

# Reduced-Complexity Adaptive Filtering Techniques for Communications Applications

Reza Arablouei

Doctoral Thesis  
in  
Telecommunications Engineering



January 2013



# Statement of originality

I declare that:

- this thesis presents work carried out by myself and does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university;
- to the best of my knowledge, it does not contain any materials previously published or written by another person except where due reference is made in the text;
- and all substantive contributions by others to the work presented, including jointly authored publications, are clearly acknowledged.



Reza Arablouei

29/Jan./2013



# Acknowledgment

I wish to express my sincere gratitude to my principle supervisor, Associate Professor Kutluyıl Doğançay, from whom not only I received high-quality supervision and guidance for my studies, I learned invaluable lessons for life.

I wish to thank Professor Alex Grant, the director of the Institute for Telecommunication Research (ITR), for his generous support throughout my candidature.

I would also like to put on record my appreciation and gratefulness to all ITR staff and students, my international collaborator, Dr Stefan Werner, and my associate supervisor, Dr Sylvie Perreau, for their kind helps and support.

Finally yet importantly, my PhD candidature was supported by an Australian Postgraduate Award and in part by a Commonwealth Scientific and Industrial Research Organisation (CSIRO) scholarship.



# Synopsis

Adaptive filtering algorithms are powerful signal processing tools with widespread use in numerous engineering applications. Computational complexity is a key factor in determining the optimal implementation as well as real-time performance of the adaptive signal processors. To minimize the required hardware and/or software resources for implementing an adaptive filtering algorithm, it is desirable to mitigate its computational complexity as much as possible without imposing any significant sacrifice of performance.

This thesis comprises a collection of thirteen peer-reviewed published works as well as an integrating material. The works are along the lines of a common unifying theme that is to devise new low-complexity adaptive filtering algorithms for communications and, more generally, signal processing applications.

The main contributions are the new adaptive filtering algorithms, channel equalization techniques, and theoretical analyses listed below under four categories:

## 1) adaptive system identification

- affine projection algorithm with selective projections
- proportionate affine projection algorithm with selective projections
- modified RLS algorithm with enhanced tracking capability

## 2) adaptive *inverse* system identification

- low-complexity adaptive decision-feedback equalization of MIMO channels
- partial-update adaptive decision-feedback equalization
- MIMO-DFE coupled with V-BLAST for adaptive equalization of wideband MIMO channels

## 3) adaptive *set-membership* system identification

- modified quasi-OBE algorithm with improved numerical properties
- low-complexity implementation of the quasi-OBE algorithm
- steady-state mean squared error and tracking performance analysis of the quasi-OBE algorithm
- tracking performance analysis of the set-membership NLMS algorithm

## 4) adaptive *linearly-constrained* system identification

- reduced-complexity constrained recursive least-squares algorithm
- linearly-constrained recursive total least-squares algorithm
- linearly-constrained line-search algorithm for adaptive filtering.

The main techniques utilized to alleviate the computational complexity in the proposed algorithms are

- adoption of a *time-varying projection order* for the affine projection algorithm
- the concept of *partial updates*
- the *dichotomous coordinate-descent* iterations for solving the associated systems of linear equations
- the *method of weighting* for linearly-constrained filters.

The proposed algorithms yield significant savings in terms of computational complexity compared with the existing algorithms and techniques. Numerical computer simulations show that, in most cases, this is realized without incurring any noticeable performance degradation. Establishing sensible trade-offs between complexity and performance is another substantial benefit offered by the proposed algorithms. Moreover, there is a good match between simulation results and theoretically predicted values of the performance metrics for which the analyses have been carried out.

The detailed account of each work together with the appropriate literature survey, theoretical and numerical analysis, discussions, and concluding remarks has been reported in the corresponding papers that are embodied in the thesis.

# Contents

Prologue.....	1
Research theme.....	1
Thesis outline.....	4
Publications.....	7
A. Affine projection algorithm with selective projections.....	7
Abstract.....	7
1. Introduction.....	7
2. Affine projection algorithm.....	9
3. Optimizing the projection order.....	10
4. APA with selective projections.....	11
4.1. Variable projection order.....	12
4.2. Prediction of the steady-state MSE.....	14
4.3. Estimation of the error variance.....	15
4.4. Estimation of the noise variance.....	16
4.5. Selection of the input regressors.....	17
5. Simulation studies and discussions.....	18
5.1. System identification.....	18
5.2. Acoustic echo cancelation.....	19
5.3. Discussion.....	20
6. Conclusion.....	21
References.....	21
B. Proportionate affine projection algorithm with selective projections for sparse system identification.....	34
Abstract.....	34
1. Introduction.....	34
2. Proportionate-type APAs.....	36
2.1. Improved proportionate APA (IPAPA).....	36
2.2. IPAPA with proportionate memory (MIPAPA).....	37
3. APA with selective projections.....	37
3.1. Variable projection order.....	37
3.2. Regressor selection.....	38
4. New Algorithm.....	39
5. Simulation results.....	40
6. Conclusion.....	41

References.....	41
C. Modified RLS algorithm with enhanced tracking capability for MIMO channel estimation .....	49
Abstract .....	49
1. Introduction .....	49
2. Proposed algorithm.....	50
3. Simulations .....	52
4. Conclusion.....	53
References.....	53
D. Low-complexity adaptive decision-feedback equalization of MIMO channels.....	56
Abstract .....	56
1. Introduction .....	56
2. System model and algorithm description.....	58
2.1. Equalization and detection .....	59
2.2. Updating the coefficients .....	60
2.3. Ordering update .....	63
3. Computational complexity.....	64
4. Convergence analysis .....	65
4.1. Theory .....	65
4.2. Comparison with experiment .....	67
5. Simulation studies.....	68
6. Conclusion.....	69
References.....	70
E. MIMO-DFE coupled with V-BLAST for adaptive equalization of wideband MIMO channels.....	80
Abstract .....	80
1. Introduction .....	80
2. Signal and system model .....	82
3. Algorithm description .....	82
3.1. Equalization and detection .....	83
3.2. Identification and tracking of the virtual flat-fading channel .....	84
3.3. Updating filter coefficients of the adaptive MIMO-DFE.....	85
4. Computational complexity.....	85
5. Simulations .....	86
6. Conclusion.....	86
References.....	87

F. Partial-update adaptive decision-feedback equalization .....	93
Abstract .....	93
1. Introduction .....	93
2. NLMS-DFE .....	94
3. Partial-update NLMS-DFE .....	95
4. Selective-partial-update NLMS-DFE .....	97
5. Combined selective-periodic partial-update NLMS-DFE .....	99
6. Complexity comparison .....	99
7. Simulations .....	99
7.1. Static channel .....	100
7.2. Time-varying channel .....	100
8. Conclusion .....	100
References .....	101
G. Modified quasi-OBE algorithm with improved numerical properties .....	107
Abstract .....	107
1. Introduction .....	107
2. Set-membership adaptive filtering .....	109
3. The QOBE Algorithm .....	110
3.1. Algorithm derivation .....	110
3.2. Numerical behavior .....	111
4. The MQOBE Algorithm .....	112
4.1. Algorithm derivation .....	112
4.2. Numerical behavior .....	114
5. Simulations .....	116
6. Conclusion .....	117
References .....	117
H. Low-complexity implementation of the quasi-OBE algorithm .....	125
Abstract .....	125
1. Introduction .....	125
2. The DCD-QOBE algorithm .....	126
3. Computational Complexity .....	128
4. Simulations .....	128
5. Conclusion .....	129
References .....	129
I. Steady-state mean squared error and tracking performance analysis of the quasi-OBE algorithm .....	131
Abstract .....	131
1. Introduction .....	131

2. The QOBE algorithm.....	133
3. Steady-state mean squared error analysis.....	135
4. Tracking analysis.....	138
5. Numerical studies .....	140
5.1. Stationary case.....	140
5.2. Nonstationary case.....	141
6. Concluding remarks .....	142
Appendix.....	143
References.....	146
J. Tracking performance analysis of the set-membership NLMS adaptive filtering algorithm.....	153
Abstract.....	153
1. Introduction .....	153
2. The Set-Membership NLMS algorithm.....	155
3. Tracking performance analysis .....	156
4. Simulations .....	159
5. Concluding Remarks .....	160
Appendix.....	161
References.....	161
K. Reduced-complexity constrained recursive least-squares adaptive filtering algorithm.....	165
Abstract.....	165
1. Introduction .....	165
2. Algorithm description .....	167
3. Computational complexity.....	170
4. Simulations .....	171
4.1. Linear-phase system identification (LPSI) .....	171
4.2. Linearly-constrained minimum-variance (LCMV) filtering.....	171
5. Concluding remarks .....	172
Appendix.....	173
References.....	174
L. Linearly-constrained recursive total least-squares algorithm.....	181
Abstract.....	181
1. Introduction .....	181
2. Algorithm description .....	182
3. Simulations .....	186
4. Conclusion.....	187
References.....	187

M. Linearly-constrained line-search algorithm for adaptive filtering.....	192
Abstract .....	192
1. Introduction .....	192
2. Algorithm description .....	193
3. Simulations .....	195
4. Conclusion.....	196
References.....	196
Epilogue .....	199
Summary .....	199
Future work .....	200
Appendices .....	201
Contributions of the co-authors to the publications.....	201
Copyright permissions .....	202
Bibliography.....	219



# Prologue

## Research theme

Adaptive filtering algorithms play a key role in adaptive signal processing, especially for applications where low-cost real-time estimation of some unknown parameters is required. The central merit of an adaptive filter is its ability to respond quickly to the variations in the sought-after parameters or statistical properties of the observed data [PD08], [AS08], [SH02].

Adaptive filters are chiefly implemented in four configurations:

- 1) system identification
- 2) inverse system identification
- 3) noise cancelation
- 4) linear prediction.

In this thesis, the first two configurations are considered. Adaptive echo cancelation and adaptive channel estimation are two prominent communications applications where adaptive filters are used in system identification configuration. Adaptive channel equalization is also an important communications application where inverse system identification is performed [AS08].

Adaptive filters are generally evaluated by their convergence rate, steady-state mean-squared error, computational complexity, and numerical stability. Widrow's least mean square (LMS) adaptive filter [BW60] is the first in chronology. It is computationally simple and numerically stable but slow in convergence, particularly when the filter input data is correlated. In the past few decades, numerous subsequent algorithms have been developed with the objective of enhancing performance of the adaptive filters in terms of convergence rate, steady-state mean-squared error, and tracking capability. However, such enhancements normally come at the expense of increased computational complexity.

Computational complexity is a crucial factor in real-time applications. When an adaptive filter is implemented in an online (real-time) signal processing system, existing hardware and power limitations may affect the performance of the system. The more the computational complexity of an algorithm is, the greater the required hardware resources and power consumption are. Therefore, there exists a great appeal in reducing the computational complexity of the adaptive filtering algorithms. Although the rapid development of silicon technology might make the concerns about the computational power seem unwarranted, in mobile communications, lower complexity still translates into the possibility of realizing smaller portable devices

with longer-running batteries as well as lower production and maintenance costs. This is particularly important when we recall that progresses in the battery technology have not been able to keep up with the advances in microelectronics.

The pivotal theme of the PhD research that has yielded this thesis has been to devise and analyse new low-complexity adaptive filtering algorithms for communications applications. To this end, several tools and techniques have been utilized to decrease the computational requirements of adaptive filtering algorithms in various applications pertaining to adaptive system identification and adaptive inverse system identification. The developed algorithms are appreciably less expensive than the existing ones. Notably, the computational savings are usually achieved with no significant loss of performance. The theoretical performance analysis of some of the low-complexity algorithms has also been carried out.

The findings of the PhD research have been reported through the following publications:

R. Arablouei and K. Doğançay, "Modified RLS algorithm with enhanced tracking capability for MIMO channel estimation," *Electronics Letters*, vol. 47, no. 19, pp. 1101-1103, 2011.

R. Arablouei and K. Doğançay, "Low-complexity adaptive decision-feedback equalization of MIMO channels," *Signal Processing*, vol. 92, pp. 1515-1524, 2012.

R. Arablouei and K. Doğançay, "Affine projection algorithm with selective projections," *Signal Processing*, vol. 92, pp. 2253-2263, 2012.

R. Arablouei and K. Doğançay, "Low-complexity implementation of the quasi-OBE algorithm," *Electronics Letters*, vol. 48, no. 11, pp. 621-623, 2012.

R. Arablouei and K. Doğançay, "Linearly-constrained line-search algorithm for adaptive filtering," *Electronics Letters*, vol. 48, no. 19, pp. 1208-1209, 2012.

R. Arablouei and K. Doğançay, "Linearly-constrained recursive total least-squares algorithm," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 821-824, 2012.

R. Arablouei and K. Doğançay, "Reduced-complexity constrained recursive least-squares adaptive filtering algorithm," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6687-6692, Dec. 2012.

R. Arablouei and K. Doğançay, "Steady-state mean squared error and tracking performance analysis of the quasi-OBE algorithm," *Signal Processing*, vol. 93, pp. 100-108, 2013.

R. Arablouei and K. Doğançay, "Modified quasi-OBE algorithm with improved numerical properties," *Signal Processing*, vol. 93, pp. 797-803, 2013.

R. Arablouei, K. Doğançay, and Sylvie Perreau, "Proportionate affine projection algorithm with selective projections for sparse system identification," in *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, Singapore, Dec. 2010, pp. 362–366.

R. Arablouei, K. Doğançay, and Sylvie Perreau, "MIMO-DFE coupled with V-BLAST for adaptive equalization of wideband MIMO channels," in *Proceedings of 19<sup>th</sup> European Signal Processing Conference*, Barcelona, Spain, Sep. 2011, pp. 639-643.

R. Arablouei, K. Doğançay, and Sylvie Perreau, "Partial-update adaptive decision-feedback equalization," in *Proceedings of 19<sup>th</sup> European Signal Processing Conference*, Barcelona, Spain, Sep. 2011, pp. 2205-2209.

R. Arablouei and K. Doğançay, "Set-membership recursive least-squares adaptive filtering algorithm," in *Proceedings of The 37<sup>th</sup> International Conference on Acoustics, Speech and Signal Processing*, Kyoto, Japan, Mar. 2012, pp. 3765-3768.

R. Arablouei and K. Doğançay, "Affine projection algorithm with variable projection order," in *Proceedings of IEEE International Conference on Communications*, Ottawa, Canada, Jun. 2012, pp. 3734-3738.

R. Arablouei and K. Doğançay, "Adaptive decision-feedback equalization of MIMO channels using coordinate descent iterations," in *Proceedings of IEEE International Conference on Communications*, Ottawa, Canada, Jun. 2012, pp. 3793-3798.

R. Arablouei and K. Doğançay, "Tracking performance analysis of the set-membership NLMS algorithm," in *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, Hollywood, USA, Dec. 2012, paper id: 200.

R. Arablouei and K. Doğançay, "Low-complexity implementation of the constrained recursive least-squares adaptive filtering algorithm," in *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, Hollywood, USA, Dec. 2012, paper id: 10.

The following manuscripts have also been submitted for publication:

R. Arablouei, S. Werner, and K. Doğançay, "Analysis of the gradient-descent total least-squares algorithm," submitted to *IEEE Transactions on Signal Processing*, 2012.

R. Arablouei, S. Werner, and K. Doğançay, "Diffusion-based distributed adaptive estimation utilizing gradient-descent total least-squares," submitted to *The 38<sup>th</sup> International Conference on Acoustics, Speech and Signal Processing*, Vancouver, Canada, 2013.

R. Arablouei, S. Werner, and K. Doğançay, "Adaptive frequency estimation of three-phase power systems with noisy measurements," submitted to *The 38<sup>th</sup> International Conference on Acoustics, Speech and Signal Processing*, Vancouver, Canada, 2013.

F. Albu, M. Rotaru, R. Arablouei, and K. Doğançay, “Intermittently-updated affine projection algorithm,” submitted to *The 38th International Conference on Acoustics, Speech and Signal Processing*, Vancouver, Canada, 2013.

## Thesis outline

This document is a *thesis containing published research* in accordance with University of South Australia’s Academic Regulations for Higher Degrees by Research, Clause 16.1.1.b. Thirteen refereed publications, denoted by Papers A to M, constitute the main body of the thesis. The permissions for reusing the papers have been acquired from the publishers and attached in an Appendix. The papers have been reformatted to enhance the presentation and clarity of the thesis. The IEEE citation style has been adopted in all the reformatted papers. The author of the thesis is the primary author of all the included publications, which have been co-authored with his PhD supervisors. The contributions of the co-authors to the publications are elaborated in an Appendix. The subject matters of all the papers are closely related to the core research theme, which is to develop and analyse new low-complexity adaptive filtering techniques for communications applications.

Based on the configuration of the associated adaptive filter(s) in each work, the included publications may be grouped as:

Papers A to C: adaptive system identification

Papers D to F: adaptive *inverse* system identification

Papers G to J: adaptive *set-membership* system identification

Papers K to M: adaptive *linearly-constrained* system identification.

In Paper A, a new affine projection (AP) algorithm [K084] is developed that has a time-varying projection order. This algorithm requires significantly less computations than other existing AP algorithms. Moreover, it outperforms its competitors in terms of steady-state mean-squared error. This algorithm is well suited for echo cancellation applications.

In Paper B, a new proportionate affine projection (PAP) algorithm [TG00] with a time-varying projection order is proposed. With much less complexity, it outperforms the existing PAP algorithms in sparse system identification applications, such as network echo cancelation.

In Paper C, a modified recursive least-squares (RLS) algorithm [SH02] is proposed that possesses an excellent tracking performance while having a computational complexity comparable to the one of the conventional RLS algorithm. This algorithm is particularly suitable for estimating and tracking fast-varying communications channels.

