moment of the signal does not exist, the NLMP algorithms still perform better in the case of heavy-tailed finite-variance signals than the NLMS algorithms. The VSS-NLMS ($p = 2$) performed worse than the VSS-NLMP with $p < 2$ in Bernoulli-Gaussian interference because the BG distribution, albeit having a finite variance, has a heavier tail than the Gaussian distribution.</span>

Due to their small differences between the original VSS algorithms and their variants, we only present detailed performance evaluation for the original VSS-NLMP algorithms and omit those of the variants in the following subsections.

*B. Convergence of Real-Coefficient Systems*

The misalignment of the VSS-NLMP algorithms for real-coefficient system was compared with the FSS-NLMP algorithm under white or colored inputs and Gaussian or BG interference scenarios, as shown in Fig. 2. In white input and Gaussian interference, the two VSS-NLMP algorithms achieved faster convergence and lower steady-state error than the FSS-NLMP for all orders of $p$, as shown in Figs. 2(a), 2(c), and 2(e). For AR(1) input and Gaussian interference, the VSS-NLMP algorithms exhibited similar initial convergence but slightly slower secondary convergence than the FSS-NLMP. On the other hand, they converged to much lower steady state error than the FSS-NLMP. Both VSS-NLMP algorithms behaved similarly in their misalignment curves, and the same was observed in their excess MSE curves (not shown). The VSS-NLMP algorithms achieved similar steady-state error and convergence rate for all $p$. In contrast, the FSS-NLMP with a larger order $p$ converged more slowly but to a lower steady-state error than that of a smaller order. This means that the selection of order for the VSS-NLMP algorithms has low importance to the performance in Gaussian interference.

In the BG interference cases as shown in Figs. 2(b), 2(d), and 2(f), the VSS-NLMP algorithms converged faster and achieved lower steady-state error than the corresponding FSS-NLMP for all orders when the input was white. The two VSS-NLMP algorithms behaved similarly when $p$ is large, but quite differently when $p = 1$, with the VSS-NLMPa achieving lower steady-state error than the VSS-NLMPb. With the AR(1) input, the two VSS-NLMP algorithms with a smaller order performed much better than their corresponding FSS-NLMP, while the VSS-NLMP algorithms with a larger order performed comparably to the corresponding FSS-NLMP. Among all algorithms, the VSS-NLMPa with $p = 1$ had the best performance in all signal and interference scenarios achieving both fast convergence and low steady state error.

Comparing the two VSS-NLMP algorithms, the VSS-NLMPa and VSS-NLMPb have similar behavior when $p$ is large, but different behavior when $p \to 1$, especially in impulsive interference. The VSS-NLMPa ($p = 1$) greatly outperformed the VSS-NLMPb in terms of steady-state error in BG interference, as shown in Fig. 2(f). These results may be explained by the analysis of the coefficient ratio $b^2(k)/[\mathcal{S}_e(k)a(k)]$ in the Appendix. When $p = 2$, the ratio is close to 1 for large filters length, which results in similar performance of the VSS-NLMPa and VSS-NLMPb. When $p \to 1$, or $L$ is small, or interference is impulsive, the ratio drops dramatically and the VSS-NLMPa is shown

to perform better than the VSS-NLMPb.

## C. Convergence of Complex-Coefficient Systems

The performance of the VSS-NLMP algorithm for the complex system was also compared with the FSS-NLMP algorithm, as shown in Fig. 3. The complex-coefficient system was generated as a sum of sinusoids [42] for each of the 128 taps. Compound K interference with shape parameter of $\nu = 0.7$ was generated using [43]. The misalignment of the VSS-NLMP algorithms performed similarly for all orders, achieving faster convergence and smaller steady-state error than the corresponding FSS-NLMP algorithm. We also verified that the excess mean square error (MSE) curves (not shown here) of the VSS-NLMP algorithms exhibited similar performance gains over the FSS-NLMP.

## D. Tracking Performance

The tracking performance of the VSS-NLMP algorithms was evaluated with a real-coefficient system. The ideal system impulse response was modified at iteration $2.5 \times 10^4$ by multiplying the weight coefficients by $-1$. The tracking performance of both VSS-NLMP algorithms was compared in BG interference with white or colored inputs, as shown in Fig 4. Both VSS-NLMP algorithms with all values of $p$ were able to track the sudden change and converged quickly to the new system response. In other input and interference scenarios, the tracking performance was similar to those in Fig. 4 with slightly different convergence rate and steady-state errors. The tracking performance of the FSS-NLMP algorithm also behaved similarly to the VSS-NLMP algorithms.

## E. Discussion on Stability

For normalized LMS, it is well known that the step size multiplier has to satisfy $0 < \mu_0 < 2$ to ensure stability [1]. This applies to the selection of variable step sizes in the VSS-NLMP algorithms with $p = 2$. The variable step size selected in (18) and (23) with $\mu_0 \leq 1$ ensures that the two VSS-NLMP algorithms satisfy the stability condition when $p = 2$, thus ensuring the stability of the VSS-NLMP algorithms. In practice we noted that the filters remain stable even for $p < 2$ with the choice $\mu_0 \leq 1$.
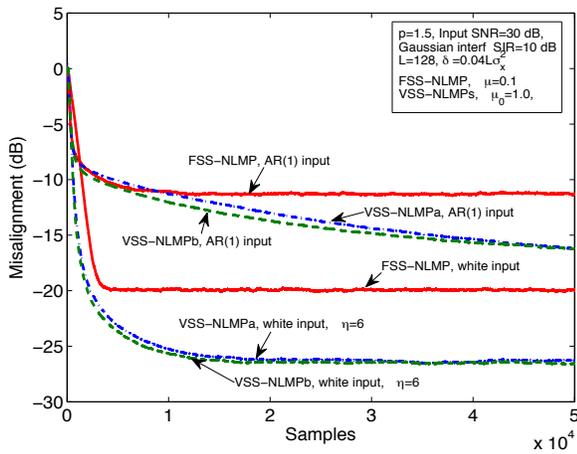
For $p = 1$, the LMP algorithm becomes a member of the $L_1$-norm (or sign algorithm, SA) family. For fixed step size sign algorithms, limited studies are devoted on the convergence analysis of the sign algorithm and its variants
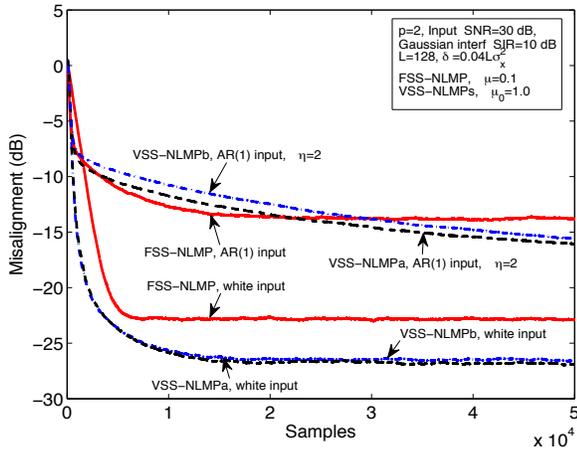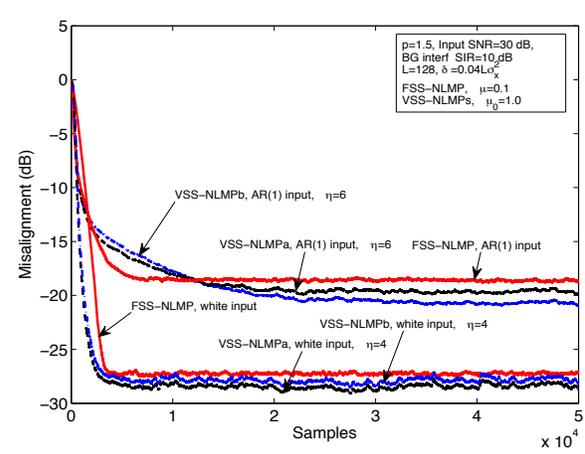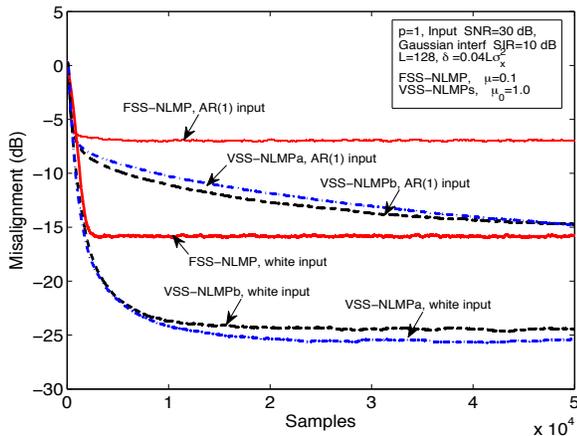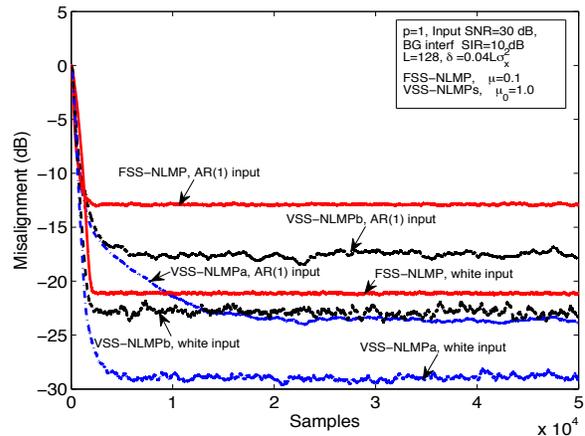
(a) p=2, Gaussian Interference
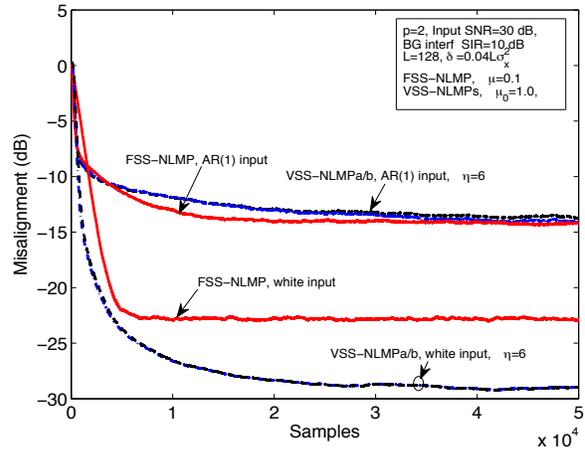
(b) p=2, BG Interference

