Summer 2011

Robust adaptive filtering algorithms for system identification and array signal processing in non-Gaussian environment

Tiange Shao

Department: Electrical and Computer Engineering

ROBUST ADAPTIVE FILTERING ALGORITHMS FOR SYSTEM IDENTIFICATION AND ARRAY SIGNAL PROCESSING IN NON-GAUSSIAN ENVIRONMENT

by

TIANGE SHAO

A DISSERTATION

Presented to the Faculty of the Graduate School of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

ELECTRICAL ENGINEERING

2011

Approved by

Yahong Rosa Zheng, Advisor
Steve Grant
Randy H. Moss
R.Joe Stanley
Xuerong Meggie Wen

PUBLICATION DISSERTATION OPTION

This dissertation consists of the following five published or submitted papers, formatted in the style used by the Missouri University of Science and Technology, listed as follows:

Paper 1, T. Shao, Y. R. Zheng, and J. Benesty, "An affine projection sign algorithm robust against impulsive interferences," has been published in IEEE Signal Processing Letters, vol.17, no.4, pp.327-330, April 2010.

Paper 2, T. Shao, Y. R. Zheng, and J. Benesty,"A variable step-size sign algorithm for acoustic echo cancelation," has been published in Proc. International Conf. on Acoustics, Speech, and Signal Processing (ICASSP10), Mar. 2010.

Paper 3, Y. R. Zheng, T. Shao, and Vitor Nascimento, "A new variable step-size fractional lower-order moment algorithm for non-Gaussian interference environments," has been submitted to Elsevier Digital Signal Processing, Dec. 2010.

Paper 4, Y. R. Zheng and T. Shao, "A Variable Step-Size LMP Algorithm for Heavy-Tailed Interference Suppression in Phased Array Radar," has been published in Proc. IEEE Aerospace Conf. (AeroConf'09), Mar. 2009.

Paper 5, Y. R. Zheng, T. Shao, and E. Blasch, "A fast-converging space-time adaptive processing algorithm for non-Gaussian clutter suppression," has been published in Elsevier Digital Signal Processing Journal, Nov. 2010.

ABSTRACT

This dissertation proposes four new algorithms based on fractionally lower order statistics for adaptive filtering in a non-Gaussian interference environment. One is the affine projection sign algorithm (APSA) based on L_1 norm minimization, which combines the ability of decorrelating colored input and suppressing divergence when an outlier occurs. The second one is the variable-step-size normalized sign algorithm (VSS-NSA), which adjusts its step size automatically by matching the L_1 norm of the *a posteriori* error to that of noise. The third one adopts the same variable-step-size scheme but extends L_1 minimization to L_p minimization and the variable step-size normalized fractionally lower-order moment (VSS-NFLOM) algorithms are generalized. Instead of variable step size, the variable order is another trial to facilitate adaptive algorithms where no *a priori* statistics are available, which leads to the variable-order least mean pth norm (VO-LMP) algorithm, as the fourth one.

These algorithms are applied to system identification for impulsive interference suppression, echo cancelation, and noise reduction. They are also applied to a phased array radar system with space-time adaptive processing (beamforming) to combat heavy-tailed non-Gaussian clutters.

The proposed algorithms are tested by extensive computer simulations. The results demonstrate significant performance improvements in terms of convergence rate, steady-state error, computational simplicity, and robustness against impulsive noise and interference.

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Dr. Yahong Rosa Zheng, for her generous support both technically and financially. I truly appreciate all her contributions of time, guidance, and funding that have made my Ph.D. experience stimulating and adventurous. She has taught me, both consciously and unconsciously, how a female Ph.D can strengthen her mind and make her way during career pursuit. Her diligent working style, zealous scientific attitude, and strict academic requirements have made her an excellent example for me as a successful scientist and professor.

I am very grateful to the members of my advisory committee, Drs. Steve L. Grant, Randy H. Moss, R. Joe Stanley, and Xuerong Meggie Wen, for their precious time in examining this dissertation and their valuable advice to my research work.

It is my pleasure to thank all my friends for their company and help during my Ph.D. years and all the colleagues at Missouri S&T for their assistance in my research work.

Unconditioned, I owe my deepest gratitude to both sides of my family for their unselfish love and everlasting support. I am indebted to my parents, for shaping my character of great virtue, for teaching me right from wrong, for encouraging me to insist the righteousness, for backing me up both spiritually and financially. I am equally indebted to my parents-in-law for their all-embracing love, understanding, and constant encouragement. My extraordinary love and gratitude go to my husband, Dr. Jian Zhang, for paying back his true love with all he has.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

1 INTRODUCTION

1.1 BACKGROUND

The adaptive filter is one of the most important operations of digital signal processing with deep theoretical challenges and immense practical relevance. In the last thirty years the adaptive filter has made significant achievements and has been used widely in many applications: radar, sonar, communication, navigation, seismology, biomedical engineering, and financial engineering [1, 2].

An adaptive filter is classified as linear if its input-output map obeys the principle of superposition when its parameters are fixed. Otherwise, the adaptive filter is said to be nonlinear. The linear adaptive filter is illustrated in Fig.1.1, where $x(k)$ denotes the input signal, $\mathbf{w}(k)$ denotes the weight vector of the taped delay line, $y(k) = \mathbf{w}^T(k)\mathbf{x}(k)$ is the filter output signal, and $d(k)$ defines the observed signal. The error signal $e(k)$ is calculated as the difference between observed signal and the actual filter output, $e(k) = d(k) - y(k)$. The objective of the adaptive filter is to extract information of interest contained in the observed noisy signal $d(k)$ according to some statistical criterion. To implement this objective, the first step of designing an adaptive filter is to choose a cost function of the error signal $e(k)$. Minimization of the cost function of error means minimization of the noise effects on the adaptive-filter output and guarantees that the adaptive-filter output signal matches the desired signal in a certain statistical sense.

The minimum mean-square error (MMSE) criterion based on the L_2 norm is well-known and serves as the fundamental and origin of the the adaptive filter. One reason of the high importance of the MMSE criterion lies in its optimality for a Gaussian distributed signal. The tails of Gaussian distribution are shown to have

decay rates with exponential square order which guarantees that the MMSE criterion is adequate to provide effective estimation under the Gaussian assumption. Another reason its feasibility of mathematical manipulations. The optimum solution is found to be the conditional expectation of variables of interest given observed ones. Such conditional expectation is generally hard to evaluate in closed-form, this is the reason that we try to find a linear estimator which can work effectively in practice. The optimum solution of MMSE for a linear filter is commonly known as the Wiener filter. Specially, we solved

$$\min_{\mathbf{w}(k+1)} E[|e(k)|^2]. \tag{1.1}$$

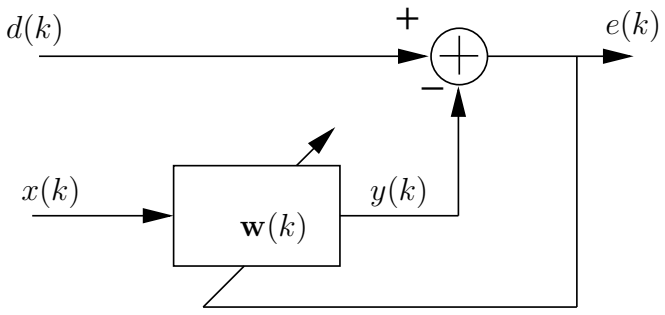


Figure 1.1. Adaptive filter

Designing and implementing a Wiener filter requires a priori information about the statistics of the input signal. However, the statistics are usually unknown and are often changing over time, which makes the computation of the Wiener solution impossible or no longer optimum. Moreover, direct implementation of Wiener filter involves matrix inversion which is often prohibitive in practical systems due to long filter length. In this situation, it becomes necessary to resort to an iterative procedure to approximate the optimum solution. The most popular adaptive algorithms based on the MMSE criterion are the least mean square (LMS) algorithm and the normalized

LMS (NLMS) algorithm. The NLMS algorithm updates its weight vector as

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu \frac{\mathbf{x}(k)e(k)}{\mathbf{x}(k)^T\mathbf{x}(k)}. \tag{1.2}$$

where $\mathbf{x}(k) = [x(k), x(k-1), \cdots, x(k-L+1)]^T$ is the input signal vector. The LMS and NLMS algorithms have the advantage of simplicity, low steady-state error, and fast tracking. However, their major drawbacks are the slow convergence with correlated input signal and performance degradation with non-Gaussian interference.

To speed up the convergence rate, the recursive least square (RLS) algorithm is proposed at the cost of high computational complexity. The affine projection algorithm (APA) provides a compromise of fast convergence and low computational complexity between the LMS and RLS algorithms. The APA updates the weight coefficients vector as

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu \mathbf{X}(k) \left[\mathbf{X}^T(k)\mathbf{X}(k) + \epsilon \mathbf{I}\right]^{-1} \mathbf{e}(k), \tag{1.3}$$

where $\mathbf{X}(k) = [\mathbf{x}(k), \mathbf{x}(k-1), \cdots, \mathbf{x}(k-M+1)]$ is the input signal matrix and ϵ is regularization parameter. Both the APA and the RLS algorithm can lead to an optimal solution in the mean-square error sense and have decorrelation properties that make them converge much faster than LMS and NLMS. Besides the APA and the RLS algorithm, variable step-size LMS-type algorithms have also been proposed to achieve both fast convergence and low steady-state errors. Nonlinear preprocessing is another methods often used to combat a correlated input signal and non-Gaussian interference. Most of the LMS-type algorithms have taken a combination of the above approaches. However, the algorithms with the MMSE criterion are based on the error $e(k)$ being Gaussian which is often disobeyed in a real-world environment.

Impulsive, non-Gaussian interference often occurs in practical applications and the signals with heavy-tailed statistics produce more large-magnitude outliers than

observed in the Gaussian model. In phased array radar systems, non-Gaussian clutters often occur in backscatters from mountain tops, dense forest canopy, rough sea surfaces, and manmade concrete objects. In acoustic echo cancelation (AEC), double-talk situations can also be viewed as an impulsive interference source. In these cases, the algorithms based on the L_2 norm, which have effective performance in a Gaussian environment, suffer serious performance degradation in non-Gaussian environments. This is because the characterization of a non-Gaussian signal by its second order moment is no longer optimal and other moment characterizations may be required.

Many studies have shown that lower order statistics can lead to improved convergence and robustness against non-Gaussian interference [3] by minimization of least mean p-th norm

$$\min_{\mathbf{w}(k+1)} E[|e(k)|^p] \tag{1.4}$$

where $1 \leq p \leq 2$. The approach using lower order statistics yields several robust algorithms against heavily-tailed interference, including the normalized sign algorithm (NSA) or least absolute deviation (LAD) algorithm and fractional lower-order moment/statistic (FLOM or FLOS) algorithm. These algorithms are based on the L_1, L_p, respectively, rather than the L_2 norm. We refer to 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 LAD algorithm. To combine the benefits of different norms, the mixed-norm algorithm is proposed based on the weighted combination of the L_1 and L_2 norms. The switched-norm algorithm which switches between L_1 and L_2 norms is proposed afterwards.

1.2 PROBLEM STATEMENT

Although lower-order moment statistics have proved to provide an robustness against impulsive environment in the general sense, its adaptive algorithms have

not been researched as rigorously as the LMS algorithm and a lot of challenging problems still remain. First, with the same step-size parameters, the LMP algorithms converge faster but have a greater steady-state error than the LMS algorithm. The fixed step size cannot provide a good compromise among fast convergence and small steady-state error and fast tracking. Second, the LMP algorithms with orders smaller than two suffers lower satiability than the LMS algorithms. Third, the fractional order increases computational cost by computing fractional statistics. Fourth, similar to the LMS algorithm, the LMP algorithms also suffer degradation with a highly correlated input signal. Fifth, convergence analysis for lower-order statistics adaptive algorithms remains scarce thus limited design guidelines are available in terms of the convergence rate, steady-state error, tracking properties, robustness to system error, and computational requirements.

Based on the discussion above, this work addresses the problems on lower-order statistic algorithms and proposes several new adaptive algorithms based on L_1 and L_p norm with $1 < p < 2$. By applying the idea of affine projection, the proposed algorithms are developed based on the cost function

$$\begin{aligned} &\min_{\mathbf{w}(k+1)} E[\|\mathbf{d}(k) - \mathbf{X}^T(k)\mathbf{w}(k+1)\|_p^p] \\ &\text{subject to } q\text{-th norm constraints.} \end{aligned} \tag{1.5}$$

where $\mathbf{d}(k)$ is the filter output vector defined as $\mathbf{y}(k) = [y(k), y(k-1), \cdots, y(k-M+1)]^T$. The value of q could be chosen as $1 \leq q \leq 2$.

My research in this dissertation yields promising results in two special cases. One is the affine projection sign algorithm (APSA) which uses L_1 minimization with

the minimum disturbance constraint, which has the cost function

$$\begin{aligned} &\min_{\mathbf{w}(k+1)} E[\|\mathbf{y}(k) - \mathbf{X}^T(k)\mathbf{w}(k+1)\|_1] \\ &\text{subject to} \quad \|\mathbf{w}(k+1) - \mathbf{w}(k)\|_2^2 \leq \delta^2. \end{aligned} \tag{1.6}$$

This constraint ensures that the updating weight coefficients vector does not change dramatically thus improving the algorithm's stability with impulsive outliers.

The second is the variable step-size LMP algorithms which are based on the cost function

$$\begin{aligned} &\min_{\mathbf{w}(k+1)} E[|y(k) - \mathbf{x}^T(k)\mathbf{w}(k+1)|^p] \\ &\text{subject to} \quad E\{|\varepsilon(k)|^p\} = E\{|v(k)|^p\}. \end{aligned} \tag{1.7}$$

where $\varepsilon(k)$ denotes the *a posteriori* error and $v(k)$ is background noise plus impulsive interference. The variable step size is chosen in the way of making the p-th norm of the *a posteriori* error equal to the p-th norm of the noise and interference in order to provide a compromise among fast convergence, small misadjustment, and good tracking ability. To avoid estimating the *a priori* information which is not available, this work mainly focusing on deriving nonparametric VSS algorithms [4].

Among all the lower-order algorithms, the family of sign algorithms based on the L_1-norm minimization has attracted more attention due to its considerably low computational cost and easy implementation. This is why the L_1 norm is sorted out and researched independently.

In addition to deriving lower-order statistic algorithms, convergence analysis is another important problem in adaptive filters [2]. Although an adaptive filter is physically implemented as a linear combiner, it is a highly nonlinear estimator in reality. For this reason, convergence analysis of adaptive algorithms is rather

challenging. The independence assumption and ordinary differential-equation method [1,2] is commonly used in convergence analysis of many LMS-type adaptive algorithms and leads to reasonable agreements between theory and practice. These assumptions and methods are also used in the analysis of the lower-order statistic algorithms in this work.

These new algorithms proposed in this work are applied to phased array antennas systems and acoustic echo cancelation (AEC). In array radar systems, heavy-tailed spiky clutters often occur in radar clutters and cause traditional methods to suffer significant performance degradation. In AEC systems, double talk also acts as a large-level uncorrelated impulsive interference source and easily causes the adaptive filter to diverge. The problem of impulsive noise shows the value of lower-order statistics methods in real applications.

1.3 SUMMARY OF CONTRIBUTIONS

This dissertation will consist of the journal publications and conference papers listed in the publication list. The published and expected contributions are:

1. Affine projection sign algorithm robust against impulsive interferences. A new affine projection sign algorithm (APSA) is proposed, which is robust against non-Gaussian impulsive interferences and has fast convergence. The conventional affine projection algorithm (APA) converges fast at a high cost in terms of computational complexity and it also suffers performance degradation in the presence of impulsive interferences. The family of sign algorithms (SAs) stands out due to its low complexity and robustness against impulsive noise. The proposed APSA combines the benefits of the APA and SA by updating its weight vector according to the L1-norm optimization criterion while using multiple projections. The features of the APA and the L1-norm minimization guarantee the APSA an excellent candidate for combatting

impulsive interference and speeding up the convergence rate for colored inputs at a low computational complexity.

2. A variable step-size sign algorithm for acoustic echo cancelation. A variable step size normalized sign algorithm (VSS-NSA) is proposed, for acoustic echo cancelation, which adjusts its step size automatically by matching the L_1-norm of the a posteriori error to that of the background noise plus near-end signal. Simulation results show that the new algorithm combined with double-talk detection outperforms the dual sign algorithm (DSA) and the normalized triple-state sign algorithm (NTSSA) in terms of convergence rate and stability.

3. Variable step-size fractional lower-order moment algorithm for system identification in non-Gaussian interference environments. A variable step-size fractional lower-order moment (VSS-FLOM) algorithm is proposed for system identification in an impulsive noise environment, which adapts the weight vector via the p-th moment of the a priori error. The step-size is automatically adjusted by matching the power of the a posteriori error to that of the background white noise. This low-complexity iterative algorithm is developed using time-averaging estimates of the second and p-th ($1 \leq p \leq 2$) error moments. The excess MSE and misalignment of the proposed VSS-FLOM algorithm are evaluated intensively by computer simulation under Gaussian and impulsive non-Gaussian interference environments, with white or colored Gaussian inputs, and for real and complex systems. The results show that the new VSS-FLOM algorithm combines the benefit of variable step size with the robustness of lower order statistics algorithms against impulsive interference.

4. Fast-converging space-time adaptive processing algorithm for non-Gaussian clutter suppression. Several new variable step-size and variable order least mean p-norm (VSS-LMP) algorithms are proposed for phased array radar application with space-time adaptive processing to combat heavy-tailed non-Gaussian clutters. The variable step-size LMP algorithms automatically change the step size according to the

estimated lower-order statistics of the error while the variable order LMP algorithm changes orders according a specific scheme for faster convergence. These algorithms are evaluated via a space-slow-time STAP example and the excess mean square error (MSE) and misalignment results show that the proposed algorithms converges fast and reach lower steady-state error than the fixed step-size or fixed order LMP algorithm. It also provides a better compromise between convergence speed and low steady state error than existing variable step-size LMS algorithms in both Gaussian and Compound K clutter environments.

1.4 REFERENCES

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

[2] Ali H. Sayed, *Fundamentals of of Adaptive Filtering*, John Wiley & Sons, Hoboken, New Jersey, 2003.

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

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

PAPER

I. AN AFFINE PROJECTION SIGN ALGORITHM ROBUST AGAINST IMPULSIVE INTERFERENCE

Tiange Shao, Yahong Rosa Zheng, and Jacob Benesty

Abstract—A new affine projection sign algorithm (APSA) is proposed, which is robust against non-Gaussian impulsive interferences and has fast convergence. The conventional affine projection algorithm (APA) converges fast at a high cost in terms of computational complexity and it also suffers performance degradation in the presence of impulsive interferences. The family of sign algorithms (SAs) stands out due to its low complexity and robustness against impulsive noise. The proposed APSA combines the benefits of the APA and SA by updating its weight vector according to the L_1-norm optimization criterion while using multiple projections. The features of the APA and the L_1-norm minimization guarantee the APSA an excellent candidate for combatting impulsive interference and speeding up the convergence rate for colored inputs at a low computational complexity. Simulations in a system identification context show that the proposed APSA outperforms the normalized least-mean-square (NLMS) algorithm, APA, and normalized sign algorithm (NSA) in terms of convergence rate and steady-state error. The robustness of the APSA against impulsive interference is also demonstrated.

1 Introduction

Adaptive filters have been commonly used in various applications of system identification, such as channel estimation, noise cancelation, echo cancelation, image

restoration, and seismic system identification [1]. The most popular adaptive filters are the least-mean-square (LMS) and normalized LMS (NLMS) algorithms due to their simplicity. However, their major drawbacks are slow convergence and performance degradation with colored input signals or in the presence of heavy-tailed impulsive interferences [2].

To overcome the deterioration of convergence performance caused by colored input signals, an affine projection algorithm (APA), which is based on affine subspace projections, has been proposed in [3]. Many variants of the APA have been developed in recent years [4]. The family of APAs updates the weight coefficients by multiple, most recent input vectors instead of a single, current data vector used in the LMS and NLMS algorithms. As the projection order of the APA increases, the convergence rate increases and so does the computational complexity. This is why computational efficient methods have also been developed to reduce the computational cost, such as the fast affine projection (FAP) algorithm [5]. In addition to the drawback of computational complexity, the APA also suffers performance degradation in non-Gaussian interference due to the nature of the L_2-norm optimization. Interfering signals with heavy-tailed distributions produce more outliers than Gaussian models and the L_2-norm minimization criterion is no longer a proper choice.

Many studies have shown that lower-order norms lead to robustness against impulsive and intensive interference. The least mean p-norm (LMP) algorithm based on the L_p norm is proposed in [2]. Among all the lower-order algorithms, the family of sign algorithms based on the L_1-norm minimization has attracted more attention due to its considerably low computational cost and easy implementation. Only the sign of the error signal is involved in the updating process. Many variants of the sign algorithm have been developed, including the normalized sign algorithm (NSA) [6], dual sign algorithm (DSA) [7], and variable step-size sign algorithm [8, 9]. The mixed-norm algorithm based on the weighted combination of the L_1 and L_2 norms

is proposed in [10]. The switched-norm algorithm is proposed in [11] which switches between L_1 and L_2 norms. Although the sign algorithms achieve good performance in many applications due to their low complexity and robustness against impulsive noise, they suffer from slow convergence rate, especially for highly correlated input signals.

We proposes an affine projection sign algorithm (APSA) which updates the weight vector with the L_1-norm optimization criterion by using multiple input vectors. The combination of the benefit of affine projection and L_1-norm minimization improves performance on combatting impulsive interference, speeding up the convergence rate with colored input signals, and lowering the computational complexity. The weight adaptation of the proposed algorithm does not involve any matrix inversion but only uses the sign operation of the error vector. The increase of the computational burden caused by high projection orders is much lower than the conventional APA.

The performance of the proposed APSA is evaluated in the context of system identification and compared with the NSA, APA, and NLMS algorithm. Simulation results with BG interference and colored input signals demonstrate the robustness of the APSA against impulsive interference, outperforming the APA and NLMS algorithm. The APSA also converges much faster and reaches a smaller steady-state misalignment than the NSA.

2 Conventional Affine Projection Algorithm

Consider a system identification problem where all signals are real. The output signal from an unknown system with a weight coefficients vector $\mathbf{w}$ is $y(k) = \mathbf{w}^T\mathbf{x}(k) + v(k)$, where $\mathbf{x}(k) = [x(k), x(k-1), \cdots, x(k-L+1)]^T$ is the input signal

vector of length L. The variable $v(k)$ represents the background noise plus interference signal. The superscript $()^T$ denotes vector transpose operation. Let $\hat{\mathbf{w}}(k)$ be an estimate of $\mathbf{w}$ at iteration k. The *a priori* is defined as $e(k) = y(k) - \hat{\mathbf{w}}^T(k)\mathbf{x}(k)$, while the *a posteriori* errors is defined as $\varepsilon(k) = y(k) - \hat{\mathbf{w}}^T(k+1)\mathbf{x}(k)$. Grouping the M recent input vectors $\mathbf{x}(k)$ together gives the input signal matrix: $\mathbf{X}(k) = [\mathbf{x}(k), \mathbf{x}(k-1), \cdots, \mathbf{x}(k-M+1)]$. We define the *a priori* and *a posteriori* error vectors as

$$\mathbf{e}(k) = [e(k), e(k-1), \cdots, e(k-M+1)]^T, \tag{1}$$

$$\mathbf{e}_{\mathrm{p}}(k) = [\varepsilon(k), \varepsilon(k-1), \cdots, \varepsilon(k-M+1)]^T, \tag{2}$$

and they can be computed as

$$\mathbf{e}(k) = \mathbf{y}(k) - \mathbf{X}^T(k)\hat{\mathbf{w}}(k), \tag{3}$$

$$\mathbf{e}_{\mathrm{p}}(k) = \mathbf{y}(k) - \mathbf{X}^T(k)\hat{\mathbf{w}}(k+1), \tag{4}$$

where $\mathbf{y}(k)$ is the output vector defined as $\mathbf{y}(k) = [y(k), y(k-1), \cdots, y(k-M+1)]^T$. The classical APA [3] is obtained by minimizing

$$\begin{aligned} &\|\hat{\mathbf{w}}(k+1) - \hat{\mathbf{w}}(k)\|_2^2 \\ &\text{subject to} \quad \mathbf{y}(k) - \mathbf{X}^T(k)\hat{\mathbf{w}}(k+1) = 0. \end{aligned} \tag{5}$$

The APA updates the weight coefficients vector as

$$\hat{\mathbf{w}}(k+1) = \hat{\mathbf{w}}(k) + \mu\mathbf{X}(k)\left[\mathbf{X}^T(k)\mathbf{X}(k) + \epsilon\mathbf{I}\right]^{-1}\mathbf{e}(k), \tag{6}$$

where the step size μ and regularization ϵ (both are positive numbers) have been added in (6) for a better control of the algorithm.