In Paper D, a new strategy for adaptive equalization of flat-fading multiple-input multiple-output (MIMO) channels is devised. Thanks to its innovative and efficient adaptive decision-feedback equalization (DFE) and adaptive ordered-successive interference cancelation, the proposed equalizer enjoys a significant reduction in computational complexity compared with the previous equalizers. Nonetheless, the complexity gain does not incur any performance loss. The theoretical performance analysis of this equalizer is also presented.

While the equalizer of Paper D is for narrowband (flat-fading) MIMO channels, the equalizer proposed in Paper E is capable of dealing with wideband MIMO channels. It is composed of an adaptive MIMO-DFE coupled with a vertical Bell Laboratories layered space-time architecture (V-BLAST) detector [GF99] and performs separate intersymbol interference (ISI) and inter-channel interference (ICI) cancelation. The computational complexity of this equalizer is lower than the existing adaptive wideband MIMO equalizers.

Utilization of the notion of partial updates [KD08] in adaptive decision-feedback equalization is studied in Paper F. New partial-update normalized least mean square algorithms are proposed, which can help system designers establish a tradeoff between computational complexity and performance.

Since the works of Papers D to F are related to channel equalization, they obviously fall into the category of adaptive inverse system identification.

In Paper G, it is shown that, despite its potential to alleviate the average computational burden as a result of being based on the concept of set-membership filtering, the quasi optimal bounding ellipsoid (quasi-OBE) algorithm [SN99] suffers from numerical instability. A no-cost but effective remedy is proposed via transformation of the internal variables of the quasi-OBE algorithm into a set of new variables.

In Paper H, a low-complexity implementation of the quasi-OBE algorithm employing the dichotomous coordinate descent (DCD) iterations [YZ04] is introduced. This implementation can be straightforwardly extended to the modified quasi-OBE algorithm proposed in Paper G. In the proposed algorithm, utilizing the DCD iterations leads to considerably reduced per-update computational complexity and provides a tradeoff between complexity and performance.

The steady-state mean-squared error and tracking performance of the quasi-OBE algorithm is analyzed in Paper I. The tracking performance of the set-membership normalized least mean square algorithm is also analyzed in Paper J. The analyses exploit the energy conservation argument [MR96] and the theoretical predictions match the simulation results well.

In Paper K, a new linearly-constrained recursive least-squares algorithm is developed that is based on the method of weighting [CV85] and the DCD iterations. The proposed algorithm has significantly less computational complexity than a previously-proposed constrained recursive least-squares (CRLS) algorithm [LR96] while delivering convergence performance on par with it.

In Paper L, a new linearly-constrained recursive *total* least-squares algorithm [SV91] is developed by applying an approach similar to the method of weighting and the line-search optimization algorithm. In addition to having appreciably lower computational complexity than the CRLS algorithm, the proposed algorithm outperforms the CRLS algorithm when both input and output data are observed with noise.

In paper M, a new linearly-constrained line-search adaptive filtering algorithm is developed by incorporating the linear constraints into the least-squares problem and searching the solution (filter weights) along the Kalman gain vector. The proposed algorithm performs close to the CRLS algorithm while having a computational complexity comparable to the constrained least mean square (CLMS) algorithm [OF72].

# Publications

## Paper A

Originally published as

R. Arablouei and K. Doğançay, "Affine projection algorithm with selective projections," *Signal Processing*, vol. 92, pp. 2253-2263, 2012.

Copyright © 2012 Elsevier. Reprinted with permission.

## Affine projection algorithm with selective projections

Reza Arablouei and Kutluyıl Doğançay

### Abstract

In the affine projection adaptive filtering algorithm, convergence is sped up by increasing the projection order but with an unwelcome consequence of increased steady-state misalignment. To address this unfavorable compromise, we propose a new affine projection algorithm with selective projections. This algorithm adaptively changes the projection order according to the estimated variance of the filter output error. The error variance is estimated using exponential window averaging with a variable forgetting factor and a simple moving averaging technique. The input regressors are selected according to two different criteria to update the filter coefficients at each iteration. Simulations, carried out for different adaptive filtering applications, demonstrate that the new algorithm provides fast initial convergence and low steady-state misalignment without necessarily trading off one for the other in addition to a significant reduction in average computational complexity.

Keywords: adaptive filtering; affine projection algorithm; selective projections; error variance estimation.

### 1. Introduction

The affine projection algorithm (APA) [1] was introduced to improve the convergence rate of the LMS-type adaptive filtering algorithms, in particular when the input signal

is highly correlated. However, it suffers from an inherent drawback that imposes a trade-off between high initial convergence rate and low steady-state mean square error (MSE). It is known that for a given adaptive filtering scenario, the larger the projection order is, the faster the APA converges but to a larger steady-state MSE (or coefficients misalignment) [2]–[7]. The improvement in convergence rate can be attributed to the decorrelating property of the APA [5]. Alternatively, it can be ascribed to the APA’s feature of producing direction vectors that are strongly correlated with the weight error vector [6]. However, these enhancements come at the expense of amplifying energy of the background noise and consequently increasing the steady-state MSE. This impairment is supposedly caused by the APA’s linear but time-varying pre-whitening filter and is aggravated by larger projection orders [5], [7].

Several approaches have been proposed to tackle the abovementioned problem. Some are based on the optimization of the step-size [8]–[11] or the regularization parameter [12], [13] of the algorithm, while others aim to optimize its projection order [14]–[18] and/or select the optimal input regressors [18]–[21]. In [14] and [15], a set-membership APA with variable data-reuse factor (SM-AP vdr algorithm) is proposed that changes the projection order during adaptation utilizing the information provided by the data-dependent step-size of the set-membership filter. The APA with data selective method proposed in [20] and [21] decides whether to use the new input vectors or not based on their impact on estimated condition number of the input correlation matrix. The APA with selective regressors (SR-APA) is proposed in [19], where a criterion is developed to decide the rank of each input regressor vector in contributing to the convergence and a fixed number of the highest rank input regressors are selected at each iteration to update the filter coefficients. This algorithm is effective when the input signal is extremely correlated and a large number of input regressors are used in the selection process. The APA with dynamic selection of input vectors (DS-APA) [18] proposes to select a dynamic number of input regressors based on a criterion devised to maximize gradient of the mean square deviation. A similar algorithm is proposed in [17] that uses two different projection orders in different stages of adaptation. The APA with evolving order (E-APA) [16] tunes the projection order relying on an approximation of the theoretical estimate given in [2] for the steady-state MSE of the APA. Both E-APA and DS-APA approximate expectations of the associated stochastic quantities with their instantaneous values.

The major contributions of the paper are summarized below:

- We propose a new affine projection algorithm with selective projections (APA-SP) that changes the projection order in accordance with the adaptation state and selects the best input regressors to update the filter coefficients at each

iteration. Our approach is different from the ones taken in the previous variable-projection-order APAs and is based on both theoretical and experimental findings. Although the proposed algorithm shares some features with the previously proposed algorithms, its performance is considerably improved while its average computational complexity is also appreciably reduced.

- We also introduce a new improved technique to estimate the variance of the filter output error, which provides improved estimation compared to the conventional ones. In fact, the introduced estimation technique plays a key role in the proposed algorithm and has a significant contribution to its success.

We review the conventional APA in Section 2. In Section 3, we pose the problem of optimizing the projection order and investigate possible theoretical approaches to its solution. In section 4, we describe the proposed algorithm. Simulation studies and discussions are presented in section 5 and the paper concludes in section 6.

## 2. Affine projection algorithm

Let us consider data  $\{d(n)\}$  that arise from the model

$$d(n) = \mathbf{w}_o^* \mathbf{x}(n) + v(n) \quad (1)$$

where  $\mathbf{w}_o$  is an unknown  $L \times 1$  column vector that is to be estimated,  $v(n)$  represents measurement noise,  $\mathbf{x}(n)$  is the  $L \times 1$  input regressor vector

$$\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-L+1)]^T,$$

and the superscripts  $T$  and  $*$  denote transpose and conjugate transpose of a matrix, respectively. The affine projection algorithm calculates  $\mathbf{w}(n)$ , an estimate for  $\mathbf{w}_o$  at iteration  $n$ , via

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \mathbf{X}(n)(\mathbf{X}^*(n)\mathbf{X}(n) + \delta \mathbf{I}_k)^{-1} \mathbf{e}^*(n) \quad (2)$$

where  $\mu$  is the step-size,  $\delta$  is the regularization parameter,  $\mathbf{I}_k$  is a  $k \times k$  identity matrix,  $k$  is the projection order,

$$\mathbf{X}(n) = [\mathbf{x}(n), \mathbf{x}(n-1), \dots, \mathbf{x}(n-k+1)],$$

$$\mathbf{d}(n) = [d(n), d(n-1), \dots, d(n-k+1)],$$

and

$$\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{w}^*(n)\mathbf{X}(n). \quad (3)$$

Defining weight error vector as  $\tilde{\mathbf{w}}(n) = \mathbf{w}_o - \mathbf{w}(n)$ , we have

$$\tilde{\mathbf{w}}(n+1) = \tilde{\mathbf{w}}(n) - \mu \mathbf{X}(n)(\mathbf{X}^*(n)\mathbf{X}(n) + \delta \mathbf{I}_k)^{-1} \mathbf{e}^*(n). \quad (4)$$

Subsequently, the gradient of the mean square deviation (MSD) can be defined and calculated as

$$\begin{aligned}\Delta(n) &= E[\|\tilde{\mathbf{w}}(n)\|^2] - E[\|\tilde{\mathbf{w}}(n+1)\|^2] \\ &= 2\mu \text{Re} \left( E \left[ \mathbf{e}(n) (\mathbf{X}^*(n) \mathbf{X}(n))^{-1} \mathbf{X}^*(n) \tilde{\mathbf{w}}(n) \right] \right) \\ &\quad - \mu^2 E \left[ \mathbf{e}(n) (\mathbf{X}^*(n) \mathbf{X}(n))^{-1} \mathbf{e}^*(n) \right]\end{aligned}\quad (5)$$

by squaring both sides of (4), taking expectations, and neglecting  $\delta \mathbf{I}_k$  [8]. Here,  $E[\cdot]$  is the expectation operator,  $\|\cdot\|$  stands for the Euclidean norm of a vector and  $\text{Re}(\cdot)$  denotes the real part of a complex quantity.

### 3. Optimizing the projection order

Fig. 1 plots the MSD gradient of the APA with different projection orders ( $k = 1, 2, 3$ , and 4) in a typical system identification set-up with  $L = 16$ ,  $\mu = 0.5$ , SNR = 10 dB, Gaussian AR1(0.95) input and ensemble-averaging over  $10^4$  independent trials. The Gaussian AR1(0.95) input signal is generated by passing white Gaussian noise with zero mean and unit variance through the first-order autoregressive system

$$H_1(z) = \frac{1}{1 - 0.95z^{-1}}.$$

It is clear from Fig. 1 that at the early stages of adaptation, larger projection orders lead to larger  $\Delta(n)$ , but near the steady state, larger projection orders produce smaller  $\Delta(n)$ . Accordingly, the circles in Fig. 1 indicate the time instances at which the use of a smaller projection order leads to a greater  $\Delta(n)$  and consequently faster convergence.

These observations corroborate that in order to achieve the fastest convergence and the best tracking capability, the projection order should be varied in a suitable manner to maximize  $\Delta(n)$  during the entire filtering process. This leads to MSD undergoing the largest descent at each iteration and also guarantees the achievement of the lowest possible steady-state MSE [4]. Moreover, optimizing the projection order in an efficient way can potentially yield significant computational saving. From an analytical standpoint, an optimization problem can be formulated as

$$\kappa_o(n) = \arg \max_{1 \leq k \leq k_{max}} \Delta(n) \quad (6)$$

where  $\kappa_o(n)$  is the optimum time-varying projection order and is bounded between 1 and a predefined maximum value  $k_{max}$ .

It is worthwhile to note that in [8] the projection order is considered fixed and a time-varying step-size rule is derived based on maximization of  $\Delta(n)$ . Here, we consider a fixed step-size and let the projection order vary. Unlike the approach of [8], a variable

projection order can provide considerable reduction in the average computational complexity of the APA. However, it results in a non-uniform complexity that may change in time and even with different statistical properties of the input signal and/or the target system while complexity of the APA with a variable step-size is constant over time and for any signal/system statistics.

Solving the optimization problem of (6) at every iteration is practically infeasible since exact calculation of  $\Delta(n)$  requires prior knowledge of  $\mathbf{w}_o$  and signal statistics to evaluate the associated expectations. If we assume that the noise sequence  $v(n)$  is independent and identically distributed (i.i.d.) with zero mean and variance  $\sigma_v^2$  and statistically independent of the regression data  $\{\mathbf{X}(n)\}$  and ignore the dependence of  $\mathbf{w}(n)$  on the past noise, we can rewrite (5) as

$$\begin{aligned} \Delta(n) \approx & \mu(2 - \mu)E \left[ \mathbf{e}(n)(\mathbf{X}^*(n)\mathbf{X}(n))^{-1} \mathbf{e}^*(n) \right] \\ & - 2\mu\sigma_v^2 \text{Tr} \left( E \left[ (\mathbf{X}^*(n)\mathbf{X}(n))^{-1} \right] \right) \end{aligned} \quad (7)$$

where  $\text{Tr}(\cdot)$  is the matrix trace operator [18]. By approximating the expectations with their instantaneous estimates in (7), we have

$$\Delta(n) = \mu(2 - \mu)\mathbf{e}(n)(\mathbf{X}^*(n)\mathbf{X}(n))^{-1} \mathbf{e}^*(n) - 2\mu\sigma_v^2 \text{Tr} \left( (\mathbf{X}^*(n)\mathbf{X}(n))^{-1} \right) \quad (8)$$

The optimization problem in (6) with the simplified  $\Delta(n)$  of (8) is still computationally prohibitive because it implies that  $k_{max}$  independent APAs are required to run in parallel during the adaptation process in order to evaluate  $\Delta(n)$  for each  $k$  and compare them to decide the optimum projection order. The DS-APA of [18] is based on further approximation of (8) by assuming that the diagonal components of  $\mathbf{X}^*(n)\mathbf{X}(n)$  are much larger than its off-diagonal components.

#### 4. APA with selective projections

As discussed above, any exact analytical solution for the optimization problem of (6) will be clearly either impracticable or too costly to implement in real-world applications. In addition, approximate approaches, despite being still expensive, usually fail when the assumptions made in their derivations do not hold. Thus, based on analytical and experimental insights gained from extensive investigations and simulations in different adaptive filtering settings, we propose a new scheme to control the projection order of the APA by monitoring variations of the filter output error, which is considered to be indicative of the adaptation state. This scheme indirectly attempts to maximize  $\Delta(n)$  at each time instant. In what follows, we explain the proposed scheme and its rationale.

#### 4.1. Variable projection order

It has been proven that the slope of the APA's learning curve (convergence rate) is directly related to the projection order [2], [3]. This means that a larger projection order leads to a faster convergence. In the presence of background noise and after the initial convergence when the variance of the filter output error becomes comparable to the background noise power, a large projection order degrades the performance and increases the steady-state MSE by amplifying the background noise. On the other hand, a small projection order drives the algorithm towards convergence slower at early stages of the adaptation, but achieves a lower steady-state MSE as a result of reduced noise amplification. Hence, in order to have a fast convergence and reach a low steady-state MSE, the projection order should be set as large as possible at the start-up and kept large as long as the variance of the filter output error is considerably higher than the noise power. Then, when the algorithm approaches its steady state and the variance of the error becomes comparable to the noise power, the projection order should be reduced to counter the noise amplification. This leads to a smaller steady-state MSE as well as reduced complexity.

It is inferred from the above discussion that to achieve a desirable performance, the projection order needs to be changed depending on the variance of the filter output error. Of course, this dependence should be formulated in a proper way. Recalling that the learning curve of the APA exhibits an exponential decay in time, evidenced by both theoretical analyses (see, e.g., [2]–[4]) and simulations (see, e.g., Fig. 5), we propose to define a linear dependence of the projection order on the logarithm of the filter output error variance. In other words, we let the projection order,  $\kappa(n)$ , take integer values between 1 and  $k_{max}$  at each time instant while depending linearly on the logarithm of the filter output error variance,  $\sigma_e^2(n)$ . The projection order takes its highest value,  $k_{max}$ , as long as the error variance is higher than a threshold, defined as  $\gamma$ , and reduces to 1 when the error variance is equal or lower than the predicted steady-state MSE,  $\eta$ . Between these thresholds, the projection order is computed depending linearly on the logarithm of the error variance. We can formulate this rule as

$$\frac{\tilde{\kappa}(n) - 1}{k_{max} - 1} = \frac{\ln \sigma_e^2(n) - \ln \eta}{\ln \gamma - \ln \eta} \quad (9)$$

or

$$\tilde{\kappa}(n) = 1 + (k_{max} - 1) \frac{\ln(\sigma_e^2(n)/\eta)}{\ln(\gamma/\eta)} \quad (10a)$$

$$\kappa(n) = \min \{k_{max}, \max \{1, [\tilde{\kappa}(n)]\}\} \quad (10b)$$

where  $\lceil x \rceil$  denotes the smallest integer greater than or equal to  $x$ . We refer to the APA with time-varying projection order given by (10) as *affine projection algorithm with selective projections* (APA-SP). The steady-state MSE of the APA,  $\eta$ , is predicted using (14)–(16) given in Section 4.2 and the error variance,  $\sigma_e^2(n)$ , is estimated by means of the technique described in Section 4.3. According to the rule of (10), the projection order is reduced when the algorithm converges (i.e.,  $\sigma_e^2(n)$  decreases) and it is increased when the algorithm moves away from the steady state and  $\sigma_e^2(n)$  increases. The latter is usually the case with time-varying target systems and/or non-stationary input signals.