(c) p=1.5, Gaussian Interference

(d) p=1.5, BG Interference

(e) p=1, Gaussian Interference

(f) p=1, BG Interference

Fig. 2.   Misalignment of the VSS-NLMP and FSS-NLMP algorithms for a real system with $L = 128$ taps. Solid lines – FSS-NLMP, Dash-dotted lines – VSS-NLMPa, Dashed lines –VSS-NLMPb. All VSS-NLMP algorithms performed better than FSS-NLMP, and VSS-NLMPa performed significantly better than VSS-NLMPb for $p = 1$. The results were the average of 100 simulations trials.
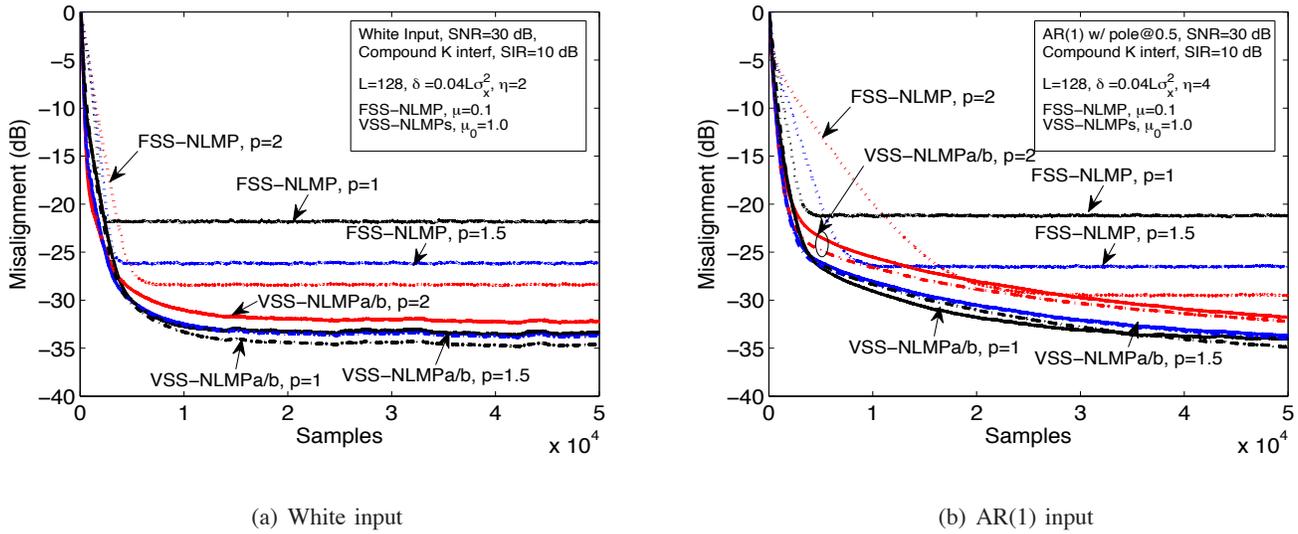
(a) White input

(b) AR(1) input

Fig. 3.    Misalignment of the VSS-NLMP and FSS-NLMP algorithms for a complex-coefficient system with $L = 128$ taps. The inputs were white or colored complex Gaussian and the interference was compound K distributed with shape parameter $\nu = 0.7$. Dotted lines – FSS-NLMP, Solid lines – VSS-NLMPa, Dash-dotted lines –VSS-NLMPb. The results were the average of 100 simulations trials.
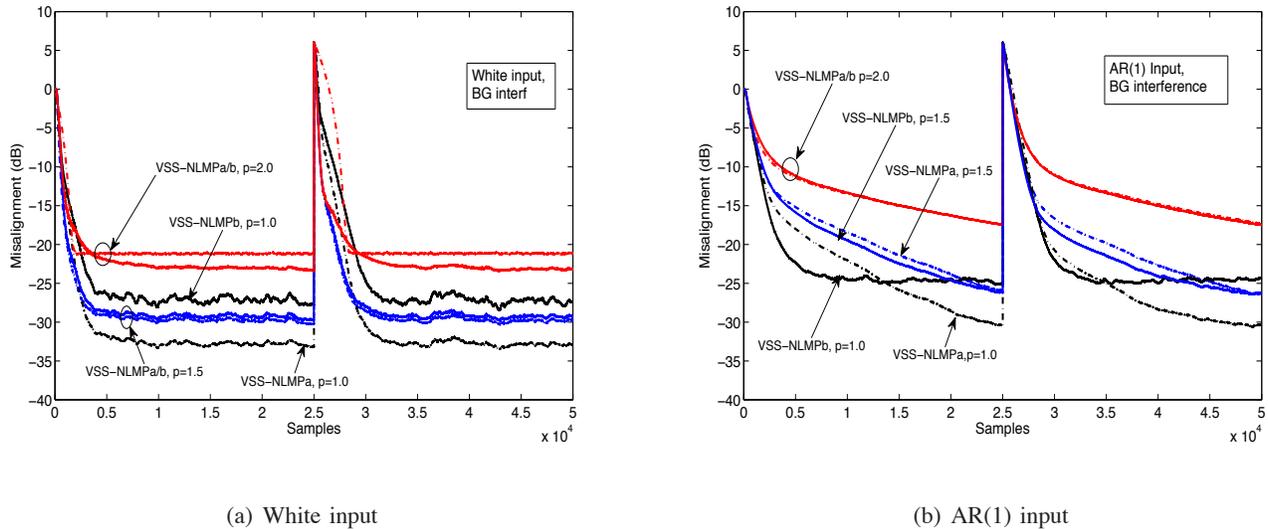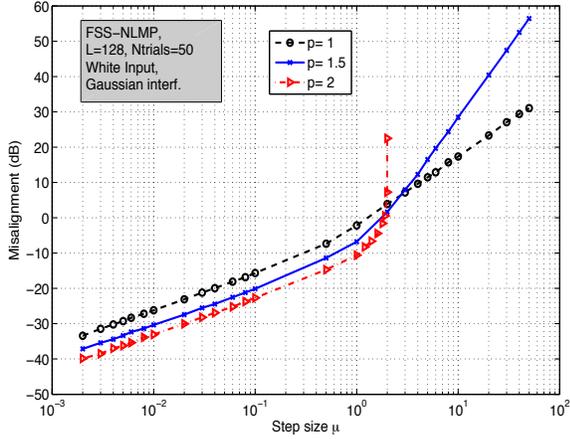


(a) White input

(b) AR(1) input

Fig. 4.    Tracking performance of the VSS-NLMP algorithms for a real system with $L = 128$ taps. The inputs were white or colored Gaussian and interference was BG. Other parameters were the same as Fig. 2. Dash-dotted lines – VSS-NLMPa. Solid lines – VSS-NLMPb. The results were the average of 100 simulations trials.

[31], [32], [44]–[48]. Attempts to finding a stability bound for the SA family have been reported in [32], [44], [45] using a second order stochastic model similar to that in the analysis of the $L_2$-norm algorithms. However, this approach is proven to be incorrect for the $L_1$-norm algorithms [31]. Instead, an interesting result for $L_1$-norm algorithms is that the sign algorithm is asymptotically bounded for any step-size greater than zero [31], [47], [48]. This property is not present in the $L_2$-norm algorithms and it proves to be a significant advantage of the sign algorithm family in terms of robustness. Upper bounds for the time-averaged mean absolute deviation (weight misalignment) and time-averaged mean square error at steady state are derived as functions of the step size in [31], [47], which give guidelines for choosing the step size in practical applications. As a variant of the sign algorithm, the FSS-NLMP with $p = 1$ also exhibits the asymptotic convergence property for step sizes greater than zero. This is verified by simulation for the asymptotic misalignment, as shown in Fig. 5. The steady-state misalignment was bounded for all step sizes $\mu > 0$ and $p = 1$.

To the best of our knowledge, a rigorous analysis for the stability of FSS-NLMP algorithms with fractional order does not exist in the literature. Our simulation results show that the FSS-NLMP with $p = 1.5$ also guaranteed asymptotic convergence for any step size greater than zero, as shown in Fig. 5. In contrast, the misalignment of the FSS-NLMP with $p = 2$ was bounded only when $\mu < 2$ for both Gaussian and BG interference scenarios. Comparing the steady-state misalignment of the three orders, the misalignment of $p = 2$ remained nearly the same for the two interference scenarios, while the FSS-NLMP with $p < 2$ performed better in BG interference when the step size was small. The FSS-NLMP with $p = 1.5$ achieved smaller steady-state misalignment than that of $p = 1$ when step size $\mu < 3$, and the FSS-NLMP with $p = 1$ performed better than that of $p = 1.5$ when $\mu > 3$. For practical applications, a small $\mu \ll 2$ is suggested for all orders of $p$ in the FSS-NLMP to ensure small steady state errors.