3 Affine Projection Sign Algorithm

The proposed affine projection sign algorithm is obtained by minimizing the L_1-norm of the *a posteriori* error vector with a constraint on the filter coefficients,

$$\min_{\hat{\mathbf{w}}(k+1)} \|\mathbf{y}(k) - \mathbf{X}^T(k)\hat{\mathbf{w}}(k+1)\|_1 \tag{7}$$

$$\text{subject to} \quad \|\hat{\mathbf{w}}(k+1) - \hat{\mathbf{w}}(k)\|_2^2 \leq \delta^2, \tag{8}$$

where δ^2 is a parameter ensuring that the updating weight coefficients vector does not change dramatically [11]. We can also view (8) as the minimum disturbance constraint. The minimum disturbance δ controls the convergence level of the algorithm and it shall be as small as possible. Using the method of Lagrange multipliers, the unconstrained cost function can be obtained by combining (7) and (8),

$$J(\hat{\mathbf{w}}(k+1)) = \|\mathbf{e}_{\mathrm{p}}(k)\|_1 + \beta \left[\|\hat{\mathbf{w}}(k+1) - \hat{\mathbf{w}}(k)\|_2^2 - \delta^2\right], \tag{9}$$

where β is a Lagrange multiplier. The derivative of the cost function (9)with respect to the weight vector $\hat{\mathbf{w}}(k+1)$ is

$$\begin{aligned}\frac{\partial J(\hat{\mathbf{w}}(k+1))}{\partial \hat{\mathbf{w}}(k+1)} &= -\sum_{m=0}^{M-1} \mathrm{sgn}(\varepsilon(k-m))\mathbf{x}(k-m) + 2\beta\left[\hat{\mathbf{w}}(k+1) - \hat{\mathbf{w}}(k)\right] \\ &= -\mathbf{X}(k)\mathrm{sgn}(\mathbf{e}_{\mathrm{p}}(k)) + 2\beta\left[\hat{\mathbf{w}}(k+1) - \hat{\mathbf{w}}(k)\right], \end{aligned} \tag{10}$$

where $\mathrm{sgn}(\cdot)$ denotes the sign function and $\mathrm{sgn}(\mathbf{e}_{\mathrm{p}}(k)) = [\mathrm{sgn}(\varepsilon(k)), \cdots, \mathrm{sgn}(\varepsilon(k-M+1))]^T$. Setting the derivative of $J(\hat{\mathbf{w}}(k+1))$ equal to zero, we get

$$\hat{\mathbf{w}}(k+1) = \hat{\mathbf{w}}(k) + \frac{1}{2\beta}\mathbf{X}(k) \cdot \mathrm{sgn}(\mathbf{e}_{\mathrm{p}}(k)). \tag{11}$$

Substituting (11) into the constraint (8), we obtain

$$\frac{1}{2\beta} = \frac{\delta}{\sqrt{\mathrm{sgn}(\mathbf{e}_\mathrm{p}^T(k))\mathbf{X}(k)\mathbf{X}^T(k)\mathrm{sgn}(\mathbf{e}_\mathrm{p}(k))}}. \tag{12}$$

Substituting (12) into (11), the update equation for the weight vector is then:

$$\hat{\mathbf{w}}(k+1) = \hat{\mathbf{w}}(k) + \frac{\delta\mathbf{X}(k)\cdot\mathrm{sgn}(\mathbf{e}_\mathrm{p}(k))}{\sqrt{\mathrm{sgn}(\mathbf{e}_\mathrm{p}^T(k))\mathbf{X}^T(k)\mathbf{X}(k)\mathrm{sgn}(\mathbf{e}_\mathrm{p}(k))}}. \tag{13}$$

Since the *a posteriori* error vector $\mathbf{e}_\mathrm{p}(k)$ depends on $\hat{\mathbf{w}}(k+1)$ which is not accessible before the current update, it is reasonable to approximate it with the *a priori* error vector $\mathbf{e}(k)$. The minimum disturbance δ controls the convergence level of the algorithm and it should be much smaller than one to guarantee convergence. It serves the similar purpose as the step-size parameter in conventional adaptive algorithms. Following the conventions, we replace δ by the step-size parameter μ. Defining $\mathbf{x}_\mathrm{s}(k) = \mathbf{X}(k)\mathrm{sgn}(\mathbf{e}(k))$ with $\mathrm{sgn}(\mathbf{e}(k)) = [\mathrm{sgn}(e(k)), \cdots, \mathrm{sgn}(e(k-M+1))]^T$, we obtain the APSA:

$$\hat{\mathbf{w}}(k+1) = \hat{\mathbf{w}}(k) + \mu\frac{\mathbf{x}_\mathrm{s}(k)}{\sqrt{\mathbf{x}_\mathrm{s}^T(k)\mathbf{x}_\mathrm{s}(k)+\epsilon}}, \tag{14}$$

where ϵ represents the regularization parameter which should be a positive number. Since μ comes from the minimum disturbance constraint δ, we should choose $0 < \mu \ll 1$ to ensure the stability of the algorithm and to achieve a small steady-state misalignment.

As shown in (14), no matrix inversion is needed for the proposed APSA and it only requires L multiplications at each iteration for the normalization. In comparison, the computational complexities of the APA and FAP algorithm are $2LM + K_\mathrm{inv}M^2$ and $2L+20M$ [5] multiplications respectively, where K_inv is the factor associated with the complexity required in matrix inversion. The proposed APSA is much simpler in implementation than the APA and even the FAP algorithm. Besides, it does not

have the numerical problems that the FAP exhibits. It is worth mentioning that the APSA with $M = 1$ reduces to a new kind of normalized sign algorithm, whose normalization is based on the Euclidean norm of the input vector. This is different from the normalized least-mean-absolute deviation (NLMAD) algorithm in [6] which is normalized by the L_1 norm of the input vector.

4 Algorithm Performance

The proposed APSA is compared to the NLMS, APA, and NSA via system identification applications. The adaptive filter has a length $L = 256$ taps. The input signal is chosen to be a colored Gaussian process. This input is generated by filtering a white Gaussian noise through a first order system with a pole at 0.8 or 0.95. An independent white Gaussian noise is added to the system background with a 30 dB signal-to-noise ratio (SNR). In addition, a strong interference signal is also added to the system output $y(k)$ with a signal-to-interference ratio (SIR) of -30 to 10 dB. The Bernoulli-Gaussian (BG) distribution [11] is used for modeling the interference signal, which is generated as the product of a Bernoulli process and a Gaussian process, *i.e.*, $z(k) = \omega(k)n(k)$, where $n(k)$ is a white Gaussian random sequence with zero mean and variance σ_n^2, and $\omega(k)$ is a Bernoulli process with the probability mass function given as $P(\omega) = 1 - Pr$ for $\omega = 0$, and $P(\omega) = Pr$ for $\omega = 1$. The average power of a BG process is $Pr \cdot \sigma_n^2$. Keeping the average power constant, a BG process is spikier when Pr is smaller. It reduces to a Gaussian process when $Pr = 1$.

The convergence is evaluated by the normalized misalignment [1] defined as $\mathcal{M}(k) = 20 \log_{10}\{||\hat{\mathbf{w}}(k) - \mathbf{w}||_2 \setminus ||\mathbf{w}||_2\}$. The ensemble average of 20 trails is used for $\mathcal{M}(k)$. The regularization parameter ϵ is set to 0.0001 for the APA and 0 for the APSA.

This work first examines the performance of the APSA with different projection orders M, as shown in Fig. 1, where $M = 1, 2, 5, 10, 20$, and the interference is a BG with SIR $= -30$ dB and $Pr = 0.001$ is used. The APSA with higher projection order achieves both faster convergence and lower misalignment for $M = 1$ to 10. When M is larger than a certain value (in this case is 10), the convergence is faster with a larger M but the steady-state misalignment level is higher. Increasing the projection order also means increased computational complexity. Therefore a proper selection of M provides good tradeoff between convergence rate and computational complexity.

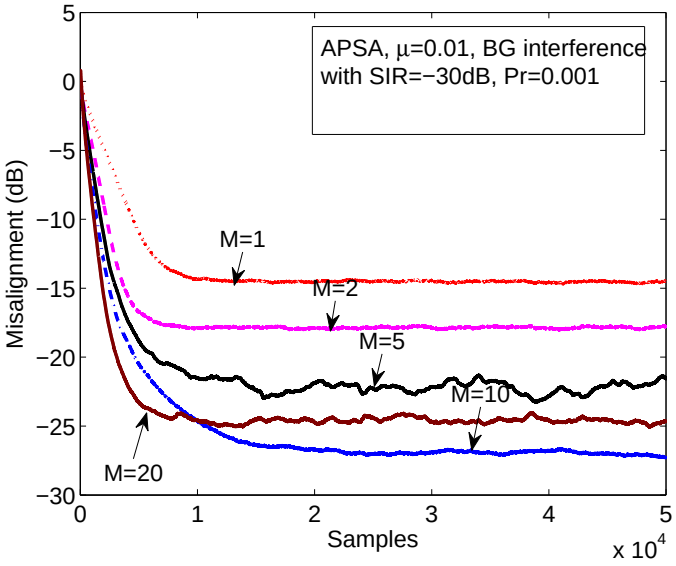


Figure 1. Misalignment of the APSA with varying values of $M = 1, 2, 5, 10, 20$ and same step size of $\mu = 0.01$. The input is an AR(1) with pole at 0.8. The background noise is Gaussian with SNR=30 dB. The interference is a BG with SIR=-30 dB and $Pr = 0.001$.

This work also examines the effect of the step size on the misalignment of the APSA, as shown in Fig. 2, where four step sizes, $\mu = 0.1, 0.01, 0.0025, 0.001$, are

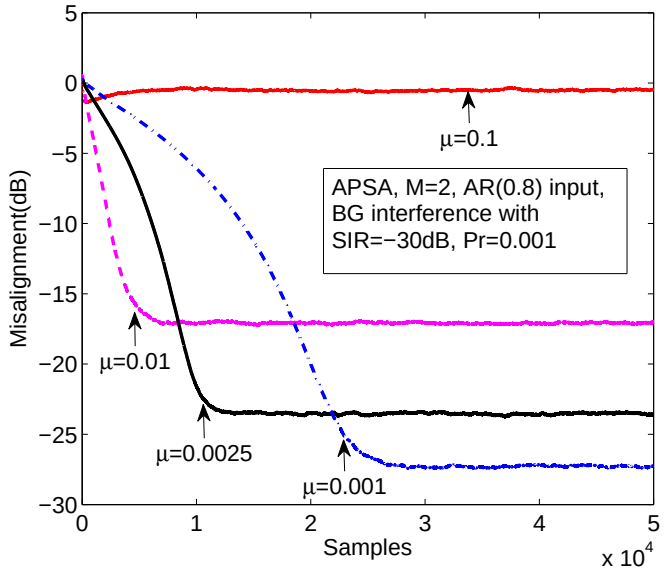


Figure 2. Misalignment of the APSA with varying step sizes of $\mu = 0.1, 0.01, 0.0025, 0.001$ and same projection order of $M = 2$. Other parameters are the same as those in Fig. 1.

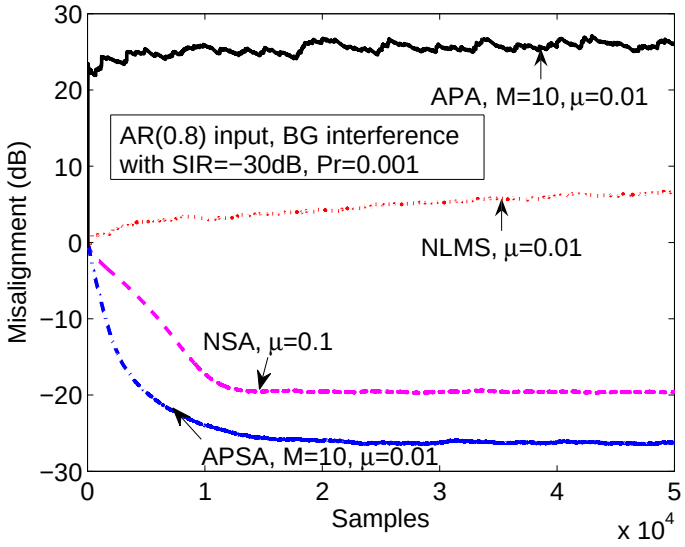


Figure 3. Misalignment comparison of the APSA, NLMS, APA, and NSA. Other parameters are the same as those in Figs 1 and 2.

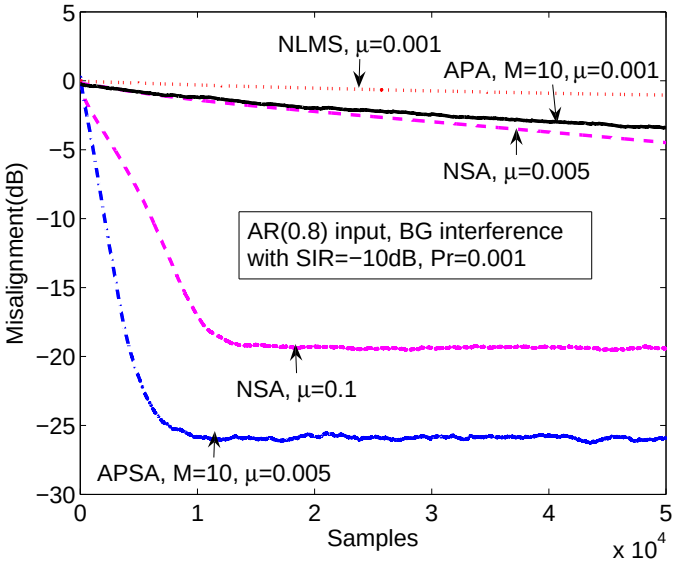


Figure 4. Misalignment comparison of the APSA, NLMS, APA, and NSA. SIR=−10 dB. Other parameters are the same as those in Fig. 3.

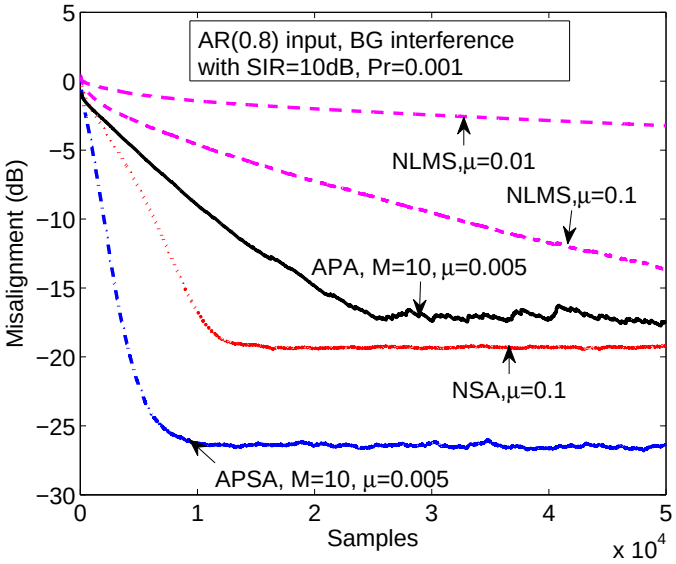


Figure 5. Misalignment comparison of the APSA, NLMS, APA, and NSA. SIR=10 dB, $Pr = 0.001$. Other parameters are the same as those in Fig. 3.

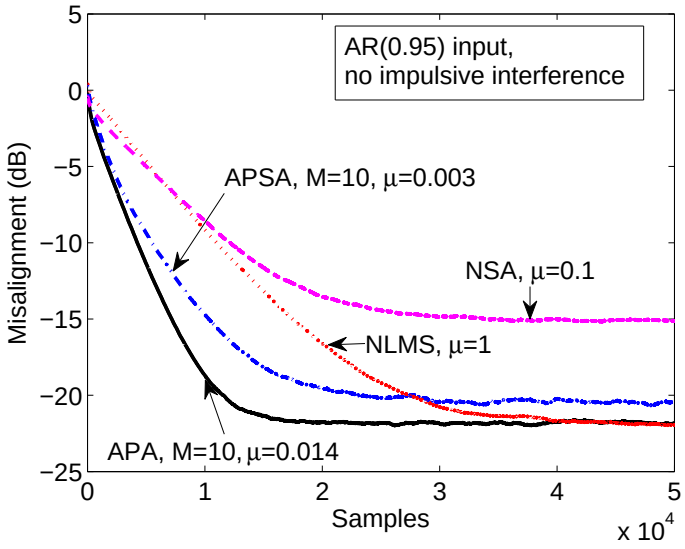


Figure 6. Misalignment comparison of APSA, NLMS, APA, and NSA. No impulsive noise. The input is an AR(1) with pole at 0.95.

used in the simulations. A small step size slows down the convergence rate but also lowers the steady-state misalignment. In contrast, a large step size speeds up the convergence rate but gives a higher steady-state misalignment.

The comparisons of the proposed APSA with the conventional NLMS, APA, and NSA are shown in Figs. 3 to 5 under a BG interference with various SIRs but with a fixed value of Pr. The SIR is chosen as -30 dB, -10 dB, and 10 dB. The value of Pr is set to 0.001. In all three cases, the APA and NLMS are more likely to diverge in more intensive interference while the two sign algorithms are robust against the impulsive interference. The APSA converges faster and achieves smaller steady-state misalignment than the NSA.

The case of no impulsive interference is shown in Fig. 6, where only background noise with an SNR of 30 dB is present. The input is a higher correlated AR(1) process with a pole at 0.95. The APA performs best and the NSA performs worst. The APSA performs in the middle, but no worse than the NLMS.

Finally, this work studies the effect of the impulsiveness of the BG interference on the performance of the APSA and NSA, as shown in Fig. 7. The SIR is fixed at -10 dB and the values of Pr are selected to be 0.001, 0.01, and 0.1. The more spikier the interference, the better the APSA and NSA perform. In all the interference scenarios, the APSA outperforms the NSA and the performance gain of the APSA are greater than that of the NSA.

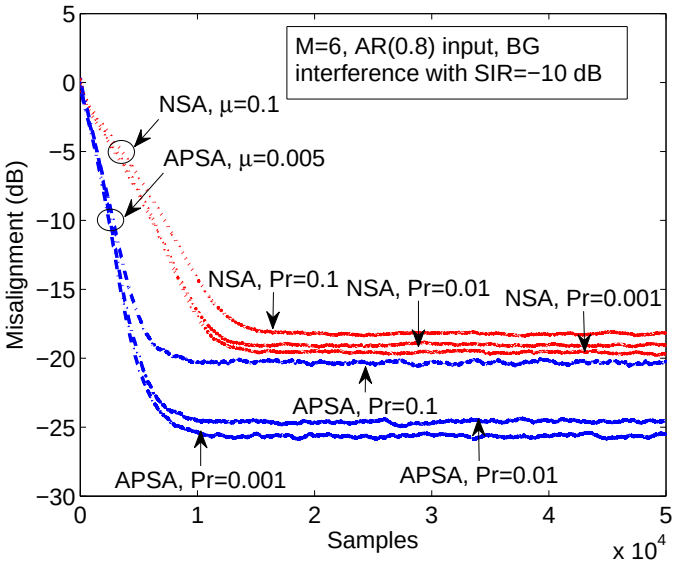


Figure 7. Misalignment comparison of APSA and NSA under a BG interference with various levels of impulsiveness. SIR$=-10$ dB, the values of Pr are chosen as $0.001, 0.01$, and 0.1. The input is an AR(1) with pole at 0.8.

5 Conclusions

This paper has proposed an affine projection sign algorithm (APSA) that updates its weight vector according to the sign of the *a priori* error vector based on

the L_1-norm optimization. A constraint is applied on variations of the weight vector, leading to normalization based on the correlation matrix of the input signal. The proposed APSA combines the benefits of the APA and sign algorithm. The affine projection makes the APSA converge fast with colored input signals while the L_1 optimization guarantees its robustness against impulsive interference. In addition, the APSA has much lower computational complexity than the conventional APA because its adaptation only involves the sign operation. As a result, a large projection order can be selected to achieve faster convergence rate with affordable computational cost. Simulations have also confirmed the APSA's improved ability to combat impulsive interferences and accelerate the convergence rate with colored input signals.

6 References

[1] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Upper Saddle River, New Jersey, 4 edition, 2002.

[2] 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, July 1993.

[3] 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.

[4] H. C. Shin, A. H. Sayed, and W. J. Song, "Variable step-size NLMS and affine projection algorithms," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 132–135, Feb. 2004.

[5] S. L. Gay and S. Tavathia, "The fast affine projection algorithm," in *Proc. IEEE ICASSP*, 1995, vol. 5, pp. 3023–3026.

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

[7] C. Kwong, "Dual sign algorithm for adaptive filtering," *IEEE Trans. Communications*, vol. 34, no. 12, pp. 1272–1275, Dec 1986.

[8] T. Shao and Y.R. Zheng, "A new variable step-size fractional lower-order moment algorithm for non-Gaussian interference environments," in *Proc. IEEE ISCAS*, May 2009, pp. 2065–2068.

[9] Y.R. Zheng and Tiange Shao, "A variable step-size LMP algorithm for heavy-tailed interference suppression in phased array radar," in *Proc. IEEE Aerospace conference*, March 2009, pp. 1–6.

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

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

II. A VARIABLE STEP-SIZE NORMALIZED SIGN ALGORITHM FOR ACOUSTIC ECHO CANCELATION

Tiange Shao, Yahong Rosa Zheng, and Jacob Benesty

Abstract—A variable step size normalized sign algorithm (VSS-NSA) is proposed, for acoustic echo cancelation, which adjusts its step size automatically by matching the L_1 norm of the *a posteriori* error to that of the background noise plus near-end signal. Simulation results show that the new algorithm combined with double-talk detection outperforms the dual sign algorithm (DSA) and the normalized triple-state sign algorithm (NTSSA) in terms of convergence rate and stability.

1 Introduction

In echo cancelation applications, the family of sign algorithms have become popular due to its simplicity and ease of implementation. Only the sign of the error signal is involved in the updating process. However, the fixed step-size normalized algorithms can not meet the conflicting goals of fast convergence and small steady-state error. A large step size leads to fast convergence but large steady-state error while a small step size yields small steady-state error but slow convergence. Another conflict is that high convergence rate is usually more sensitive to near-end disturbances, especially accompanied by high divergence rate in the presence of double talk.

Several variable step-size sign algorithms have been proposed in the literatures [1, 2, 3, 4] to overcome these conflicts. The dual sign algorithm (DSA) [1] operates as if two sign algorithms with different step-size parameters are working in cooperation. It transits from a large step size to a small one at the presence of double talk, thus reducing the divergence rate and improving stability. However,

switching between two step-size parameters does not ensure non-divergence during double talk, especially at sharp and large transitions between non-speech and speech at near-end. The normalized triple-state sign algorithm (NTSSA) [2] improves upon the DSA by inserting a third step size to provide a better trade-off between stability and convergence. Unlike the hard-switching of the DSA, the NTSSA which involves three-state step size ensures soft transition from one step size to another. The design and performance of the DSA and NTSSA are determined by the values of transition thresholds, hangover times and the selections of two (for DSA) or three (for NTSSA) step-size parameters. Three parameters are involved in the DSA and a rough rule is provided in [1, 3] for the selection of these parameters. In contrast, the NTSSA has to choose 13 parameters including 3 step-size parameters, 5 thresholds, 5 hangover times. The selection and coordination of these parameters are critical to the performance of the algorithm and they are dependent on near-end/far-end signals and background noises. Unfortunately, no clear guidance has been provided for parameter selection of the NTSSA and the selection is done by trial and error, making it very difficult to implement in practical applications.

In this paper, we propose a novel variable step-size normalized sign algorithm which adjusts its time-varying step size automatically, according to input and error statistics. It avoids complicated, manual selection of parameters and ensure automatic change of the step size. It achieves both fast convergence and small steady-state error. The proposed VSS-NSA is combined with Geigel double-talk detection algorithm [5, 6, 7] to ensure stability at simultaneous present of far-end and near-end signals.

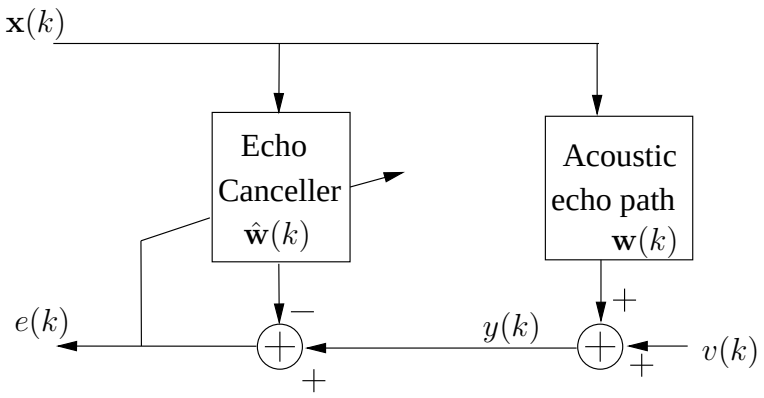


Figure 1. Block diagram of an echo canceler. $\mathbf{x}(k)$–far-end signal vector, $v(k)$–near-end signal plus background noise.