The threshold  $\gamma$  determines at what level of  $\sigma_e^2(n)$  the algorithm starts switching the projection order. With a lower  $\gamma$ , during the initial convergence, the algorithm starts to switch the projection order later. Consequently, the algorithm runs with a projection order of  $k_{max}$  for a longer time. This ensures a fast convergence but results in less computational saving. Conversely, a higher  $\gamma$  means the algorithm starts to switch the projection order earlier, hence providing more saving in the computations. When  $k_{max}$  is small, it is beneficial to have a small  $\gamma$  so that the algorithm maintains a fast initial convergence. On the other hand, with a relatively large  $k_{max}$ , a larger  $\gamma$  is preferable since it leads to more reduction in complexity without sacrificing the convergence performance. Accordingly, defining  $\sigma_d^2$  as the power of the reference signal  $d(n)$ <sup>1</sup>, we compute  $\gamma$  in such a way that it takes values between

$$\gamma_{max} = \sqrt{\eta\sigma_d^2} \quad \text{or} \quad \ln \gamma_{max} = \frac{\ln \sigma_d^2 + \ln \eta}{2}$$

for  $k_{max} \geq 8$  and  $\gamma_{min} = \eta$  for  $k_{max} \leq 2$  while its logarithm depends linearly on  $k_{max}$  for  $2 < k_{max} < 8$ . This linear relationship between  $\ln \gamma$  and  $k_{max}$  can be expressed as

$$\frac{\ln \gamma - \ln \eta}{\frac{\ln \sigma_d^2 + \ln \eta}{2} - \ln \eta} = \frac{k_{max} - 2}{8 - 2}, \quad 2 \leq k_{max} \leq 8. \quad (11)$$

Subsequently, the threshold  $\gamma$  is computed in the linear scale via

$$\gamma = \eta \left( \frac{\sigma_d^2}{\eta} \right)^z \quad (12a)$$

where

---

<sup>1</sup> In fact,  $\sigma_d^2$  is the initial value for the error variance estimate,  $\sigma_e^2(0)$ , since we initialize the algorithm to  $\mathbf{w}(0) = \mathbf{0}$  so  $e(0) = d(0) - \mathbf{w}^*(0)\mathbf{x}(0) = d(0)$ . The reference signal power, if not known *a priori*, can be estimated using first  $k_{max}$  samples of  $d(n)$  during initial population of the input regression matrix.

$$z = \max\left\{0, \min\left\{\frac{1}{2}, \frac{k_{max} - 2}{12}\right\}\right\}. \quad (12b)$$

Note that with the chosen  $\gamma_{max}$ , the algorithm starts switching the projection order when the estimated error variance is halfway between its initial and final values in the logarithmic scale.

Fig. 2 illustrates the proposed rule for controlling the projection order by depicting the time-varying continuous projection order,  $\tilde{\kappa}(n)$ , as a linear function of the logarithm of the error variance,  $\ln\sigma_e^2(n)$ , together with the bounded and quantized integer projection order,  $\kappa(n)$ .

From a practical point of view, we may plan to adjust the variable projection order in a manner that it takes power-of-two values, i.e., 1, 2, 4, 8, ... Therefore, we can replace (10b) with

$$\kappa(n) = \min\{k_{max}, 2^{\lceil \log_2 \max\{1, \tilde{\kappa}(n)\} \rceil}\}. \quad (13)$$

This equation resembles a quantization function and can be evaluated by using a lookup table with a very low computational overhead. As an example, for  $k_{max} = 28$ , in accordance with the adaptation state,  $\kappa(n)$  can take on values of 28, 16, 8, 4, 2, and 1.

#### 4.2. Prediction of the steady-state MSE

In order to predict the steady-state MSE of the APA, two expressions have been reported in [2] and [22, eq. 40]. The former is

$$\begin{aligned} \eta' &= \sigma_v^2 \left( 1 + \frac{\mu}{2 - \mu} \text{Tr}(\mathbf{R}_x) E \left[ \frac{k}{\|\mathbf{x}\|^2} \right] \right) \\ &\approx \sigma_v^2 \left( 1 + \frac{\mu k}{2 - \mu} \right) \end{aligned} \quad (14)$$

where  $\mathbf{R}_x$  is the input autocorrelation matrix and the latter is

$$\begin{aligned} \eta'' &= \sigma_v^2 \left( 1 + \frac{\mu k L}{L - 2\mu k + \mu k L} \right) \\ &\approx \sigma_v^2 \left( \frac{1 + 2\mu k}{1 + \mu k} \right). \end{aligned} \quad (15)$$

Comparison of  $\eta'$  and  $\eta''$  against simulation results reveals that, in most cases, one overestimates the actual value of the steady-state MSE while the other one underestimates it. Thus, taking the geometric mean of these two values, i.e.,

$$\eta = \sqrt{\eta' \eta''} \quad (16)$$

results in a much better prediction of the steady-state MSE. This can be interpreted in view of the fact that  $\eta'$  and  $\eta''$  are derived using different assumptions, e.g., white *a priori* error vector assumption in [2, A.2] versus white input signal assumption in [22].

In Fig. 3, the experimental normalized steady-state MSE of the APA along with  $\eta'$ ,  $\eta''$ , and  $\sqrt{\eta'\eta''}$  are depicted for  $k = 2, 3, \dots, 35$  and  $\mu = 0.01$  and  $0.1$  for a typical system identification scenario with  $L = 100$  randomly-generated taps. The input signal is Gaussian AR1(0.95) and the experimental steady-state MSE is measured by time-averaging the instantaneous squared errors over  $10^3$  time instants at the steady state and ensemble-averaging over  $10^4$  independent trials. It is seen that  $\sqrt{\eta'\eta''}$  is more accurate than both  $\eta'$  and  $\eta''$ .

### 4.3. Estimation of the error variance

The most commonly used method for estimating the error variance is the well-known exponential window approach [8]–[10]:

$$\sigma_e^2(n) = \lambda\sigma_e^2(n-1) + (1-\lambda)e^2(n) \quad (17)$$

where  $\lambda$  is the forgetting factor and the initial value is  $\sigma_e^2(0) = \sigma_d^2$ . Although this method is widely used, its accuracy is very sensitive to the value of  $\lambda$  and it may fail to yield a reliable estimate in all stages of adaptation mainly because of using a fixed  $\lambda$  [12], [23]. A smaller value of  $\lambda$  allows better tracking during the transient state and a larger value of  $\lambda$  results in smoother estimation close to and at the steady state. Hence, we propose to control  $\lambda$  using a rule similar to (10), viz., to reduce it at the transient state and increase it when approaching the steady state. Therefore, the time-varying forgetting factor is adapted via

$$\tilde{\alpha}(n) = \alpha_{min} + (\alpha_{max} - \alpha_{min}) \frac{\ln(\sigma_e^2(n-1)/\sigma_d^2)}{\ln(\eta/\sigma_d^2)} \quad (18a)$$

$$\alpha(n) = \min \{ \alpha_{max}, \max \{ \alpha_{min}, \tilde{\alpha}(n) \} \} \quad (18b)$$

where  $\alpha_{max}$  and  $\alpha_{min}$  are maximum and minimum allowed values for  $\alpha(n)$ , respectively. Experiments suggest that selecting suitable values for  $\alpha_{max}$  and  $\alpha_{min}$  is a more tractable task than choosing an appropriate fixed  $\lambda$  for (17). Furthermore, incorporating a simple moving average filtering scheme with the exponential window approach improves the estimation accuracy appreciably at the expense of a slight increase in computations. Therefore, we modify (17) to

$$\sigma_e^2(n) = \alpha(n)\sigma_e^2(n-1) + (1-\alpha(n)) \cdot \text{SMA}\{e^2(n), N_m\} \quad (19)$$

where the simple moving averaging (SMA) is implemented as

$$\text{SMA}\{e^2(n), N_m\} = (e^2(n) + e^2(n-1) + \dots + e^2(n - N_m + 1))/N_m. \quad (20)$$

Here,  $N_m$  is the moving average window size that is usually set to  $L/2$ . It is easy to show that (19) simplifies to

$$\begin{aligned} \sigma_e^2(n) = & \frac{1 - \alpha(n)\alpha(n-1)}{1 - \alpha(n-1)} \sigma_e^2(n-1) + \frac{\alpha(n-1)(\alpha(n) - 1)}{1 - \alpha(n-1)} \sigma_e^2(n-2) \\ & + \frac{(1 - \alpha(n))}{N_m} (e^2(n) - e^2(n - N_m)). \end{aligned} \quad (21)$$

Fig. 4 shows the curves of estimated error variance employing different estimation techniques for a single realization of the APA in a typical system identification experiment. The mean square error obtained by ensemble-averaging over  $10^4$  independent trials is also shown in Fig. 4. Both the target system and the adaptive filter are assumed to have 32 taps, step-size is  $\mu = 0.5$ , the fixed projection order is  $k = 16$ , noise power is  $\sigma_v^2 = 10^{-3}$ , and input signal is Gaussian AR1(0.95). Simulated estimation techniques are the exponential window approach (EWA) with two different values for the forgetting factor ( $\lambda = 0.8$  and  $0.95$ ), EWA with a variable forgetting factor (EWA+VA) where  $0.2 \leq \alpha(n) \leq 0.99$  and EWA+VA incorporating a simple moving averaging (EWA+VA+SMA). As seen in Fig. 4, with a large value of  $\lambda$ , the EWA is biased and overestimates the error variance at the transient state. On the other hand, with a small value of  $\lambda$ , the EWA yields a good estimation during the transient state but its estimation has large fluctuations around the true MSE at the steady state. A variable forgetting factor in the EWA improves the estimation accuracy. Moreover, incorporating the SMA into the EWA+VA provides further improvement by smoothing the variance estimate.

#### 4.4. Estimation of the noise variance

Thus far, we have assumed that the background (or measurement) noise has a fixed and known variance. This is a realistic assumption in certain applications such as acoustic or network echo cancelation where the noise power can be measured during silent periods. However, if the noise power varies in time, it can be estimated adaptively using the method proposed in [9]. Squaring (1) and taking the expectation of both sides gives

$$E[v^2(n)] = E[d^2(n)] - E[y^2(n)] \quad (22)$$

where  $y(n) = \mathbf{w}_o^* \mathbf{x}(n)$  is the optimal filter output and we assume that  $y(n)$  and  $v(n)$  are uncorrelated. Supposing that the algorithm has reached a certain level of convergence, we can approximate (22) by

$$E[v^2(n)] = E[d^2(n)] - E[\hat{y}^2(n)] \quad (23)$$

where  $\hat{y}(n) = \mathbf{w}^*(n)\mathbf{x}(n)$  is the actual filter output. Consequently, the noise variance can be estimated via

$$\sigma_v^2(n) = |\sigma_d^2(n) - \sigma_{\hat{y}}^2(n)| \quad (24)$$

while using the exponential window approach to estimate  $\sigma_d^2(n)$  and  $\sigma_{\hat{y}}^2(n)$ . Note that, to avoid negative values of  $\sigma_v^2(n)$  that can be the case when the estimates are poor and  $\sigma_{\hat{y}}^2(n) > \sigma_d^2(n)$ , we take the absolute value in (24). Moreover, since  $\sigma_d^2(n)$  and  $\sigma_{\hat{y}}^2(n)$  are both initialized to the same value at the beginning of the estimation, e.g.,  $\sigma_d^2(0) = 0$  and  $\sigma_{\hat{y}}^2(0) = 0$ , the estimated value of  $\sigma_v^2(n)$  is smaller than its steady-state value during the initial convergence. In the other words,  $\sigma_v^2(n)$  may be underestimated when the algorithm is far from convergence. Under-estimation of the noise variance leads to under-prediction of the steady-state MSE, which can result in delayed down-switching of the projection order. This may increase the complexity but will not affect the performance significantly.

#### 4.5. Selection of the input regressors

When the input signal is extremely correlated (e.g., speech signal), the orthogonality of the immediate past input regressors can deteriorate tremendously. In such cases, it is useful to select and utilize a subset of input regressors that most constructively contribute to maximize the convergence rate. In this regard, two methods have been proposed in [18] and [19] for input regressor selection. In [18], the helpful input regressors are distinguished according to the following criterion

$$|\varepsilon_i(n)| > \sqrt{2/(2-\mu)} \sigma_v, \quad 0 \leq i \leq k_{max} - 1 \quad (25)$$

where  $\varepsilon_i(n)$  is the  $i$ th element of the error vector,  $\mathbf{e}(n)$ , and the input regressor corresponding to this element is  $\mathbf{x}(n-i)$ . Alternatively, one can rank the values of  $|\varepsilon_i(n)|$ :

$$\text{sort} \{|\varepsilon_i(n)|, 0 \leq i \leq k_{max} - 1\} \quad (26)$$

and choose the regressors associated with the largest values for updating the filter coefficients. The other method proposed in [19] ranks the values of  $\varepsilon_i^2(n)/\|\mathbf{x}(n-i)\|^2$ :

$$\text{sort} \{\varepsilon_i^2(n)/\|\mathbf{x}(n-i)\|^2, 0 \leq i \leq k_{max} - 1\} \quad (27)$$

and selects the regressors associated with the largest ones.

We integrate both techniques with the new variable projection order scheme in such a way that the number of used input regressors at each iteration,  $\kappa(n)$ , is determined by the rule of (10) and the  $\kappa(n)$  highest-rank regressors are selected according to (26) or (27). This results in two variants of APA-SP which we call APA-SP1 and APA-

SP2, corresponding to the input regressor selection criteria of (26) and (27), respectively.

## 5. Simulation studies and discussions

In this section, we present simulation results to examine the performance of the proposed algorithm and compare it with the most relevant previously proposed algorithms.

### 5.1. System identification

In the first part of the simulations, we examine identification of an unknown system with 32 taps. We generate the taps randomly and normalize the unknown system to unit energy. We also assume the same number of taps for the corresponding adaptive filters ( $L = 32$ ) and initialize them to zero. It should be noted that, in practice, it is rare to know the exact number of taps in the target system and the adaptive filter performance can degrade if the system order is significantly under- or overestimated. The maximum projection order for the algorithms with variable projection order is  $k_{max} = 16$ , the noise variance  $\sigma_v^2 = 10^{-3}$ , the regularization parameter  $\delta = 10^{-5}$ , step-size  $\mu = 0.5$ , minimum and maximum values for the forgetting factor  $\alpha_{min} = 0.2$  and  $\alpha_{max} = 0.99$ , and the moving average window size  $N_m = 16$ . The ensemble-averaged results are obtained by averaging over  $10^4$  independent trials. Both white and correlated input signals are considered. Correlated input signals are generated by passing a zero-mean unit-variance Gaussian white noise through an autoregressive moving average process ARMA(2,2) with transfer function

$$H_2(z) = \frac{1 - z^{-2}}{1 - 1.70233z^{-1} + 0.71902z^{-2}}.$$

The interesting property of  $H_2(z)$  is that it generates highly correlated signals with the associated autocorrelation matrix having a condition number larger than  $10^5$  [12].

Fig. 5 plots the learning curves of the APA-SP1, the NLMS algorithm (APA with  $k = 1$ ), and the APA with different projection orders ( $k = 2, 4, 8, \text{ and } 16$ ) for white Gaussian input signal with zero mean and unite variance. To highlight the time indices and the MSE levels at which the projection order is switched by APA-SP1, its learning curve is shown in different colors for different values of its time-varying projection order,  $\kappa(n)$ . It is clear that APA-SP1 converges as fast as the APA with the highest projection order and reaches the smallest steady-state MSE which is the same as that of the NLMS algorithm. It is noteworthy that the experiment of Fig. 5 is not intended to imitate a real-world application. It only serves as an example to illustrate how the process of switching the projection order works in the proposed algorithm and how effective it can be. Therefore, the chosen value of the maximum projection order

facilitates the illustration, even though it seems rather unrealistic while the input signal is white.

In Fig. 6, both variants of APA-SP are compared with DS-APA, E-APA, and SR-APA proposed in [16], [18], and [19], respectively. It is claimed in [16] that E-APA outperforms the SM-AP vdr algorithm [14], [15]. However, since set-membership adaptive filtering algorithms are derived using a bounded-error criterion and feature data-discerning parameter updates, comparing them with conventional adaptive filtering algorithms is not straightforward. Therefore, we do not include SM-AP vdr algorithm here. To obtain the results of Fig. 6, the Gaussian ARMA(2,2) signal is used as input, APA's projection order is  $k = 16$ , and the number of input regressors that are selected out of last  $k_{max}$  regressors for SR-APA is set to  $k_{SR} = 12$ . Moreover, in order to compare the capability of the algorithms in tracking sudden system variations, the unknown target system is altered with another randomly generated system, half way through the simulations [24], [25]. Fig. 7 shows the ensemble-averaged time-varying projection order of APA-SP1, APA-SP2, DS-APA, and E-APA for the experiment of Fig. 6. The corresponding curves of projection order versus time for one realization are depicted in Fig. 8. Fig. 9 also shows the time evolution of the variable forgetting factor associated with both variants of APA-SP in the experiment of Fig. 6.