It is worth noting that the guaranteed stability of the FSS-NLMP with $p < 2$ and $\mu > 0$ does not mean that the VSS-NLMP algorithms are also guaranteed to be stable with the step size multiplier $\mu_0 > 0$ for $p < 2$. The stability analysis for the fractional-order VSS-NLMP algorithms is even more difficult than that of $p = 1$ or $p = 2$. We resort to simulation results to show that $\mu_0 = 1$ guarantees the stability of the VSS-NLMP algorithms for $1 \le p \le 2$. However, $\mu_0 > 1$ may cause the VSS-NLMP to diverge for all $1 \le p \le 2$, as illustrated in Fig. 6,

(a) Gaussian interference with SIR= 10 dB
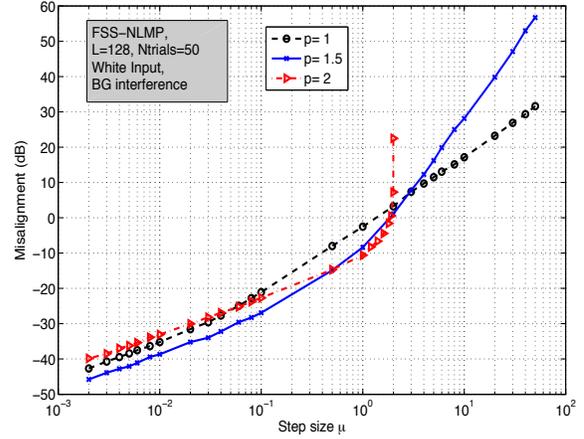
(b) BG interference with SIR= 10 dB and $Pr = 0.01$

Fig. 5. Steady-state misalignment of fixed step-size LMP algorithms as a function of the step-size $\mu$. The input was a real-coefficient white Gaussian signal with SNR= 30 dB. The interference was Gaussian or BG. The results were the average of 100 simulations trials.

where the step size multiplier used was $\mu_0 = 2$. The two VSS-NLMP algorithms with $p = 1$ and the VSS-NLMPa with $p = 2$ diverged in both white and AR(1) input scenarios with BG interference.
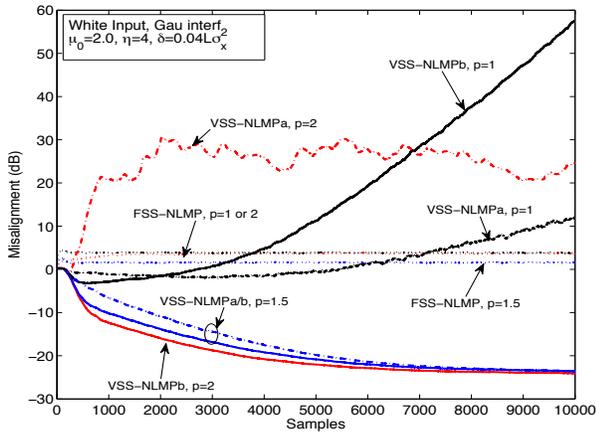


(a) White input
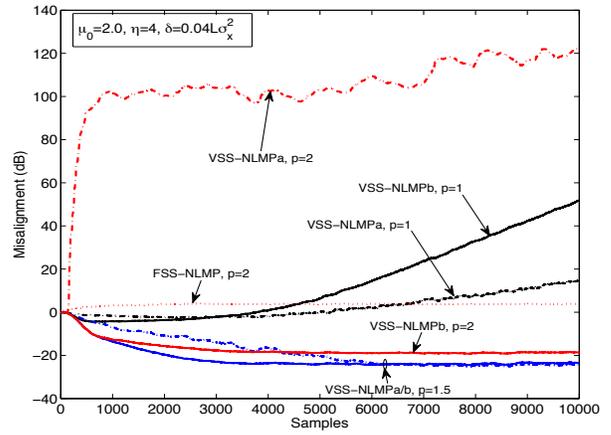
(b) AR(1) input

Fig. 6. Diverging misalignment curves for VSS-NLMP with the step size multiplier $\mu_0 = 2$. Other parameters were the same as Fig. 2. Misalignment was averaged over 100 trials.

## IV. STEADY-STATE ANALYSIS OF EXCESS MEAN SQUARE ERROR

In this section, a theoretical model is derived for the steady-state excess mean-square error (EMSE) performance of the VSS-NLMP algorithms using the energy conservation method of [36] for an adaptive filter with recursion

in (5). First we define the weight error vector

$$\tilde{\mathbf{w}}(k) = \mathbf{w}_o - \hat{\mathbf{w}}(k), \tag{28}$$

where $\mathbf{w}_o$ is the optimum weight vector from the model (1). Define also the *a priori* and *a posteriori* excess errors

$$e_a(k) = \tilde{\mathbf{w}}^H(k-1)\mathbf{x}(k), \qquad\qquad e_p(k) = \tilde{\mathbf{w}}^H(k)\mathbf{x}(k). \tag{29}$$

Denote the steady-state EMSE by $\zeta$:

$$\zeta \triangleq \lim_{k\to\infty} E\left\{|e_a(k)|^2\right\}. \tag{30}$$

Assuming that the recursion (5) converges, the energy conservation method [36] shows that

$$\lim_{k\to\infty} 2\mathcal{R}e\left\{E\left\{\mu(k)e_a(k)g(e_k)\right\}\right\} =$$
$$\lim_{k\to\infty} E\left\{\mu^2(k)\|\mathbf{x}(k)\|^2 |g(e_k)|^2\right\}, \tag{31}$$

where $\mathcal{R}e\{\cdot\}$ denotes the real part of a complex quantity.

Assuming that the noise plus interference signal $v(k)$ is independent and identically distributed (iid) and is independent of $\mathbf{x}(k)$, the EMSE $\zeta$ is related with the MSE $E\{|e_k|^2\}$ as [36]:

$$\mathcal{S}_e \triangleq \lim_{k\to\infty} E\left\{|e_k|^2\right\} = \zeta + \mathcal{S}_v. \tag{32}$$

We can use (31) and (32) to find an approximation for $\zeta$. The model described below is based on (15), but the same analysis may be modified for the other VSS-NLMP algorithms without difficulty.

### A. Model for Gaussian noise

For the Gaussian noise model, we first make some assumptions and approximations:

1) The signals $v(k)$ and $\mathbf{x}(k)$ are circular Gaussian, jointly stationary, with zero mean. The variance of $v(k)$ is $\mathcal{S}_v$, and the autocorrelation matrix of $\mathbf{x}(k)$ is $\mathbf{R}_{xx} = E\{\mathbf{x}(k)\mathbf{x}^H(k)\}$. The restriction of Gaussian $v(k)$ will be relaxed in Section IV-B.

2) The value of the forgetting factor $\lambda$ is close to one, so that quantities depending on it (such as $\hat{a}(k)$ and $\hat{\mathcal{P}}_{xx}(k)$) may be approximated by their mean values in the expressions for $\mu(k)$. Under these conditions, the step-size $\mu(k)$ is approximately independent of $\mathbf{x}(k)$, $v(k)$ and the errors $e_a(k)$ and $e_p(k)$. For $\hat{\mathcal{S}}_e(k)$ a more precise model using a Taylor series is used.

3) The estimate of noise plus interference variance $\hat{S}_v$ does not change during the period of analysis. This assumption is consistent with the fact that $\hat{S}_v$ is computed either off-line, or in periods in which the filter is not adapting, thus we can assume that its rate of change is much slower than that of the other variables.

4) We disregard the mechanism for zeroing $\mu(k)$ when $\hat{S}_e(k) < \hat{S}_v$.

5) The *a priori* excess error $e_a(k)$ is approximately Gaussian in steady-state. This is a common assumption widely used in theoretical analysis of adaptive filters [36].

6) In steady-state, $\mathbf{x}(k)$ is approximately independent of $e_a(k)$ and consequently, also of $e_k$ due to the previous assumptions. This is the so-called separation principle invoked in [36].

7) It follows from Assumptions 1 and 5 that the *a priori* error $e_k$ is approximately circular Gaussian in steady-state. This will no longer be true in non-Gaussian noise as shown in Section IV-B.

This last assumption follows directly from the first, which does not hold for impulsive noise. However, it allows us to evaluate some higher-order moments that arise in the analysis of the VSS-NLMP algorithms and it leads to models that correctly depict the dependence of the algorithms on its several parameters.

Under these assumptions, the expected values in (31) can be approximated by

$$\lim_{k\to\infty} 2\mathcal{R}e\left\{E\left\{\mu(k)e_a(k)g(e_k)\right\}\right\} \approx$$

$$\lim_{k\to\infty} 2E\{\mu(k)\}\mathcal{R}e\left\{E\left\{e_a(k)g(e_k)\right\}\right\} \tag{33}$$

$$\lim_{k\to\infty} E\left\{\mu^2(k)\|\mathbf{x}(k)\|^2\,|g(e_k)|^2\right\} \approx$$

$$\lim_{k\to\infty} E\left\{\mu^2(k)\right\} E\left\{\|\mathbf{x}(k)\|^2\right\} E\left\{|g(e_k)|^2\right\}. \tag{34}$$

We now must evaluate the expectations $E\{\mu(k)\}$, $E\{\mu^2(k)\}$, $E\left\{e_a(k)g(e_k)\right\}$ and $E\left\{|g(e_k)|^2\right\}$. Starting with the first two terms, we are to obtain $E\{\mu(k)\}$ and $E\{\mu^2(k)\}$ for the VSS-NLMPa algorithm. Taking the expectation of $\mu(k)$ in (15), we have

$$E\{\mu(k)\} = E\left\{\sqrt{\frac{\hat{S}_e(k)}{\hat{a}(k)}}\left[1-\sqrt{\frac{\hat{S}_v}{\hat{S}_e(k)}}\right]\right\}, \tag{35}$$

where we assume that the estimated parameters $\hat{S}_v, \hat{S}_e(k)$, and $\hat{a}(k)$ are used. From (35) and Assumption 2, we have

$$E\{\mu(k)\} \approx \sqrt{\frac{1}{E\{\hat{a}(k)\}}}E\left\{\sqrt{\hat{S}_e(k)}-\sqrt{\hat{S}_v}\right\} \tag{36}$$

$$E\{\mu^2(k)\} \approx \frac{1}{E\{\hat{a}(k)\}}E\left\{\left[\sqrt{\hat{S}_e(k)}-\sqrt{\hat{S}_v}\right]^2\right\}, \tag{37}$$

Now, approximations for all the expected values in (36) and (37) are to be evaluated at the limit of $k \to \infty$.