2 The Proposed Variable step-size Normalized Sign Algorithm

The echo cancelation system can be modeled as a system identification problem, as shown in Fig. 1. The echo canceler's goal is to detect and remove echo, thereby enhancing voice quality of the near-end speech. The echo is generated by filtering the far-end speech $x(k)$ by the echo path vector $\mathbf{w}$ of length L. The microphone signal $y(k)$ is the echo plus background noise, including the near-end speech when double talk happens, which is expressed as

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

where $v(k)$ is the background noise plus near-end speech. The superscript $()^T$ denotes transpose. Let $\hat{\mathbf{w}}(k)$ be an estimate for the true echo path vector $\mathbf{w}$ at iteration k. The cost function used here is the mean absolute value:

$$J(\hat{\mathbf{w}}) = E\left\{|y(k) - \hat{\mathbf{w}}^T\mathbf{x}(k)|\right\}, \tag{2}$$

where $E\{\cdot\}$ is the expectation operator. The sign algorithm updates the filter coefficients along the steepest descent of the cost function in (2). Using the stochastic gradient approach, the filter coefficients are solved iteratively by [3]:

$$\hat{\mathbf{w}}(k) = \hat{\mathbf{w}}(k-1) + \mu(k)\text{sign}(e(k))\mathbf{x}(k), \tag{3}$$

where $\mu(k)$ is the variable step size.

The *a priori* and *a posteriori* errors are defined respectively as

$$\begin{aligned} e(k) &= y(k) - \hat{\mathbf{w}}^T(k-1)\mathbf{x}(k), & (4)\\ \varepsilon(k) &= y(k) - \hat{\mathbf{w}}^T(k)\mathbf{x}(k). & (5) \end{aligned}$$

Substituting (2) into (8) and substituting (5) yields

$$\begin{aligned} \varepsilon(k) &= e(k) + [\hat{\mathbf{w}}(k-1) - \hat{\mathbf{w}}(k)]^T \mathbf{x}(k) \\ &= e(k) - \mu(k)\text{sign}(e(k))\mathbf{x}^T(k)\mathbf{x}(k). \end{aligned} \tag{6}$$

In the absence of noise, a reasonable method for selecting a variable step size is to set $\varepsilon(k)$ equal to 0. However, in the presence of noise, a better criterion [8] is to set $[\mathbf{w} - \hat{\mathbf{w}}(k)]^T\mathbf{x}(k)$ equal to 0 for all k. This implies, based on (8), that the variable step size is selected to ensure

$$E\{|\varepsilon(k)|\} = E\{|v(k)|\}. \tag{7}$$

Substituting (6) into (7) , the *a posteriori* error in terms of $\mu(k)$ can be expressed as

$$\begin{aligned} E\left[|\varepsilon(k)|\right] &= E\left[|e(k)-\mu(k)\text{sign}(e(k))\mathbf{x}^T(k)\mathbf{x}(k)|\right] \\ &= E\left[||e(k)|-\mu(k)\mathbf{x}^T(k)\mathbf{x}(k)|\right] \\ &= E\left[|v(k)|\right]. \end{aligned} \tag{8}$$

Assume $\mu(k)$ is much smaller than 1 and the power of the input signal is normalized to 1, approximation can be made as follow by first-order Taylor expansion:

$$\begin{aligned} E\left[|\varepsilon(k)|\right] &\approx E\left[|e(k)|\right]-\mu(k)E\left[\mathbf{x}^T(k)\mathbf{x}(k)\right] \\ &= E\left[|v(k)|\right]. \end{aligned} \tag{9}$$

The variable step size $\mu(k)$ is directly obtained from (9):

$$\mu(k)=\frac{L_1(e(k))-L_1(v(k))}{E\left[\mathbf{x}^T(k)\mathbf{x}(k)\right]}, \text{ if } L_1(e(k))\geq L_1(v(k)) \tag{10}$$

where $L_1(e(k)) \doteq E[|e(k)|]$ and $L_1(v(k)) \doteq E[|v(k)|]$. In practical implementation, the expectation $E\left[\mathbf{x}^T(k)\mathbf{x}(k)\right]$ can be replaced by the instantaneous signal energy $\mathbf{x}^T(k)\mathbf{x}(k)$. The function $L_1(e(k))$ can be estimated by time averaging

$$\hat{L}_1(e(k)) = \lambda\hat{L}_1(e(k-1))+(1-\lambda)|e(k)|, \tag{11}$$

where λ is the forgetting factor. This yields a variable step-size normalized sign algorithm

$$\hat{\mathbf{w}}(k+1)=\hat{\mathbf{w}}(k)+\frac{\hat{L}_1(e(k))-L_1(v(k))}{\mathbf{x}^T(k)\mathbf{x}(k)+\delta}\text{sign}(e(k))\mathbf{x}(k), \tag{12}$$

if $L_1(e(k))\geq L_1(v(k))$. Otherwise, the algorithm stops updating. The value of δ is a regulation parameter.

3 Geigel Double-Talk Detection

The performance of an echo canceler during double talk is an important measurement because near-end speech often causes divergence, especially at high convergence rate. A double-talk detector (DTD) is a good method to meet the contradictory requirement of low divergence rate and fast convergence in echo cancelation. It inhibits updates while the far- and near-end speeches are present simultaneously. To ensure the stability of the algorithm, the proposed VSS-NSA is combined with a simple DTD algorithm, the Geigel DTD algorithm [7]. The Geigel DTD detects the near-end signals by comparing the magnitude of current far-end sample and the maximum magnitude of the recent past samples of the near-end signals, which means declaring double-talk when

$$|y(k)| > T \max\{|x(k)|, |x(k-1)|, \cdots, |x(k-L+1)|\}. \tag{13}$$

The factor of T is usually set to 0.5 based on the assumption of 6 dB hybrid attenuation. Once the double talk is declared, the updates is inhibited for some hangover time in order to reduce the miss of detection.

4 Algorithm Performances

The proposed VSS-NSA was compared to the DSA and NTSSA via an echo cancelation application. The echo path $\mathbf{w}$ was taken from an acoustic impulse response of a room, which was generated according to the image model [9] and truncated to $L = 256$ taps. The far-end and near-end speech was sampled at 8 kHz. The power of near-end signal was 10 dB less than the far-end signal. An independent white

Gaussian noise was added as system background noise with a 30 dB signal-to-noise ratio (SNR). We used two sets of near-end speeches, as shown in Fig. 2(b) and Fig. 4(b) to test the algorithms. The far-end speech remained the same, as shown in Fig. 2(a) and Fig. 4(a). The Geigel double-talk detection was used with the assumption of 0 dB hybrid attenuation and the threshold T was set to 1.2. The detection results of the two cases are shown in Fig. 2(c) and Fig. 4(c), respectively, where the value of 1 stands for a double-talk declaration while the value of 0 stands for no near-end speech.

For both the DSA and NTSSA, parameter selection affected the algorithms directly. Although the study of the step size for the DSA had been carried out in [1,3] and offered methods of calculating the appropriate parameters, those methods only yielded a rough range of the parameter estimation based on the statistics of the input signal. Manual adjustment of each parameter was needed to achieve good performance. As for the NTSSA proposed in [2], which involved more parameters (3 step sizes, 5 transition thresholds, and 5 handover times), there were no general rules and the selection was done by trial and error. Besides, the statistics of speech signals had great impact on the parameter selection and consequently on the algorithm performance. Once speech signals were changed, the old parameters might not work any more and they needed to be re-tuned. In contrast, the VSS-NSA updated its step size automatically without off-line calculation of statistics. The step size became small when DTD declared double talk and large when DTD declared no double talk.

In case 1, we chose 0.03 and 0.3 for the two step sizes and 80 for the transition threshold in simulations of the DSA based on the rough guide of [1]. For the NTSSA, we chose 2^{-2}, 2^{-3}, 2^{-6} for the three step sizes, other requiring parameters were the same as in [2]. The forgetting factor λ was set to 0.976 for both the DSA and NTSSA while set to 0.998 for the VSS-NSA. The initial step size of the VSS-NSA was set to

1. The hangover time for the Geigel DTD was set to 200 samples equivalent to 25 ms.

In case 2, in order to get better performance, the two step-size parameters of the DSA had to be changed to 0.01 and 0.5 while the three step-size parameters of the NTSSA had to be changed to 2^{-3}, 2^{-4}, 2^{-6} and its forgetting factor λ was changed to 0.99. No changes were needed for any parameters of the VSS-NSA.

The convergence performances were evaluated by the normalized misalignment $\mathcal{M}(k)$ defined as [10]

$$\mathcal{M}(k) \quad = \quad 20\log_{10}\frac{||\hat{\mathbf{w}}(k)-\mathbf{w}||}{||\mathbf{w}||}. \tag{14}$$

The comparisons of the three sign algorithms for case 1 and case 2 are shown in Fig. 3 and Fig. 5, respectively. In both the two cases, the NTSSA was superior to the DSA. The VSS-NSA outperformed both the NTSSA and DSA in terms of convergence rate. In addition, the VSS-NSA was robust against the change of the near-end speeches and did not need recalculating the parameters like the DSA and NTSSA.

Actually, the Geigel DTD was not very effective on acoustic signal detection. There were a lot of misses and false alarms. When a miss happened, the step size that should had been frozen to zero increased greatly and caused a little divergence, as shown in Fig. 3. Fortunately, the VSS-NSA was robust enough to prevent the divergence in a short time. As a result, the VSS-NSA ensured the system stability with the help of Geigel DTD. The combination of the VSS-NSA and Geigel DTD served as a robust algorithm for acoustic echo cancelation.

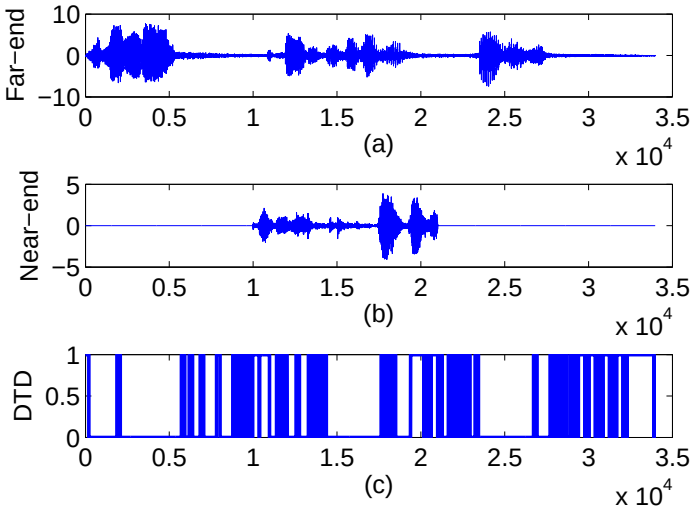


Figure 2. Case 1 of acoustic echo cancelation. (a) the far-end speech; (b) the near-end speech; (c) results of the Geigel DTD.

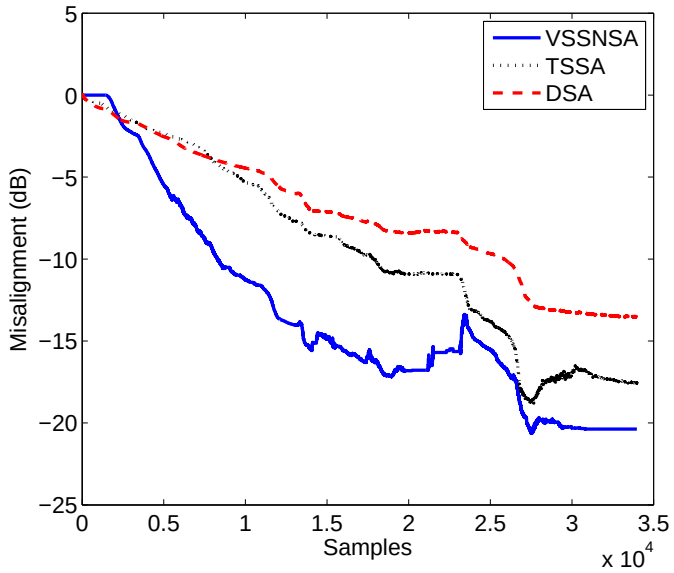


Figure 3. Misalignment of the VSS-NSA, DSA, and NTSSA for case 1. The VSS-NSA converged faster than the DSA and NTSSA even though the DTD had lots of false alarms and the VSS-NSA froze adaptation at each declared double talk.

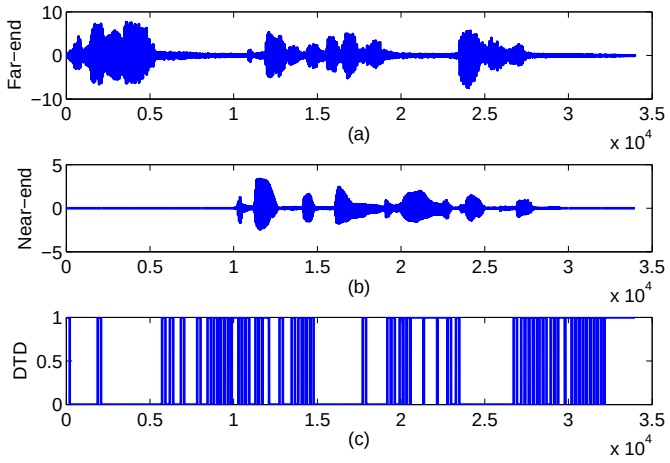


Figure 4. Case 2 of acoustic echo cancelation. (a) the far-end speech; (b) the near-end speech; (c) results of the Geigel DTD.

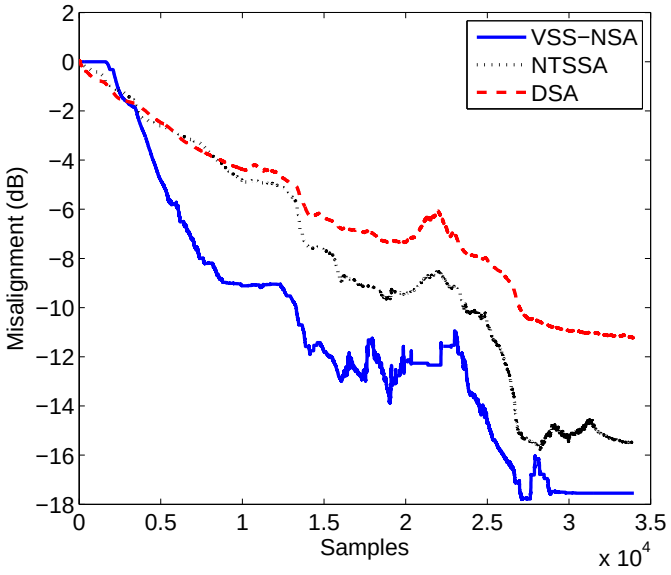


Figure 5. Misalignment of the VSS-NSA, DSA, and NTSSA for case 2. The VSS-NSA had the same parameters as those in case 1 but the DSA and NTSSA needed to change parameters to get good performance.

5 Conclusions

In this paper, a variable step-size normalized sign algorithm (VSS-NSA) has been proposed and compared with other popular sign algorithms such as the dual sign algorithm (DSA) and normalized triple-state sign algorithm (NTSSA) for application of acoustic echo cancelation. The DSA and NTSSA involve several step-size parameters and transition thresholds based on off-line calculation of signal statistics. In addition, the two kinds of sign algorithms are affected greatly by practical parameters selection. Different from the DSA and NTSSA, the VSS-NSA is much more intelligent in that it automatically adjusts the step size by matching the L_1 norm of the *a posterior* error to that of the unwanted noise. The proposed VSS-NSA improves convergence rate while reduces the steady-state error. However, the fast convergence is also accompanied with high divergence rate during double-talk periods. In this paper, we use the Geigel double-talk detector to help the VSS-NSA combat the double talk. Simulations demonstrate that the proposed VSS-NSA combined with Geigel double-talk detection outperforms other sign algorithms in terms of both convergence rate and system stability.

6 Acknowledgement

The authors wish to thank Dr. Steve Grant for his suggestion and discussion about the double-talk detection problem.

7 References

[1] C. Kwong, "Dual sign algorithm for adaptive filtering," *IEEE Trans. Communications*, vol. 34, no. 12, pp. 1272–1275, Dec 1986.

[2] S. B. Jebara and H. Besbes, "Variable step size filtered sign algorithm for acoustic echo cancellation," *Electron. Lett.*, vol. 39, no. 12, pp. 936–938, June 2003.

[3] N. Verhoeckx and T. Claasen, "Some considerations on the design of adaptive digital filters equipped with the sign algorithm," *IEEE Trans. Communications*, vol. 32, no. 3, pp. 258–266, Mar 1984.

[4] Y. R. Zheng and T. Shao, "A variable step-size lmp algorithm for heavy-tailed interference suppression in phased array radar," in *Proc. IEEE AeroConf09*, Big sky, MT, Mar. 2009.

[5] J. Benesty, T. Gansler, D.R. Morgan, M.M. Sondhi, and S.L. Gay, *Advances in Network and Acoustic Echo Cancellation*, Springer-Verlag, Berlin, Germany, 2001.

[6] T. Gansler, S.L. Gay, M.M. Sondhi, and J. Benesty, "Double-talk robust fast converging algorithms for network echo cancellation," *IEEE Trans. Speech, Audio Processing*, vol. 8, no. 6, pp. 656–663, Nov 2000.

[7] D. Duttweiler, "A twelve-channel digital echo canceler," *IEEE Trans. Communications*, vol. 26, no. 5, pp. 647–653, May 1978.

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

[9] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Acoustical Society of America Journal*, vol. 65, pp. 943–950, Apr. 1979.

[10] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Upper Saddle River, New Jersey, 4 edition, 2002.

III. A VARIABLE STEP-SIZE NORMALIZED FRACTIONALLY LOW-ORDER MOMENT ALGORITHMS FOR NON-GAUSSIAN INTERFERENCE ENVIRONMENT

Yahong Rosa Zheng, Tiange Shao, and Vitor Nascimento

Abstract—Two variable step-size normalized fractionally lower-order moment (VSS-NFLOM) algorithms are proposed for system identification in a 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 $1 \leq p \leq 2$. The proposed VSS-NFLOM 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-NFLOM algorithms with a smaller p converges faster and achieves lower steady-state error in impulsive interference and/or a colored input environment. The VSS-NFLOM algorithms also perform better than the Fixed Step-Size (FSS) NFLOM in both Gaussian and impulsive interference environments.

1 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, low steady-state error, and fast tracking. However, their major drawbacks are slow convergence [8] and performance degradation in colored input or non-Gaussian interference [9].

Over the past two decades, many variants of LMS have been proposed to overcome these problems [10, 11, 12, 13, 14, 15, 16, 17, 18, 5, 19, 9, 20]. 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 [11, 10, 16, 12, 13, 14, 15]; 2) using a small-order affine projection [17, 16] to compromise fast convergence and low computational complexity for colored inputs; 3) using nonlinear filtering such as median filtering, higher or lower order moment algorithms to combat non-Gaussian interference [5, 18, 9, 21, 22, 19, 23]; and 4) using combinations of algorithms with different properties [24, 25].

Impulsive, non-Gaussian interference often occurs in practical applications and the LMS algorithm, as an optimal method 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. The approach using higher or lower order statistics yields several families of algorithms including the normalized sign algorithms (NSA) or least absolute deviation (LAD) algorithms [26, 5, 27], fractional lower-order moment/statistic (FLOM or FLOS) algorithms [9, 21], and least mean fourth-moment (LMF) algorithms [28]. 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 to 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 LAD algorithm.

In particular, the FLOM algorithms (or LMP with $1 \leq p < 2$) have been developed for very impulsive models. They were first proposed for systems with alpha-stable distributed inputs [9], and later reformulated to the normalized FLOM

algorithm [19]. It has been shown [4, 5, 29] that the NFLOM algorithms perform better than the NLMS in heavy-tailed non-Gaussian interference, but slightly worse in a Gaussian interference environment. When the order p is smaller, the NFLOM algorithm achieves faster convergence but with 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 NFLOM 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, 11, 16, 30, 15, 14, 12, 13]. 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 [30, 16] 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 [24, 25] 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) algorithms for NFLOM adaptive filters, thus combining the benefits of variable step sizes with the robustness of the lower order statistics algorithms against 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]. The proposed VSS-NFLOM 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-NFLOM algorithms. In contrast, the exact solution to the quadratic equation does not lead to a good VSS algorithm.

The proposed VSS-NFLOM algorithms are 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-NFLOM algorithms achieve faster or comparable convergence rate and smaller steady-state error than the fixed step-size NFLOM (FSS-NFLOM) algorithms in all signal scenarios. The VSS-NFLOM 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.

2 The Proposed VSS-NFLOM 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

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

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

respectively. The adaptive filter must minimize the cost function selected as the p-th order moment of ε_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 FLOM algorithm provides an approach similar to the LMS algorithm for updating the filter coefficients along the steepest descent of the cost function (2). Using the stochastic gradient approach, the filter coefficients are solved iteratively by [9]

$$\hat{\mathbf{w}}(k) = \hat{\mathbf{w}}(k-1) + \mu(k) e_k^{<p-1>} \mathbf{x}(k) \tag{5}$$

where $\mu(k)$ is the step size, the operation $z^{<p>}$ is defined as $z^*|z|^{p-1}$, with superscript * denoting 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 ε_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 (8), 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 (8) and substituting (5), the *a posteriori* error in terms of $\mu(k)$ can be expressed as,

$$\begin{aligned} \varepsilon_k &= e_k + [\hat{\mathbf{w}}(k-1) - \hat{\mathbf{w}}(k)]^H\mathbf{x}(k), \\ &= e_k - \mu(k)[e_k^{<p-1>}]^*\mathbf{x}^H(k)\mathbf{x}(k). \end{aligned} \tag{7}$$

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 $\mu(k)$:

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

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

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

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

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

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

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

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 only approximately true after the filter converged; however, as our simulations will show, the alterative filters also show good performance in the transient.

The quadratic function (9) can be easily solved using the 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 (see Appendix). 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. Instead, we use two approximate solutions to (9) to derive two good variable step-size algorithms.

Replacing $b(k)/\mathcal{S}_e(k)$ by $\sqrt{a(k)/\mathcal{S}_e(k)}$ yields the VSS-NFLOMa algorithm

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

The variable step size is then formulated as

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

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

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

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

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

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

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

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

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

An alternative approximation to (14) is to replace $a(k)/\mathcal{S}_e(k)$ by $[b(k)/\mathcal{S}_e(k)]^2$

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

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

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

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

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

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-NFLOM, (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 $\|\mathbf{x}(k)\|^2$.

We show in the Appendix that the ratio $a(k)\mathcal{S}_e(k)/b^2(k)$ is very close to one for $p = 2$ and large filter length L, thus in this case the VSS-NFLOMa is equivalent to the VSS-NFLOMb. However, for small L and $p < 2$, the two algorithms are significantly different, and Section 4 will show that VSS-NFLOMa performs better than VSS-NFLOMb for small p.

The proposed two VSS-NFLOM algorithms and their variants are summarized in Table 1, where δ and ϵ are small positive constants to avoid division by zero, and μ_0 is a step size multiplier often selected as $\mu_0 \leq 1$. The left column gives the VSS-NFLOMa algorithms and the right column shows VSS-NFLOMb.

3 Performance Analysis

The proposed VSS-NFLOM algorithms were evaluated through computer simulation 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

Table 1. The VSS-NFLOM algorithms

Parameters	$\lambda = 1 - 1/(\eta L)$, forgetting factor, with $2 \leq \eta \leq 10$ $\mathcal{S}_v$= variance of background noise and interference $\delta = CL\sigma_x^2$, C = constant, σ_x^2 = power of input signal $\epsilon > 0$, small constant to avoid division by zero μ_0= step size multiplier	
Initialization	$\hat{\mathbf{w}}(0) = \mathbf{0}$, $\hat{\mathcal{S}}_e(0) = 0$	
Init (original) Init. (variant)	original $\hat{a}(0) = 0$ $\hat{\mathcal{L}}_{2p-2}(e_0) = 0$, $\hat{\mathcal{P}}_{xx}(0) = 0$	$\hat{b}(0) = 0$ $\hat{\mathcal{L}}_p(e_0) = 0$, $\hat{\mathcal{S}}_{xx}(0) = 0$
Error	$e_k = y(k) - \hat{\mathbf{w}}^H(k-1)\mathbf{x}(k)$	
Estimates	$\hat{\mathcal{S}}_e(k) = \lambda\hat{\mathcal{S}}_e(k-1) + (1-\lambda)\lvert e_k\rvert^2$ $\gamma(k) = \left[1 - \sqrt{\frac{\mathcal{S}_v}{\hat{\mathcal{S}}_e(k)+\epsilon}}\right]$ $\mathcal{L}_x(k) = \mathbf{x}^H(k)\mathbf{x}(k)$	
Est. (original) Est. (variant)	$\hat{a}(k) = \lambda\hat{a}(k-1) + (1-\lambda)\lvert e_k\rvert^{2p-2}\mathcal{L}_x^2(k)$ $\hat{\mathcal{P}}_{xx}(k) = \lambda\hat{\mathcal{P}}_{xx}(k-1) + (1-\lambda)\mathcal{L}_x^2(k)$ $\hat{\mathcal{L}}_{2p-2}(e_k) = \lambda\hat{\mathcal{L}}_{2p-2}(e_{k-1}) + (1-\lambda)\lvert e_k\rvert^{2p-2}$	$\hat{b}(k) = \lambda\hat{b}(k-1) + (1-\lambda)\lvert e_k\rvert^p\mathcal{L}_x(k)$ $\hat{\mathcal{S}}_{xx}(k) = \lambda\hat{\mathcal{S}}_{xx}(k-1) + (1-\lambda)\mathcal{L}_x(k)$ $\hat{\mathcal{L}}_p(e_k) = \lambda\hat{\mathcal{L}}_p(e_{k-1}) + (1-\lambda)\lvert e_k\rvert^p$
Update (original) Update (variant)	$\alpha(k) = \sqrt{\frac{\hat{\mathcal{S}}_e(k)}{\hat{a}(k)+\delta^2}}$ $\alpha(k) = \sqrt{\frac{\hat{\mathcal{S}}_e(k)}{\hat{\mathcal{L}}_{2p-2}(e_k)\hat{\mathcal{P}}_{xx}(k)+\delta^2}}$	$\beta(k) = \frac{\hat{\mathcal{S}}_e(k)}{\hat{b}(k)+\delta}$ $\beta(k) = \frac{\hat{\mathcal{S}}_e(k)}{\hat{\mathcal{L}}_p(e_k)\hat{\mathcal{S}}_{xx}(k)+\delta}$
Step-size	$\mu(k) =$ $\begin{cases} \mu_0\alpha(k)\gamma(k), & \text{if } \mathcal{S}_v \leq \hat{\mathcal{S}}_e(k) \\ 0, & \text{otherwise} \end{cases}$	$\mu(k) =$ $\begin{cases} \mu_0\beta(k)\gamma(k), & \text{if } \mathcal{S}_v \leq \hat{\mathcal{S}}_e(k) \\ 0, & \text{otherwise} \end{cases}$
Weight vector	$\hat{\mathbf{w}}(k) = \hat{\mathbf{w}}(k-1) + \mu(k)\mathbf{x}(k)e^{<p-1>}(k)$	

complex-coefficient system, often found in communications and radar signal processing, such as channel estimation and adaptive beamforming [1]. 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 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. 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

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

where $N(k)$ was a white Gaussian random sequence with zero mean and variance σ_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, 31, 32].