To assess the capability of the new algorithms to adapt to time-varying noise power, the simulations of Fig. 6 are repeated by changing the noise power instead of the system impulse response. Fig. 10 shows the learning curves of different algorithms when SNR is reduced by 10 dB between iterations 800 and 1600 in the experiment of Fig. 6 with a fixed target system. The corresponding curves of ensemble-averaged variable projection order are also shown in Fig. 11. It is observed in Fig. 10 that the gap in MSE between APA-SP and its contenders increases during the change in the noise power since, unlike the other algorithms, APA-SP can adaptively track changes in the background noise power. Moreover, Fig. 11 shows that the projection order of APA-SP is switched up when the noise power suddenly changes as a result of interim inaccuracy in noise power estimation. The projection order is then switched down as the accuracy of the adaptive noise power estimation improves quickly.

## 5.2. Acoustic echo cancellation

In this part, we consider an acoustic echo cancellation problem where the objective of adaptive filtering is to estimate the acoustic echo path. The considered echo path is a measured car echo impulse response of length 256 [26, Fig. 1]. Accordingly, the associated adaptive filter also has  $L = 256$  taps. The input is a real speech signal sampled at 8 kHz and quantized to 16 bits. The other simulated parameters are  $\mu = 0.5$ ,  $k_{max} = 16$  (for APA-SP, E-APA, DS-APA, and SR-APA),  $k = 16$  (for APA),  $k_{SR} = 12$  (for SR-APA),  $\delta = 10^{-6}$ , SNR = 40 dB,  $N_m = 128$ ,  $\alpha_{min} = 0.1$ , and

$\alpha_{max} = 0.95$ . Figs. 12 and 13 compare normalized misalignment, i.e.,  $\|\mathbf{w}_o - \mathbf{w}(n)\|^2 / \|\mathbf{w}_o\|^2$ , and segmental *echo return loss enhancement* (ERLE) of different algorithms, respectively. The input signal is also depicted on top of Figs. 12 and 13. The ERLE is defined as

$$\text{ERLE} = 10 \log_{10} \frac{E[d^2(n)]}{E[e^2(n)]}$$

and the segmental ERLE estimates are obtained by averaging  $d^2(n)$  and  $e^2(n)$  over 80 ms segments (640 samples). The time-varying projection orders of different algorithms are shown in Fig. 14. In the presented results, both variants of APA-SP stand out with their superior performance. Moreover, it is seen that the performance of the algorithms might seriously deteriorate when the input signal is not persistently exciting.

To examine the tracking performance of the proposed algorithm, we assume a time-varying echo path impulse response. To represent the time variations, we employ the modified first-order Markov model [27]

$$\mathbf{w}_o(n) = \zeta \mathbf{w}_o(n-1) + \sqrt{1 - \zeta^2} \mathbf{q}(n)$$

where the parameter  $\zeta$  determines the rate of the variations and  $\mathbf{q}(n)$  is a vector of Gaussian random variables with zero mean and unit variance. As reported in [22], a source moving at 0.2 m/s for acoustic impulse responses generated using the method of images [28] approximately corresponds to  $\zeta = 0.9999$  in the above model. Fig. 15 shows the normalized misalignment curves of different algorithms when the input is a stationary signal with a speech-like spectrum and the echo path impulse response varies in time with a rate of  $\zeta = 0.9999$ . The other simulated parameters are  $\mu = 0.5$ ,  $k_{max} = 8$ ,  $k = 8$ ,  $k_{SR} = 6$ ,  $\delta = 10^{-4}$ ,  $\text{SNR} = 30$  dB,  $N_m = 64$ ,  $\alpha_{min} = 0.1$ , and  $\alpha_{max} = 0.99$ . Fig. 16 also shows time variations of the projection order for different algorithms.

### 5.3. Discussion

It is clearly seen in the simulations that both variants of the new algorithm, APA-SP1 and APA-SP2, outperform their contenders in terms of convergence rate and steady state error. In addition, they tune the projection order in a better way than E-APA and DS-APA. In particular, they reduce to the NLMS algorithm at the steady state whereas the other algorithms lack this desirable property. Reducing to the NLMS algorithm, in addition to decreasing the steady-state misalignment, alleviates the average computational load dramatically since the NLMS algorithm needs no matrix inversion (or solution of any linear system of equations). This superior performance emerges prominently for highly-correlated and non-stationary excitations.

Bearing in mind that E-APA relies on an approximation of the steady-state MSE that loses its accuracy in practical applications with highly-correlated and non-stationary input signals and the fact that DS-APA is built on weak assumptions, which do not hold for realistic scenarios, we attribute the superiority of the proposed algorithm to two aspects:

- the underpinning idea of tuning the projection order in a sensible manner by monitoring the adaptation state and
- attaining a good estimate of the error variance,  $E[e^2(n)]$ , which faithfully represents the adaptation state, rather than relying on the instantaneous value of the squared error,  $e^2(n)$ , which is the case in E-APA and DS-APA.

There is another interesting observation that, in most of the simulations, APA-SP2 performs only slightly better than APA-SP1, though it imposes more computational overhead. For stationary signals and fixed target systems, both variants of the proposed algorithm exhibit virtually the same performance.

The complexity of the algorithms considered in the paper is compared in Table 1 in terms of the number of required multiplications per iteration. The average required number of multiplications per iteration in the experiments of Figs. 12 and 15 is also presented in Table 1.

## 6. Conclusion

A new affine projection algorithm was proposed with a variable projection order that is controlled by the estimated error variance. The error variance is estimated using a new method, which takes advantage of both exponential window and simple moving average schemes in addition to utilizing a variable forgetting factor. Two regressor selection criteria were also adopted to facilitate selective regressor updating. In addition to its superior performance in terms of high initial convergence rate and low steady-state misalignment, the new algorithm is able to track system variations effectively. Reduced average computational burden is one of the most significant benefits of the proposed algorithm as it intends to minimize the projection order without sacrificing the convergence performance. This reduction is more noticeable in applications, such as acoustic echo cancelation, where long adaptive filters with large projection orders are required. In addition, the proposed algorithm does not rely on the *a priori* knowledge of the background noise since it can be estimated adaptively.

## References

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

- [2] H.-C. Shin and A. H. Sayed, "Mean-square performance of a family of affine projection algorithms," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 90–102, Jan. 2004.
- [3] S. G. Sankaran and A. A. (L.) Beex, "Convergence behavior of affine projection algorithms," *IEEE Trans. Signal Process.*, vol. 48, no. 4, pp. 1086–1096, Apr. 2000.
- [4] S. J. M. de Almeida, J. C. M. Bermudez, N. J. Bershad, and M. H. Costa, "A statistical analysis of the affine projection algorithm for unity step size and autoregressive inputs," *IEEE Trans. Circuits Syst. I*, vol. 52, pp. 1394–1405, Jul. 2005.
- [5] M. Rupp, "A family of adaptive filter algorithms with decorrelating properties," *IEEE Trans. Signal Process.*, vol. 46, no. 3, pp. 771–775, Mar. 1998.
- [6] A. A. (L.) Beex and S. G. Sankaran, "On affine projection direction vector (non-)whiteness," in *Proc. 38th Asilomar Conf.*, Pacific Grove, CA, 2004, vol. 1, pp. 251–258.
- [7] D. T. M. Slock, "The block underdetermined covariance (BUC) fast transversal filter (FTF) algorithm for adaptive filtering," in *Proc. 26th Asilomar Conf.*, Pacific Grove, CA, 1992, pp. 550–554.
- [8] H.-C. Shin, A. H. Sayed, and W.-J. Song, "Variable step-size NLMS and affine projection algorithm," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 132–135, Feb. 2004.
- [9] C. Paleologu, J. Benesty, and S. Ciochină "A variable step-size affine projection algorithm designed for acoustic echo cancellation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 16, no.8, pp. 1466–1478, Nov. 2008.
- [10] L. Ligang, M. Fukumoto, and Z. Shiyong, "A variable step-size proportionate affine projection algorithm for network echo cancellation," in *Proc. 16th Int. Conf. on Digital Signal Process.*, Santorini, Greece, Jul. 2009, pp. 1–6.
- [11] C. Paleologu, S. Ciochină, and J. Benesty, "An efficient proportionate affine projection algorithm for echo cancellation," *IEEE Signal Process. Lett.*, vol. 17, no. 2, pp. 165–168, Feb. 2010.
- [12] H. Rey, L. Rey Vega, S. Tressens, and J. Benesty, "Variable explicit regularization in affine projection algorithm: Robustness issues and optimal choice," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2096–2108, May 2007.
- [13] Y.-S. Choi, H.-C. Shin, and W.-J. Song, "Adaptive regularization matrix for affine projection algorithms," *IEEE Trans. Circuits Syst. II*, vol. 54, no. 12, pp. 1087–1091, Dec. 2007.

- [14] S. Werner, P. S. R. Diniz, and J. E. W. Moreira, "Set-membership affine projection algorithm with variable data-reuse factor," in *Proc. 2006 IEEE Int. Symp. Circuits Syst.*, Island of Kos, Greece, May 2006, pp. 261–264.
- [15] S. Werner, J. A. Apolinario Jr., and P. S. R. Diniz, "Set-membership proportionate affine projection algorithm," *EURASIP J. Audio Speech Music Process.*, vol. 2007, Article ID 34242, 10 pages, 2007.
- [16] S.-E. Kim, S.-J. Kong, and W.-J. Song, "An affine projection algorithm with evolving order," *IEEE Signal Process. Lett.*, vol. 16, no. 11, pp. 937-940, Nov. 2009.
- [17] N.-W. Kong, M.-S. Chang, P.-G. Park, and S.-W. Kim, "An affine projection algorithm with two numbers of input vectors," in *Proc. Fifth Int. Conf. Fuzzy Syst. & Knowledge Discovery*, vol. 2, Jinan, China, Oct. 2008, pp. 272–275.
- [18] S.-J. Kong, K.-Y. Hwang, and W.-J. Song, "An affine projection algorithm with dynamic selection of input vectors," *IEEE Signal Process. Lett.*, vol. 14, no. 8, pp. 529–532, Aug. 2007.
- [19] K.-Y. Hwang and W.-J. Song, "An affine projection adaptive filtering algorithm with selective regressors," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 54, no. 1, pp. 43–46, Jan. 2007.
- [20] C.-W. Lee, Y.-K. Lee, M.-S. Chang, S.-W. Kim, and P.-G. Park, "A new affine projection algorithm with data selective method," in *Proc. 2006 SICE-ICASE Int. Joint Conf.*, Busan, Korea, Oct. 2006, pp. 1701–1704.
- [21] C.-W. Lee, Y.-K. Lee, and S.-W. Kim, "An affine projection algorithm with a data-selective method of using the condition number," *Signal Processing*, vol. 88, no. 5, pp. 1289–1296, May 2008.
- [22] A. W. H. Khong and P. A. Naylor, "Selective-tap adaptive filtering with performance analysis for identification of time-varying systems," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 15, no.5, pp. 1681–1695, Jul. 2007.
- [23] Y. Fan and J. Zhang, "Variable step-size affine projection algorithm with exponential smoothing factors," *Electron. Lett.*, vol. 45, no. 17, pp. 911–913, Aug. 2009.
- [24] C. Breining, P. Dreiscitel, E. Hänslér, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, "Acoustic echo control. An application of very-high-order adaptive filters," *IEEE Signal Process. Magazine*, vol. 16, no. 4, pp. 42–69, Jul. 1999.

- [25] E. Hänsler and G. Schmidt, eds., *Topics in Acoustic Echo and Noise Control*, Berlin: Springer, 2006.
- [26] K. Doğançay and O. Tanrikulu, "Adaptive filtering algorithms with selective partial updates," *IEEE Trans. Circuits Syst. II*, vol. 48, no. 8, pp. 762–769, Aug. 2001.
- [27] N. J. Bershad, S. McLaughlin, and C. F. N. Cowan, "Performance comparison of RLS and LMS algorithms for tracking a first order Markov communication channel," in *Proc. IEEE Int. Symp. Circuits Syst.*, 1990, vol. 1, pp. 266-270.
- [28] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943-950, Apr. 1979.

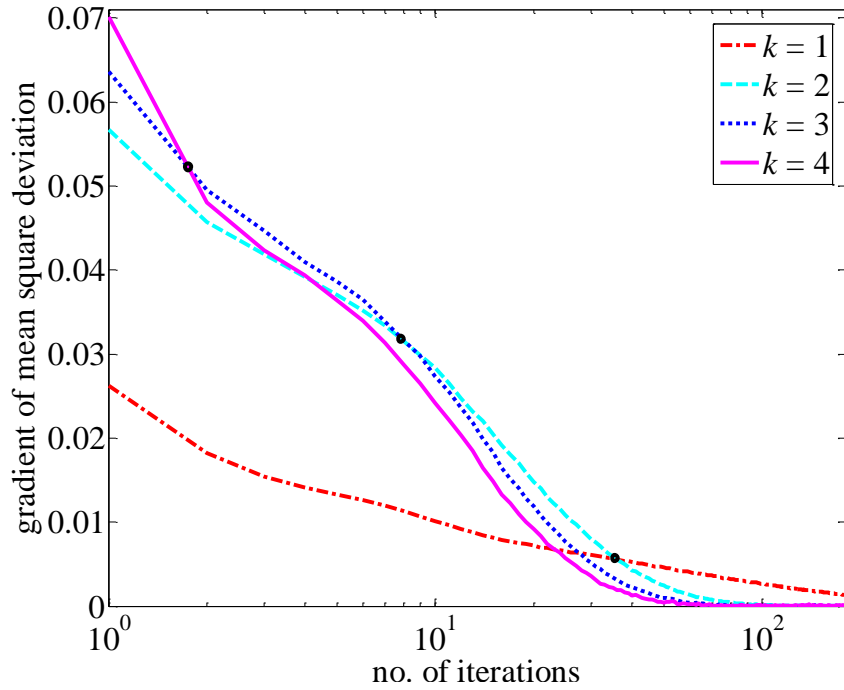


Fig. 1, MSD gradient of the APA with different projection orders in a typical system identification set-up with  $L = 16$ ,  $\mu = 0.5$ , SNR = 10 dB, Gaussian AR1(0.95) input signal, and ensemble-averaged over  $10^4$  independent trials.

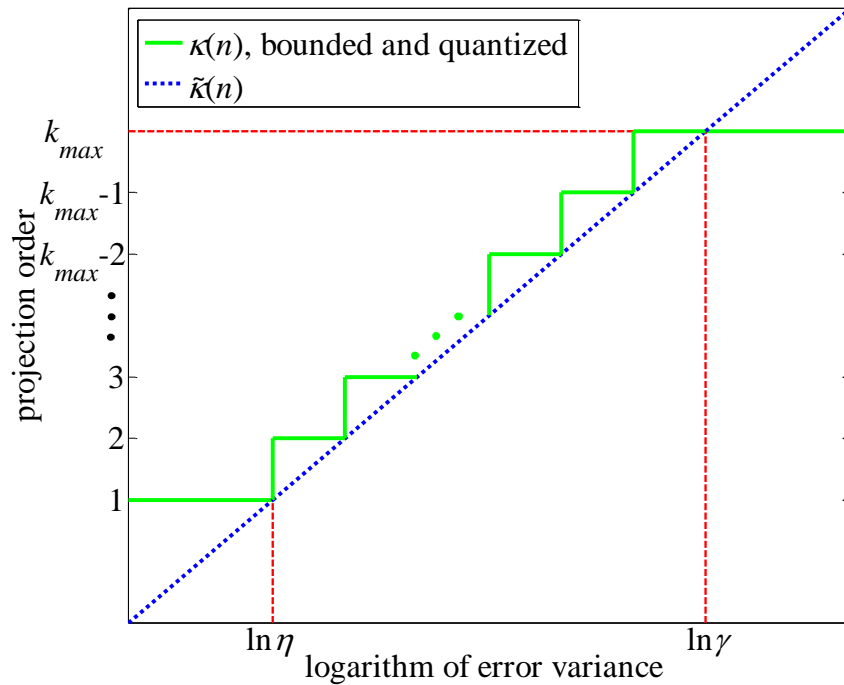
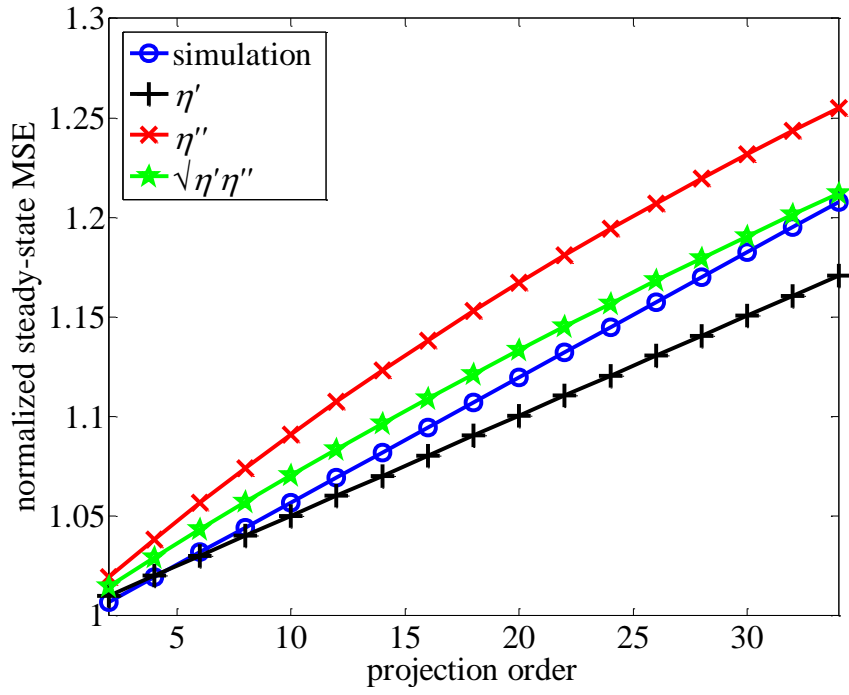
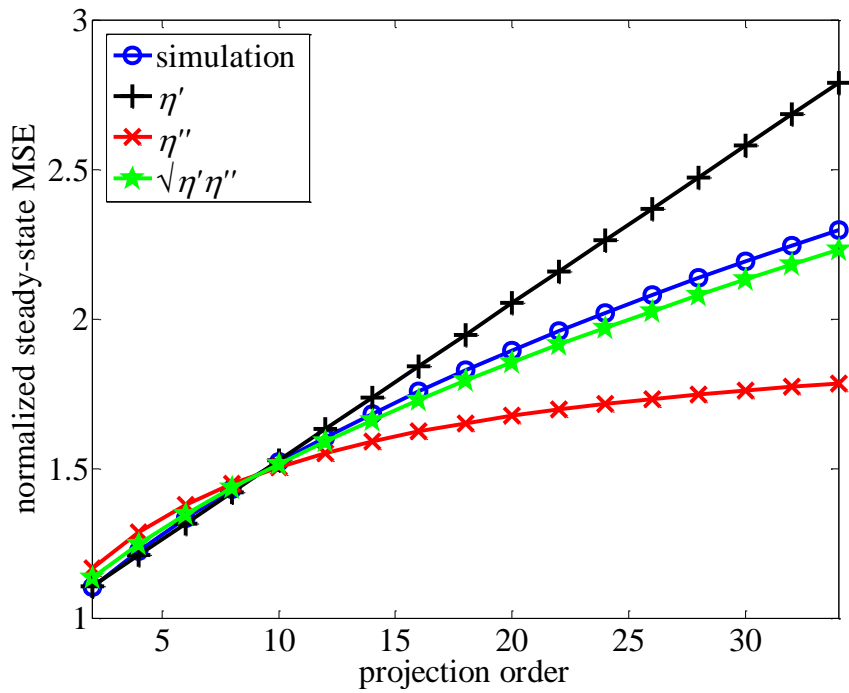


Fig. 2, Time-varying continuous projection order,  $\tilde{\kappa}(n)$ , as a linear function of the logarithm of the error variance,  $\ln \sigma_e^2(n)$ , together with the bounded and quantized integer projection order,  $\kappa(n)$ . The projection order depends on the error variance according to the rule of (10).