For $E\{\hat{a}(k)\}$: Taking expectations on both sides of (17), we obtain,

$$E\{\hat{a}(k)\} = \lambda E\{\hat{a}(k-1)\} + (1-\lambda)E\{|e_k|^{2p-2}\left|\mathbf{x}^H(k)\mathbf{x}(k)\right|^2\}.$$

At the limit of $k \to \infty$, we have, using Assumption 6

$$\lim_{k \to \infty} E\{\hat{a}(k)\} = \lim_{k \to \infty} E\{|e_k|^{2p-2}\left|\mathbf{x}^H(k)\mathbf{x}(k)\right|^2\}$$

$$\approx \lim_{k \to \infty} E\{|e_k|^{2p-2}\}E\{\left|\mathbf{x}^H(k)\mathbf{x}(k)\right|^2\}. \tag{38}$$

Under Assumption 7, $e_k$ is a zero-mean Gaussian variable with variance $\mathcal{S}_e$ for large $k$. This means that the limit

of $E\{|e_k|^{2p-2}\}$ reduces to the $(2p-2)$-th moment of a circular complex Gaussian variable.

Let $z$ be a zero-mean circularly complex Gaussian variable with variance $\sigma_z^2$. Since $z$ is circular, the variances of

its real and imaginary parts $z_r$ and $z_i$ are equal to $\sigma_z^2/2$. Therefore, $t = 2|z|^2/\sigma_z^2 = 2(z_r^2 + z_i^2)/\sigma_z^2$ is a $\chi^2$ variable

with two degrees of freedom, which reduces to an exponential probability density function

$$f_t(t) = \frac{1}{2}e^{-t/2}, \quad t \geq 0.$$

Since $E\{|z|^{2p}\} = \sigma_z^{2p} E\{t^p\}/(2^p)$, we obtain by direct computation

$$\lim_{k \to \infty} E\{|e_k|^{2p-2}\} = \Gamma(p)\mathcal{S}_e^{p-1}, \tag{39}$$

where $\Gamma(\cdot)$ is the gamma function.

On the other hand, under Assumption 1, $E\{|\mathbf{x}^H(k)\mathbf{x}(k)|^2\}$ is the trace of $E\{\mathbf{x}(k)\mathbf{x}^H(k)\mathbf{x}(k)\mathbf{x}^H(k)\}$. This can

be evaluated as described for example in [36], resulting

$$E\{|\mathbf{x}^H(k)\mathbf{x}(k)|^2\} = \text{Tr}(\mathbf{R}_{xx}^2) + \text{Tr}^2(\mathbf{R}_{xx}), \tag{40}$$

where $\text{Tr}(\mathbf{R}_{xx})$ is the trace of $\mathbf{R}_{xx}$.

For $E\left\{\sqrt{\hat{\mathcal{S}}_e(k)} - \sqrt{\hat{\mathcal{S}}_v}\right\}$ and $E\left\{\left[\sqrt{\hat{\mathcal{S}}_e(k)} - \sqrt{\hat{\mathcal{S}}_v}\right]^2\right\}$: Since we are interested in the steady-state, we simplify the

notation, writing $\hat{\mathcal{S}}_e$ instead of $\hat{\mathcal{S}}_e(k)$ and define $h(\hat{\mathcal{S}}_e) \triangleq \sqrt{\hat{\mathcal{S}}_e} - \sqrt{\hat{\mathcal{S}}_v}$. To evaluate $E\{h(\hat{\mathcal{S}}_e)\}$ and $E\{h^2(\hat{\mathcal{S}}_e)\}$ in

(36) and (37), we expand $h(\hat{\mathcal{S}}_e)$ and $h^2(\hat{\mathcal{S}}_e)$ in Taylor series around their means. From the recursion for $\hat{\mathcal{S}}_e(k)$ in

(16), we have

$$E\{\hat{\mathcal{S}}_e(k)\} = \lambda E\{\hat{\mathcal{S}}_e(k-1)\} + (1-\lambda)E\{|e_k|^2\},$$

therefore, $E\{\hat{\mathcal{S}}_e\} = \lim_{k\to\infty} E\{\hat{\mathcal{S}}_e(k)\} = \mathcal{S}_e$.

Letting then $\hat{\mathcal{S}}_e = E\{\hat{\mathcal{S}}_e\} + \Delta = \mathcal{S}_e + \Delta$, we obtain

$$h(\mathcal{S}_e + \Delta) \approx h(\mathcal{S}_e) + \frac{dh}{d\Delta}\Delta + \frac{1}{2}\frac{d^2h}{d\Delta^2}\Delta^2, \tag{41}$$

$$h^2(\mathcal{S}_e + \Delta) \approx h^2(\mathcal{S}_e) + \frac{d(h^2)}{d\delta}\Delta + \frac{1}{2}\frac{d^2(h^2)}{d\Delta^2}\Delta^2. \tag{42}$$

The derivatives are easily computed, resulting in

$$\left.\frac{dh}{d\Delta}\right|_{\Delta=0} = \frac{1}{2\sqrt{\mathcal{S}_e}}, \qquad\qquad \left.\frac{d^2h}{d\Delta^2}\right|_{\Delta=0} = -\frac{1}{4\sqrt{\mathcal{S}_e^3}}, \tag{43}$$

$$\left.\frac{d(h^2)}{d\Delta}\right|_{\Delta=0} = 1 - \sqrt{\frac{\hat{\mathcal{S}}_v}{\mathcal{S}_e}}, \qquad\qquad \left.\frac{d^2(h^2)}{d\Delta^2}\right|_{\Delta=0} = \frac{1}{2}\sqrt{\frac{\hat{\mathcal{S}}_v}{\mathcal{S}_e^3}}. \tag{44}$$

In the steady-state, $\Delta$ is a zero-mean random variable. Its variance $\sigma_\Delta^2$ may be obtained by squaring both sides of

(16):

$$E\{\hat{\mathcal{S}}_e^2(k)\} = \lambda^2 E\{\hat{\mathcal{S}}_e^2(k-1)\} + 2\lambda(1-\lambda)E\{\hat{\mathcal{S}}_e(k-1)|e_k|^2\}$$

$$+ (1-\lambda)^2 E\{|e_k|^4\}.$$

Since $\hat{\mathcal{S}}_e(k)$ varies much more slowly than $e_k$ for $\lambda \approx 1$, we can approximate $E\{\hat{\mathcal{S}}_e(k-1)|e_k|^2\} \approx E^2\{\hat{\mathcal{S}}_e\}$. For

circular complex variables under Assumption 7 we have $E\{|e_k|^4\} = 2\mathcal{S}_e^2$, and thus

$$\lim_{k\to\infty} E\{\hat{\mathcal{S}}_e^2(k)\} = \frac{2}{1+\lambda}\mathcal{S}_e^2. \tag{45}$$

Subtracting $\mathcal{S}_e^2$ from (45), we obtain the variance of $\Delta$:

$$\sigma_\Delta^2 = \frac{1-\lambda}{1+\lambda}\mathcal{S}_e^2, \tag{46}$$

If we ignore the mechanism to zero $\mu(k)$ when $\hat{\mathcal{S}}_e(k) < \hat{\mathcal{S}}_v$, then $E\{\Delta\} = 0$. Letting $C = (1-\lambda)/(8(1+\lambda))$,

we have

$$E\{h(\hat{\mathcal{S}}_e)\} \approx \sqrt{\mathcal{S}_e} - \sqrt{\hat{\mathcal{S}}_v} - C\sqrt{\mathcal{S}_e}, \tag{47}$$

$$E\{h^2(\hat{\mathcal{S}}_e)\} \approx \left[\sqrt{\mathcal{S}_e} - \sqrt{\hat{\mathcal{S}}_v}\right]^2 + 2C\sqrt{\hat{\mathcal{S}}_v\mathcal{S}_e}. \tag{48}$$

Substituting these expectations in (36) and (37), we obtain

$$\lim_{k \to \infty} E\{\mu(k)\} \approx \frac{\sqrt{\mathcal{S}_e} - \sqrt{\hat{\mathcal{S}}_v} - C\sqrt{\mathcal{S}_e}}{\sqrt{\Gamma(p)\mathcal{S}_e^{p-1}\left(\mathrm{Tr}(\mathbf{R}_{xx}^2) + \mathrm{Tr}^2(\mathbf{R}_{xx})\right)}}, \tag{49}$$

$$\lim_{k \to \infty} E\{\mu^2(k)\} \approx \frac{\left[\sqrt{\mathcal{S}_e} - \sqrt{\hat{\mathcal{S}}_v}\right]^2 + 2C\sqrt{\hat{\mathcal{S}}_v \mathcal{S}_e}}{\Gamma(p)\mathcal{S}_e^{p-1}\left(\mathrm{Tr}(\mathbf{R}_{xx}^2) + \mathrm{Tr}^2(\mathbf{R}_{xx})\right)}. \tag{50}$$

Next, we evaluate the remaining expectations $E\{|g(e_k)|^2\}$ and $E\{e_a(k)g(e_k)\}$ in (33) and (34) in the limit of $k \to \infty$. Recalling $g(e_k) = e_k^*|e_k|^{p-2}$, we have, as in (39),

$$\lim_{k \to \infty} E\{|g(e_k)|^2\} = \lim_{k \to \infty} E\{|e_k|^{2p-2}\} = \Gamma(p)\mathcal{S}_e^{p-1} \tag{51}$$