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 [33]. 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 [33]

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

where $K_\nu(\cdot)$ is the modified Bessel function of the second kind and with order ν, ν is the shape parameter, and β the scale parameter. The average power of the compound K process is $\beta \cdot \nu$. With a fixed average power, the compound K distribution with a smaller ν 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 [33, 34, 35].

For all simulation studies in both the real and complex systems, the VSS-NFLOM 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-NFLOM was $\mu_0 = 1$ and the step size for FSS-NFLOM was $\mu = 0.1$. The multipliers were chosen so that the initial rate of convergence of all filters is the same.

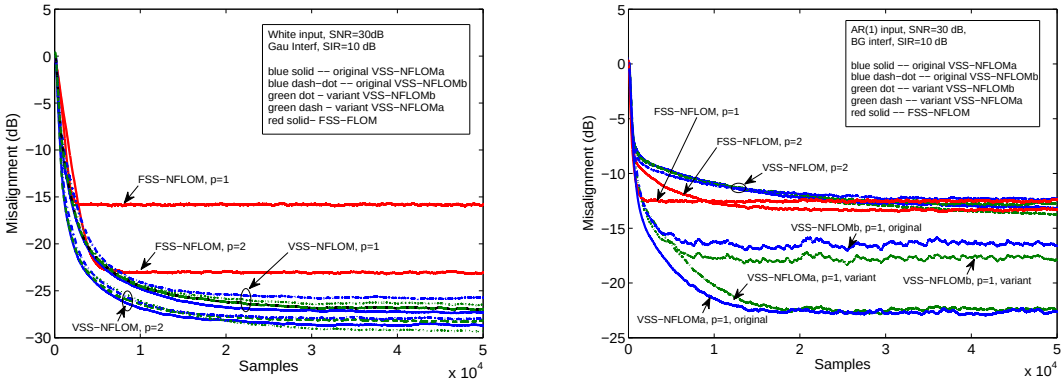
The convergence performance was evaluated by the normalized misalignment $\mathcal{M}(k)$ defined as [1]

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

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

3.1 Comparison of VSS-NFLOM algorithms and variants

The performance of the proposed two VSS-NFLOM algorithms and their variants were compared with the FSS-NFLOM algorithms using the example of $L = 128$ real-coefficient system. With white input and Gaussian interference, all VSS-NFLOM



(a) White input, Gaussian interference

(b) AR(1) input, BG interference

Figure 1. Misalignment of VSS-NFLOM algorithms in a real-coefficient system. SNR=30 dB. SIR=10 dB. $\delta = 0.04L\sigma_x^2, \eta = 4$. The two VSS-NFLOM algorithms had very similar performance as their variants for all orders of p.

algorithms performed better than the FSS-NFLOM, as shown in Fig. 3.1(a). The two VSS-NFLOM 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-NFLOM algorithms and their variants performed similarly, as shown in Fig. 3.1(b). Their convergence rate was slightly inferior to the FSS-NFLOM, but the steady-state errors were slightly lower than that of the FSS-NFLOM. For $p = 1$, the VSS-NFLOMa performed better than the VSS-NFLOMb with more than 5 dB lower steady-state error. Both VSS-NFLOM algorithms performed better than the FSS-NFLOM. The variants of the VSS-NFLOMb algorithm exhibited 2 dB lower steady-state error than the original VSS-NFLOMb algorithm; while the original VSS-NFLOMa converged slightly faster than its variant but achieved the same steady-state error.

In the following subsections, we only present detailed performance evaluation for the original VSS-NFLOM algorithms and omit those of the variants due to their small differences.

3.2 Convergence of Real-Coefficient Systems

The misalignment of the VSS-NFLOM algorithms for a real-coefficient system was compared with the FSS-NFLOM 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-NFLOM algorithms achieved faster convergence and lower steady-state error than the FSS-NFLOM for all orders of p, as shown in Figs. 3.2(a), 3.2(c), and 3.2(e). For AR(1) input and Gaussian interference, the VSS-NFLOM algorithms exhibited similar initial convergence but slightly slower secondary convergence than the FSS-NFLOM. On the other hand, they converged to much lower steady state error than the FSS-NFLOM. Both VSS-NFLOM algorithms behaved similarly in their misalignment curves, and the same was observed in their excess MSE curves (not shown). The VSS-NFLOM algorithms achieved similar steady-state error and convergence rate for all p. In contrast, the FSS-NFLOM 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-NFLOM algorithms has low importance to the performance in Gaussian interference.

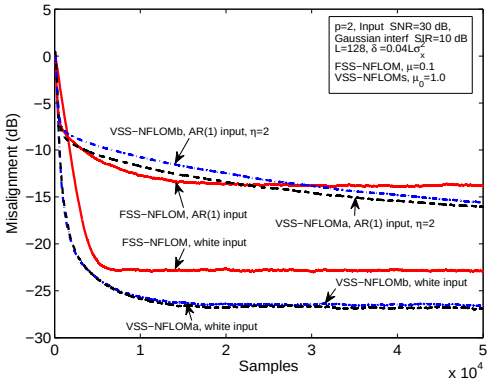
In the BG interference cases as shown in Figs. 3.2(b), 3.2(d), and 3.2(f), the VSS-NFLOM algorithms converged faster and achieved lower steady-state error than the corresponding FSS-NFLOM for all orders when the input was white. The two VSS-NFLOM algorithms behaved similarly when p is large, but quite differently when $p = 1$, with the VSS-NFLOMa achieving lower steady-state error than the VSS-NFLOMb. With the AR(1) input, the two VSS-NFLOM algorithms with a smaller

order performed much better than their corresponding FSS-NFLOM, while the VSS-NFLOM algorithms with a larger order performed comparably to the corresponding FSS-NFLOM. Among all algorithms, the VSS-NFLOMa with $p = 1$ had the best performance in all signal and interference scenarios achieving both fast convergence and low steady state error.

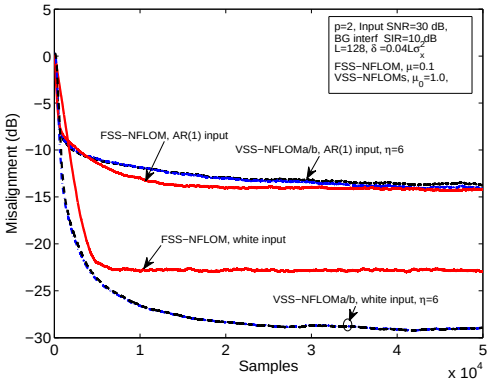
Comparing the two VSS-NFLOM algorithms, the VSS-NFLOMa and VSS-NFLOMb have similar behavior when p is large, but different behavior when $p \to 1$, especially in impulsive interference. The VSS-NFLOMa ($p = 1$) greatly outperformed the VSS-NFLOMb in terms of steady-state error in BG interference, as shown in Fig. 3.2(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-NFLOMa and VSS-NFLOMb. When $p \to 1$, or L is small, or interference is impulsive, the ratio drops dramatically and the VSS-NFLOMa is shown to perform better than the VSS-NFLOMb.

3.3 Convergence of Complex-Coefficient Systems

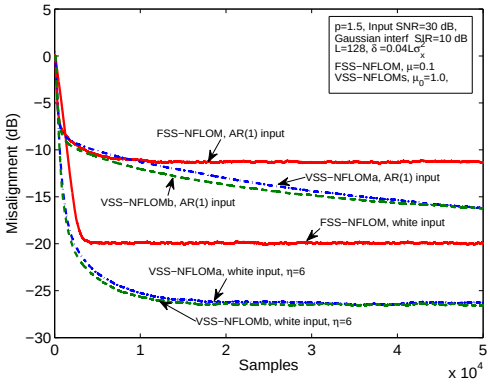
The performance of the VSS-NFLOM algorithm for the complex system was also compared with the FSS-NFLOM algorithm, as shown in Fig. 3. The complex-coefficient system was generated as a sum of sinusoids [36] for each of the 128 taps. Compound K interference with shape parameter of $\nu = 0.7$ was generated using [37]. The misalignment of the VSS-NFLOM algorithms performed similarly for all orders, achieving faster convergence and smaller steady-state error than the corresponding FSS-NFLOM algorithm. We also verified that the excess mean square error (MSE) curves (not shown here) of the VSS-NFLOM algorithms exhibited similar performance gains over the FSS-NFLOM.



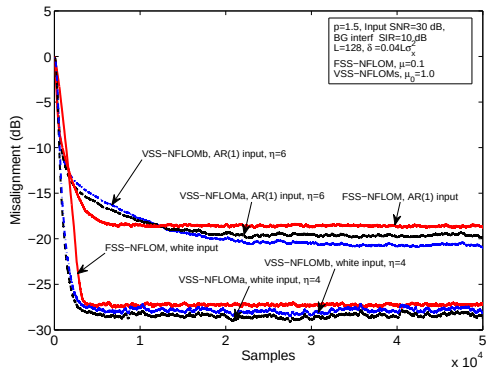
(a) p=2, Gaussian Interference



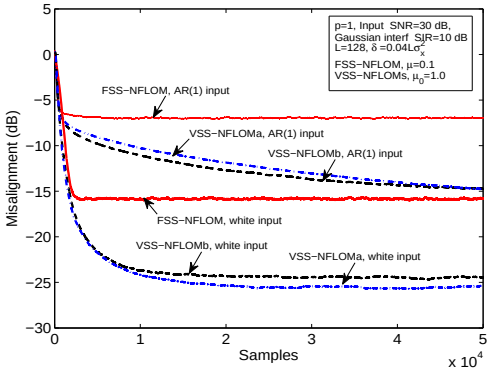
(b) p=2, BG Interference



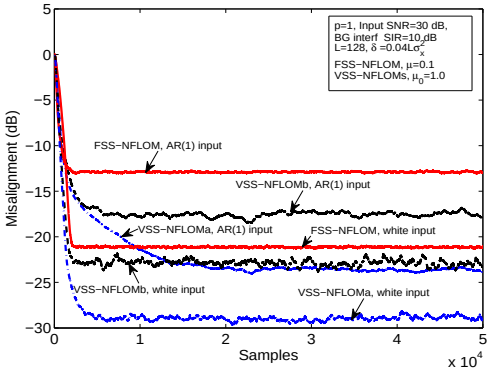
(c) p=1.5, Gaussian Interference



(d) p=1.5, BG Interference

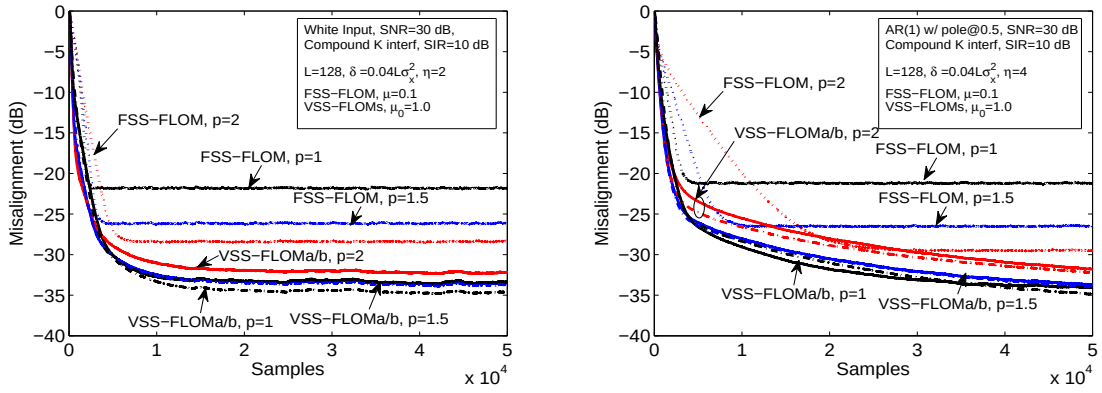


(e) p=1, Gaussian Interference



(f) p=1, BG Interference

Figure 2. Misalignment of the VSS-NFLOM and FSS-NFLOM algorithms for a real system with $L = 128$ taps. Solid lines – FSS-NFLOM, Dash-dotted lines – VSS-NFLOMa, Dashed lines –VSS-NFLOMb. All VSS-NFLOM algorithms performed better than FSS-NFLOM, and VSS-NFLOMa performed significantly better than VSS-NFLOMb for $p = 1$.



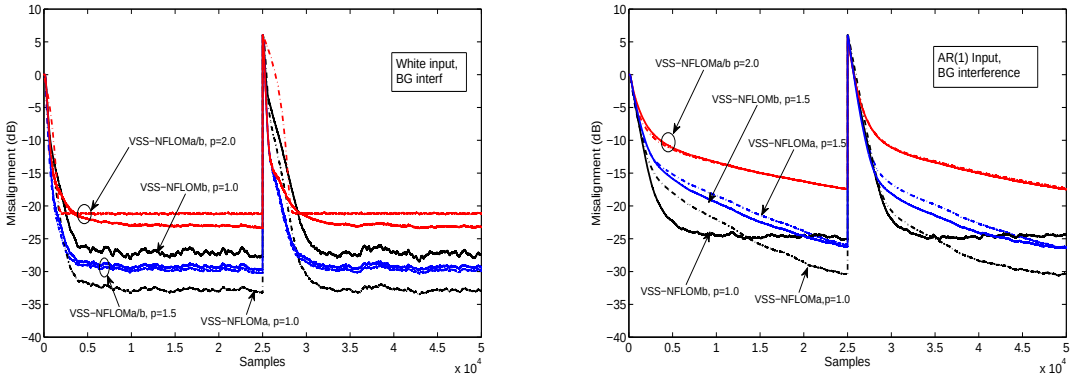
(a) White input

(b) AR(1) input

Figure 3. Misalignment of the VSS-NFLOM and FSS-NFLOM 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-NFLOM, Solid lines – VSS-NFLOMa, Dash-dotted lines –VSS-NFLOMb.

3.4 Tracking Performance

The tracking performance of the VSS-NFLOM algorithms was evaluated with a real-coefficient system. The ideal system impulse response was modified at iteration 2.5×10^4 by multiplying the weight coefficients by -1. The tracking performance of both VSS-NFLOM algorithms was compared in BG interference with white or colored inputs, as shown in Fig 4. Both VSS-NFLOM 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-NFLOM algorithm also behaved similarly to the VSS-NFLOM algorithms.



(a) White input

(b) AR(1) input

Figure 4. Tracking performance of the VSS-NFLOM 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-NFLOMa. Solid lines – VSS-NFLOMb.

3.5 Discussion on Stability

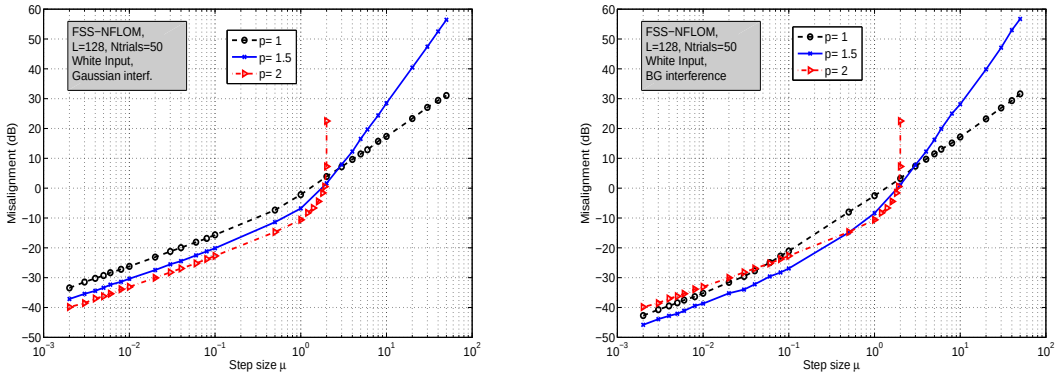
For normalized LMS, it is well known that the step size multiplier has to satisfy $0 < \mu_0 < 2$ to ensure stability [1]. This applies to the selection of variable step sizes in the VSS-NFLOM algorithms with $p = 2$. The variable step size selected in (18) and (23) with $\mu_0 \leq 1$ ensures that the two VSS-NFLOM algorithms satisfy the stability condition when $p = 2$, thus ensuring the stability of the VSS-NFLOM 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 FLOM algorithm becomes a member of the L_1-norm (or sign algorithm) family. For fixed step size sign algorithms, limited studies are devoted on the convergence analysis of the sign algorithm and its variants [27, 38, 39, 26, 40, 41, 42]. Attempts to finding a stability bound for the SA family have been reported in [27, 38, 39] 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 [26]. Instead, an interesting result for L_1-norm algorithms is that the sign algorithm is asymptotically bounded for any step-size greater than zero [26, 41, 42]. 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 [26, 41], which give guidelines for choosing the step size in practical applications. As a variant of the sign algorithm, the FSS-NFLOM 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-NFLOM algorithms with fractional order does not exist in the literature. Our simulation results show that the FSS-NFLOM 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-NFLOM 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-NFLOM with $p < 2$ performed better in BG interference when the step size was small. The FSS-NFLOM with $p = 1.5$ achieved smaller steady-state misalignment than that of $p = 1$ when step size $\mu < 3$, and the FSS-NFLOM 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-NFLOM to ensure small steady state errors.

It is worth noting that the guaranteed stability of the FSS-NFLOM with $p < 2$ and $\mu > 0$ does not mean that the VSS-NFLOM algorithms are also guaranteed to

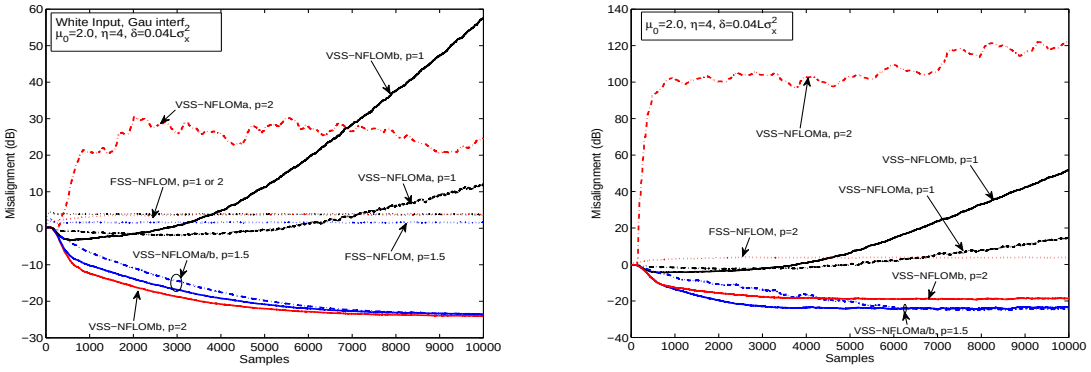


(a) Gaussian interference with SIR= 10 dB

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

Figure 5. Steady-state misalignment of fixed step-size FLOM algorithms as a function of the step-size μ. The input was a real-coefficient white Gaussian signal with SNR= 30 dB. The interference was Gaussian or BG.

be stable with the step size multiplier $\mu_0 > 0$ for $p < 2$. The stability analysis for the fractional-order VSS-NFLOM 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-NFLOM algorithms for $1 \leq p \leq 2$. However, $\mu_0 > 1$ may cause the VSS-NFLOM to diverge for all $1 \leq p \leq 2$, as illustrated in Fig. 6, where the step size multiplier used was $\mu_0 = 2$. The two VSS-NFLOM algorithms with $p = 1$ and the VSS-NFLOMa with $p = 2$ diverged in both white and AR(1) input scenarios with BG interference.



(a) White input

(b) AR(1) input

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

4 Conclusions

Two variable step-size normalized fractionally lower-order moment (VSS-NFLOM) algorithms have been proposed for system identification applications, which 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 VSS-NFLOM algorithm. The proposed VSS-NFLOM 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-NFLOM algorithms combine the benefits of variable step sizes with the robustness of the NFLOM algorithms against impulsive interference, thus achieving better tradeoff between fast convergence and small steady-state error

than the FSS-NFLOM and the VSS-NLMS [11]. The proposed VSS-NFLOM 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 $0 < \mu_0 \leq 1$.

5 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 σ_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 $R_{xx}(k) = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^*$, 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

$$\mathcal{S}_{xx}(k) = E[\mathcal{Z}] = \sum_{l=1}^{L} \sigma(l) E[|h(l)|^2] = \sum_{l=1}^{L} \sigma(l); \tag{28}$$

$$\begin{aligned} \mathcal{P}_{xx}(k) &= E[|\mathcal{Z}|^2] = E\left[\sum_{l_1=1}^{L} \sigma(l_1)|h(l_1)|^2 \sum_{l_2=1}^{L} \sigma(l_2)|h(l_2)|^2\right] \\ &= \sum_{l=1}^{L} \sigma^2(l) E\left[|h(l)|^4\right] + 2\sum_{l_1=1}^{L-1} \sum_{l_2=l_1+1}^{L} \sigma(l_1)\sigma(l_2) E\left[|h(l_2)|^2|h(l_2)|^2\right] \\ &= 3\sum_{l=1}^{L} \sigma^2(l) + 2\sum_{l_1=1}^{L-1} \sum_{l_2=l_1+1}^{L} \sigma(l_1)\sigma(l_2) \end{aligned} \tag{29}$$

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

$$e_k = [\mathbf{w}_o - \hat{\mathbf{w}}(k-1)]^H \mathbf{x}(k) + v(k). \tag{30}$$

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]

$$E[|\mathcal{N}(0, \sigma_e^2)|^p] = \frac{2^{p/2}\Gamma(\frac{p+1}{2})}{\sqrt{\pi}} \sigma_e^p, \quad p > 0, \tag{31}$$

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

Using (28), (29), and (31), we have for Gaussian inputs and Gaussian interference

$$\frac{b(k)}{\mathcal{S}_e(k)} = \frac{\mathcal{L}_p(e_k)}{\mathcal{S}_e(k)}\mathcal{S}_{xx}(k) = \frac{2^{p/2}\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}\sigma_e^{p-2}\sum_{l=1}^{L}\sigma(l) \tag{32}$$

$$\frac{a(k)}{\mathcal{S}_e(k)} = \frac{\mathcal{L}_{2p-2}(e_k)}{\mathcal{S}_e(k)}\mathcal{P}_{xx}(k) \tag{33}$$

$$= \frac{2^{(2p-2)/2}\Gamma(\frac{2p-1}{2})}{\sqrt{\pi}}\sigma_e^{2p-4}\left[\sum_{l=1}^{L}3\sigma^2(l) + 2\sum_{l_1=1}^{L-1}\sum_{l_2=l_1+1}^{L}\sigma(l_1)\sigma(l_2)\right], \quad p > 1.$$

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. 7, 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. 8. 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. 5.8(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-NFLOM algorithms are similar for p large, but they are quite different for p small.