(a)



(b)

Fig. 3, Predicted and experimentally measured normalized steady-state MSE of the APA in a system identification scenario with  $L = 100$ , Gaussian AR1(0.95) input, and (a)  $\mu = 0.01$ , and (b)  $\mu = 0.1$ . Measured data is time-averaged over  $10^3$  steady-state values and ensemble-averaged over  $10^4$  runs.

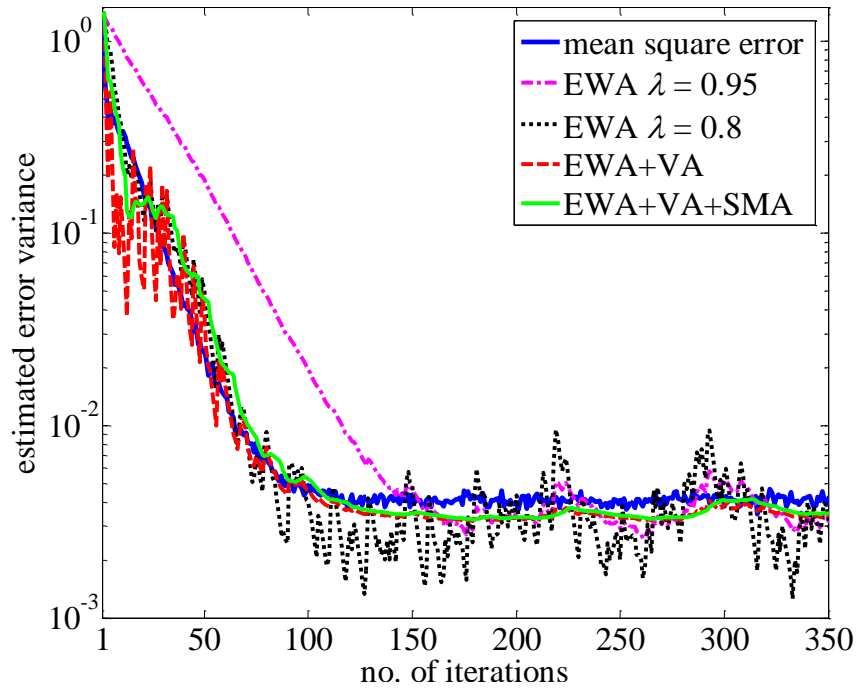


Fig. 4, The curves of estimated error variance employing different estimation techniques for a single run of the APA together with its mean square error ensemble-averaged over  $10^4$  independent runs in a typical system identification scenario with  $L = 32$ ,  $\mu = 0.5$ ,  $k = 16$ ,  $\sigma_v^2 = 10^{-3}$ , and Gaussian AR1(0.95) input.

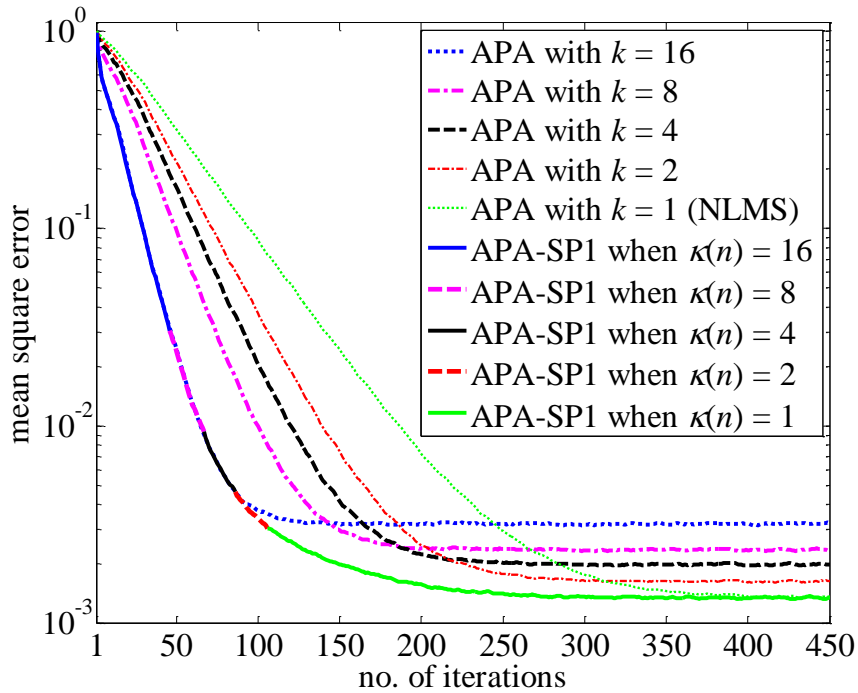


Fig. 5, Learning curves of APA-SP1, NLMS, and APA with different projection orders ranging from 2 to 16. Input is zero-mean unit-variance Gaussian and  $\mu = 0.5$ .

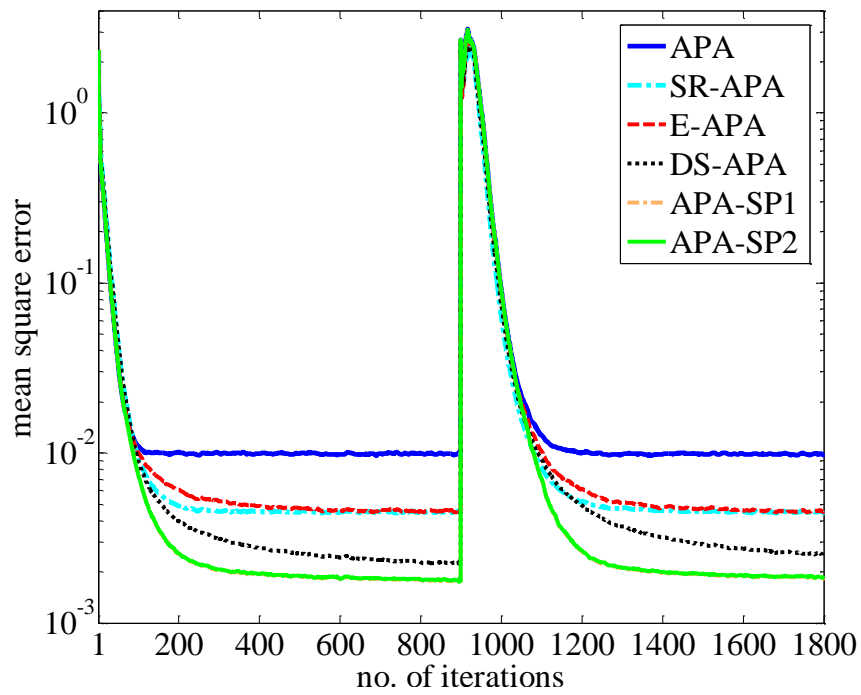


Fig. 6, Comparing performance of different algorithms for Gaussian ARMA(2,2) input when a sudden change in the impulse response of the unknown target system occurs.

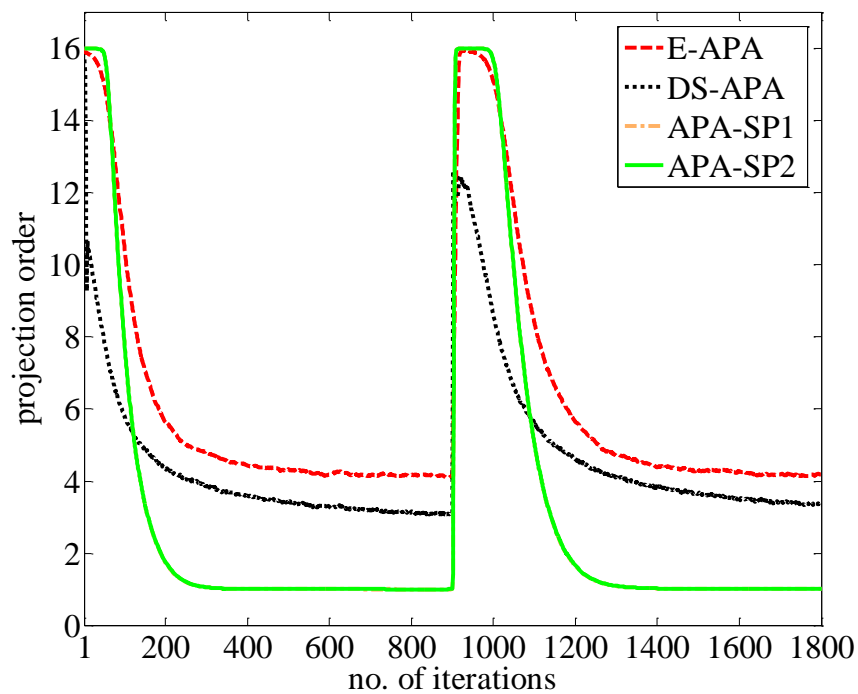


Fig. 7, Time-varying projection order of different algorithms for the experiment of Fig. 6.

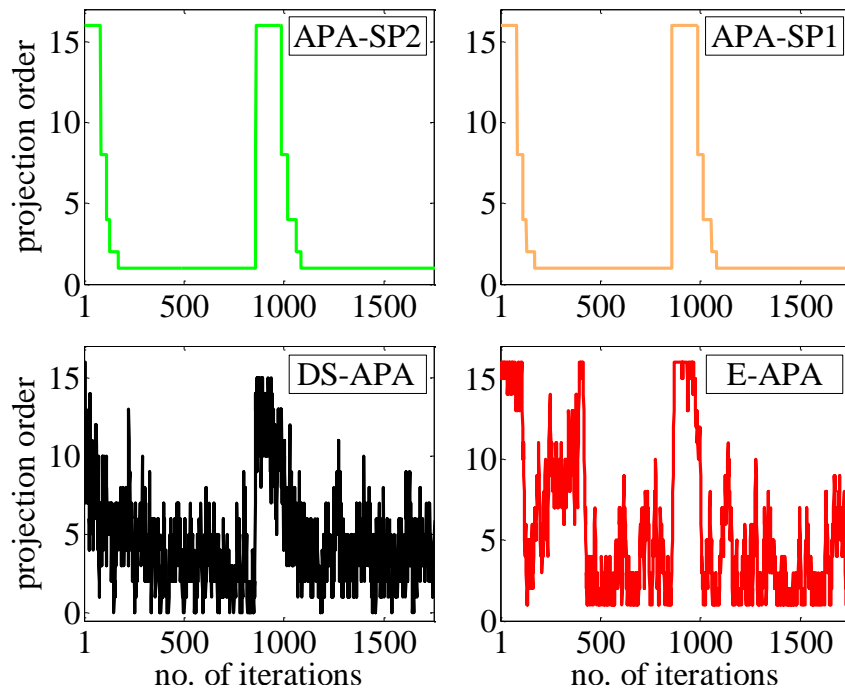


Fig. 8, Projection order versus time for one realization of the experiment of Fig. 6.

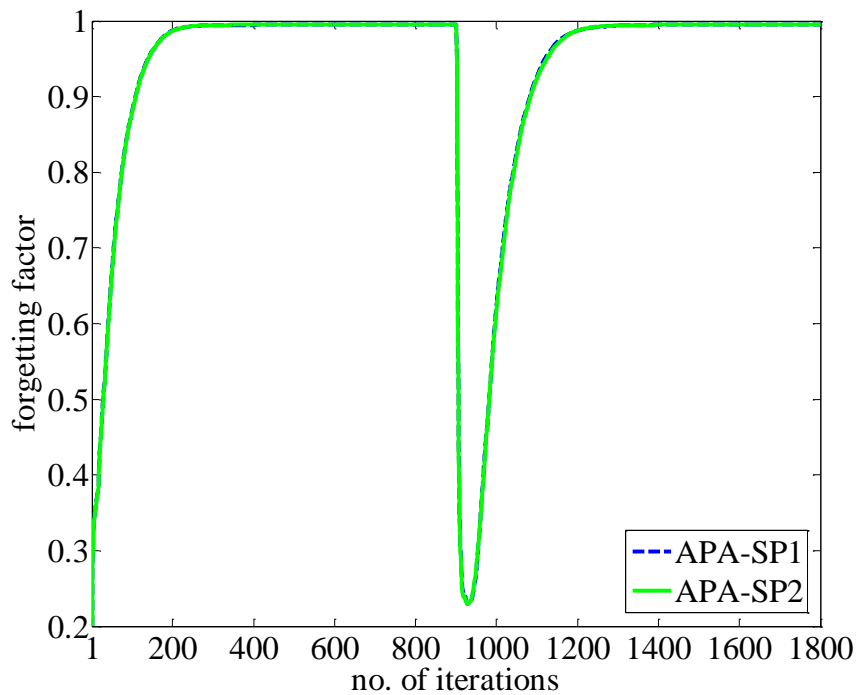


Fig. 9, Time-varying forgetting factor of both variants of the new algorithm for the experiment of Fig. 6 (the curves overlap).

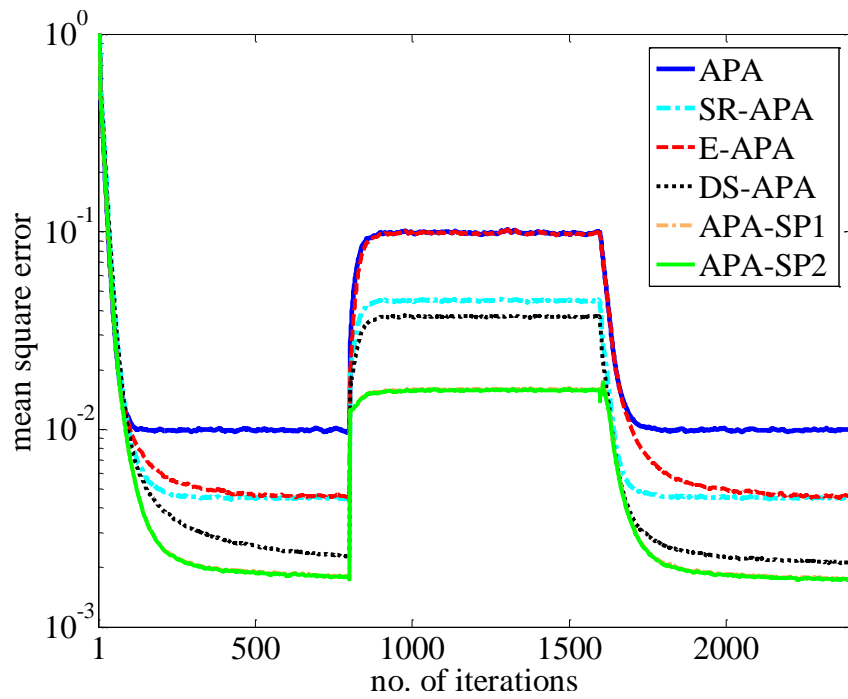


Fig. 10, Same as the experiment of Fig. 6 when the noise power instead of the system impulse response is abruptly changed, i.e. SNR is reduced by 10 dB between iterations 800 and 1600.