To evaluate $\lim_{k \to \infty} E\{e_a(k)g(e_k)\}$, we have

$$E\{e_a(k)g(e_k)\} = E\{e_a(k)g(e_a(k) + v(k))\}$$

$$= E\{e_a(k)\left(e_a(k) + v(k)\right)^* |e_a(k) + v(k)|^{p-2}\}. \tag{52}$$

This expected value can be computed using the complex form of Price's Theorem [49]. For two circularly complex variables $z_1$ and $z_2$, it can be shown from Price's theorem that [36]

$$E\{z_1 g(z_2)\} = \frac{E\{z_1 z_2^*\}}{E\{|z_2|^2\}} E\{z_2 g(z_2)\}.$$

In our case, we use $z_1 \leftarrow e_a(k)$, $z_2 \leftarrow e_k = e_a(k) + v(k)$, and $g(z_2) = z_2^*|z_2|^{p-2}$ so that

$$\lim_{k \to \infty} E\{e_a(k)g(e_k)\}$$

$$= \lim_{k \to \infty} E\{e_a(k)(e_a^*(k) + v^*(k))|e_a(k) + v(k)|^{p-2}\}$$

$$= \lim_{k \to \infty} \frac{E\{e_a(k)(e_a^*(k) + v^*(k))\}}{E\{|e_a(k) + v(k)|^2\}} E\{|e_a(k) + v(k)|^p\}$$

$$= \zeta\Gamma\left(\frac{p}{2} + 1\right)\mathcal{S}_e^{p/2-1}. \tag{53}$$

Finally, substituting (49), (50), (51), and (53) into (31), we obtain

$$2\zeta\left\{(1 - C)\sqrt{\mathcal{S}_e} - \sqrt{\hat{\mathcal{S}}_v}\right\} =$$

$$\frac{\sqrt{\Gamma(p)}}{\Gamma\left(\frac{p}{2} + 1\right)} \frac{\sqrt{\mathcal{S}_e}\left\{\left[\sqrt{\mathcal{S}_e} - \sqrt{\hat{\mathcal{S}}_v}\right]^2 + 2C\sqrt{\hat{\mathcal{S}}_v \mathcal{S}_e}\right\}}{\sqrt{\mathrm{Tr}(\mathbf{R}_{xx}^2) + \mathrm{Tr}^2(\mathbf{R}_{xx})}/\mathrm{Tr}(\mathbf{R}_{xx})}. \tag{54}$$

(a) $p = 1.0$  (b) $p = 1.5$  (c) $p = 2.0$
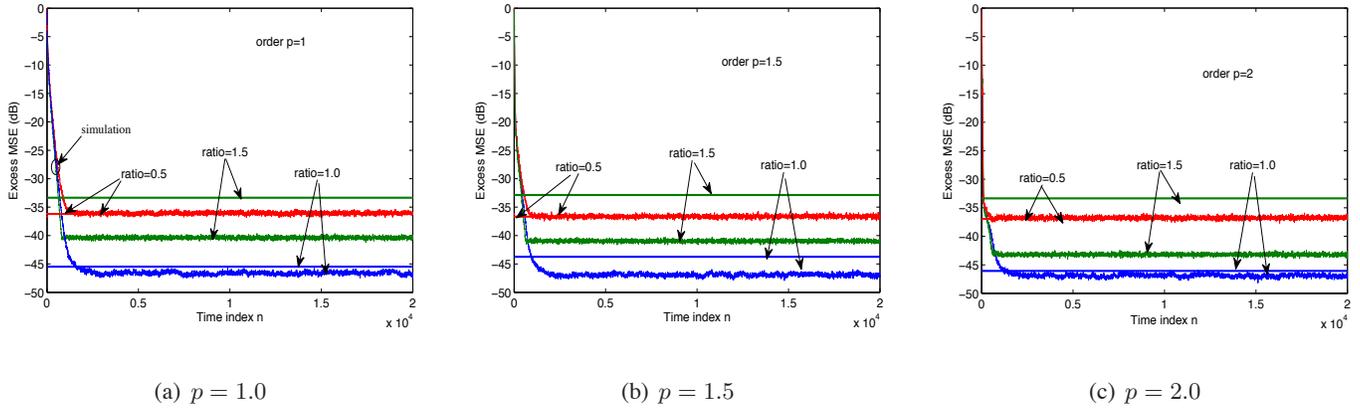
Fig. 7.   Theoretical and experimental EMSE for white Gaussian inputs with $\lambda = 0.99$, $\sigma_x^2 = 1$, $\mathcal{S}_v = 10^{-3}$, and ratio $= \hat{\mathcal{S}}_v/\mathcal{S}_v$. The simulation results were obtained as an average of 100 trials. The dark dashed lines are the solutions of the theoretical steady-state EMSE model (54).

Recalling that $\mathcal{S}_e = \zeta + \mathcal{S}_v$, the solution of $\zeta$ to (54) approximates well the steady-state EMSE, especially for $\hat{\mathcal{S}}_v < \mathcal{S}_v$, a situation in which the probability of $\hat{\mathcal{S}}_e < \hat{\mathcal{S}}_v$ is small, and therefore Assumption 4 is closer to reality.

The model (54) was verified by simulations with three cases: ratio $= \hat{\mathcal{S}}_v/\mathcal{S}_v = 0.5, 1.0$, and $1.5$, with results shown in Fig. 7. The simulation was run with white Gaussian complex inputs and noise and with parameters $\lambda = 0.99$, $\sigma_x^2 = 1$, and $\mathcal{S}_v = 10^{-3}$. The simulated steady-state EMSE was an average of 100 trials. Figure 7(a) shows the case of $p = 1.0$, where the time average EMSE of the last $1.5 \times 10^4$ iterations in steady-state was $\zeta = -36.15, -46.30$, and $-40.18$ dB for ratio $=0.5, 1.0$, and $1.5$ respectively; while the solution of (54) was $-36.20, -45.50$, and $-33.35$ dB for the three ratios, respectively. For other orders, similar results were obtained, as shown in Fig. 7(b) and 7(c). The model matches the simulation very well when the estimate $\hat{\mathcal{S}}_v \leq \mathcal{S}_v$. As expected, the theoretical approximations become worse as the probability of $\hat{\mathcal{S}}_e < \hat{\mathcal{S}}_v$ increases, which happens as $\hat{\mathcal{S}}_v$ is increased.

## B. Model for Bernoulli-Gaussian noise

The derivation in the previous section relied in the assumption of Gaussian noise. Although it is difficult to extend the model to general non-Gaussian noise, the extension to Bernoulli-Gaussian noise is achieved with a few changes in the derivation. First, we modify Assumptions 1 and 7 as follows:

1') The noise $\{v(k)\}$ sequence is iid and independent of the regressor sequence $\{\mathbf{x}(n)\}$. The regressor follows

a Gaussian distribution, whereas the noise follows a complex-valued Bernoulli-Gaussian distribution, that is, $v(k) = \omega(k)N(k)$, where now $N(k)$ is a zero-mean circular Gaussian variable with variance $\sigma_N^2$, and $\omega(k)$ is Bernoulli process with probability $P(\omega(k) = 1) = P_r$.

7') The *a priori* error $e_k = e_a(k) + v(k)$ is now the sum of a Gaussian and a Bernoulli-Gaussian variables, independent of each other.

We must re-evaluate all expected values related to $v(k)$ to take into account the new distribution. Note that now, conditioned on $\omega(k) = 1$, $v(k)$ is Gaussian with variance $\sigma_N^2$. The expected value of $|e_k|^p$ can be evaluated as follows.

$$
\begin{aligned}
E\{|e_k|^r\} = & P_r E\{|e_a(k) + v(k)|^r |\omega(k) = 1\} \\
& + (1 - P_r)E\{|e_a(k) + v(k)|^r |\omega(k) = 0\}.
\end{aligned}
\tag{55}
$$

Given that $\omega(k) = 1$, $e_a(k) + v(k)$ is a Gaussian variable with zero mean and variance $\zeta(k) + \sigma_N^2$. Therefore, by (39), at steady-state we have

$$
E\{|e_k|^r |\omega(k) = 1\} = \Gamma(r/2 + 1)(\zeta + \sigma_N^2)^{r/2}.
$$

Similarly, if $\omega(k) = 0$, $e_k = e_a(k)$ follows a Gaussian distribution with variance $\zeta$, so

$$
E\{|e_k|^r |\omega(k) = 0\} = \Gamma(r/2 + 1)\zeta^{r/2}.
$$

We conclude that

$$
\lim_{k \to \infty} E\{|e_k|^r\} = \Gamma(\frac{r}{2} + 1)\left(P_r(\zeta + \sigma_N^2)^{\frac{r}{2}} + (1 - P_r)\zeta^{\frac{r}{2}}\right).
\tag{56}
$$

In particular, for $r = 2$, we have $\lim_{k \to \infty} E\{|e_k|^2\} = 2\mathcal{S}_e$. Applying (56) with $r = 4$ we can evaluate $E\{\hat{\mathcal{S}}_e^2(k)\}$, so (45) and (46) become

$$
\begin{aligned}
\lim_{k \to \infty} E\{\hat{\mathcal{S}}_e^2(k)\} &= \frac{2}{1 + \lambda}\mathcal{S}_e^2 + 2\frac{1 - \lambda}{1 + \lambda}P_r(1 - P_r)\sigma_N^4, \\
\sigma_\Delta^2 &= \frac{1 - \lambda}{1 + \lambda}\mathcal{S}_e^2 + 2\frac{1 - \lambda}{1 + \lambda}P_r(1 - P_r)\sigma_N^4.
\end{aligned}
\tag{57}
$$