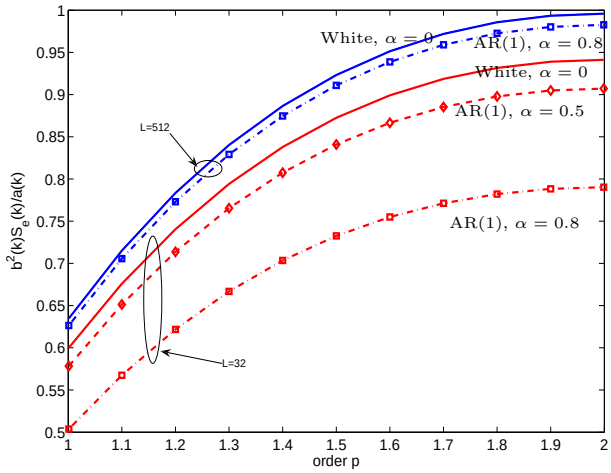
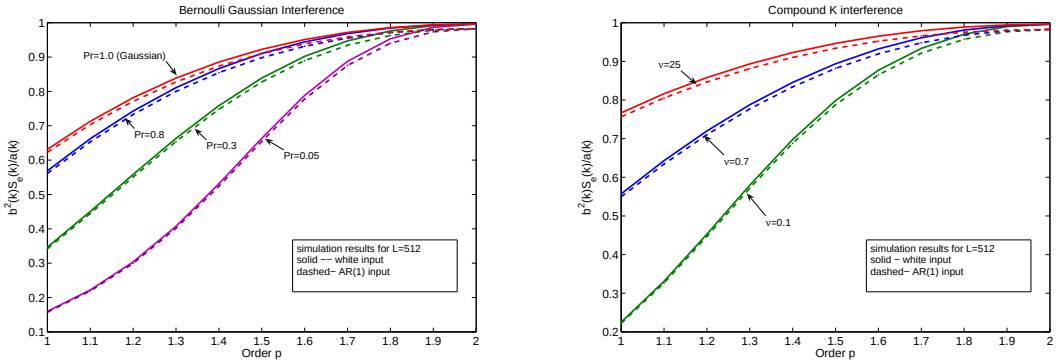


Figure 7. The ratio of $b^2(k)\mathcal{S}_e(k)/a(k)$ for white or AR(1) inputs with Gaussian interference. Results were computed based on (32) and (34). Parameter L was the number of coefficient taps and the AR(1) input was assumed to be a filtered white Gaussian by an IIR filter with a pole at α.



(a) BG Interference, simulation

(b) Compound K Interference, simulation

Figure 8. 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. 7.

6 References

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

[2] B. Widrow, Jr. J.R. Glover, J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Zeidler, Jr. Eugene Dong, and R.C. 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*, John Wiley, Inc., NY, 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*, vol.3, pp. 1441 -1446, Fort Lauderdale, FL, USA, Sep. 1996.

[8] R. Sharma, W.A. Sethares, and J.A. 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.R. 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.A. 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] Hyun-Chool Shin and A.H. 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 α-stable random processes," *IEEE Trans. Circuits Syst. II*, vol. 46, no. 2, pp. 198–202, Feb. 1999.

[20] 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.

[21] 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.

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

[23] 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.

[24] Magno T. M. Silva and Vítor H. Nascimento, "Improving the tracking capability of adaptive filters via convex combination," vol. 56, no. 7, pp. 3137–3149, July 2008.

[25] Jernimo Arenas-Garca, Anbal R. Figueiras-Vidal, and Ali H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," vol. 54, no. 3, pp. 1078–1090, Mar. 2006.

[26] 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.

[27] V. Mathews and S.H. 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.

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

[29] Y.R. Zheng, T.G. Shao, and E. P. Blasch, "A fast-converging space-time adaptive processing algorithm for non-Gaussian clutter suppression," *Elsevier J. Digital Signal Processing*, vol. in press, Nov. 2010.

[30] A. Mader, H. Puder, and G.H. Schmidt, "Step-size control for acoustic echo cancellation filters — an overview," *Signal Processing*, vol. 80, pp. 1697–1719, Sep. 2000.

[31] 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.

[32] 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.

[33] 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.

[34] W.B. Yang and T.C. 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.

[35] 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.

[36] 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.

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

[38] 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.

[39] 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.

[40] 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.

[41] 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.

[42] Han-Fu 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.

IV. A VARIABLE STEP-SIZE LMP ALGORITHM FOR HEAVY-TAILED INTERFERENCE SUPPRESSION IN PHASED ARRAY RADAR

Yahong Rosa Zheng, Tiange Shao

Abstract—A new variable step-size Least Mean p-norm (VSS-LMP) algorithm is proposed for phased array radar application with space-time adaptive processing to combat heavy-tailed non-Gaussian clutters. The algorithms automatically change the step size according to the estimated p-th and $(2p-2)$-th moments of the error, where $1 \leq p \leq 2$. The algorithm is evaluated via a space-slow-time STAP example and the excess Mean Square Error (MSE) and misadjustment results show that the proposed VSS-LMP converges fast and reaches lower steady-state error than the fixed step-size LMP. It also provides a better compromise between convergence speed and low steady state error than existing VSS Least Mean Square (LMS) algorithms in both Gaussian and Compound K clutter environments.

1 Introduction

In Radar and Sonar applications, adaptive filters are commonly used for space-time adaptive processing (STAP) [1,2] which combines spatial beamforming with temporal Doppler filtering, thus provides significant performance gain in clutter suppression and target detection. Special challenges are encountered in STAP applications, including non-Gaussian, heavy-tailed clutter distribution, high computational complexity of large array size, and slow convergence of adaptive algorithms. Especially, heavy-tailed clutters, often modeled as the compound-K distribution [3], degrade the

conventional STAP and adaptive algorithms significantly, resulting in low signal-to-interference-noise-ratio (SINR) at filter output and high false alarm rate of target detectors [2, 4, 5, 6, 7].

Conventionally, optimal adaptive filters are designed using the minimum mean square error (MMSE) criterion which leads to simple, closed-form solutions requiring the inverse of the covariance matrix of the array input signal. In practice, the Sample Matrix Inversion (SMI) method is often employed [4] and iterative adaptation algorithms, such as the least mean square (LMS) algorithm, can provide an alternative, low-complexity, real-time solution. The advantage of the LMS algorithms are its simplicity, low steady-state error, and fast tracking property. However, its major drawback is its slow convergence [8] and degraded robustness against non-Gaussian interference [5]. Over the past two decades, many variants of LMS have been proposed to overcome these problems, including the variable step size approach [9, 10], the affine projection algorithms (APA) [10], and the higher or lower order statistics method [5]. The third approach yields many robust algorithms with improved convergence and robustness against non-Gaussian interference, including the normalized sign algorithm (NSA) or Least Absolute Deviation (LAD) [8], least mean p-norm (LMP) or fractional lower-order moment/statistic (FLOM or FLOS) algorithm [5, 7], and least mean fourth-moment (LMF) algorithm. However, most NSA and LMP algorithms use a fixed step size for adaptation and the VSS approach has not been fully investigated for LMP algorithms.

In this paper, a new variable step-size Least Mean p-norm (VSS-LMP) algorithm is proposed to combat heavy-tailed non-Gaussian clutters. The algorithms automatically change the step size according to the estimated p-th and $(2p-2)$-th moments of the error, where $1 \leq p \leq 2$. As its two special cases, the proposed VSS-LMP algorithm becomes a new VSS-SNA and a new VSS-LMS when $p = 1$ and

$p = 2$, respectively. The algorithm is evaluated via a space-slow-time STAP example and the excess Mean Square Error (MSE) and misadjustment results show that the proposed VSS-LMP converges fast and reaches lower steady-state error than the fixed step-size LMP. It also provides a better compromise between convergence speed and low steady state error than existing VSS-LMS algorithms in both Gaussian and impulsive clutter environments.

2 STAP and NLMP Algorithm

Consider an arbitrary radar array antenna consisting of M elements with the m-th element located at $\mathbf{\Theta}_m$ in a spherical coordinate system. Coherent bursts of K pulses are transmitted at a constant pulse repetition frequency (PRF) $f_r = 1/T_r$, where T_r is the pulse repetition interval (PRI). Radar returns are collected over a coherent processing interval (CPI) of length KT_r. Within each PRI, there are N time (range) samples collected to cover the range interval. This multidimensional data set can be visualized as a $M \times K \times N$ cube of complex samples [2]. For STAP performed in the space-slow-time domain, let $L = M \times K$, then the concatenated space-time sample vector is an $L \times 1$ column vector $\mathbf{u}(k)$ with k being the time index.

The radar return vector $\mathbf{u}(k)$ is a mixture of the target echo ($\mathbf{u}_s$) with the uncorrelated jammer ($\mathbf{u}_J$), uncorrelated clutters ($\mathbf{u}_c$), and background noise ($\mathbf{u}_n$):

$$\mathbf{u}(k) \quad = \quad \mathbf{u}_s(k) + \mathbf{u}_J(k) + \mathbf{u}_c(k) + \mathbf{u}_n(k), \tag{1}$$

where

$$\begin{aligned}
\mathbf{u}_s(k) &= S(k)\mathbf{b}(\omega_s) \otimes \mathbf{a}(\mathbf{\Theta}_s), \\
\mathbf{u}_J(k) &= \sum_{i=1}^{N_J} S_{J_i}\mathbf{b}_{J_i} \otimes \mathbf{a}(\mathbf{\Theta}_{J_i}) \\
\mathbf{u}_c(k) &= \sum_{i=1}^{N_c} S_{c_i}\mathbf{b}(\omega_{c_i}) \otimes \mathbf{a}(\mathbf{\Theta}_{c_i}).
\end{aligned}$$

where the point target $S(k)$ is located at $\mathbf{\Theta}_s$ and with a Doppler frequency f_0. The operator $\otimes$ denotes the Kronecker matrix product, the temporal steering vector $\mathbf{b}(\omega_s) = [1, \cdots, e^{-jk\omega_s}, \cdots, e^{-j(K-1)\omega_s}]^H$ with the normalized Doppler frequency $\omega_s = 2\pi f_s/f_r$. The spatial steering vector $\mathbf{a}(\mathbf{\Theta}_s) = \left[1, e^{-j\Omega(\tau_{2s}-\tau_{1s})}, \cdots, e^{-j\Omega(\tau_{Ms}-\tau_{1s})}\right]^H$ for location $\mathbf{\Theta}_s$, where $\tau_{ms} = |\mathbf{\Theta}_m - \mathbf{\Theta}_s|/c$ is the propagation delay from the signal source to the m-th array element, c is the wave propagation speed, and Ω the operating frequency. The N_J jammers S_{J_i} are at locations $\mathbf{\Theta}_{J_i}$ with gain vectors $\mathbf{b}_{J_i}$. The N_c independent clutter patches are uniformly distributed in a circular ring/sphere around the radar platform [2] with the i-th patch at $\mathbf{\Theta}_{c_i}$ and having a Doppler frequency ω_{c_i} proportional to its angular location. The receiver noise $\mathbf{u}_n$ appears as a uniform noise floor on the angle-Doppler plane.

The STAP system consists of a tapped-delay-line attached to each array element. Let $\mathbf{w}$ be the concatenated weight vector of the STAP processor, then the output of the STAP $y(k)$ can be expressed in a matrix form as $y(k) = \mathbf{w}^H\mathbf{u}(k)$. where superscript $()^H$ denotes conjugate transpose. The least mean p-norm adaptive method is to minimize the p-th order statistic of the output with the weight vector

$$\min_{\mathbf{w}} E\left\{|y(k)|^p\right\}, \quad \text{subject to} \quad \mathbf{C}^H\mathbf{w} = \mathbf{h}, \tag{2}$$

where $E\{\cdot\}$ is the expectation operator and $|\cdot|$ is the absolute value operator. The matrix $\mathbf{C}$ is a set of linear constraints and $\mathbf{h}$ is the desired response vector. For

example, a simple point constraint [1, Chapter 5] may be chosen as $\mathbf{C} = \mathbf{b}(\omega_s) \otimes \mathbf{a}(\mathbf{\Theta}_s)$ and $\mathbf{h} = 1$, which enforces a unit gain response at the target location $\mathbf{\Theta}_s$ and its Doppler frequency f_s.

For the special cases of $p = 2$ and $p = 1$, the constrained minimization problem (2) reduces to the Minimum Mean Square Error (MMSE) and Minimum Mean Absolute Error (MMAE) solutions. For a fractional order p, however, there is no closed-form solution and an iterative least mean p-norm (LMP) adaptation algorithm has been derived as [6]. Here we present a Generalized Sidelobe Canceller (GSC) implementation of the LMP. The weight vector is first decomposed into two orthogonal components:

$$\mathbf{w}(k) = \mathbf{w}_q - \mathbf{C}_a \mathbf{w}_a(k), \tag{3}$$

where $\mathbf{w}_a(k)$ is the unconstrained adaptive weight vector, the signal blocking matrix $\mathbf{C}_a$ is orthogonal to $\mathbf{C}$ satisfying $\mathbf{C}^H \mathbf{C}_a = \mathbf{0}$. The transient vector $\mathbf{w}_q$ is a fixed beamformer determined by

$$\mathbf{w}_q = \mathbf{C}(\mathbf{C}^H \mathbf{C})^{-1} \mathbf{h}, \tag{4}$$

The LMP adaptation algorithm for $\mathbf{w}_a(k)$ estimation is then

$$\hat{\mathbf{w}}_a(k+1) = \hat{\mathbf{w}}_a(k) + \mu y^{<p-1>}(k) \mathbf{x}(k), \tag{5}$$

where μ is the step size, the operator $z^{<p>} = z^* |z|^{p-1}$ with the superscript * denoting the conjugate, and $\mathbf{x}(k) = \mathbf{C}_a^H \mathbf{u}(k)$.

3 The New VSS-NLMP Algorithms for STAP

To derive the variable step size LMP algorithm, we first decompose the STAP output into two components

$$y(k) = d(k) - \mathbf{w}_a^H \mathbf{x}(k), \tag{6}$$

where $d(k) = \mathbf{w}_q^H \mathbf{u}(k)$. For the k-th iteration, we define the *a priori* and *a posteriori* errors as

$$\begin{aligned} e(k) &= d(k) - \hat{\mathbf{w}}_a^H(k-1)\mathbf{x}(k) && (7) \\ \varepsilon(k) &= d(k) - \hat{\mathbf{w}}_a^H(k)\mathbf{x}(k) && (8) \end{aligned}$$

Substituting (7) into (8) yields

$$\begin{aligned} \varepsilon(k) &= e(k) + \left[\hat{\mathbf{w}}_a(k-1) - \hat{\mathbf{w}}_a(k)\right]^H \mathbf{x}(k) \\ &= e(k) - \mu(k)[e^{<p-1>}(k)]^* \mathbf{x}^H(k)\mathbf{x}(k). \end{aligned} \tag{9}$$

The second equality makes use of (5) and $y(k) = e(k)$. For a small $\mu \ll 1$, the p-th moment of the posteriori error can be approximated by

$$\begin{aligned} & E\left[|\varepsilon(k)|^p\right] = E\left[|e(k)|^p |1 - \mu |e(k)|^{p-2} \mathbf{x}^H(k)\mathbf{x}(k)|^p\right] \\ & \approx E\left[|e(k)|^p\right] - p\mu(k) E\left[|e(k)|^{2p-2} \mathbf{x}^H(k)\mathbf{x}(k)|^p\right] \end{aligned} \tag{10}$$

Assume that the error $e(k)$ and the input signal $\mathbf{x}(k)$ are uncorrelated and define $S_{xx}(k) = E\left[\mathbf{x}^H(k)\mathbf{x}(k)\right]$. For STAP applications, the desired output is the signal

target $S(k)$. Hence, we choose the step size such that

$$E\left[|\varepsilon(k)|^p\right] = E\left[|S(k)|^p\right] \triangleq \mathcal{L}_p(S) \tag{11}$$

and the solution to $\mu(k)$ is then

$$\mu(k) = \mu_0 \frac{\mathcal{L}_p(e_k) - \mathcal{L}_p(S)}{p\mathcal{L}_q(e_k)S_{xx}(k)}, \text{ if } \mathcal{L}_p(e_k) \geq \mathcal{L}_p(S), \tag{12}$$

where μ_0 is a small constant to ensure that $\mu(k)$ is small, $\mathcal{L}_p(e_k) \triangleq E\left[|e(k)|^p\right]$, and $\mathcal{L}_q(e_k) \triangleq E\left[|e(k)|^{<2p-2>}\right]$. In practice, the average input power $S_{xx}(k)$ can be replaced by $\mathbf{x}^H(k)\mathbf{x}(k)+\delta$ with δ being a regularization parameter. The p-th and q-th moments of the error may be estimated by time averaging

$$\hat{\mathcal{L}}_p(e_k) = \lambda\hat{\mathcal{L}}_p(e_{k-1}) + (1-\lambda)|e(k)|^p \tag{13}$$

$$\hat{\mathcal{L}}_q(e_k) = \lambda\hat{\mathcal{L}}_q(e_{k-1}) + (1-\lambda)|e(k)|^{2p-2} \tag{14}$$

where λ is the forgetting factor. This results in a normalized LMP algorithm for the unconstrained weight vector

$$\hat{\mathbf{w}}_a(k+1) = \hat{\mathbf{w}}_a(k) + \mu_{nlmp}(k)\frac{e^{<p-1>}(k)\mathbf{x}(k)}{\mathbf{x}^H(k)\mathbf{x}(k)+\delta}, \tag{15}$$

with a variable step-size

$$\mu_{nlmp}(k) = \mu_0 \frac{\hat{\mathcal{L}}_p(e_k) - \hat{\mathcal{L}}_p(S)}{p\hat{\mathcal{L}}_q(e_k)}, \tag{16}$$

if $\hat{\mathcal{L}}_p(e_k) \geq \hat{\mathcal{L}}_p(S)$. Otherwise, the step size is set to zero.

4 Simulation Results

The proposed VSS-NLMP algorithm was evaluated via computer simulation of a linear phased array example. The array consisted of $M = 10$ equally spaced elements at half wavelength of the operation frequency. The coherent pulse interval (CPI) was $K = 7$. The target signal had a power of 0 dB with respect to the background noise and was located at 20° angle of arrival (AoA) with a normalized Doppler frequency of 0.25. The noises were independent white Gaussian among antenna elements and CPI taps. Two wideband jammers presented at AoA of $-20°$ and $+50°$, respectively. Each jammer had a full Doppler spectrum and with 15 dB power. In addition, many clutters impinged on the array from different AoAs which were uniformly distributed between $-180°$ and 180°. The total clutter power was 30 dB above the background noise. Three types of clutter distribution were used: a complex Gaussian (Rayleigh envelop) and two compound K distributions with shape parameters $\nu = 2.0$ and $\nu = 0.7$, respectively. The complex Gaussian clutter was generated by sum of sinusoids (SoS) with proper clutter ridge/Doppler spectrum. The compound K clutters were generated by multiplying a correlated gamma texture with a complex Gaussian speckle [7]. The Matlab code from RadarWorks website [11] was used to generate the correlated gamma texture. The probability density function of the compound-K distribution is given by [3]

$$f_{_R}(r) = \frac{4}{\sqrt{\beta}\Gamma(\nu)} \left(\frac{r}{\sqrt{\beta}}\right)^{\nu} K_{\nu-1}\left(\frac{2r}{\sqrt{\beta}}\right), \quad r \geq 0 \tag{17}$$

where K_ν is the modified Bessel function of the second kind and with order ν. It exhibits heavier tail than complex Gaussian/Rayleigh distribution but is lighter than the alpha-stable interference. A smaller shape parameter ν indicates a more impulsive

interference and the Rayleigh distribution is its special case with $\nu = \infty$. Compound-K clutter with ν in the range of 0.3 to 5 are commonly encountered in practice [3].

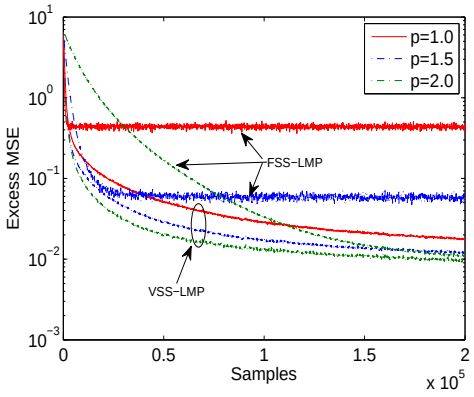
The convergence performances were evaluated by the excess MSE $J_{ex}(k)$ and the normalized misadjustment $\mathcal{M}(k)$ defined as [1]

$$J_{ex}(k) = E\left[|\hat{\mathbf{w}}^H(k)\mathbf{u}(k)|^2\right] - J_{min} \quad (18)$$

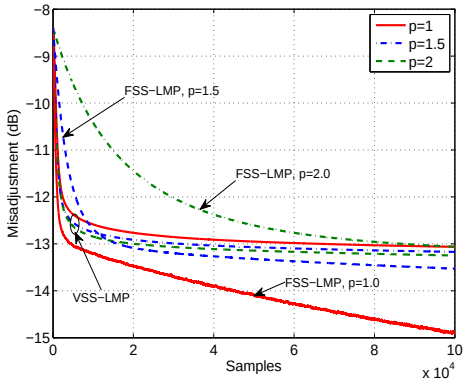
$$\mathcal{M}(k) = 20\log_{10}\frac{||\hat{\mathbf{w}}(k) - \mathbf{w}_{opt}||}{||\mathbf{w}_{opt}||} \quad (19)$$

where $J_{min} = E[|\mathbf{w}_{opt}^H\mathbf{u}(k)|^2]$. The ensemble average of 100 trails was used for $J_{ex}(k)$ and $\mathcal{M}(k)$ calculation. For all simulation studies, the regularization constant was set to $\delta = 0.4L*$SNR. The forgetting factor was chosen as $\lambda = 1-\eta/L$ with $L = MK = 70$ and $\eta = 0.1$ for all cases.

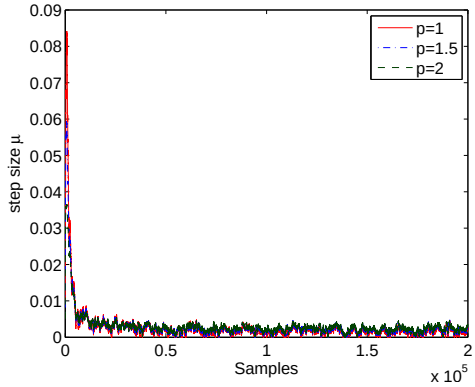
The convergence curves of the new VSS-LMP algorithm are plotted in Fig. 1 and Fig. 2 for Gaussian and K clutters, respectively. The excess MSE and misadjustment for order $p = 1, 1.5, 2.0$ are compared with the fixed step size (FSS) NLMP of $\mu = 0.002$. In Gaussian clutters, as shown in Fig. 4.1(a), the excess MSE curves of the FSS-NLMP varied widely with the $p = 2$ curve converges slow but to a low steady-state MSE, the $p = 1$ curve converged very fast but leveled off at a high MSE, and the $p = 1.5$ curve performed in between. In contrast, the VSS-LMP curves for all p had a similar fast initial convergence and then slowed down and converged to very low MSE levels. The $p = 2$ curve of the VSS-LMP achieved the same level of steady-state MSE as its FSS-NLMP counter part but with much faster convergence speed, and the $p = 1.5$ and $p = 1.0$ curves of the VSS-LMP converged to a lower level steady-state MSE than that of the corresponding FSS-NLMP curves. The improvement of the VSS-LMP algorithms over the FSS-NLMP is clearly demonstrated.



(a) Excess MSE



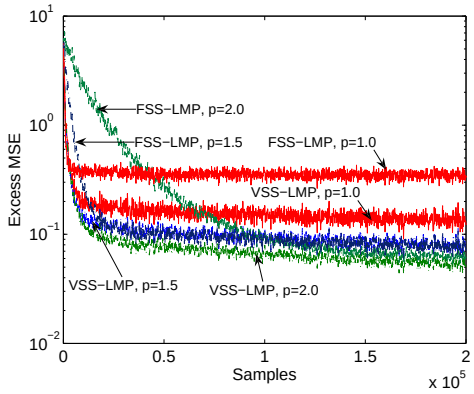
(b) Misadjustment



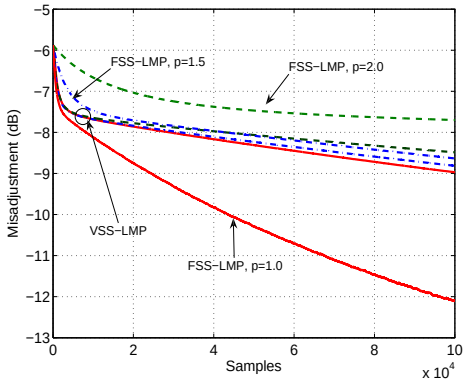
(c) Step size change of VSS-LMP

Figure 1. Performances of the VSS-LMP algorithm in Gaussian clutters. Each $J_{ex}(t)$ and $\mathcal{M}(t)$ curve was an ensemble average of 100 trials. Step size curves were of one trial. The FSS-NLMP used $\mu = 0.002$.

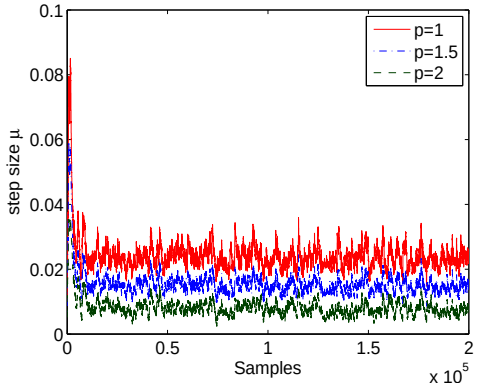
For misadjustment in Gaussian clutters, as shown in Fig. 4.1(b), the FSS-NLMP curves also varied widely: the $p = 2$ curve converged slow and leveled off high; the $p = 1$ curve converged fast and continued dropping, and the $p = 1.5$ curve performed in between. Although the misadjustment of the $p = 1$ curve of the FSS-NLMP seemed to perform the best from system identification standpoint, its excess MSE was the worst for STAP applications. This is due to the property that the $p = 1$



(a) Excess MSE



(b) Misadjustment



(c) Step size change of VSS-LMP

Figure 2. Performances of the VSS LMP algorithm in compound-K clutters with $\nu = 0.7$. Each $J_{ex}(t)$ and $\mathcal{M}(t)$ curve was an ensemble average of 100 trials. Step size curves were of one trial. The FSS-NLMP used $\mu = 0.002$.