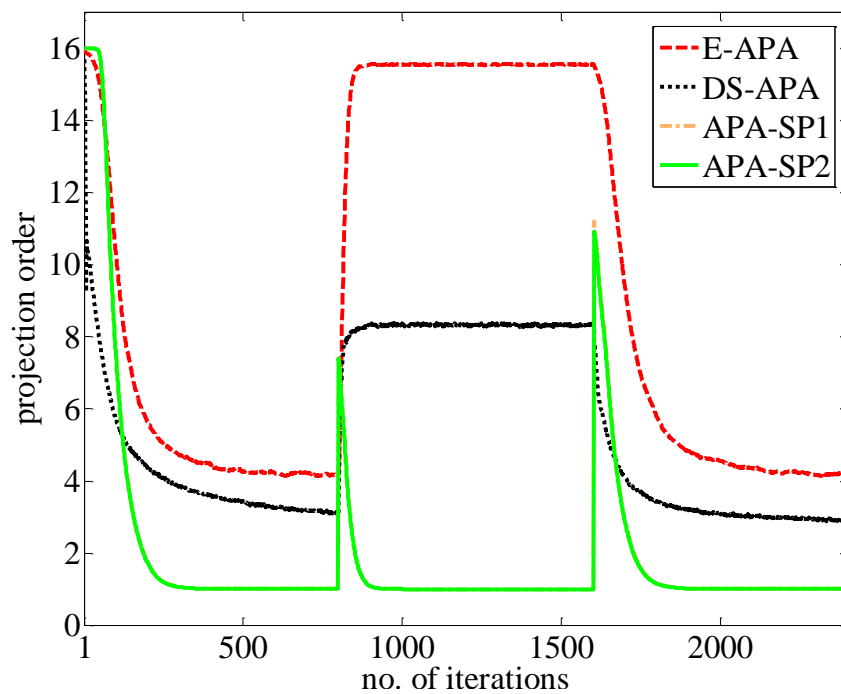


Fig. 11, Time-varying projection order of different algorithms for the experiment of Fig. 10.

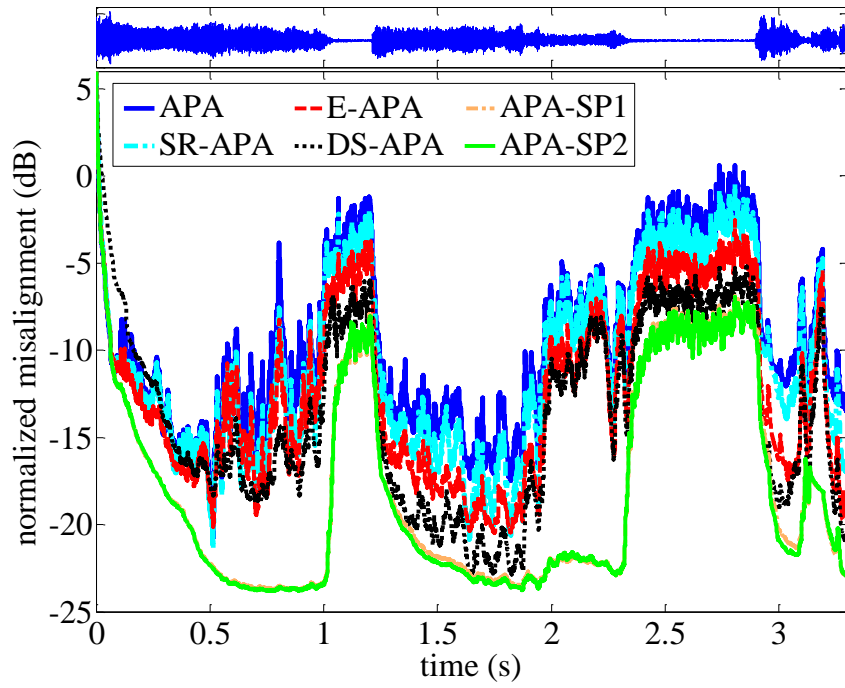


Fig. 12, Normalized misalignment of different algorithms in an acoustic echo cancellation scenario. The input voice signal is also shown on top.

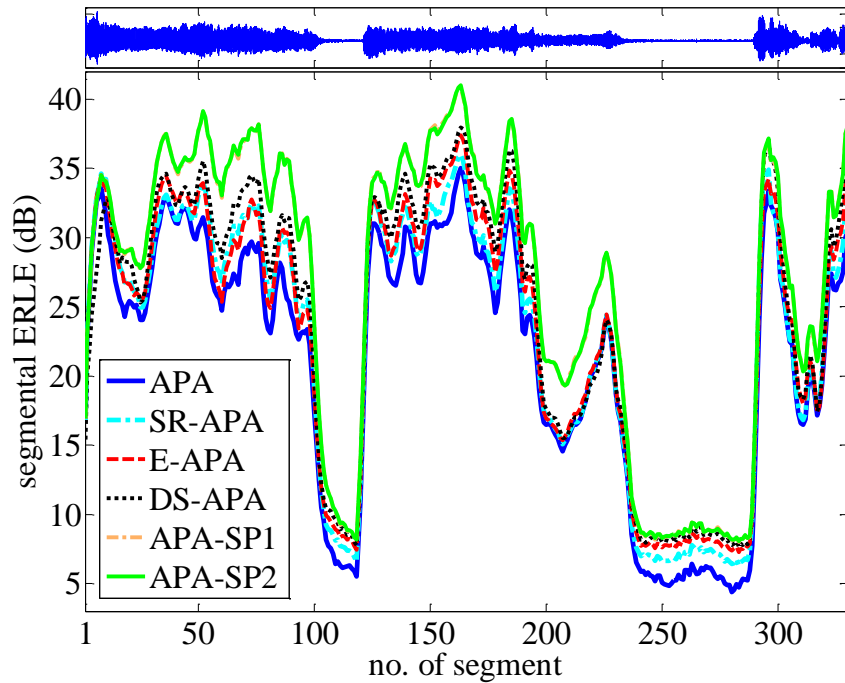


Fig. 13, Segmental ERLE of different algorithms for the experiment of Fig. 12.

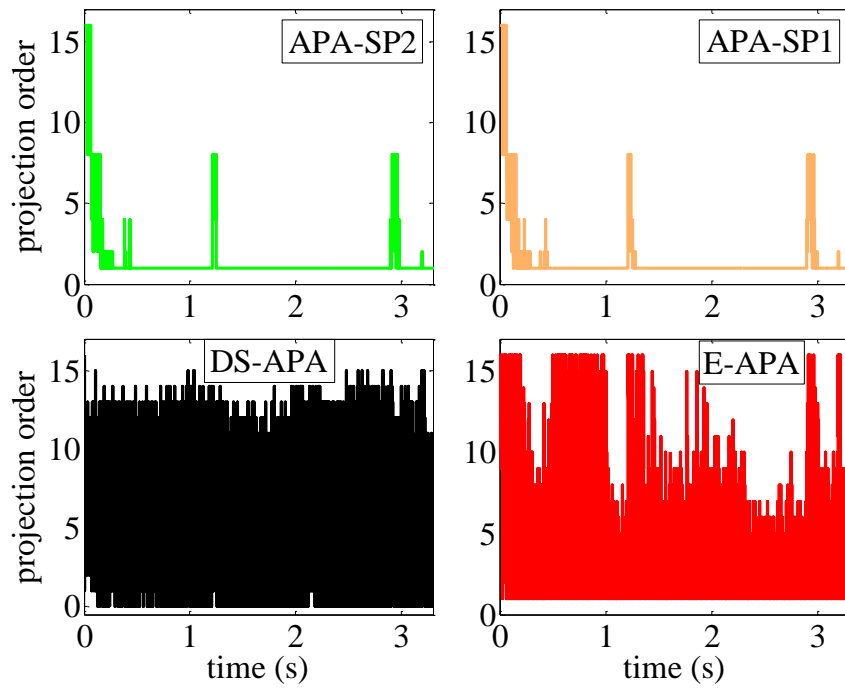


Fig. 14, Time-varying projection order of different algorithms for the experiment of Fig. 12.

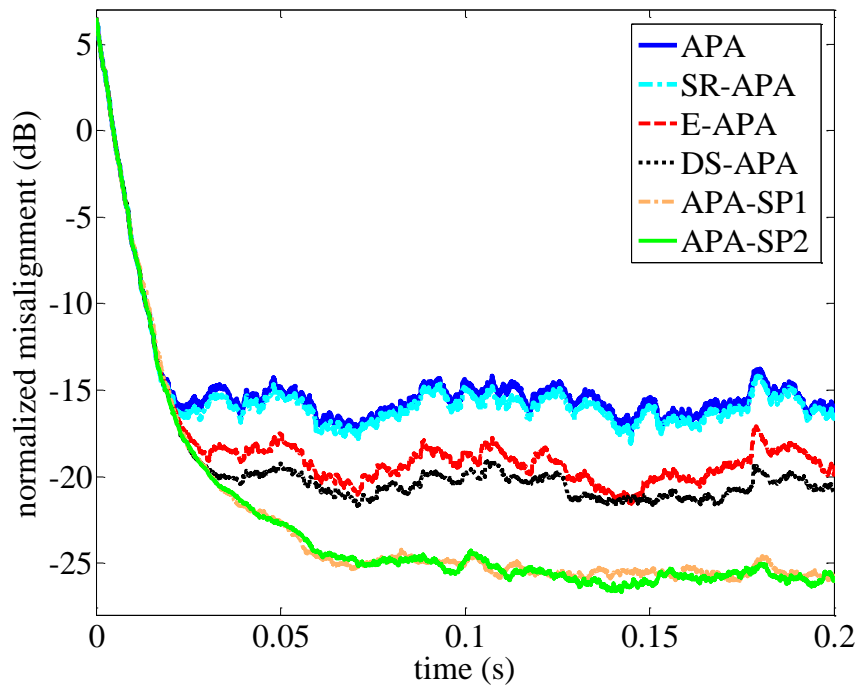


Fig. 15, Comparing tracking performance of different algorithms in an acoustic echo cancellation scenario with a time-varying echo path impulse response when the input is a stationary signal with a speech-like spectrum.

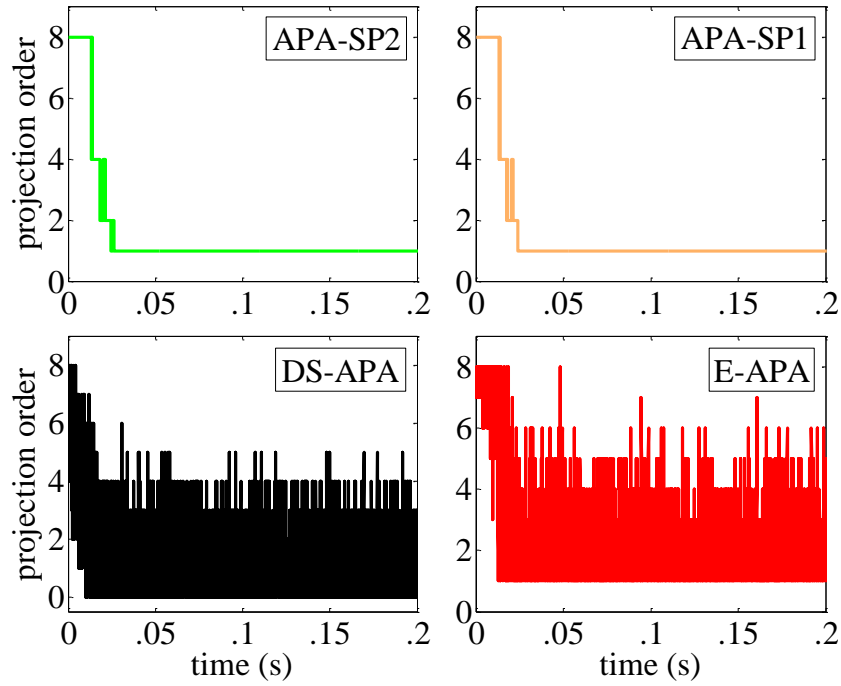


Fig. 16, Time-varying projection order of different algorithms for the experiment of Fig. 15.

Table 1, Complexity comparison of different algorithms in terms of required number of multiplications per iteration together with the average required number of multiplications per iteration for all the algorithms in the experiments of Figs. 12 and 15.

		Fig. 12	Fig. 15
APA	$(k^2 + 2k)L + k^3 + k^2$	78080	21056
SR-APA	$(k_{SR}^2 + k_{max} + k_{SR})L + k_{SR}^3 + k_{SR}^2 + k_{max}$	45920	13060
DS-APA	$(\kappa^2(n) + k_{max} + \kappa(n))L + \kappa^3(n) + \kappa^2(n)$	14280	4375
E-APA	$(\kappa^2(n) + 2\kappa(n))L + \kappa^3(n) + \kappa^2(n)$	8225	4408
APA-SP1	$(\kappa^2(n) + k_{max} + \kappa(n))L + \kappa^3(n) + \kappa^2(n)$	6378	3062
APA-SP2	$(\kappa^2(n) + k_{max} + \kappa(n))L + \kappa^3(n) + \kappa^2(n) + k_{max}$	6387	3089

# Paper B

Originally published as

R. Arablouei, K. Doğançay, and Sylvie Perreau, “Proportionate affine projection algorithm with selective projections for sparse system identification,” in *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, Singapore, Dec. 2010, pp. 362–366.

Copyright © 2010 APSIPA. Reprinted with permission.

## Proportionate affine projection algorithm with selective projections for sparse system identification

Reza Arablouei, Kutluyıl Doğançay, and Sylvie Perreau

### Abstract

Proportionate-type adaptive filtering algorithms are powerful tools specifically designed for sparse system identification. Among them, proportionate affine projection algorithms are well known for their capability in providing satisfactory convergence and tracking performance with affordable complexity. In this paper, we develop a new proportionate affine projection algorithm that employs a variable projection order and takes into account the proportionate history, resulting in a significant improvement in convergence performance as well as appreciable reduction in average computational complexity. Simulation results corroborate the superior performance of the developed algorithm with respect to previously proposed proportionate affine projection algorithms.

### 1. Introduction

In certain system identification problems such as network echo cancelation, the system impulse response can be modeled as sparse with many coefficients close to or equal to zero. When using adaptive filtering to estimate and track sparse systems, it is inevitable to choose a sufficiently high-order FIR filter so that all possible flat delays preceding the nonzero part of the system response are accommodated. It is well known that using proportionate updating rather than conventional schemes in LMS-type adaptive filtering algorithms leads to a considerable performance improvement

for the specific problem of sparse system identification [1]–[3]. The basic idea behind proportionate adaptive filters is to update each coefficient independently by assigning a step size in proportion to estimated magnitude of the filter coefficient.

The normalized least mean square (NLMS) algorithm [4] and its proportionate-type variants are the most commonly used algorithms in the realm of sparse system identification, mainly because of NLMS's relative simplicity and ease of implementation. However, it suffers from slow convergence, particularly with long systems (echo paths) and consequently long adaptive filters, usually comprising 512 to 2048 taps in order to deal with total delay greater than 64 ms [5]. In addition, colored input signal is known to influence convergence rate of the NLMS-based adaptive filters adversely. An attractive solution was proposed in [6] to cope with these impairments, which was named affine projection algorithm (APA). It is based on projections of the input regressor vectors onto an affine subspace of past input regressors to decorrelate the adaptive filter excitation and produce semi-orthogonal direction vectors for the adaptation of the algorithm. Appropriate selection of the projection order can lead to only a graceful increase in the computational cost compared to the NLMS algorithm, while delivering considerably improved performance. However, the value of the projection order impinges a crucial compromise upon the APA, where using a larger projection order leads to faster initial convergence while producing a higher steady-state mean square error (MSE) or misalignment [7].

One of the first proportionate-type algorithms was the one proposed in [8], which is called proportionate NLMS (PNLMS). This algorithm offers improved convergence rate over NLMS for sparse systems but its performance is not better than NLMS for dispersive systems. Several improved versions of this algorithm have been developed [9]–[11]. Among them are the  $\mu$ -law PNLMS [10] and the improved PNLMS [11]. The former uses logarithm of the magnitude as the step gain of each coefficient instead of using magnitude itself and the latter proposes a less complex rule for individual coefficient gain calculation that improves performance regardless of the system impulse response's nature.

Proportionate-type APAs are mainly derived based on a straightforward extension of the proportionate-type NLMS algorithms to the APA. In [12], a new framework for developing proportionate APAs by taking into account history of the proportionate factors is proposed based on which memory improved proportionate APA (MIPAPA) is introduced. In addition to superior performance, this algorithm is realized with less complexity compared to its classical counterparts thanks to the recursive implementation of the proportionate history.

In this paper, we propose a new proportionate APA that employs a variable projection order along with a simple but efficient criterion for input regressor

selection. Utilization of this selective projections method results in significant performance enhancement in terms of steady-state misalignment as well as noticeable reduction of average computational complexity as the algorithm reduces to NLMS at the steady state. In addition, the new algorithm takes into account the proportionate history, yielding further improvement in performance and alleviation in computations. The sparsity measure of the system is another tool that can be exploited in the new algorithm to efficiently deal with systems having different degrees of sparsity.

Section 2 provides a review of the important proportionate-type APAs and section 3 describes the APA with selective projections. The proposed algorithm is presented in section 4, while sections 5 and 6 provide simulation results and concluding statements, respectively.