Denoting $D = P_r(1 - P_r)\sigma_N^4$, we can update the expressions (49) and (50) for $E\{\mu(k)\}$ and $E\{\mu^2(k)\}$, resulting

$$\lim_{k \to \infty} E\{\mu(k)\} \approx \frac{(1 - C)\sqrt{\mathcal{S}_e} - \sqrt{\hat{\mathcal{S}}_v} - \frac{2CD}{\mathcal{S}_e^{3/2}}}{\sqrt{\Gamma(p)\mathcal{S}_e^{p-1} \left(\text{Tr}(\mathbf{R}_{xx}^2) + \text{Tr}^2(\mathbf{R}_{xx})\right)}}, \tag{58}$$

$$\lim_{k \to \infty} E\{\mu^2(k)\} \approx \frac{\left[\sqrt{\mathcal{S}_e} - \sqrt{\hat{\mathcal{S}}_v}\right]^2 + 2C\sqrt{\hat{\mathcal{S}}_v\mathcal{S}_e} + \frac{4CD\sqrt{\hat{\mathcal{S}}_v}}{\mathcal{S}_e^{3/2}}}{\Gamma(p)\mathcal{S}_e^{p-1} \left(\text{Tr}(\mathbf{R}_{xx}^2) + \text{Tr}^2(\mathbf{R}_{xx})\right)}. \tag{59}$$

Using again (56) and conditional expectation, (51) and (53) become

$$\lim_{k \to \infty} E\{|g(e_k)|^2\} = \Gamma(p) \left[P_r \left(\zeta + \sigma_N^2\right)^{p-1} + (1 - P_r)\zeta^{p-1}\right], \tag{60}$$

$$\lim_{k \to \infty} E\{e_a(k)g(e_k)\} = \Gamma\left(\frac{p}{2} + 1\right) \left[P_r\zeta\mathcal{S}_e^{\frac{p}{2}-1} + (1 - P_r)\zeta^{\frac{p}{2}}\right]. \tag{61}$$

Substituting (58)–(61) into (31), we obtain a modified version of (54):

$$2\Gamma\left(\frac{p}{2} + 1\right) \left[(1 - C)\sqrt{\mathcal{S}_e} - \sqrt{\hat{\mathcal{S}}_v} - \frac{2CD}{\mathcal{S}_e^{3/2}}\right]$$

$$\times \left[P_r\zeta\mathcal{S}_e^{p/2-1} + (1 - P_r)\zeta^{p/2}\right]$$

$$\approx \frac{\Gamma^{1/2}(p)\text{Tr}(\mathbf{R}_{xx})}{\sqrt{\mathcal{S}_e^{p-1}\left(\text{Tr}(\mathbf{R}_{xx}^2) + \text{Tr}^2(\mathbf{R}_{xx})\right)}}$$

$$\times \left\{\left[\sqrt{\mathcal{S}_e} - \sqrt{\hat{\mathcal{S}}_v}\right]^2 + 2C\sqrt{\hat{\mathcal{S}}_v\mathcal{S}_e} + \frac{4CD\sqrt{\hat{\mathcal{S}}_v}}{\mathcal{S}_e^{3/2}}\right\}$$

$$\times \left[P_r \left(\zeta + \sigma_N^2\right)^{p-1} + (1 - P_r)\zeta^{p-1}\right] \tag{62}$$

As in the Gaussian case, the solution $\zeta$ to (62) approximates well the steady-state EMSE when $\hat{S}_v < S_v$, since in this case the probability of $\hat{S}_e < \hat{S}_v$ is small, and Assumption 4 is approximately true. For larger values of $\hat{S}_v$, the theoretical predictions are farther from simulation results than in the Gaussian case. We repeated the simulations presented in the previous section for Bernoulli-Gaussian noise, always keeping $S_v = 10^{-3}$, but in two conditions: $P_r = 0.1$ and $\sigma_N^2 = 10^{-2}$, and $P_r = 10^{-2}$ and $\sigma_N^2 = 0.1$. The other parameters are as in Figure **??**, except that the experimental EMSE was computed with the average of 100 curves. The results are given in Table III. The table also includes the previous results for Gaussian noise, in the rows with $P_r = 1$.

## V. CONCLUSION

Two variable step-size (VSS) adaptive algorithms have been proposed using normalized fractionally lower-order moment minimization for system identification applications. The proposed adaptive VSS-NLMP algorithms

TABLE III

SIMULATION AND THEORETICAL RESULTS FOR BERNOULLI-GAUSSIAN NOISE.

| $\hat{S}_v/S_v$ | $P_r$ | $\sigma_N^2$ | $\zeta$ (simulations) | $\zeta_{\text{th}}$ from (62) |
|---|---|---|---|---|
| 0.5 | 1 | $10^{-3}$ | $2.17 \times 10^{-4}$ | $2.14 \times 10^{-4}$ |
| 1 | 1 | $10^{-3}$ | $2.04 \times 10^{-5}$ | $4.23 \times 10^{-5}$ |
| 1.5 | 1 | $10^{-3}$ | $7.60 \times 10^{-5}$ | $5.16 \times 10^{-4}$ |
| 0.5 | 0.1 | 0.01 | $9.5 \times 10^{-5}$ | $8.4 \times 10^{-5}$ |
| 1 | 0.1 | 0.01 | $2.4 \times 10^{-5}$ | $1.2 \times 10^{-4}$ |
| 1.5 | 0.1 | 0.01 | $1.8 \times 10^{-5}$ | $5.5 \times 10^{-4}$ |
| 0.5 | 0.01 | 0.1 | $2.3 \times 10^{-4}$ | $2.5 \times 10^{-4}$ |
| 1 | 0.01 | 0.1 | $1.2 \times 10^{-4}$ | $4.7 \times 10^{-4}$ |
| 1.5 | 0.01 | 0.1 | $7.5 \times 10^{-5}$ | $8.4 \times 10^{-4}$ |

automatically adjust the step size by approximating the power of the *a posteriori* error to that of the background noise. Variants of implementation using time-averaged estimates of error and signal statistics are also developed for each adaptive VSS-NLMP algorithm. The proposed VSS-NLMP algorithms have been evaluated extensively by computer simulations under Gaussian or heavy-tailed non-Gaussian interference signals with white or colored inputs, and for real- and complex-coefficient systems. The results have shown that the new VSS-NLMP algorithms combine the benefits of variable step sizes with the robustness of the NLMP algorithms against impulsive interference, thus achieving better tradeoff between fast convergence and small steady-state error than the FSS-NLMP. In heavy-tailed interference scenarios, the proposed VSS-NLMP algorithms with $p < 2$ also achieve better steady-state performance and faster convergence than the VSS-NLMS [11]. The proposed VSS-NLMP algorithms with order $p = 1$ exhibit best performance in both Gaussian and impulsive interference environments and its asymptotic convergence is guaranteed for all step sizes $\mu(k) > 0$ when the step size multiplier satisfies $0 < \mu_0 \leq 1$. The steady-state excess mean square error (EMSE) of the proposed algorithm has also been analyzed via the energy conservation method for Gaussian and Bernoulli-Gaussian noise.

## VI. ACKNOWLEDGEMENT

## APPENDIX

We show the relationship between the coefficients of the quadratic equation (9) under the assumption that the *a priori* error is independent from the input signal. First let us consider the statistics of $\mathbf{x}^H(k)\mathbf{x}(k)$ and assume that the input signal $x(k)$ is white or colored (real/complex) Gaussian with zero mean and average power $\sigma_x^2$. Denote the correlation matrix of the input vector $\mathbf{x}(k)$ as $R_{xx}(k) = E[\mathbf{x}(k)\mathbf{x}^H(k)]$ whose eigenvalue decomposition is $R_{xx}(k) = \mathbf{U\Sigma U}^*$, where $\mathbf{U}$ is a unitary matrix and $\mathbf{\Sigma}$ is a diagonal matrix of nonnegative eigenvalues, $\sigma(l), l = 1, \cdots, L$. The input vector can be linearly transformed from $L$ independent, identically distributed Gaussian variables $h(l) \sim \mathcal{N}(0,1), l = 1, \cdots, L$. Therefore, the distribution of $\mathbf{x}^H(k)\mathbf{x}(k)$ is the same as the distribution of $\mathcal{Z} = \sum_{l=1}^{L} \sigma(l)|h(l)|^2$. The statistics of $\mathcal{Z}$ are

$$\mathcal{S}_{xx}(k) = E[\mathcal{Z}] = \sum_{l=1}^{L} \sigma(l)E[|h(l)|^2] = \sum_{l=1}^{L} \sigma(l); \tag{63}$$

$$\begin{aligned}
\mathcal{P}_{xx}(k) &= E[|\mathcal{Z}|^2] = E\left[\sum_{l_1=1}^{L} \sigma(l_1)|h(l_1)|^2 \sum_{l_2=1}^{L} \sigma(l_2)|h(l_2)|^2\right] \\
&= \sum_{l=1}^{L} \sigma^2(l)E\left[|h(l)|^4\right] + 2\sum_{l_1=1}^{L-1}\sum_{l_2=l_1+1}^{L} \sigma(l_1)\sigma(l_2)E\left[|h(l_2)|^2|h(l_2)|^2\right] \\
&= 3\sum_{l=1}^{L} \sigma^2(l) + 2\sum_{l_1=1}^{L-1}\sum_{l_2=l_1+1}^{L} \sigma(l_1)\sigma(l_2) \tag{64}
\end{aligned}$$

The last equality uses the 4-th moment of a zero-mean Gaussian distribution. If the input signal is white, then the singular values $\sigma(l)$ are all equal to one and $\mathcal{Z}$ is Chi-square distributed with $E[\mathcal{Z}] = L\sigma(1)$ and $E[|\mathcal{Z}|^2] = (L^2 + 2L)\sigma(1)$. If the input is colored Gaussian, then $\mathcal{S}_{xx}(k)$ remains the same as that of white Gaussian input because $\sum_{l=1}^{L} \sigma(l) = \text{trace}[R_{xx}(k)] = L\sigma_x^2$, but $\mathcal{P}_{xx}(k)$ is usually different from that of the white input.