NLMP algorithm is less sensitive to the changes of inputs and thus yielding larger excess MSE than the $p = 2$ FSS-NLMP. The misadjustment of the VSS-LMP for all p behaved similar to that of the $p = 1.5$ FSS-NLMP.

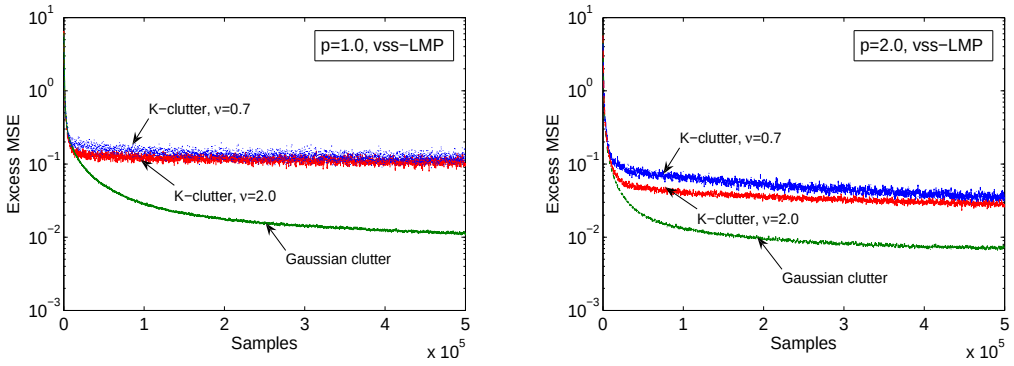
The step sizes of the VSS-LMP algorithm are shown in Fig. 4.1(c) for one trial in Gaussian clutters. The step size of the VSS-LMP behaved very similar with all three p cases. It started at a large μ at around 0.08 and dropped quickly to below

0.01 and stayed in the neighborhood of 0.002. This behavior matched their excess MSE curves.

For K clutters with $\nu = 0.7$, as shown in Fig. 2, the excess MSE and misadjustment of the FSS-NLMP and VSS-LMP exhibited similar behaviors as those in Gaussian clutters, but with much higher steady-state errors and much higher step sizes for all p values. The performance gains of the VSS-LMP over the FSS-NLMP were similar to those in Gaussian clutters. The robustness of the NLMP algorithms against impulsive clutter is demonstrated.

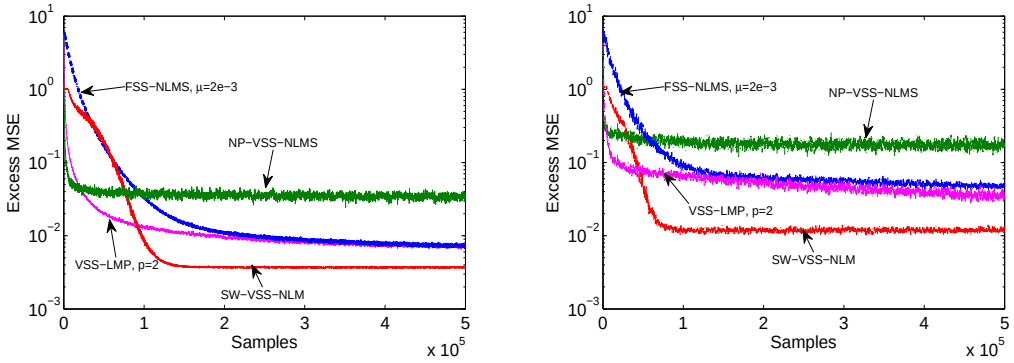
The excess MSE curves of the VSS-LMP algorithm are compared for different clutter environments in Fig. 3. Comparing the curves of $p = 1$ in Fig. 4.3(a) with those of $p = 2$ in Fig. 4.3(b), both VSS-LMP algorithms achieved the same low excess MSE in Gaussian clutters but the $p = 1$ VSS-LMP converged slightly slower than the $p = 2$ VSS-LMP. In the K clutters, the $p = 1$ VSS-LMP converged faster but leveled off at a higher excess MSE than the $p = 2$ VSS-LMP. The performance difference between two types of K clutters was small with the $p = 1$ VSS-LMP, while the $p = 2$ VSS-LMP was more sensitive to the degree of clutter impulsiveness. With the same order p, the initial convergence speed was similar, but the steady-state errors were different for different clutter types.

It is also interesting to note that the new VSS-LMP algorithm reduces to a VSS-LMS algorithm when $p = 2$. Therefore, we also compare the new $p = 2$ VSS-LMP with other two VSS-NLMS algorithms — the non-parametric (NP) VSS-NLMS [9] and the switched mode (SM) VSS-NLMS [12]. The excess MSE curves are shown in Fig. 4 for Gaussian and K ($\nu = 0.7$) clutters. The FSS-NLMS curve with $\mu = 0.002$ is also plotted for comparison. In both clutter environments, the NP-VSS-NLMS algorithm converged fast but leveled off at a high excess MSE, while the SM-VSS-NLMS converged slower and achieved a lower excess MSE. The new VSS-LMP algorithm converged slightly slower than the NP-VSS-NLMS and achieved



(a) p=1.0 (b) p=2.0

Figure 3. Convergence of the new VSS NLMS algorithm in different clutter environments. With the same order p, the initial convergence speed was similar, but the steady-state errors were different for different clutters.



(a) Gaussian Clutters (b) K-Clutters, $\nu = 0.7$

Figure 4. Convergence of the three VSS NLMS algorithms in comparison with the fixed step-size NLMS. The proposed new VSS-LMP with $p = 2$ becomes a VSS-NLMS algorithm and it provides a good compromise between fast convergence and low steady-state error.

a medium excess MSE as that of the FSS-NLMS ($\mu = 0.002$). It is clear that the new VSS-LMP provides a better compromise between fast convergence and low steady-state error than the two VSS-NLMS algorithms.

5 Conclusions

A new variable step-size Least Mean p-norm (VSS-LMP) algorithm has been proposed for phased array radar application with space-time adaptive processing to combat heavy-tailed non-Gaussian clutters. The algorithm automatically changes the step size according to the estimated p-th and $(2p-2)$-th moments of the error, where $1 \leq p \leq 2$. It has been evaluated via a space-slow-time STAP example in both Gaussian and compound K clutters. The excess Mean Square Error (MSE) and misadjustment results show that the proposed VSS-LMP converges fast and reaches lower steady-state error than the fixed step-size LMP.It also provides a better compromise between convergence speed and low steady state error than existing VSS Least Mean Square (LMS) algorithms in both Gaussian and impulsive clutter environments .

6 Acknowledgment

This work is supported by AFOSR under grant FA9550-07-1-0336.

7 References

[1] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Upper Saddle River, New Jersey, 07458., 4 edition, 2002.

[2] M. A. Richards, *Fundamentals of Radar Signal Processing*, McGraw-Hill,, New York, 2005.

[3] 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.

[4] M. C. Wicks, M. Rangaswamy, R. Adve, and T. B. Hale, "Space-time adaptive processing: A knowledge-based perspective for airborne radar," *IEEE Signal Processing Magazine*, vol. 23, no. 1, pp. 51–65, Jan. 2006.

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

[6] 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.

[7] Yahong Rosa Zheng, Genshe Chen, and Erik Blasch, "A normalized fractionally lower-order moment algorithm for space-time adaptive processing," Orlando, FL, Oct. 2007, IEEE Military Communications Conference.

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

[9] J. Benesty, H. Rey, L.R. Vega, and S. Tressens, "A nonparametric vss nlms algorithm," *IEEE SP Letters*, vol. 13, no. 10, pp. 581–584, Oct. 2006.

[10] H.-C. Shin, A.H. Sayed, and W.-J. Song, "Variabl step-size nlms and affine projection algorithms," *IEEE SP Letters*, vol. 11, no. 2, pp. 132–135, Feb. 2004.

[11] Glen Davidson, *Matlab and C Radar Toolbox*, The Japan Society for the Promotion of Science (JSPS), http://www.radarworks.com,2008.

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

V. A FAST-CONVERGING SPACE-TIME ADAPTIVE PROCESSING ALGORITHM FOR NON-GAUSSIAN CLUTTER SUPPRESSION

Yahong Rosa Zheng, Tiange Shao, and Erik Blasch

Abstract—The normalized fractionally-lower order moment (NFLOM) algorithm differs from the normalized least mean square (NLMS) algorithm in that it minimizes the lower order moment ($p < 2$) of the error rather than the variance ($p = 2$). This paper first evaluates the performances of the NFLOM for space-time adaptive processing in heavy-tailed compound K clutters in terms of the excess mean square error (MSE), misalignment, beampatterns, and output signal-to-interference-and-noise-ratio (SINR). The results show that the MSE curve of a small-order NFLOM exhibits faster convergence but higher steady-state error than a large-order NFLOM. Second, this paper proposes a new variable-order FLOM algorithm to dynamically change the order during adaptation, thus achieving both fast initial convergence and low steady-state error. The new algorithm is applied to STAP for Gaussian and non-Gaussian clutter suppression. The simulation results show that it achieves the best compromise between fast convergence and low steady-state error in both types of clutters.

1 Introduction

Space-time adaptive processing (STAP) refers to combined spatial beamforming and temporal filtering of radar/sonar returns in phased array systems. It uses multiple antenna elements followed by tapped-delay-lines to coherently process multiple pulses, thus providing superior ability to suppress jammers and clutters while preserving desired signal target [1]. Since its introduction, STAP has been rigorously researched and has been shown to provide significant performance gains in interference suppression and target detection [2]. Many STAP algorithms deal with common

scenarios where clutters and noises are complex Gaussian, which leads to mathematically tractable solutions [1, 2, 3]. However, recent studies and field measurements have found [4, 5, 6, 7, 8, 9, 10] that heavy-tailed non-Gaussian clutters often occur in backscatters from mountain tops, dense forest canopy, rough sea surfaces, and man-made concrete objects, etc. These radar clutters are spiky, impulsive in nature and can cause significant performance degradation in STAP and target detection.

Several statistical models have been used to describe the impulsive non-Gaussian clutter environment including the compound complex Gaussian [4], the generalized complex Gaussian [5], and the complex alpha-stable [10, 11]. The compound complex Gaussian model is often used in practice [4], where the clutter/noise process is modeled as the product of two random processes: $X = \sqrt{\tau} \cdot G$, with τ being the texture and G the speckle . If τ follows the Gamma distribution and G the complex Gaussian, then the envelop of X, defined as the magnitude $R = |X|$, will be the compound K distribution with the probability density function (pdf) given as [4, 5]

$$f_{_R}(r) = \frac{2}{\sigma\Gamma(\nu)} \left(\frac{r}{2\sigma}\right)^{\nu} K_{\nu-1}\left(\frac{r}{2\sigma}\right) \tag{1}$$

where ν is the shape parameter, $K_\nu()$ is the modified Bessel Function of the second kind of order ν, and $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ is the Gamma function. The pdf's of the compound K distribution with different shape parameters are plotted in Fig. 1, where all envelop distributions have a normalized second moment $E(R^2) = 2$, and the simulated compound K clutters are generated by the nonlinear memoryless transform (NMLT) method [8, 9]. The tails of the compound K pdfs are much higher than the Rayleigh distribution which is the envelope pdf of the complex Gaussian process.

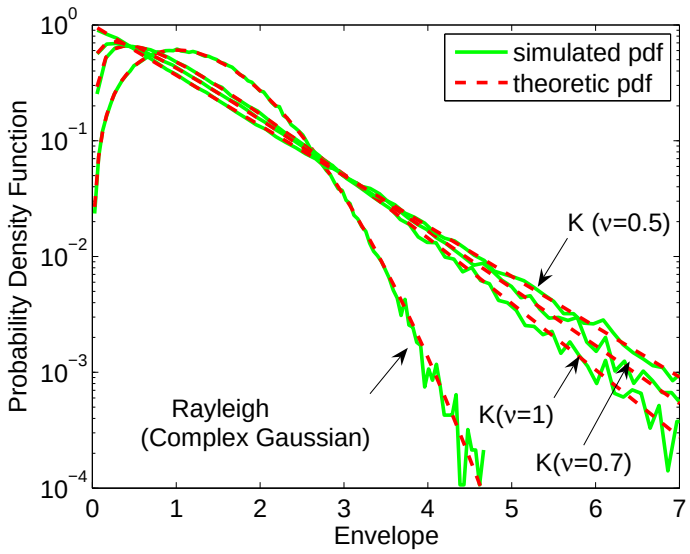


Figure 1. The envelop Probability Density Function (pdf) of the compound K clutters, in comparison to complex Gaussian clutters whose envelop is Rayleigh (a special case of compound K with $\nu = \infty$). The compound K distributions exhibit much heavier tails than Rayleigh clutter.

Besides the heavy-tailed distribution, non-Gaussian clutters also exhibit different auto-covariance properties. The Rayleigh distributed speckle component, modeling the short-term fluctuation of the scatters, has a very fast-decaying auto-covariance function (ACF) spanning only a few slow-time samples. The Gamma distributed texture component, on the other hand, represents modulations of the local power of the clutters and often has a slowly-decaying ACF. The ACF of the Gamma texture is commonly modeled as $R_G(t) = \exp(-t/\lambda)$, where λ is on the order of several hundred samples, as shown in Fig. 2 in comparison with the ACFs of the Rayleigh and compound K clutters.

For Gaussian clutter suppression, the Minimum Variance Distortionless Response (MVDR) method is commonly employed to minimize the output power with constraints at target location and Doppler frequency. In practice, the Sample Matrix Inversion (SMI) or Normalized Least Mean Square (NLMS) algorithm is often used

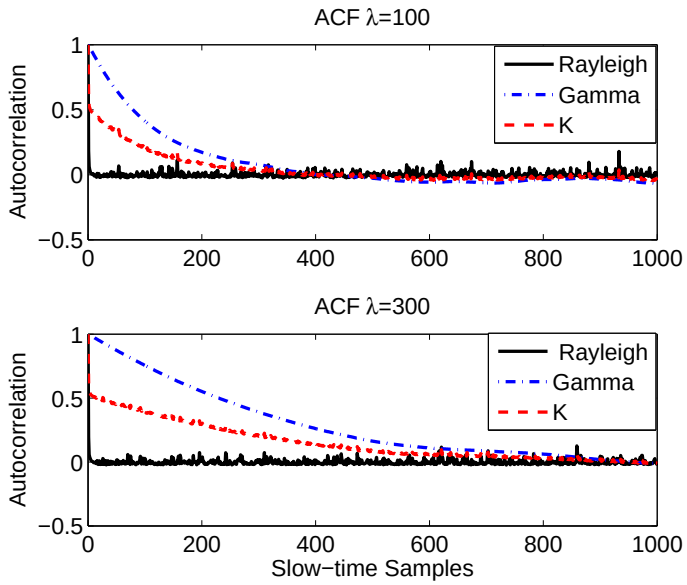


Figure 2. The Auto-Covariance Functions (ACF) of Rayleigh, Gamma texture, and compound K clutters. A larger λ indicates that the ACF decays slower.

for iterative adaptation. In contrast, a fractionally lower-order moments (FLOM) adaptive algorithm and its normalized version (NFLOM) have been proposed in [10] and [18] for suppressing heavy-tailed non-Gaussian clutters. The FLOM and NFLOM algorithms differ from the MVDR and NLMS schemes in that they minimize the p-th order moment ($0 < p \leq 2$) of the output signal rather than its variance ($p = 2$), thus reducing the detrimental effects of spiky clutter samples.

The FLOM algorithm, also called the Least Mean p-norm (LMP) algorithm in the adaptive filtering literature [13, 14], has been shown to be effective in many non-Gaussian interference environments. It has been applied to seismic system identification, synthetic aperture radar (SAR), magnetoencephalography, and medical ultrasonic imaging, etc., where Bernoulli-Gaussian models [15, 16] or symmetrical alpha-stable models [13, 14, 17] are often used for impulsive interference with an algebraic tail. It has been shown that the FLOM algorithm performs significantly better than the conventional LMS (2nd-order norm) algorithm. The effectiveness of

the FLOM algorithm in less heavy-tailed compound K models remains unclear. Furthermore, the majority of the existing results focus on the deviation of the adaptive weights from the system impulse response, which is of the primary concern in system identification type of applications [13,14,15,16]. The output Mean Square Error (MSE) and Signal-to-Interference-and-Noise-Ratio (SINR) are largely ignored, which are more important measures in STAP or other phased array applications [18,19,20].

In this paper, we first evaluate the NFLOM algorithm for STAP applications where compound K clutters are present simultaneously with high-power, wideband jammers. We measure the performances in terms of excess MSE, misalignment (of weights), beampatterns, and output SINR. We also investigate the effect of the autocovariance function (ACF) of compound K clutters on these performances. Our results show several new aspects of the NFLOM algorithm:

1. For the NFLOM with a smaller order p close to 1, the misalignment exhibits faster initial convergence as well as lower steady-state error than that of a larger-p NFLOM. This result is in agreement with the ones reported for system identification applications in heavy-tailed interference [21].

2. The excess MSE of a small-p NFLOM converges much faster than that of a larger order NFLOM, but the steady-state error is slightly higher too, resulting in slightly lower average SINR in the converged STAP output. This result is true for both Gaussian and non-Gaussian clutters.

3. The ACF of the compound K clutter has strong impact on the convergence of the large-p NFLOM algorithms. For misalignment, an higher ACF leads to faster initial convergence but quicker slowing down than a lower ACF, resulting in a higher steady-state error. The high ACF result has not been reported elsewhere, even for system identification applications. For the excess MSE, a lower ACF slows the convergence speed significantly than a higher ACF, and

the converged steady-state error is also higher, especially for large-p NFLOM algorithms.

4. The ACF of the compound K clutter has less impact on the performances of the NFLOM with $p = 1$ than on a larger-p NFLOM. With $p = 1$ (*i.e.* the normalized sign algorithm (NSA)), the steady-state error of the excess MSE remains almost the same for different ACF and different impulsiveness (shape parameters). This indicates strong robustness of the NSA algorithm against changing, impulsive clutters.

5. The beampatterns of the NFLOM and NLMS algorithms differs significantly only at time instants when very large, spiky clutter samples occur. The NLMS algorithm uses more degrees of freedom to track the impulsive clutter samples and leaks more power of jammers and noise to the output. The NFLOM ($p < 2$) algorithm places less emphasis on the spiky clutter components and achieve slightly better SINR. However, the spiky clutter components are suppressed less severely and may require a nonlinear preprocessor before matched filter detection [18].

Based on the observation that the NFLOM algorithm with a smaller order p converges faster but to a lower steady-state MSE than that of a large-order p, we propose a new variable-order (VO) FLOM algorithm to solve the conflicting goals of fast convergence and low steady-state error. The new VO-FLOM algorithm uses a small order p at the beginning of the adaptation and gradually increases the order to a large p to achieve both fast convergence and low steady-state MSE. The proposed algorithm is also evaluated in both complex Gaussian and compound K clutter scenarios. The results show that the new VO-FLOM algorithm outperforms the plain NFLOM and NLMS algorithms in both MSE and misalignment.

2 Conventional STAP and its NLMS algorithm

Consider an arbitrary radar array antenna consisting of M elements with the m-th element located at $\mathbf{\Theta}_m$ in a spherical coordinate system. Coherent bursts of K pulses are transmitted at a constant pulse repetition frequency (PRF) $f_r = 1/T_r$, where T_r is the pulse repetition interval (PRI). Radar returns are collected over a coherent processing interval (CPI) of length KT_r. Within each PRI, there are L time (range) samples collected to cover the range interval. This multidimensional data set can be visualized as a $M \times K \times L$ cube of complex samples [1]. For STAP performed in the space-slow-time domain, denote the received samples at range bin l as $u_{k,m}(t)$ with slow-time index $k = 1, 2, \cdots, K$, and array element index $m = 1, 2, \cdots, M$, and the sampling time index t. Let $N = M \times K$, then the $N \times 1$ concatenated space-time sample vector is

$$\mathbf{U}(t) = [\mathbf{u}_1^T(t), \cdots, \mathbf{u}_k^T(t), \cdots, \mathbf{u}_K^T(t)]^T, \tag{2}$$

$$\mathbf{u}_k(t) = [u_{k,1}(t), u_{k,2}(t), \cdots, u_{k,M}(t)]^T, \tag{3}$$

where the superscript $()^T$ denotes transpose.

The radar return vector $\mathbf{U}(t)$ is a mixture of the target echo ($\mathbf{U}_s$) with the uncorrelated jammer ($\mathbf{U}_J$), uncorrelated clutters ($\mathbf{U}_c$), and background noise ($\mathbf{U}_n$):

$$\mathbf{U}(t) = \mathbf{U}_s(t) + \mathbf{U}_J(t) + \mathbf{U}_c(t) + \mathbf{U}_n(t), \tag{4}$$

where

$$\begin{aligned}
\mathbf{U}_s(t) &= S(t)\mathbf{b}(\omega_s) \otimes \mathbf{a}(\mathbf{\Theta}_s), \\
\mathbf{U}_J(t) &= \sum_{i=1}^{N_J} S_{Ji}\mathbf{g}_{Ji} \otimes \mathbf{a}(\mathbf{\Theta}_{Ji}) \\
\mathbf{U}_c(t) &= \sum_{i=1}^{N_c} S_{ci}\mathbf{b}(\omega_{ci}) \otimes \mathbf{a}(\mathbf{\Theta}_{ci}).
\end{aligned}$$

where the point target $S(t)$ is located at the spherical coordinate $\mathbf{\Theta}_s = (r_s, \theta_s, \phi_s)$ with r_s, θ_s, and ϕ_s denoting the radial distance, azimuth angle, and elevation angle, respectively. The normalized angular Doppler frequency is $\omega_s = 2\pi f_s/f_r$. The operator $\otimes$ denotes the Kronecker matrix product. The spatial and temporal steering vectors are, respectively,

$$\mathbf{a}(\mathbf{\Theta}_s) = \left[1, e^{-j\Omega(\tau_{2s}-\tau_{1s})}, \cdots, e^{-j\Omega(\tau_{Ms}-\tau_{1s})}\right]^H \quad (5)$$

$$\mathbf{b}(\omega_s) = [1, \cdots, e^{-jk\omega_s}, \cdots, e^{-j(K-1)\omega_s}]^H \quad (6)$$

where , Ω the operating frequency of the phased array, and $\tau_{ms} = |\mathbf{\Theta}_m - \mathbf{\Theta}_s|/c$ is the propagation delay from the signal source to the m-th array element with c being the wave propagation speed.

The N_J jammers S_{Ji} are at locations $\mathbf{\Theta}_{Ji}$ with gains $\mathbf{g}_{Ji} = [g_{Ji}(1), \cdots, g_{Ji}(k), \cdots, g_{Ji}(K)]^T$. The N_c independent clutter patches are uniformly distributed in a circular ring/sphere around the radar platform [1] with the i-th patch at $\mathbf{\Theta}_{ci}$ and having a Doppler frequency ω_{ci} proportional to its angular location. The receiver noise $\mathbf{U}_n$ appears as a uniform noise floor throughout the angle-Doppler plane.

The STAP system consists of a tapped-delay-line attached to each array element. Let $\mathbf{W}$ be the concatenated weight vector of the STAP processor, then the

output of the STAP $y(t)$ can be expressed in a matrix form as $y(t) = \mathbf{W}^H\mathbf{U}(t)$, where the superscript $()^H$ denotes conjugate transpose (or Hermitian transpose).