## 2. Proportionate-type APAs

### 2.1. Improved proportionate APA (IPAPA)

Improved proportionate APA (IPAPA) is composed by straightforwardly extending the improved proportionate normalized least mean squares (IPNLMS) algorithm to the APA case in view of the fact that APA is a generalized form of the NLMS algorithm. IPAPA updates filter coefficients via

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \mathbf{G}(n) \mathbf{X}(n) (\delta \mathbf{I}_k + \mathbf{X}^T(n) \mathbf{G}(n) \mathbf{X}(n))^{-1} \mathbf{e}(n) \quad (1)$$

where  $\mathbf{w}(n) = [w_1(n), w_2(n), \dots, w_L(n)]^T$  and  $\mathbf{e}(n) = [e(n), \varepsilon_1(n), \dots, \varepsilon_{k-1}(n)]^T$  are  $L \times 1$  and  $k \times 1$  vectors of filter coefficients and error signals, respectively. The error vector is calculated as

$$\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{X}^T(n) \mathbf{w}(n) \quad (2)$$

where  $\mathbf{d}(n) = [d(n), d(n-1), \dots, d(n-k+1)]^T$  is the vector of past  $k$  desired responses and  $\mathbf{X}(n) = [\mathbf{x}(n), \mathbf{x}(n-1), \dots, \mathbf{x}(n-k+1)]$  is the  $L \times k$  input signal regression matrix while  $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-L+1)]^T$  is the  $L \times 1$  input signal vector. In addition,  $\mathbf{I}_k$  is a  $k \times k$  identity matrix and  $\mu, \delta, k, L$  and  $n$  denote step size, regularization parameter, projection order, filter length, and time index, respectively. As it is clear from (1), IPAPA differs from APA because of the presence of  $\mathbf{G}(n)$  that is a  $L \times L$  diagonal matrix assigning an individual step size to each filter coefficient. The diagonal elements of  $\mathbf{G}(n)$  are calculated as

$$g_l(n) = \frac{1-\alpha}{2L} + (1+\alpha) \frac{|w_l(n)|}{2 \sum_{i=1}^L |w_i(n)| + \epsilon} \quad (3)$$

where  $-1 \leq \alpha < 1$  is a parameter that controls behavior of the algorithm and enables it to handle systems with different degrees of sparsity. Good choices for  $\alpha$  are 0 and

$-0.5$  [11]. Also,  $\epsilon$  is a small positive constant to avoid division by zero, especially at the beginning of the adaptation when all filter taps are initialized to zero.

## 2.2. IPAPA with proportionate memory (MIPAPA)

As APA takes into account history of the last  $k$  time indices, it can be considered as an algorithm with memory. In [12], a modified approach for proportionate-type APA is proposed, which allows for taking advantage of proportionate memory of the algorithm. In this approach, the proportionate history of each coefficient,  $w_l(n)$   $1 \leq l \leq L$ , in terms of its proportionate factors from the last  $k - 1$  time indices are taken into account. This algorithm, named memory-IPAPA (MIPAPA), updates the coefficients via

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \mathbf{P}(n) (\delta \mathbf{I}_k + \mathbf{X}^T(n) \mathbf{P}(n))^{-1} \mathbf{e}(n) \quad (4)$$

and

$$\mathbf{P}(n) = [\mathbf{g}(n) \odot \mathbf{x}(n), \mathbf{g}(n-1) \odot \mathbf{x}(n-1), \dots, \mathbf{g}(n-k+1) \odot \mathbf{x}(n-k+1)] \quad (5)$$

where  $\mathbf{g}(n) = [g_1(n), g_2(n), \dots, g_L(n)]^T$  contains diagonal elements of  $\mathbf{G}(n)$  and  $\odot$  denotes Hadamard product, i.e.  $\mathbf{a} \odot \mathbf{b} = [a(1)b(1), a(2)b(2), \dots, a(L)b(L)]^T$  for two vectors  $\mathbf{a}$  and  $\mathbf{b}$  of size  $L$ .

The matrix  $\mathbf{P}(n)$  can be computed recursively as

$$\mathbf{P}(n) = [\mathbf{g}(n) \odot \mathbf{x}(n), \mathbf{P}_{-1}(n-1)] \quad (6)$$

where

$$\mathbf{P}_{-1}(n-1) = [\mathbf{g}(n-1) \odot \mathbf{x}(n-1), \dots, \mathbf{g}(n-k+1) \odot \mathbf{x}(n-k+1)] \quad (7)$$

contains the first  $k - 1$  columns of  $\mathbf{P}(n-1)$ .

In this algorithm, since only the latest input regressor vector is multiplied by the new proportionate gains at each time instant, computational complexity is decreased at the cost of increased memory requirement.

## 3. APA with selective projections

### 3.1. Variable projection order

In order to achieve high convergence rate and tracking ability at the transition state together with low misalignment and complexity at the steady state, the projection order should adaptively change according to the algorithm's adaptation state. We propose to control the projection order using

$$\tilde{k}(n) = 1 + (k_{max} - 1) \frac{\ln(\sigma_e^2(n)/\eta)}{\ln(\gamma/\eta)} \quad (8)$$

$$k(n) = \min(k_{max}, \max(1, \lceil \tilde{k}(n) \rceil)) \quad (9)$$

where the time-varying projection order,  $k(n)$ , is changed adaptively between 1 and  $k_{max}$  according to the error variance  $\sigma_e^2(n)$ , which is estimated using the exponential window approach [13] given by

$$\sigma_e^2(n) = \lambda \sigma_e^2(n-1) + (1-\lambda)e^2(n). \quad (10)$$

Here,  $\lambda$  is the forgetting factor and the initial value is set to  $\sigma_e^2(0) = \sigma_d^2$ , which is the variance of the desired response. In (8),  $\eta$  is the steady-state MSE of APA and  $\gamma$  is a threshold that determines at what level of  $\sigma_e^2(n)$  the algorithm starts switching  $k(n)$ . In order to attain the fastest convergence in the new algorithm, it is beneficial to decrease  $\gamma$  for small values of  $k_{max}$ . On the other hand, with relatively large values of  $k_{max}$ , a large  $\gamma$  results in more computational saving with no sacrifice of the performance. Accordingly, we compute  $\gamma$  as

$$\gamma = \eta \left( \frac{\sigma_d^2}{\eta} \right)^z \quad (11)$$

where the parameter  $z$  is tuned in such a way that  $\gamma$  gets its maximum value of  $\gamma_{max} = \sqrt[2]{\eta \sigma_d^2}$  for  $k_{max} \geq 8$  and minimum value of  $\gamma_{min} = \eta$  for  $k_{max} = 1$ . This can be realized by defining a linear dependence on  $k_{max}$  for  $z$  as

$$z = \max\left(0, \min\left(0.5, \frac{k_{max} - 2}{12}\right)\right). \quad (12)$$

The steady-state MSE,  $\eta$ , can also be approximated using the equation given in [7] as

$$\eta = \sigma_v^2 \left( 1 + \frac{\mu}{2-\mu} \text{Tr}(R_x) E \left[ \frac{k}{\|\mathbf{x}\|^2} \right] \right) \approx \sigma_v^2 \left( 1 + \frac{\mu k}{2-\mu} \right) \quad (13)$$

where  $\sigma_v^2$  denotes variance of the measurement noise,  $\text{Tr}(R_x)$  stands for trace of the input signal autocorrelation matrix and  $E[\cdot]$  is the expectation operator.

### 3.2. Regressor selection

When the input signal is extremely correlated like speech, orthogonality of the immediate past input regressors can tremendously deteriorate. In such cases, it is very useful to select and utilize past input regressors that maximize the convergence rate. In this regard, two methods of near-optimal selection have been introduced in [14] and [15]. In [14], input regressors are distinguished according to the following criterion

$$\begin{aligned} |\varepsilon_i(n)| &> \sqrt{2/(2-\mu)} \sigma_v \\ 0 &\leq i \leq k-1 \end{aligned} \quad (14)$$

where  $\varepsilon_i(n)$  is the  $i$ th element of the error vector,  $\mathbf{e}(n)$ , and the input regressor corresponding to this element is  $\mathbf{x}(n-i)$ . Alternatively, one can rank the values of  $|\varepsilon_i(n)|$  with  $0 \leq i \leq k-1$  and choose the regressors associated with the largest values for update.

Utilizing the variable projection order of subsection III.A together with the abovementioned input regressor selection criterion, we develop an algorithm called *affine projection algorithm with selective projections* (APA-SP).

#### 4. New Algorithm

To construct our new proposed algorithm, we take advantage of both selective projections and proportionate history. In order to exploit sparsity of the system and enable the algorithm to maintain its good performance for systems with different degrees of sparsity, we can also make use of the system sparsity measure proposed in [16] that is defined as

$$\xi(n) = \frac{L}{L - \sqrt{L}} \left( 1 - \frac{\|\mathbf{w}(n)\|_1}{\sqrt{L}\|\mathbf{w}(n)\|_2} \right) \quad (15)$$

where  $\|\cdot\|_l$  denotes  $l$ -norm of a vector. This value is in fact instantaneous estimate of the adaptive filter sparsity and to obtain a more reliable estimate, we use an exponential window averaging with a forgetting factor  $\rho$  similar to (10) as

$$\Xi(n) = \rho\Xi(n-1) + (1-\rho)\xi(n). \quad (16)$$

Therefore, the parameter  $\alpha$  in (3) is adapted, similar to the method of [17], as

$$\alpha(n) = 2\Xi(n) - 1. \quad (17)$$

Hence, (3) can be rewritten as

$$g_l(n) = \frac{1-\alpha(n)}{2L} + (1+\alpha(n)) \frac{|w_l(n)|}{2\|\mathbf{w}(n)\|_1 + \epsilon} \quad (18)$$

The proposed algorithm is summarized as follows:

- 1) calculate  $g_l(n)$  for  $1 \leq l \leq L$  using (15)–(18),
- 2) calculate  $\mathbf{e}(n)$  using (2),
- 3) calculate  $k(n)$  using (8)–(10),
- 4) select input regressors corresponding to the largest elements of the error vector for update,
- 5) calculate  $\mathbf{P}(n)$  using (6),

6) update filter coefficients using (4).

The proposed algorithm is called *memory improved proportionate affine projection algorithm with selective projections* (MIPAPA-SR). Memoryless version of this algorithm is also called IPAPA-SP.

## 5. Simulation results

In order to examine performance of the new algorithm, we present computer simulation results carried out in the context of system identification for echo cancelation. Two echo path impulse responses with different degrees of sparsity, shown in Fig. 1, are used. The echo paths are supposed to have 512 taps. The same length is also assumed for the corresponding FIR adaptive filters. Two different stationary input signals are applied, viz. zero-mean unit-variance white Gaussian and first-order auto-regressive Gaussian (AR1). The latter is obtained by filtering the former through a low-pass filter with a transfer function of  $1/(1 - 0.9z^{-1})$ . The normalized misalignment (in dB) is the performance measure in our simulations. It is defined as

$$10 \log_{10} \left( \frac{\|\mathbf{w}^o - \mathbf{w}(n)\|_2^2}{\|\mathbf{w}^o\|_2^2} \right),$$

where  $\mathbf{w}^o$  denotes the true impulse response of the echo path. The simulated parameters are: step size  $\mu = 0.5$ , signal to noise ratio SNR = 10 dB for stationary input signals and SNR = 30 dB for voice input signal, forgetting factors  $\lambda = 0.99$  and  $\rho = 0.9$ , regularization parameters  $\delta = \epsilon = 10^{-5}$ ,  $\alpha = 0$ , maximum projection order  $k_{max} = 8$  for the algorithms with selective projections and fixed projection order  $k = 8$  for the other algorithms. The results were acquired by averaging over 100 independent trials. In order to compare tracking capabilities of the algorithms, an abrupt change of the echo paths was introduced at certain iterations by shifting the impulse responses to the right by 10 samples.

Figs. 2 and 3 show the learning curves of different algorithms when the echo paths of Fig. 1(a) and (b) are applied with white Gaussian excitation and abrupt change of the echo path at iteration 3,000. Figs. 4 and 5 depict the results for the same scenarios with AR1 excitation and abrupt change of the echo path at iteration 10,000. In Figs. 6 and 7, ensemble-averaged projection order of APA-SP, IPAPA-SP, and MIPAPA-SP algorithms are shown for the scenarios of Figs. 2 and 5, respectively.

In Fig. 8, performance of different algorithms is compared using a real recorded voice signal as input and the echo path of Fig. 1(a) with an abrupt change at iteration 12,500. The corresponding time-varying projection order curves are also presented in Fig. 9.

It is clear from the misalignment learning curves that the new algorithm outperforms all the other algorithms in terms of both convergence rate and steady-state misalignment. Moreover, the projection order curves reveal that the new algorithm reduces to the NLMS algorithm at the steady state. This alleviates the average computational load dramatically as the number of required arithmetic operations, i.e. multiplications and additions, is much less for the NLMS algorithm and more importantly, it needs no matrix inversion.

There is an interesting observation in Fig. 8 that in the associated scenario, the algorithms with fixed projection order, despite their good initial convergence, exhibit a poor behavior afterwards while the algorithms with selective projections maintain their comparatively superior performance throughout the simulation period. Reducing the projection order and/or the step size can lessen the observed problem with fixed-projection-order algorithms but will sacrifice their initial convergence rate and tracking abilities.

## 6. Conclusion

A new proportionate APA was proposed that utilizes selective input regressors to update filter coefficients. This is carried out by adaptively changing the projection order respecting the adaptation state and choosing the input regressors according to a simple but effective criterion. The algorithm also takes into consideration history of the proportionate factors. Exploiting both the selective projections and the proportionate memory reduces the algorithm's average computational complexity compared to the conventional proportionate-type APAs. Besides, computer simulations substantiate that the proposed algorithm is capable of providing both faster convergence and lower misadjustment.

## References

- [1] E. Haensler and G. Schmidt, Eds., *Topics in Acoustic Echo and Noise Control*, Berlin, Germany, Springer-Verlag, 2006.
- [2] Z. Chen, S. L. Gay, and S. Haykin, *Proportionate Adaptation: New Paradigms in Adaptive Filters*, in S. Haykin and B. Widrow Eds. *Advances in LMS Filters*, ch. 8, Wiley, 2005.
- [3] Y. Huang, J. Benesty, and J. Chen, *Acoustic MIMO Signal Processing (Signals and Communication Technology)*, Secaucus, NJ, USA: Springer-Verlag, 2006.
- [4] S. Haykin, *Adaptive Filter Theory, 4th ed.*, Upper Saddle River, NJ: Prentice-Hall, 2003.
- [5] J. Benesty, Y. A. Huang, J. Chen, and P. A. Naylor, "Adaptive algorithms for the identification of sparse impulse responses," in *Selected methods for acoustic echo*

- and noise control*, E. Hänsler and G. Schmidt, Eds. Springer, 2006, ch. 5, pp. 125–153.
- [6] K. Ozeki and T. Umeda, “An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties,” *Electron. Commun. Jpn.*, vol. 67-A, no. 5, pp. 19–27, May 1984.
  - [7] H.-C. Shin and A. H. Sayed, “Mean-square performance of a family of affine projection algorithms,” *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 90–102, Jan. 2004.
  - [8] D. L. Duttweiler, “Proportionate normalized least-mean-squares adaptation in echo cancellers,” *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 508–518, Sep. 2000.
  - [9] H. Deng and M. Doroslovački, “Proportionate adaptive algorithms for network echo cancellation,” *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1794–1803, May 2006.
  - [10] H. Deng and M. Doroslovački, “Improving convergence of the PNLMS algorithm for sparse impulse response identification,” *IEEE Signal Process. Lett.*, vol. 12, no. 3, pp. 181–184, 2005.
  - [11] J. Benesty and S. L. Gay, “An improved PNLMS algorithm,” *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2002, pp. II-1881–II-1884.
  - [12] C. Paleologu, S. Ciochina, J. Benesty, “An efficient proportionate affine projection algorithm for echo cancellation,” *IEEE Signal Process. Lett.*, vol. 17, no. 2, pp. 165–168, Feb. 2010.
  - [13] H.-C. Shin, A. H. Sayed, and W.-J. Song, “Variable step-size NLMS and affine projection in algorithm,” *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 132–135, Feb. 2004.
  - [14] S.-J. Kong, K.-Y. Hwang, and W.-J. Song, “An affine projection algorithm with dynamic selection of input vectors,” *IEEE Signal Process. Lett.*, vol. 14, no. 8, pp. 529–532, Aug. 2007.
  - [15] K.-Y. Hwang and W.-J. Song, “An affine projection adaptive filtering algorithm with selective regressors,” *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 54, no. 1, pp. 43–46, Jan. 2007.
  - [16] P. O. Hoyer, “Non-negative matrix factorization with sparseness constraints,” *Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, Nov. 2004.

- [17] L. Liu, M. Fukumoto, and S. Saiki, "An improved mu-law proportionate NLMS algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2008, pp. 3797–3800.

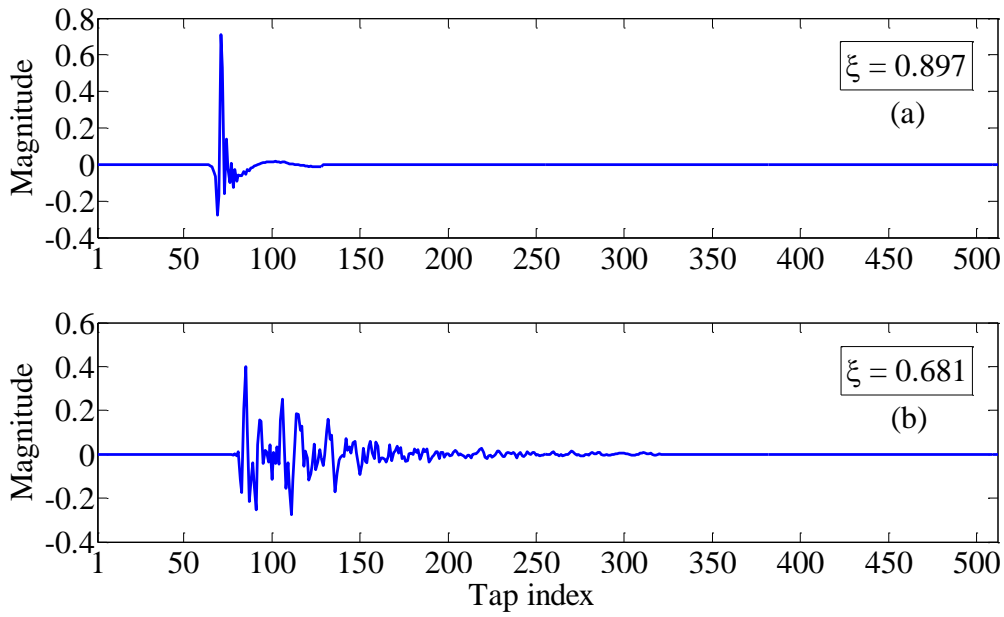


Fig. 1, Two echo paths with different degrees of sparsity used in the simulations.