Next, consider the statistics of the *a priori* error. Based on (1) and (2), we have

$$e_k = [\mathbf{w}_o - \hat{\mathbf{w}}(k-1)]^H \mathbf{x}(k) + v(k). \tag{65}$$

In the steady state, the first term on the right-hand side becomes very small and the statistics of $e_k$ are dominated by $v(k)$. Analytical results for the $p$-th moment of the error can be obtained for $e_k$ being Gaussian; while numerical evaluation of the error statistics is used for non-Gaussian distributions.

For a zero-mean Gaussian distribution, the $p$-th moment is [9]

$$E[|\mathcal{N}(0, \sigma_e^2)|^p] = \frac{2^{p/2}\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}\sigma_e^p, \quad p > 0, \tag{66}$$

where $\sigma_e^2$ is the second order moment (which is double the dispersion parameter $\gamma$ in [9]), and the Gamma function is defined as $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$.

Using (63), (64), and (66), we have for Gaussian inputs and Gaussian interference

$$\frac{b(k)}{\mathcal{S}_e(k)} = \frac{\mathcal{L}_p(e_k)}{\mathcal{S}_e(k)}\mathcal{S}_{xx}(k) = \frac{2^{p/2}\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}\sigma_e^{p-2}\sum_{l=1}^L \sigma(l) \tag{67}$$

$$\frac{a(k)}{\mathcal{S}_e(k)} = \frac{\mathcal{L}_{2p-2}(e_k)}{\mathcal{S}_e(k)}\mathcal{P}_{xx}(k) = \frac{2^{(2p-2)/2}\Gamma(\frac{2p-1}{2})}{\sqrt{\pi}}\sigma_e^{2p-4}\left[\sum_{l=1}^L 3\sigma^2(l) + 2\sum_{l_1=1}^{L-1}\sum_{l_2=l_1+1}^L \sigma(l_1)\sigma(l_2)\right], \quad p > 1. \tag{68}$$

The ratios of the two coefficients, $b^2(k)\mathcal{S}_e(k)/a(k)$, for different order, $p$, were computed based on the theoretical analysis for Gaussian interference, as shown in Fig. 8, with both white and colored input signals. When the tap length was large ($L = 512$), the two curves for white or AR(1) inputs were very close, ranging from 0.631 (for $p = 1$) to 0.996 (for $p = 2$) with white input, and from 0.625 (for $p = 1$) to 0.985 (for $p = 2$) with AR(1) inputs. When the tap length was small ($L = 32$), the distance between curves with white and AR(1) inputs became large. In all cases, the ratio was greater than 0.5 for Gaussian interference.

The analysis for other types of interference is more involved and it is difficult to exhaust all types. We use simulated signals to compute the ratios and results are shown in Fig. 9. The BG interference with $P_r = 1$ corresponds to Gaussian interference and the curves based on simulated signals matched the theoretical analysis. The ratios for small $p$ were reduced significantly for more impulsive BG interference, as shown in Fig. 9(a), with the smallest ratio being 0.167 for $p = 1$. The ratios for compound K interference exhibited a similar trend, with the smallest ratio being 0.22 for the most impulsive interference and $p = 1$. This implies that the two VSS-NLMP algorithms are similar for $p$ large, but they are quite different for $p$ small.

## REFERENCES

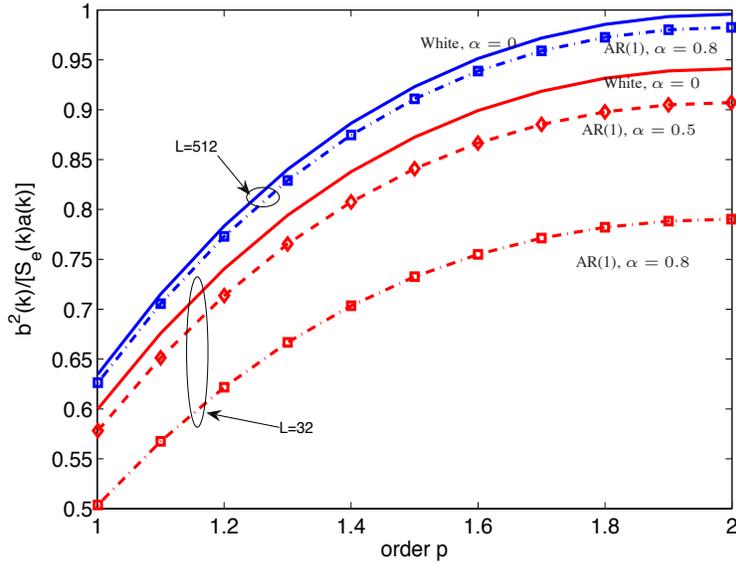[1] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, New Jersey, 07458.: MPrentice Hall, 2002.

Fig. 8. The ratio of $b^2(k)\mathcal{S}_e(k)/a(k)$ for white or AR(1) inputs with Gaussian interference. Results were computed based on (67) and (68). Parameter $L$ was the number of coefficient taps and the AR(1) input was assumed to be a filtered white Gaussian by an IIR filter with a pole at $\alpha$.
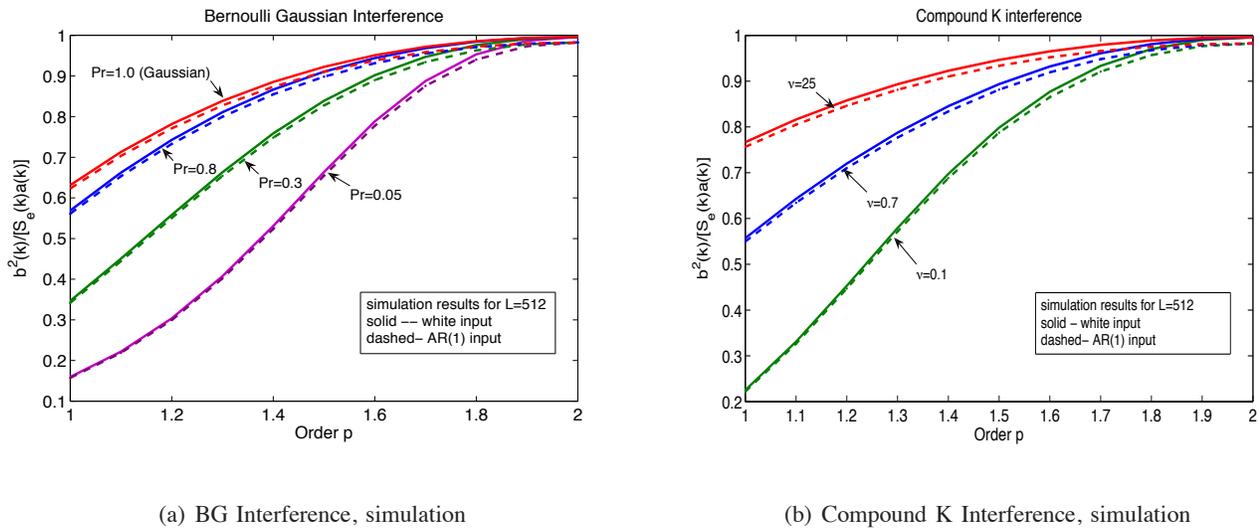


(a) BG Interference, simulation

(b) Compound K Interference, simulation

Fig. 9. The ratio of $b^2(k)\mathcal{S}_e(k)/a(k)$ for white or AR(1) inputs with BG or compound K interference. Results were computed by simulated data with $L = 512$ and $\alpha = 0.8$ for AR(1) input. Simulation results for Gaussian interference matched the theoretical analysis of Fig. 8.

[2] B. Widrow, J. Glover, Jr., J. McCool, J. Kaunitz, C. Williams, R. Hearn, J. Zeidler, E. Dong, Jr., and R. Goodlin, "Adaptive noise cancelling: Principles and applications," *Proc. IEEE*, vol. 63, no. 12, pp. 1692–1716, Dec. 1975.

[3] K. Murano, S. Unagami, and F. Amano, "Echo cancellation and applications," *IEEE Communications Magazine*, vol. 28, no. 1, pp. 49–55, Jan 1990.

[4] B. Weng and K. E. Barner, "Nonlinear system identification in impulsive environments," *IEEE Trans. Signal Processing*, vol. 53, no. 7, pp. 2588–2594, Jul. 2005.

[5] G. Arce, *Nonlinear Signal Processing*. NY: John Wiley, INC, 2005.

[6] J. Idier and Y. Goussard, "Stack algorithm for recursive deconvolution of Bernoulli-Gaussian processes," *IEEE Trans. Geoscience and Remote Sensing*, vol. 28, no. 5, pp. 975–978, Sep. 1990.

[7] M. Boujida, J. M. Boucher, B. Marsset, G. Lericolais, and J. Meunier, "Deconvolution of marine seismic signal using higher order statistics and Bernoulli-Gaussian model," in *Proc. OCEANS '96*, Fort Lauderdale, FL, USA, Sep. 1996.

[8] R. Sharma, W. Sethares, and J. Bucklew, "Asymptotic analysis of stochasic gradient-based adaptive filtering algorithms with general cost functions," *IEEE Trans. Signal Processing*, vol. 44, no. 9, pp. 2186–2194, Sep. 1996.

[9] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: stable processes and their applications," *Proc. IEEE*, vol. 81, no. 7, pp. 986–1010, Jul. 1993.

[10] T. Aboulnasr and K. Mayyas, "A robust variable step-size LMS-type algorithm: analysis and simulations," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 631–639, Mar. 1997.

[11] J. Benesty, H. Rey, L. Vega, and S. Tressens, "A nonparametric VSS NLMS algorithm," *IEEE Signal Processing Lett.*, vol. 13, no. 10, pp. 581 –584, Oct. 2006.