For the Gaussian clutter environment, the Minimum Variance Distortionless Response (MVDR) method is commonly used for adapting the weight vector $\mathbf{W}$. That is to minimize the second-order moment of the output signal subject to steering constraints

$$\min_{\mathbf{W}} E\left\{|y(t)|^2\right\}, \quad \text{subject to} \quad \mathbf{C}^H\mathbf{W} = \mathbf{h}, \tag{7}$$

where $E\{\cdot\}$ is the expectation operator, $E\left\{|y(t)|^2\right\} = \mathbf{W}^H\mathbf{R}_{\mathbf{uu}}\mathbf{W}$, and $\mathbf{R}_{\mathbf{uu}}$ is the covariance matrix of the concatenated input vector $\mathbf{U}$. The matrix $\mathbf{C}$ is a set of linear constraints and $\mathbf{h}$ is the desired response vector. For example, a simple point constraint [18] may be chosen as $\mathbf{C} = \mathbf{b}(\omega_s) \otimes \mathbf{a}(\mathbf{\Theta}_s)$ and $\mathbf{h} = 1$, which enforces a unit gain response at the target location $\mathbf{\Theta}_s$ and the Doppler frequency f_s. The optimal solution to the constrained minimization problem (7)is well-known assuming that the covariance matrix $\mathbf{R}_{\mathbf{uu}}$ has a full rank [3]:

$$\mathbf{W}_{opt} = \mathbf{R}_{\mathbf{uu}}^{-1}\mathbf{C}(\mathbf{C}^H\mathbf{R}_{\mathbf{uu}}^{-1}\mathbf{C})^{-1}\mathbf{h} \tag{8}$$

Direct implementation of (8) requires the knowledge of the covariance matrix of the array input vector and the Sample Matrix Inversion (SMI) method is often employed in practice [2]. Alternatively, the weight vector $\mathbf{W}_{opt}$ can be decomposed into two orthogonal components: a fixed beamformer $\mathbf{W}_q$ and an unconstrained adaptive weight vector $\mathbf{W}_a$. They are determined by

$$\mathbf{W}_q = \mathbf{C}(\mathbf{C}^H\mathbf{C})^{-1}\mathbf{h}, \tag{9}$$

$$\mathbf{W}_a^{opt} = (\mathbf{C}_a^H\mathbf{R}_{\mathbf{uu}}\mathbf{C}_a)^{-1}\mathbf{C}_a^H\mathbf{R}_{\mathbf{uu}}\mathbf{W}_q, \tag{10}$$

where $\mathbf{C}_a$ is termed the signal blocking matrix. It is orthogonal to $\mathbf{C}$ satisfying $\mathbf{C}^H\mathbf{C}_a = \mathbf{0}$. This decomposition, as shown in Fig. 3, is known as the Generalized Sidelobe Canceller (GSC) and $\mathbf{W}_a$ can be iteratively adapted by the Normalized Least Mean Square (NLMS) algorithm as

$$\mathbf{W}_a(t+1) = \mathbf{W}_a(t) + \mu_a \frac{\mathbf{x}(t)e^*(t)}{\mathbf{x}^H(t)\mathbf{x}(t) + \delta}, \tag{11}$$

where $\mathbf{x}(t) = \mathbf{C}_a^H\mathbf{U}(t)$, the error signal is defined by $e(t) = y(t) = [\mathbf{W}_q - \mathbf{C}_a\mathbf{W}_a]^H\mathbf{U}(t)$. The step size μ_a controls the rate of convergence and the regularization parameter δ prevents the numerical instability when the inputs are small [3].

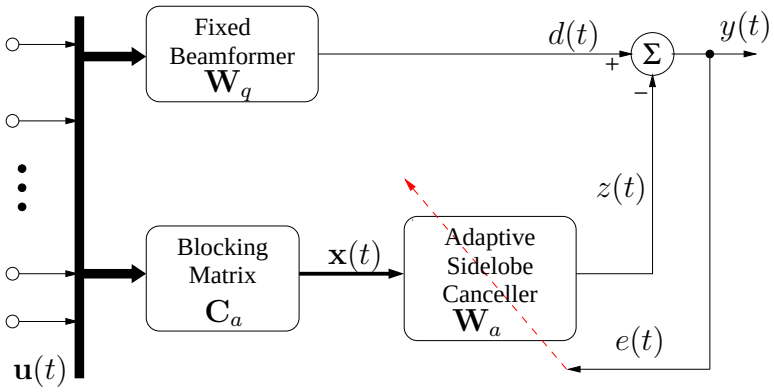


Figure 3. The Generalized Sidelobe Canceller (GSC) implementation of STAP systems.

Equivalently, the NLMS algorithm in (11) is the same as

$$\begin{aligned} \mathbf{W}(0) &= \mathbf{W}_q = \mathbf{C}(\mathbf{C}^H\mathbf{C})^{-1}\mathbf{h}, \\ \mathbf{W}(t+1) &= \mathbf{B}\left[\mathbf{W}(t) - \mu\frac{y^*(t)\mathbf{U}(t)}{\mathbf{U}^H(t)\mathbf{B}\mathbf{U}(t) + \delta}\right] + \mathbf{W}_q. \end{aligned} \tag{12}$$

where μ is the step size and

$$\mathbf{B} = \mathbf{I} - \mathbf{C}[\mathbf{C}^H\mathbf{C}]^{-1}\mathbf{C}^H. \tag{13}$$

3 The Proposed NFLOM Algorithms for STAP

In the severe, impulsive clutter environment, the conventional STAP algorithm suffers from performance loss due to two reasons: one is the high probability of outliers in the received samples; another is the large eigenvalue spread of the sample covariance matrix. An approach to combat these problems is the FLOM algorithm which minimizes the p-th order moment rather than the variance of the STAP output [10, 18]

$$\min_{\mathbf{W}} E\left\{|y(t)|^p\right\}, \quad \text{subject to} \quad \mathbf{C}^H\mathbf{W} = \mathbf{h}, \tag{14}$$

There is no closed-form solution for the optimal coefficients that minimizes the cost function, but a gradient descent method is available. Similar to the NLMS algorithm, the NFLOM algorithm is iteratively adaptive as

$$\begin{aligned} \mathbf{W}_a(0) &= 0, \\ \mathbf{W}_a(t+1) &= \mathbf{W}_a(t) + \mu_a \frac{|e(t)|^{p-2}e^*(t)\mathbf{x}(t)}{\sum_i |x_i(t)|^p + \delta}. \end{aligned} \tag{15}$$

or equivalently

$$\begin{aligned} \mathbf{W}(0) &= \mathbf{W}_q = \mathbf{C}(\mathbf{C}^H\mathbf{C})^{-1}\mathbf{h}, \\ \mathbf{W}(t+1) &= \mathbf{B}\left[\mathbf{W}(t) - \mu \frac{|y(t)|^{p-2}y^*(t)\mathbf{U}(t)}{\sum_i |x_i(t)|^p + \delta}\right] + \mathbf{W}_q \end{aligned} \tag{16}$$

where $x_i(t)$ are the elements of the blocking matrix output $\mathbf{x}(t) = \mathbf{C}_a^H \mathbf{U}(t)$. Other parameters are the same as those in the NLMS algorithm (12).

The NFLOM algorithm reduces to the NLMS algorithm when $p = 2$ and to the normalized sign algorithm (NSA) when $p = 1$ [11, 12]. Our numerical analysis has found [18,20] that, when the order p is smaller, the NFLOM algorithm converges faster but exhibits larger steady-state mean square errors (MSE). This phenomenon is observed in both Gaussian and heavy-tailed clutters, as will be shown in Section 4.

The conflicting goals of fast convergence and low steady-state error of the NFLOM algorithms motivates a new variable-order NFLOM algorithm to achieve both fast convergence and low steady-state MSE by varying the order p. Intuitively, the variable order NFLOM algorithm shall start with a small order, for example $p = 1$, and then gradually increases to $p = 2$. A straightforward approach for order switching is to estimate the excess MSE of the NFLOM algorithm in windows of size D and then compare the MSE to the previous window. If the difference exceeds a threshold, the order p is increased. The proposed variable-order NFLOM follows the procedures:

1. Choose $\mathbf{P} = \{P_l\} = [P_{min} : \Delta P : P_{max}]$. Set $l = 1$ and the initial order as $p = P_l$;

2. Select the estimation window size D and the threshold T_h; Set the output energy of the previous window $\mathcal{E}_0 = DP_U$, where P_U is the total power of the input signal $\mathbf{U}(t)$;

3. Adapt the filter coefficients $\mathbf{W}(t)$ based on (16) using the current order p. Estimate the output energy of the current window as $\mathcal{E}_1 = \sum_{i=1}^{D} |y(i)|^2$;

4. Compare $\mathcal{E}_0$ to $\mathcal{E}_1$. If $\mathcal{E}_1 - \mathcal{E}_0 > DT_h^p$, then increment l and update the order p to P_l.

5. Set $\mathcal{E}_0 = \mathcal{E}_1$ and repeat Step 3 - 4 until $p = P_{max}$.

The parameter selection of the algorithm determines the convergence rate and the steady-state MSE. The threshold T_h can be set at the 1% to 10% of the signal-to-noise-ratio (SNR) or clutter-to-noise-ratio (CNR) level. The window size is normally chosen at several hundred to several thousand samples. The selection of $\mathbf{P} = \{P_l\} = [P_{min} : \Delta P : P_{max}]$ is rather flexible with $P_{min} \geq 1$ and $P_{max} = 2$ for complex Gaussian clutters. For heavy-tailed clutters, slightly smaller P_{min} and P_{max} normally provide better results.

4 Performances Analysis

A linear phased array example was used to demonstrate the performances of the NFLOM and the VO-FLOM algorithms. The array consisted of $M = 10$ equally spaced elements at half wavelength of the operation frequency. The coherent pulse interval (CPI) was $K = 7$ and a fixed range bin was used for the STAP. The target signal had a power of 0 dB with respect to the background noise and its angle of arrival (AoA) was 20° with respect to the axis of the array. The normalized Doppler frequency of the target was fixed at 0.25. The noises were independent among antenna elements and CPI taps with white Gaussian spectrum. Two wideband jammers presented at AoA of $-20°$ and $+50°$, respectively. Both jammers had a full Doppler spectrum and a total power of 30 dB. In addition, many clutters impinged on the array from different AoAs which were uniformly distributed between $-180°$ and 180°. The Doppler frequencies of the clutters depended on their AoAs. The envelop of the clutters was either Rayleigh (complex Gaussian clutter) or compound K with a total average power of 30 dB.

The compound K clutters were simulated using the nonlinear memoryless transformation (NMLT) method [9, 22] and the auto-covariance function (ACF) of the gamma texture was $R_G(t) = \exp(-t/\lambda)$ with a large λ indicating high ACF. For performance comparison, the fixed order NFLOM used a step size $\mu = 0.002$ and the regulation parameter $\delta = 20$CNR.

The STAP algorithms can be evaluated by the output beampattern defined as

$$\Psi(\mathbf{\Theta}, f_d) = |W_{opt}^H \mathbf{b}(f_d) \otimes \mathbf{a}(\mathbf{\Theta})|^2 \tag{17}$$

The convergence performance is commonly evaluated by the excess MSE $J_{ex}(t)$ and misalignment $\mathcal{M}(t)$ defined as [3]

$$J_{ex}(t) = E[|\mathbf{W}^H(t)\mathbf{U}(t)|^2] - J_{min} \tag{18}$$

$$\mathcal{M}(t) = 20 \log_{10} \frac{|\mathbf{W}(t) - \mathbf{W}_{opt}|}{|\mathbf{W}_{opt}|} \tag{19}$$

and $J_{min} = E[|\mathbf{W}_{opt}^H \mathbf{U}(t)|^2]$. The MVDR optimal solutions were used as the common base for comparison for all clutter scenarios, although the NFLOM algorithms are designed to minimize the lower order moments. This means that, if measured in terms of L_p norms or p-th error moments, the NFLOM performance would be better than these second-order performance measures.

4.1 Performances of the NFLOM Algorithm

The beampatterns of the MVDR, NLMS, and NFLOM schemes are plotted in Fig. 4, which were obtained in impulsive K clutters with shape parameter $\nu = 0.5$ and an ACF function $R_G(t) = \exp(-t/\lambda)$ with $\lambda = 100$. The beampattern of the MVDR scheme was computed by the weight vector optimized over all clutter samples thus providing the best performance with deep nulls placed at at both jammer locations

and at the clutter ridge. It passed the target signal with unit gain and achieved high SINR. The beampatterns of the NLMS and NFLOM were computed by their weights $\mathbf{W}(t)$ when a large spiky clutter component was present at t. The NLMS used many degrees of freedom on suppressing the impulsive clutter components, but let the jammers and other clutter components leak through. The NFLOM with $p = 1.5$ and $p = 1.7$ both maintained deep nulls at jammer locations by placing less emphasis on the impulsive clutter components, thus achieving better output SINR than the NLMS algorithm.

The convergence curves of the NFLOM algorithms in complex Gaussian clutters and compound K clutters are plotted in Fig. 5, where the MSE and misalignment curves are the ensemble average over 100 independent trials. In Gaussian clutters, the MSE curves (Fig. 4.5(a)) show that a smaller order p NFLOM converges faster, but to a larger steady-state error. The misalignment curves (Fig. 4.5(b)) show that a smaller order p NFLOM converges faster and achieves lower error norm For compound K clutters, the MSE curves (Fig. 4.5(c)) show that, when p is close to 1, the curves are very similar to those in Gaussian clutters, exhibiting high robustness against impulsive clutters; when p is close to 2, the initial convergence speed is similar to those in Gaussian clutters, but the steady-state errors are higher. The misdajustment curves (Fig. 4.5(b)) behave similar to those in Gaussian clutters but with slightly higher error norms.

It is found that the ACF of the texture component of compound K clutters has significant impacts on the convergence of the NFLOM with $p > 1$. The MSE and misalignment curves in K-clutters with three ACF parameters $\lambda = 10, 100$, and 300 are compared in Fig. 6 and Fig. 7, respectively. In terms of MSE, the NFLOM converges faster and to a lower steady-state error in clutters with higher auto-covariance (larger λ) for all orders, but the effects of ACF on $p = 1$ is the smallest. In terms of misalignment, a larger λ leads to a very fast initial convergence but slows down

significantly afterwards. A small λ leads to a smooth convergence and to a lower steady-state error. A smaller p NFLOM has a lower steady-state error of misalignment than a larger p. Effects of the ACF on the performance of the NFLOM with $p = 1.5$ is between those of $p = 1.25$ and $p = 1.75$, and the curves are omitted here.

The output SINR was also evaluated for orders $p = [0.9 : 2]$ in compound K clutters with different shape parameters. If the SINR is computed on output samples from $t = 5 \times 10^4$ to $t = 10 \times 10^4$, the best SINR was achieved by the NFLOM with $1.5 \leq p \leq 1.8$ with a few dB better performance than the NLMS ($p = 2$) algorithm, as shown in Figs. 4.8(a) and 4.8(b). In comparison, if the SINR was computed after all algorithms have converged, the NFLOM with $1.5 \leq p \leq 1.8$ still performed well and the NLMS algorithm also improved significantly, as shown in Figs. 4.8(c) and 4.8(d). This is consistent with the MSE performances in that the NFLOM with $p = 1.5 \sim 1.8$ provides the best compromise between fast convergence and low steady-state error. All NFLOM algorithms achieved better SINR in clutters with a higher auto-covariance ($\lambda = 300$) than that of a lower ACF ($\lambda = 100$).

4.2 Performances of the VO-FLOM Algorithm

The conflicting goals of fast convergence and low steady-state error with the fixed order NFLOM can be met simultaneously by the variable-order FLOM algorithm. The effectiveness of the VO-FLOM algorithm is illustrated in Fig. 9 for three clutter scenarios: a complex Gaussian and two compound K scenarios with $\nu = 2$ and $\nu = 0.7$ respectively. Both compound K scenarios had an ACF parameter $\lambda = 100$. The MSE and misalignment curves are also the ensemble average of 100 trials for three clutter scenarios: Gaussian and two compound K cases with $\nu = 0.7$ and $\nu = 2$, respectively, both with $\lambda = 100$. The parameters for the VO-FLOM algorithm were: $P_{min} = 1$, $P_{max} = 2$, $\Delta P = 0.1$, $D = 1000$, and the threshold $T_h = 0.01$SNR. For the excess MSE, the VO-FLOM algorithm converges very fast at around $t = 2 \times 10^4$

sample iterations; while the NLMS algorithm does not converge at 10^5 sample iterations. The steady-state errors of the VO-FLOM algorithm were the same as those achievable by the fixed-order NFLOM. For the misalignment, the VO-FLOM converges even faster than the MSE, at $t = 1 \times 10^4$ sample iterations. The steady-state errors of the misalignment were comparable with those achieved by the NLMS algorithm. The change of the order p may vary from trial to trial and the results of a representative trial are shown in Fig. 4.9(c), indicating the initial time indexes when the corresponding order is adapted. It is observed that the change of the orders is faster in more impulsive clutters than that in Gaussian clutters. This behavior very well matches the convergence behavior of the excess MSE curves.

With the parameter $P_{max} = 2$, the VO-FLOM algorithm exhibit much higher excess MSE in impulsive clutter scenarios than that in Gaussian clutters. If this parameter is selected slightly smaller than 2, the converged VO-FLOM can achieve slightly higher performances in terms of higher output SINR, lower MSE, and better robustness against impulsive clutters. This is illustrated in Fig. 10 with the example of $P_{max} = 1.8$. The MSE and beampattern of the VO-FLOM algorithm in compound K clutters are plotted in Fig. 10. The proposed algorithm can effectively suppress the clutters and jammers by placing deep nulls at jammer locations and the clutter ridge, as shown in Fig. 4.10(b). The VO-FLOM also converges to a lower steady-state error than the fixed-order NFLOM, as shown in Fig. 4.10(a).

5 Conclusion

We have evaluated the excess MSE, misalignment, and output SINR performances of the NFLOM algorithms for STAP applications in compound K clutters. The results show that the excess MSE is a better performance measure for phased array applications than the misalignment which is used more often in system identification applications, because the MSE is tightly related with output SINR while the deviation of the weight vector from the MVDR optimal weights plays a less important role for STAP applications in heavy-tailed clutter environment. A variable-order FLOM adaptive algorithm has also been proposed for phased array signal processing, which starts with a small order for fast convergence and increases the order after the excess MSE stops decreasing. The VO-FLOM algorithm improves upon the NFLOM and NLMS algorithm in that the weight adaptation is proportional to a variable p-order moment of the error rather than a fixed order moment or the mean square error. The proposed algorithm achieves an excellent compromise between fast initial convergence and low steady-state errors by taking the advantages of small and large order NFLOM algorithms. The excess mean squared error (MSE) curves have been evaluated for both Gaussian clutter and non-Gaussian, heavy-tailed clutter scenarios. The results show that the proposed VO-FLOM converges much faster than the plain NFLOM and NLMS algorithms and achieves the same steady-state error as the NLMS algorithm.

6 Acknowledgments

This work was supported by Air Force Office of Scientific Research (AFOSR) under Grant FA9550-07-1-0336. The authors wish to thank Drs. Jon A. Sjogren,

Genshe Chen, and Robert S. Lynch for their technical inputs and suggestions on the work. The authors also wish to thank Drs. Maria Greco and Muralidhar Rangaswamy for providing help on Radar data statistics.

7 References

[1] M. A. Richards, *Fundamentals of Radar Signal Processing*, McGraw-Hill, New York, 2005. Chapter 9.

[2] M. C. Wicks, M. Rangaswamy, R. Adve, and T. B. Hale, "Space-Time adaptive Processing: A knowledge-based perspective for airborne radar," *IEEE Signal Processing Magazine*, vol. 23, no.1, pp. 51–65, Jan. 2006.

[3] S. Haykin, *Adaptive Filter Theory*, 4th Edition, Prentice Hall: Upper Saddle River, New Jersey, 07458. Chapter 5 and 6.

[4] S. Haykin, R. Bakker, and B. W. Currie, "Uncovering nonlinear dynamics – the case study of sea clutter," *Proc. IEEE*, pp.860 – 881, May 2002.

[5] 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.

[6] M. Greco, F. Gini, A. Younsi, M. Rangaswamy and A. Zoubir, "Non-stationary sea clutter: impact on disturbance covariance matrix estimate and detector CFAR," *IEEE Int'l Conf. Radar*, Adelaide, Australia, Sep. 2-5, 2008. pp. 558 - 562.

[7] E. P. Blasch and M. Hensel, "Fusion of distributions for radar clutter modeling," *IEEE Intl. Conf. Information Fusion*, 2004. [On-line] Available. http://www.fusion2004.foi.se/papers/IF04-0629.pdf,2004.

[8] D.A. Abraham and A.P. Lyons, "Simulation of non-Rayleigh reverberation and clutter," *IEEE J. Oceanic Eng.*, vol.29, No.2, pp.347-354, Apr.2004.

[9] R.J.A. Tough and K.D. Ward, "The correlation properties of gamma and other non-Gaussian processes generated by memoryless nonlinear transformation," J. Phys.32, pp.3075–3083, 1999.

[10] 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, April 1999.

[11] C. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*, New York, John Wiley, 1995.

[12] G. R. Arce, *Nonlinear Signal Processing: a statistical approach.* John Wiley, NY, 2005.

[13] G. Aydin, O. Arikan, and A.E. Cetin, "Robust adaptive filtering algorithms for alpha-stable random processes," *IEEE Trans. Circuits and Systems II: Analog and Digital Signal Processing,* vol. 46, no. 2, pp.198 - 202, Feb. 1999.

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

[15] Y. Goussard and G. Demoment, "Recursive deconvolution of Bernoulli-Gaussian processes using a MA representation," *IEEE Trans. Geoscience and Remote Sensing,* vol. 27, no. 4, pp.384 - 394, July 1989.

[16] H. M. Bücker and R. Beucker, "Using automatic differentiation for the solution of the minimum p-norm estimation problem in magnetoencephalography," *J. of Simulation Modeling Practice and Theory,* vol 12, 2004. On-line Available: http://www.elsevier.com/locate/simpat, 2008.

[17] G. Tzagkarakis, B. Beferull-Lozano, and P. Tsakalides, "Rotation-Invariant Texture Retrieval via Signature Alignment Based on Steerable Sub-Gaussian Modeling," *IEEE Trans. Image Processing,* vol. 17, no. 7, pp.1212 - 1225, July 2008.

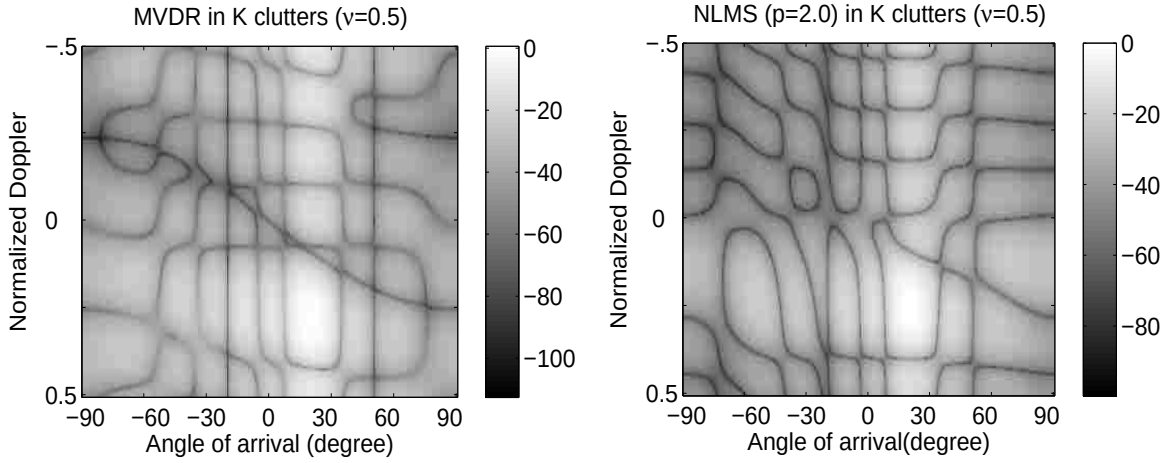
[18] Y. R. Zheng, G. Chen, and E. Blasch, "A normalized fractionally lower-order moment algorithm for space-time adaptive processing," *IEEE conf. Military Communications*, Milcom'07, Orlando, FL, Oct. 2007.

[19] T. G. Shao, Y. R. Zheng, G. Chen, and E. Blasch, "Fast-converging adaptive STAP for non-Rayleigh clutter suppression," *Sensor Information, and Signal Processing workshop*, SENSIP'08, Sedona, AZ, May 2008.

[20] Y. R. Zheng, "A variable-order FLOM algorithm for heavy-tailed clutter suppression," *IEEE conf. Military Communications*, Milcom'08, San Diego, CA, Nov. 2008.

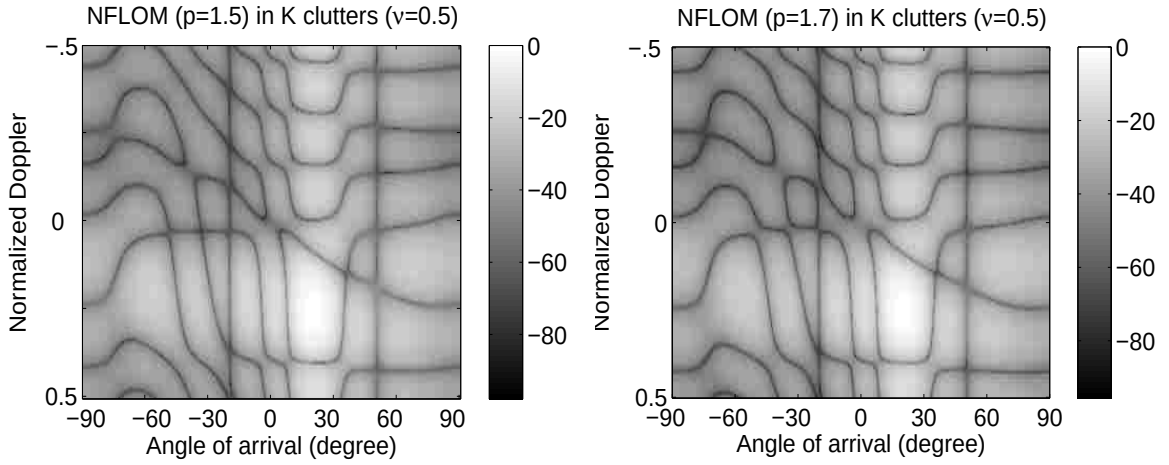
[21] Tiange Shao and Y. R. Zheng, "A New Variable Step-Size Fractional Lower-Order Moment Algorithm for Non-Gaussian Interference Environments," *IEEE Int'l Symposium Circuits & Systems* (ISCAS'09), Taipei, Taiwan, May 24-27, 2009.

[22] Glen Davidson, *Matlab and C Radar Toolbox*, The Japan Society for the Promotion of Science (JSPS), http://www.radarworks.com, 2008.



(a) MVDR $\mathbf{W}_{opt}$

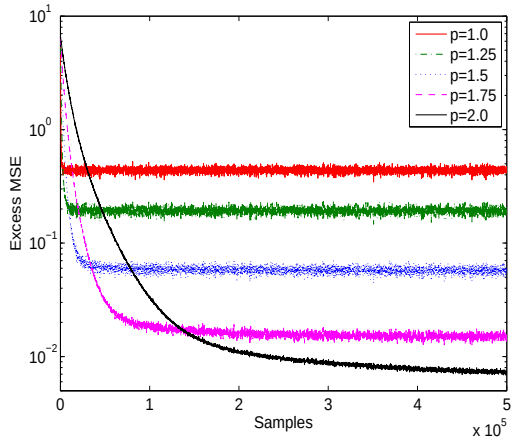
(b) NLMS algorithm



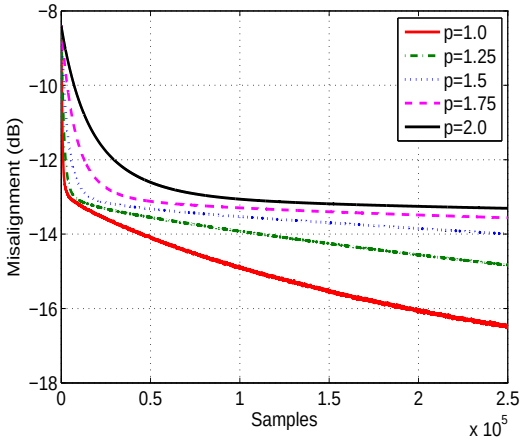
(c) NFLOM algorithm $p = 1.5$

(d) NFLOM algorithm $p = 1.7$

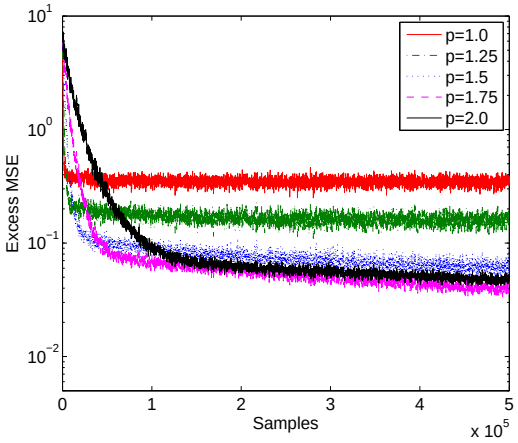
Figure 4. Beampatterns of the NLMS and NFLOM algorithms in compound K clutters ($\nu = 0.5, \lambda = 100$) when impulsive clutter samples were encountered. In comparison, the beampattern of the MVDR scheme was computed by the weight vector optimized over all clutter samples. (a) MVDR: placed deep nulls at both jammer locations and at the clutter ridge therefore passing the target signal with high SINR; (b) NLMS: used many degrees of freedom on suppressing the impulsive clutter components but let the jammers and other clutter components leak through; (c) and (d) NFLOM with $p = 1.5$ and $p = 1.7$: maintained deep nulls at jammer locations by placing less emphasis on the impulsive clutter components, thus achieving better output SINR.



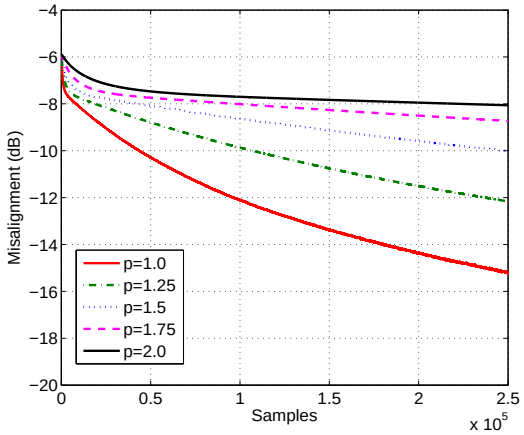
(a) Mean Square Error, Gaussian Clutters



(b) Misalignment, Gaussian Clutters

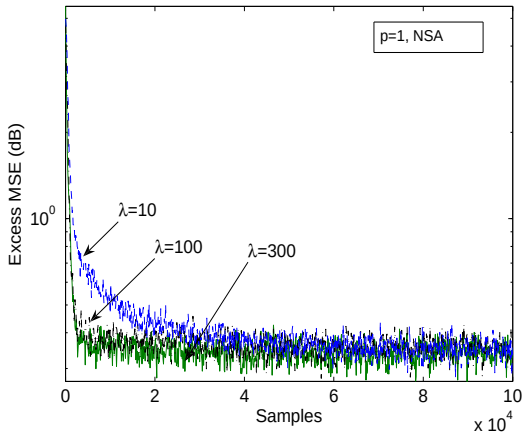


(c) Mean Square Error, K Clutters

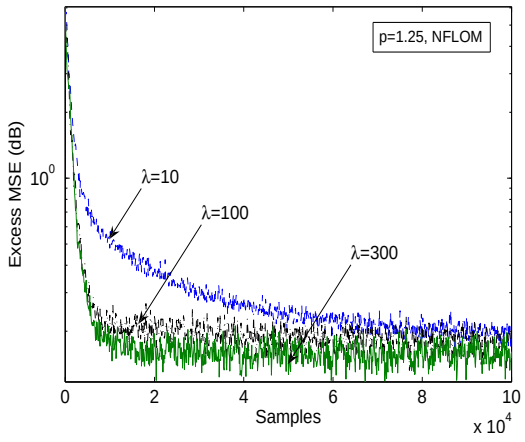


(d) Misalignment, K Clutters

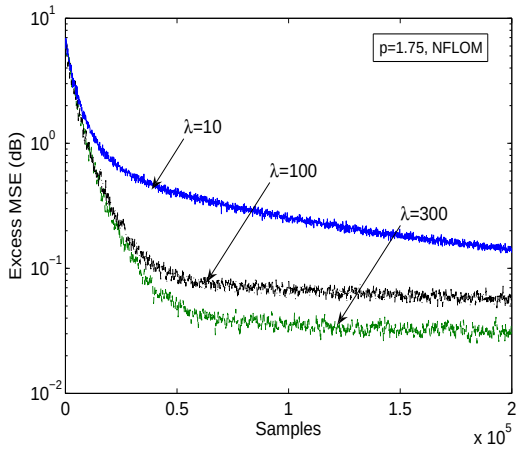
Figure 5. The convergence curves of the NFLOM algorithms in complex Gaussian clutters and compound K clutters ($\nu = 0.7, \lambda = 100$). Two wideband jammers and background noises were also present. The total powers of clutters and jammers were 30 dB above the background noise, respectively. Ensemble average of 100 trials is used in all curves. (a) MSE in Gaussian clutter: a smaller order NFLOM converges faster, but to a larger steady-state error. (b) misalignment in Gaussian clutter: a smaller p NFLOM converges faster and achieves lower error norm. (c) MSE in K clutters: orders close to 1 have the similar convergence as those in Gaussian clutter, exhibiting robustness against impulsive clutters; orders close to 2 have the similar initial convergence speed as those in Gaussian clutters, but higher steady-state errors. (d) misdajustment in K clutters: similar to those in Gaussian clutters but with slightly higher error norms.



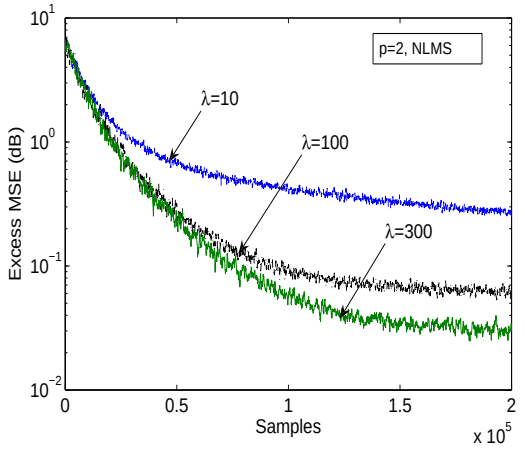
(a) Excess MSE for $p = 1$



(b) Excess MSE for $p = 1.25$

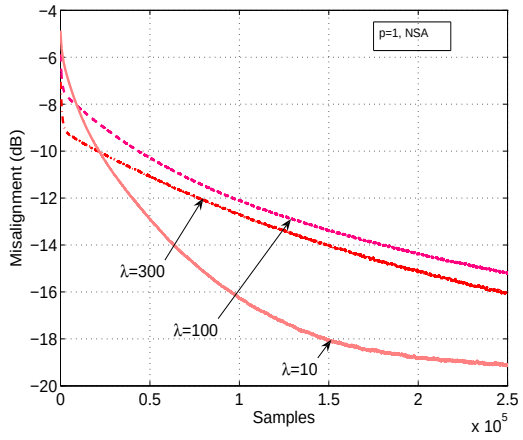


(c) Excess MSE for $p = 1.75$

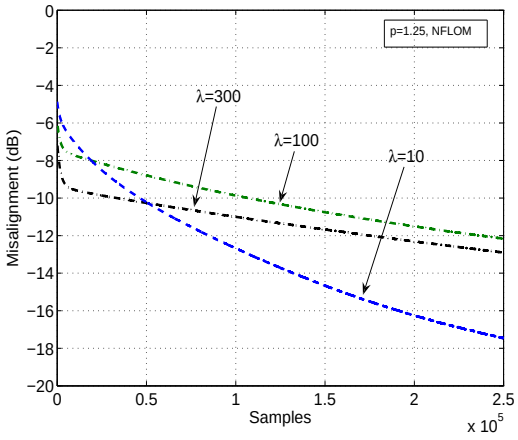


(d) Excess MSE for $p = 2$

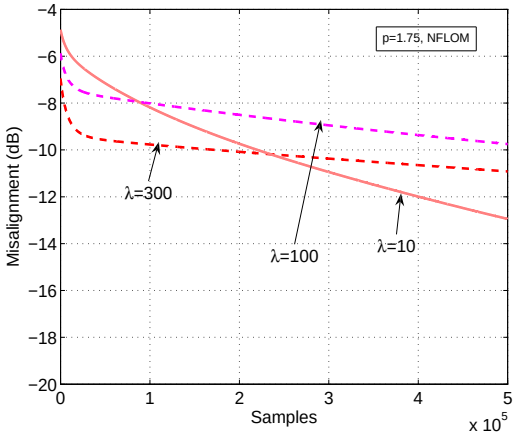
Figure 6. Effects of ACF of K-clutter ($\nu = 0.7$) on mean square error of the NFLOM algorithm. The ACF of the gamma texture is $R_G(t) = \exp(-t/\lambda)$ with $\lambda = 10, 100,$ and 300, respectively. The MSE of the NFLOM converges faster and to a lower steady-state error in clutters with higher auto-covariance (larger λ) for all orders of p. The effects of ACF on the convergence of the NFLOM with $p = 1$ is the smallest. Note the change of x-axis scale in different sub-figures.



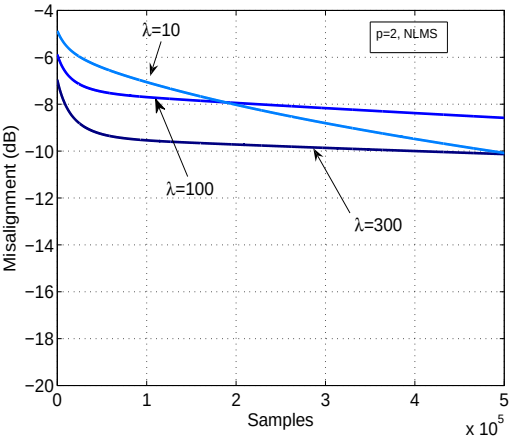
(a) Misalignment for $p = 1$



(b) Misalignment for $p = 1.25$

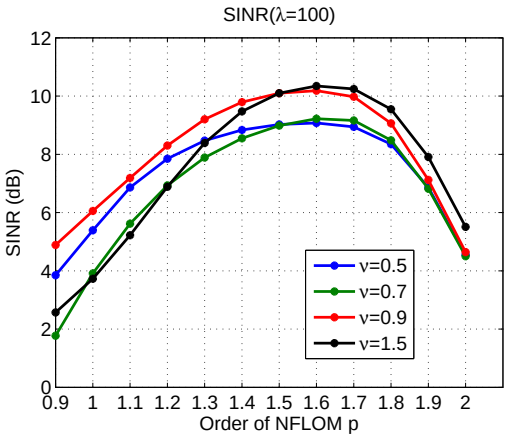


(c) Misalignment for $p = 1.75$

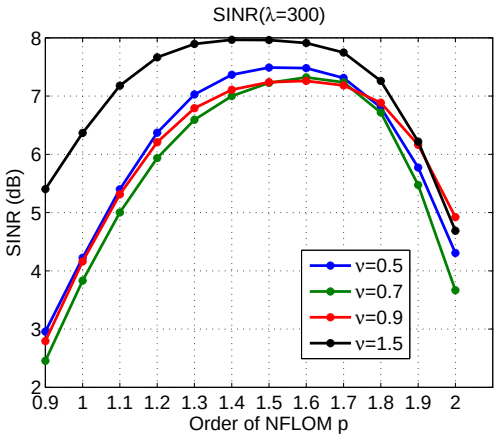


(d) Misalignment for $p = 2$

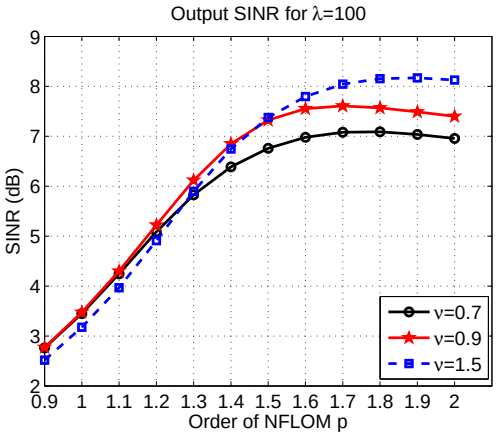
Figure 7. Effects of ACF of K-clutter ($\nu = 0.7$) on misalignment of the NFLOM algorithm. The ACF of the gamma texture is $R_G(t) = \exp(-t/\lambda)$ with $\lambda = 10, 100$, and 300, respectively. In terms of misalignment, a larger λ leads to a very fast initial convergence but slows down significantly afterwards. A small λ leads to a smooth convergence and to a lower steady-state error. A smaller p NFLOM has a lower steady-state error of misalignment than a larger p. Note the change of x-axis scale in different sub-figures.



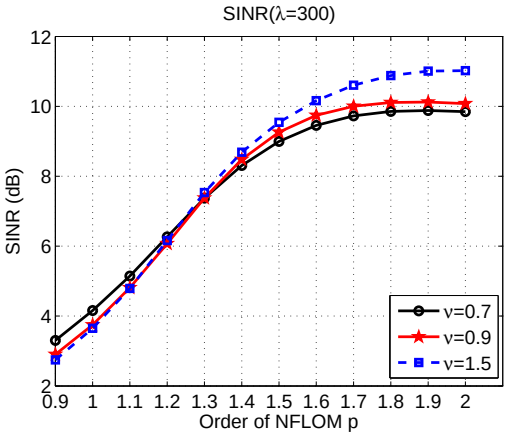
(a) $\lambda = 100$, $t = (5:10) \times 10^4$



(b) $\lambda = 300$, $t = (5:10) \times 10^4$

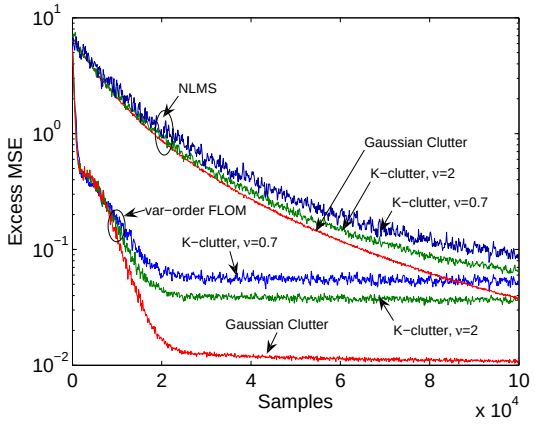


(c) $\lambda = 100$, after convergence

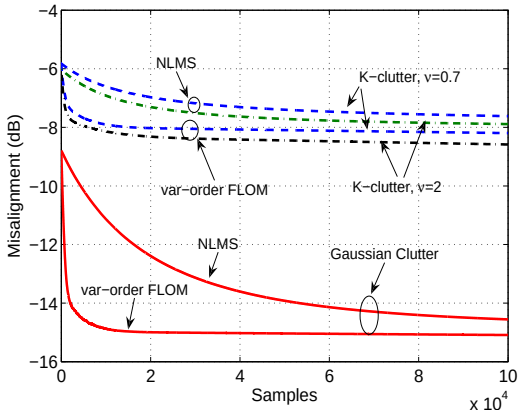


(d) $\lambda = 300$, after convergence

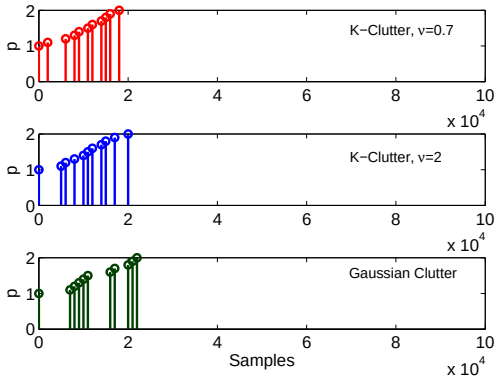
Figure 8. Output SINR as a function of fractional order p in compound K clutters with different shape parameters and ACF parameters. (a) and (b) SINR computed at iterations $t = (5:10) \times 10^4$. The best SINR was achieved by the NFLOM with $1.5 \leq p \leq 1.8$, a few dB better than the NLMS ($p = 2$) algorithm. (c) and (d) SINR computed after all algorithms converged. The NFLOM with $p > 1.5$ performed well. The NFLOM algorithms achieved better SINR in clutters that have higher auto-covariance. The NLMS algorithm performed better than the lower-order NFLOM if converged.



(a) Mean Square Error

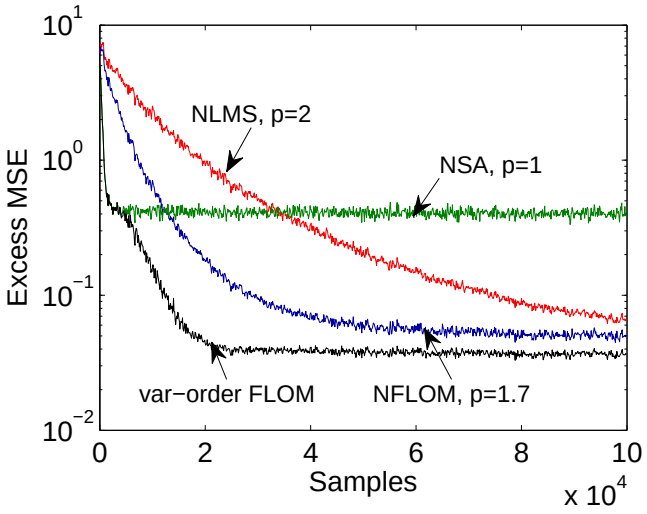


(b) Misalignment

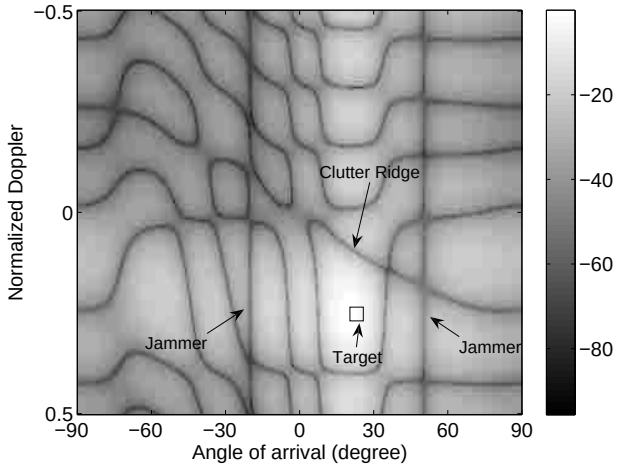


(c) Varying orders

Figure 9. Convergence curves of the VO-FLOM algorithm in comparison to the NLMS algorithm. Note that the x-axis is reduced to show significantly faster convergence of the VO-FLOM than that of the NLMS. The VO-FLOM also achieves similar steady-state error as that of the NLMS. Parameters of the VO-FLOM: $p = [1 : 0.1 : 2], D = 1000$, and $T_h = 0.01\text{SNR}$. Ensemble average of 100 trials are used in (a) and (b), while a single trial is used for (c).



(a) Mean Square Error



(b) Beampattern

Figure 10. The MSE and beampattern of the VO-FLOM with $P_{max} = 1.8$ in compound K clutters. The K-clutters are with $\nu = 2$ and $\lambda = 100$.

SECTION

2 CONCLUSIONS

This dissertation has applied the least mean pth norm estimation theory to the design of adaptive filtering, by the motivation of breaking the limitations of the least-mean-square estimation theory and improving the performance of traditional adaptive filtering algorithms.

The L_1 norm minimization is investigated particularly due to its simplicity and robustness against the impulsive noise. The affine projection sign algorithm (APSA) and variable step-size normalized sign algorithm (VSS-NSA)have been proposed according to L_1 minimization. The APSA updates its weight vector by multiple input vectors and the sign of the a priori error vector, which combines the benefits of the affine projection algorithm (APA) and normalized sign algorithm (NSA). The VSS-NSA intelligently adjusts the step size by matching the L_1 norm of the a posterior error to that of the unwanted noise, which improves convergence rate while reduces the steady-state error.

In the estimation theory, the optimum order statistics is determined by the tail-heaviness of signal distribution, which motivate us to extend L_1 norm minimization to L_p norm minimization. Based on L_p norm ($1 < p < 2$), the variable step-size normalized fractionally lower-order moment (VSS-NFLOM) algorithms have been proposed, which automatically adjust the step size by approximating the pth norm of the *a posteriori* error to that of the background noise. Furthermore, The variable order fractionally lower-order moment (VO-FLOM) algorithm has been proposed, which improves upon the fractionally lower-order moment (FLOM) and Normalized least mean square (NLMS) algorithm in that the weight adaptation is proportional

to a variable pth order moment of the error rather than a fixed order moment or the mean square error.

The four proposed adaptive algorithms of this dissertation target at difficulties of robustness against impulsive interference, convergence rate, stableness, steady-state errors, computational complexity, and tracking ability. The three main applications of this research is acoustic echo cancelation, system identification, and radar phased array clutter suppression.

The contributions of my PhD research work are summarized in three journal papers and four conference papers, among which, three journal papers and two conference paper are included in this dissertation.

Future work lies in the following aspects: 1) complement the study of convergence analysis with resealable assumptions and apply them to all the proposed algorithms. 2) Develop variable step size for affine-projection-type algorithms and also extend real-value to complex-value adaptive filter. 3) Extend the variable step size algorithm to a proportionate algorithm that employs a diagonal matrix rather than a scaler, which ensures updating each element of the weight vector separately, thus improving performance in terms of convergence and steady-state error.

3 PUBLICATIONS

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

[2] Y. R. Zheng, T. Shao, and E. Blasch, "A fast-converging space-time adaptive processing algorithm for non-Gaussian clutter suppression," Elsevier *Digital Signal Processing*, In Press, Corrected Proof, Nov. 2010.

[3] Y. R. Zheng, T. Shao, and Vitor Nascimento, "A new variable step-size fractional lower-order moment algorithm for non-Gaussian interference environments," submitted to Elsevier *Digital Signal Processing*.

[4] T. Shao, Y. R. Zheng, and J. Benesty, "A variable step-size sign algorithm for acoustic echo cancelation," in *Proc. IEEE ICCASP10*, Dallas, Texas, Mar. 2010.

[5] T. Shao, 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.

[6] Y. R. Zheng and T. Shao, "A variable step-size lmp algorithm for heavy-tailed interference suppression in phased array radar," in *Proc. IEEE AeroConf09*, Big Sky, MT, Mar. 2009.

[7] T. G. Shao, Y. R. Zheng, G. Chen, and E. Blasch, "Fast-converging adaptive STAP for non-Rayleigh clutter suppression," *Sensor Information, and Signal Processing workshop*, SENSIP'08, Sedona, AZ, May 2008.

VITA

Tiange Shao was born March 1, 1982 in Chengdu, Sichuan Province, China. She received her B.S. degree in 2004 and M.S. degree in 2007, both in Electrical Engineering from University of Electronic Science and Technology of China, Chengdu, Sichuan Province, China. She began her Ph.D. study in August 2007 at the Department of Electrical and Computer Engineering at Missouri University of Science and Technology. Her research interests include radar array signal processing, adaptive filtering, acoustic signal processing. She is expected to receive her Ph.D. degree in Electrical Engineering from Missouri University of Science and Technology in August 2011.