[12] E. V. Papoulis and T. Stathaki, "A normalized robust mixed-norm adaptive algorithm for system identification," *IEEE Signal Process. Lett.*, vol. 11, no. 1, pp. 56–59, Jan. 2004.

[13] L. R. Vega, H. Rey, J. Benesty, and S. Tressens, "A new robust variable step-size NLMS algorithm," *IEEE Trans. Signal Processing*, vol. 56, no. 5, pp. 1878 –1893, May 2008.

[14] O. Hoshuyama, R. Goubran, and A. Sugiyama, "A generalized proportionate variable step-size algorithm for fast changing acoustic environments," in *Proc. ICASSP'04*, Montreal, Quebec, Canada, May 2004.

[15] S. Ben Jebara and H. Besbes, "Variable step size filtered sign algorithm for acoustic echo cancellation," *Electronics Letters*, vol. 39, no. 12, pp. 936–938, Jun 2003.

[16] H.-C. Shin and A. Sayed, "Mean-square performance of a family of affine projection algorithms," *IEEE Trans. Signal Processing*, vol. 52, no. 1, pp. 90–102, Jan. 2004.

[17] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electron. Comm. Jpn.*, vol. 67-A, no. 5, pp. 19–27, May 1984.

[18] S. R. Kim and A. Efron, "Adaptive robust impulse noise filtering," *IEEE Trans. Signal Processing*, vol. 43, no. 8, pp. 1855–1866, Aug. 1995.

[19] G. Aydin, O. Arikan, and A. E. Cetin, "Robust adaptive filtering algorithms for $\alpha$-stable random processes," *IEEE Trans. Circuits Syst. II*, vol. 46, no. 2, pp. 198–202, Feb. 1999.

[20] P. Tsakalides and C. L. Nikias, "Robust space-time adaptive processing (STAP) in non-Gaussian clutter environments," *IEE Proceedings - Radar, Sonar and Navigation*, vol. 146, no. 2, pp. 84–93, Apr. 1999.

[21] E. Masry, "Alpha-stable signals and adaptive filtering," *IEEE Trans. Signal Processing*, vol. 48, no. 11, pp. 3011–3016, Nov. 2000.

[22] O. Arikan, A. Enis Cetin, and E. Erzin, "Adaptive filtering for non-Gaussian stable processes," *IEEE Signal Processing Lett.*, vol. 1, no. 11, pp. 163–165, Nov. 1994.

[23] E. Kuruoglu, J. Rayner, and W. Fitzgerald, "Least lp-norm impulsive noise cancellation with polynomial filters1," *Signal Processing*, vol. 69, no. 1, pp. 1–14, Aug. 1998.

[24] T. S. D. Singh and A. Chatterjee, "A comparative study of adaptation algorithms for nonlinear system identification based on second order volterra and bilinear polynomial filters," *Measurement*, vol. 44, no. 10, pp. 1915 – 1923, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0263224111002910

[25] J. Chambers and A. Avlonitis, "A robust mixed-norm adaptive filter algorithm," *IEEE Signal Processing Lett.*, vol. 4, no. 2, pp. 46–48, Feb. 1997.

[26] M. T. M. Silva and V. H. Nascimento, "Improving the tracking capability of adaptive filters via convex combination," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3137–3149, Jul. 2008.

[27] J. Arenas-Garca, A. R. Figueiras-Vidal, and A. H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1078–1090, Mar. 2006.

[28] V. H. Nascimento and A. H. Sayed, "$\ell_p$-robustness and convergence of estimation schemes in adaptive control," in *Proc. of the American Control Conference*, vol. 1, Albuquerque, NM, Jun. 1997, pp. 338–342.

[29] P. Petrus, "Robust huber adaptive filter," *IEEE Trans. Signal Process.*, vol. 47, no. 4, pp. 1129–1133, 1999.

[30] Y. Zou, S.-C. Chan, and T.-S. Ng, "Least mean m-estimate algorithms for robust adaptive filtering in impulse noise," *IEEE Trans. Circuits Syst. II*, vol. 47, no. 12, pp. 1564–1569, 2000.

[31] E. Masry and F. Bullo, "Convergence analysis of the sign algorithm for adaptive filtering," *IEEE Trans. Inform. Theory*, vol. 41, no. 2, pp. 489–495, Mar. 1995.

[32] V. Mathews and S. Cho, "Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, no. 4, pp. 450–454, Apr. 1987.

[33] P. Hubscher, J. Bermudez, and V. Nascimento, "A mean-square stability analysis of the least mean fourth adaptive algorithm," *Signal Processing, IEEE Transactions on*, vol. 55, no. 8, pp. 4018 –4028, Aug. 2007.

[34] Y. Zheng, T. Shao, and E. P. Blasch, "A fast-converging space-time adaptive processing algorithm for non-Gaussian clutter suppression," *Elsevier J. Digital Signal Processing*, vol. 22, no. 1, p. 7486, Jan. 2012.

[35] A. Mader, H. Puder, and G. Schmidt, "Step-size control for acoustic echo cancellation filters — an overview," *Signal Processing*, vol. 80, no. 4, pp. 1697–1719, Sep. 2000.

[36] A. Sayed, *Adaptive filters*. Wiley-IEEE Press, 2008.

[37] T. Shao, Y. R. Zheng, and J. Benesty, "An affine projection sign algorithm robust against impulsive interferences," *IEEE Signal Processing Lett.*, vol. 17, no. 4, pp. 327–330, Apr. 2010.

[38] T. Shao and Y. R. Zheng, "A new variable step-size fractional lower-order moment algorithm for Non-Gaussian interference environments," in *Proc. ISCAS09*, Taipei, Taiwan, May 2009.

[39] M. Greco, F. Gini, and M. Rangaswamy, "Statistical analysis of measured polarimetric clutter data at different range resolutions," *IEE Proc. Pt-F: Radar Sonar Navig.*, vol. 153, no. 6, pp. 473–481, Dec. 2006.

[40] W. Yang and T. Yang, "M-ary frequency shift keying communications over an underwater acoustic channel: performance comparison of data with models," *J. Acoust. Soc. Am.*, vol. 120, pp. 2694 – 2701, Nov. 2010.

[41] J. Zhang, J. Cross, and Y. R. Zheng, "Statistical channel modeling of wireless shallow water acoustic communications from experiment data," in *Proc. IEEE Milcom10*, San Jose, CA, Nov. 2010.

[42] Y. R. Zheng and C. Xiao, "Simulation models with correct statistical properties for rayleigh fading channels," *IEEE Transactions on Communications*, vol. 51, no. 6, pp. 920 – 928, jun. 2003.

[43] G. Davidson, *Matlab and C Radar Toolbox*, The Japan Society for the Promotion of Science (JSPS), http://www.radarworks.com.

[44] V. J. Mathews, "Performance analysis of adaptive filters equipped with the dual sign algorithm," *IEEE Trans. Signal Processing*, vol. 39, no. 1, pp. 85–91, Jan. 1991.

[45] C. B. Seung, S. Ann, and I. Song, "Performance analysis of the dual sign algorithm for additive contaminated-Gaussian noise," *IEEE Signal Processing Lett.*, vol. 1, no. 12, pp. 196–198, Dec. 1994.

[46] J. A. Bucklew, T. G. Kurtz, and W. A. Sethares, "Weak convergence and local stability properties of fixed step size recursive algorithms," *IEEE Trans. Information Theory*, vol. 39, no. 3, pp. 966 –978, May 1993.

[47] E. Eweda, "Convergence analysis of the sign algorithm without the independence and Gaussian assumptions," *IEEE Trans. Signal Processing*, vol. 48, no. 9, pp. 2535 –2544, Sep. 2000.

[48] H.-F. Chen and G. Yin, "Asymptotic properties of sign algorithms for adaptive filtering," *IEEE Trans. Automatic Control*, vol. 48, no. 9, pp. 1545 – 1556, Sep. 2003.

[49] W. McGee, "Circularly complex gaussian noise–a price theorem and a mehler expansion (corresp.)," *IEEE Trans. Inf. Theory*, vol. 15, no. 2, pp. 317–319, 1969.

PLACE
PHOTO
HERE

**Yahong Rosa Zheng** received the B.S. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 1987, and the M.S. degree from Tsinghua University, Beijing, China, in 1989, both in electrical engineering. She received the Ph.D. degree from the Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada, in 2002. She was an NSERC Postdoctoral Fellow from Jan. 2003 to April 2005 at the University of Missouri-Columbia. Since 2005, she has been assistant then associate professor with the Department of Electrical and Computer Engineering at the Missouri University of Science and Technology (formerly the University of Missouri-Rolla). Her research interests include array signal processing, space-time adaptive processing, wireless communications, and wireless sensor networks. She has served as associate editor for *IEEE Transactions on Wireless Communications* 2006–2008 and is currently associate editor for *IEEE Transactions on Vehicular Technology*. She is the recipient of an NSF CAREER award in 2009.

PLACE
PHOTO
HERE

**Vítor H. Nascimento** was born in São Paulo, Brazil. He obtained the B.S. and M.S. degrees in Electrical Engineering from the University of São Paulo, in 1989 and 1992, respectively, and the Ph.D. degree from the University of California, Los Angeles, in 1999. From 1990 to 1994 he was a Lecturer at the Univ. of São Paulo, and in 1999 he joined the faculty at the same school, where he is now an Associate Professor. In 2002, he received a IEEE SPS Best Paper Award. He served as an Associate Editor for Signal Processing Letters from 2003 to 2005, for the Transactions on Signal Processing from 2005 to 2008 and for the EURASIP Journal on Advances in Signal Processing from 2006 to 2009. He is currently a member of the IEEE-SPS Signal Processing Theory and Methods Technical Committee. His research interests include adaptive filtering theory and applications, robust and nonlinear estimation, and applied linear algebra.