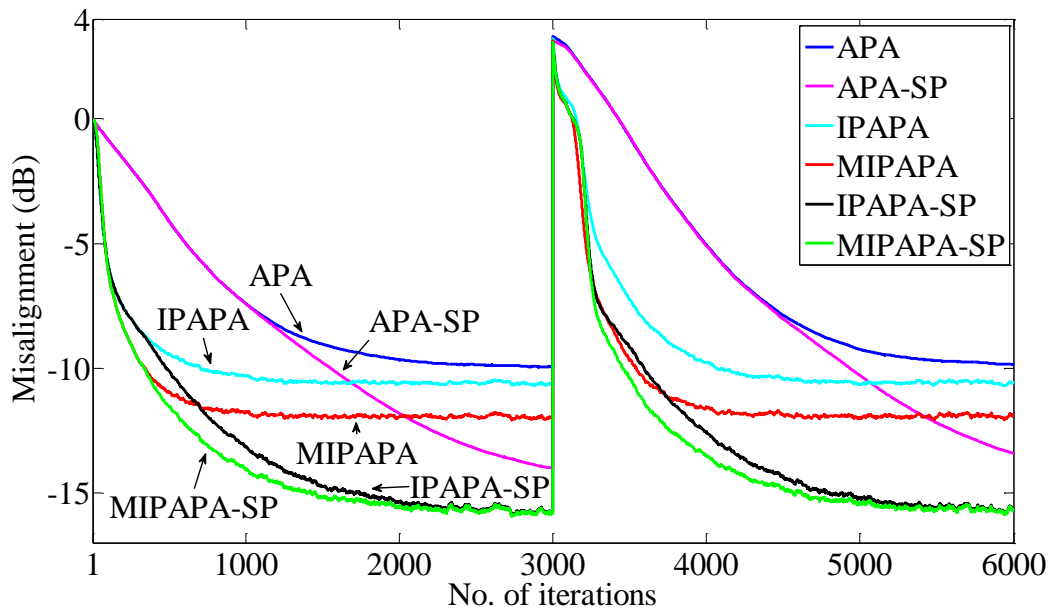


Fig. 2, Learning curves of different algorithm with the echo path of Fig. 1(a) and white Gaussian excitation.

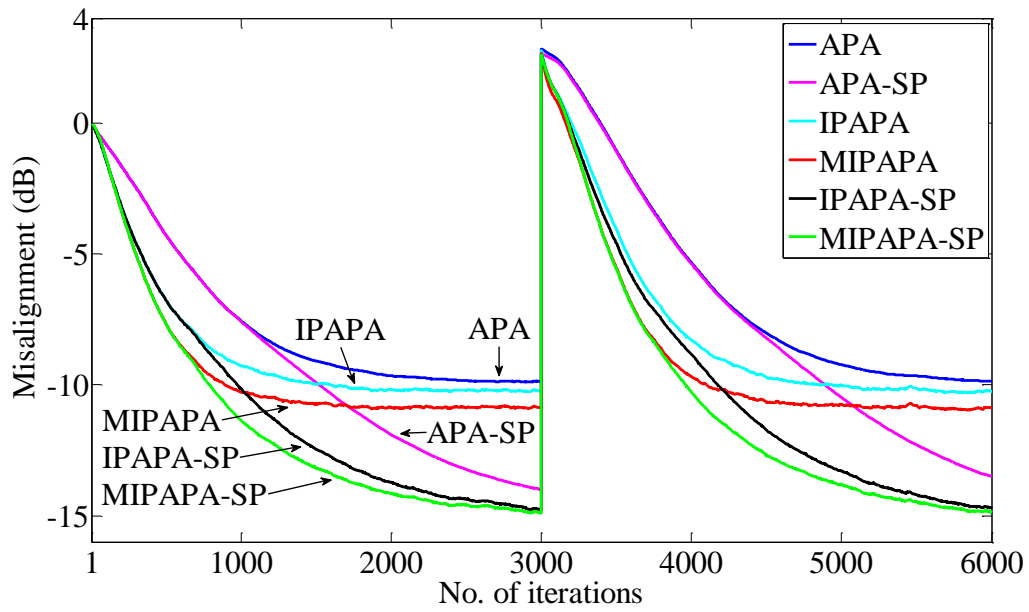


Fig. 3, Learning curves of different algorithm with the echo path of Fig. 1(b) and white Gaussian excitation.

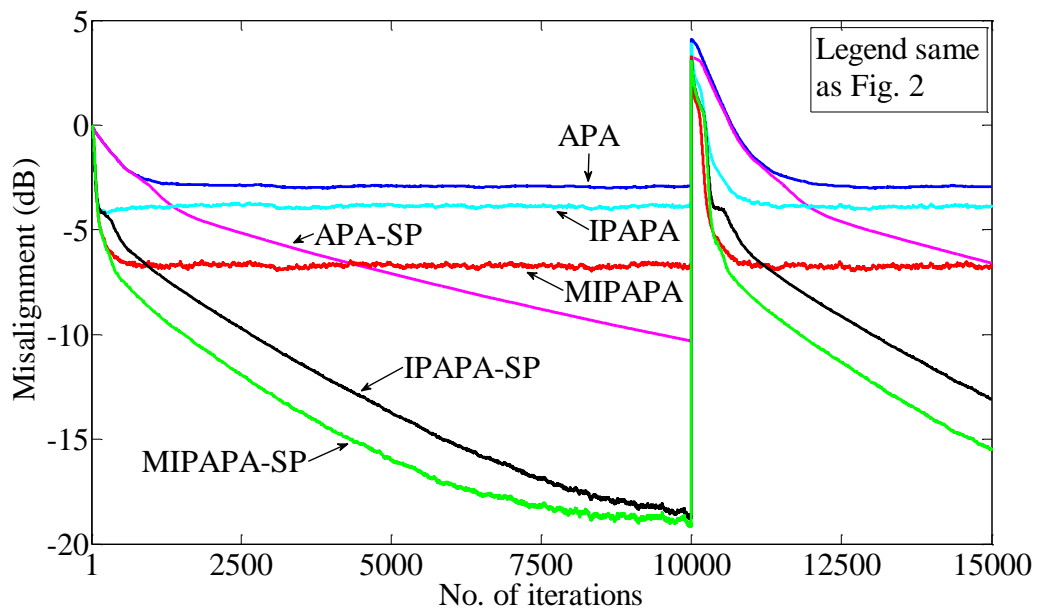


Fig. 4, Learning curves of different algorithm with the echo path of Fig. 1(a) and AR1 excitation.

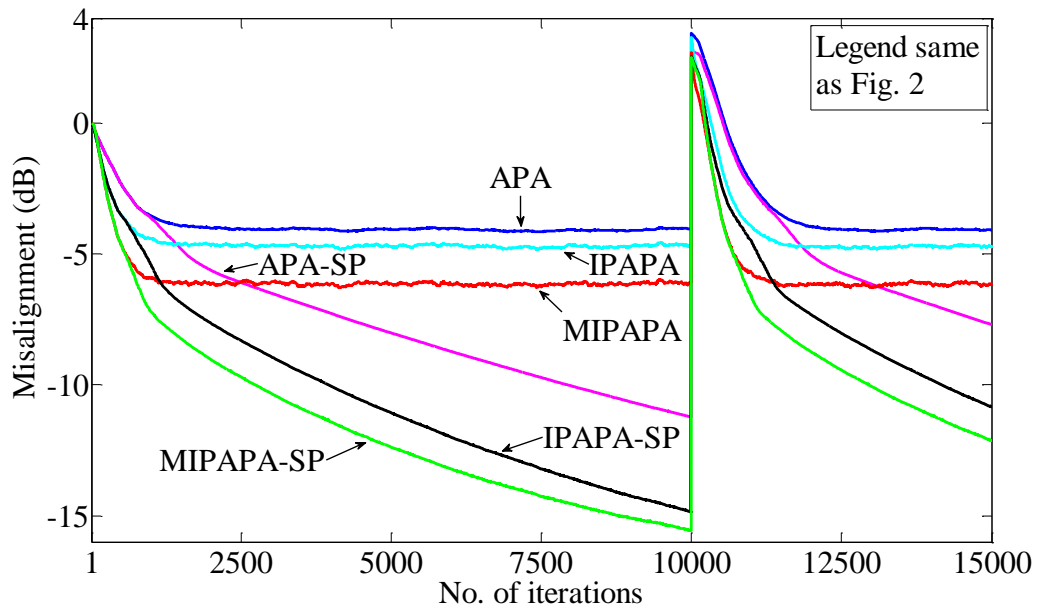


Fig. 5, Learning curves of different algorithm with the echo path of Fig. 1(b) and AR1 excitation.

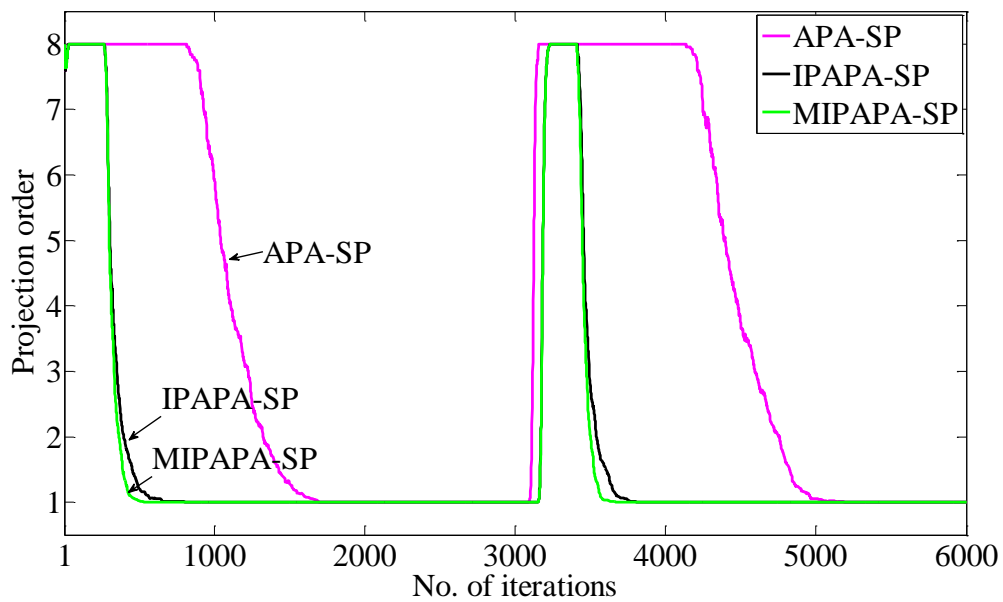


Fig. 6, Ensemble-averaged projection order of different algorithms for the scenario of Fig. 2.

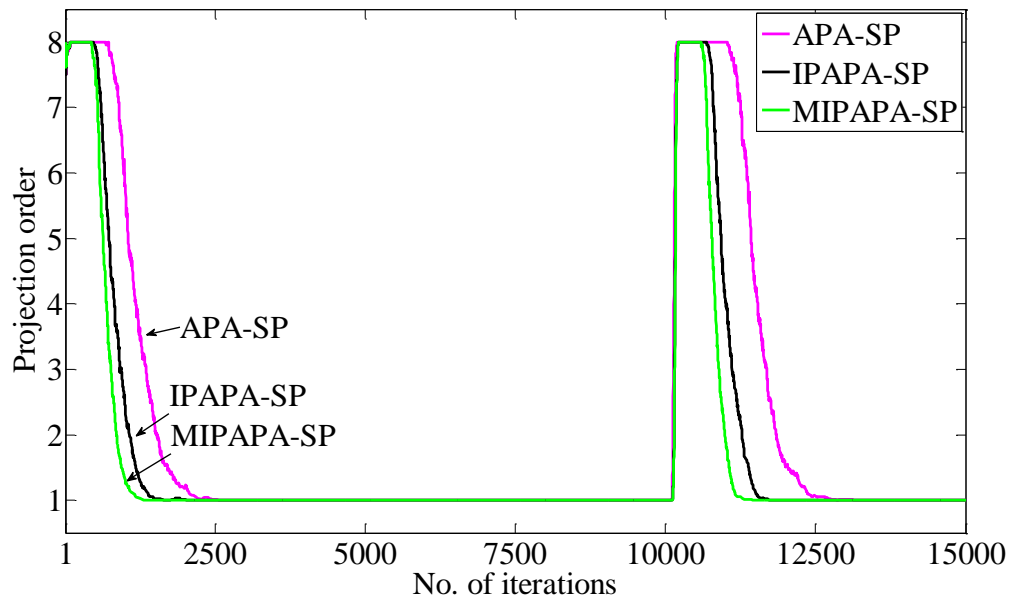


Fig. 7, Ensemble-averaged projection order of different algorithms for the scenario of Fig. 5.

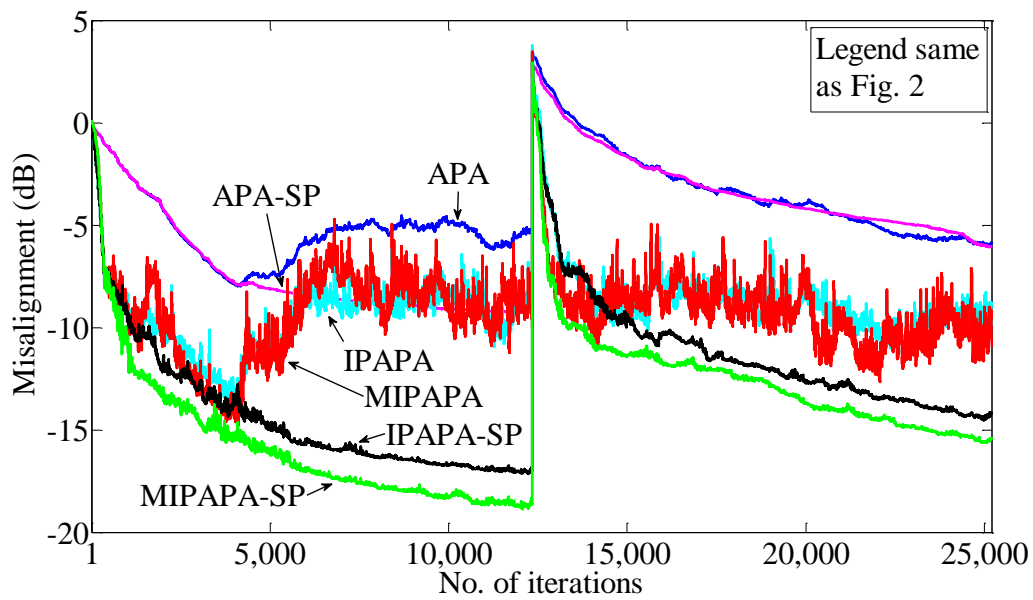


Fig. 8, Learning curves of different algorithm with the echo path of Fig. 1(a) and voice excitation.

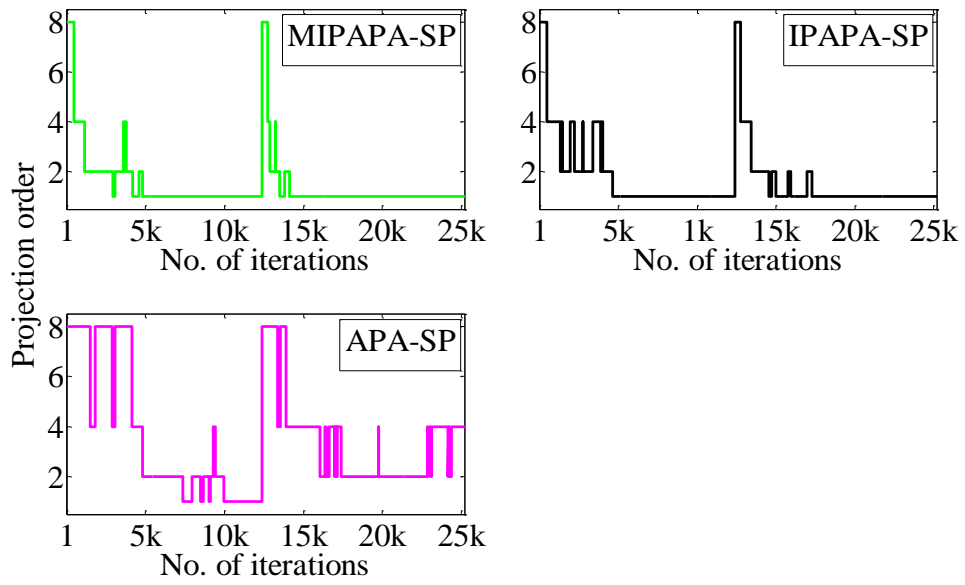


Fig. 9, Projection order of different algorithms for the scenario of Fig. 8.

# Paper C

Originally published as

R. Arablouei and K. Doğançay, "Modified RLS algorithm with enhanced tracking capability for MIMO channel estimation," *Electronics Letters*, vol. 47, no. 19, pp. 1101-1103, 2011.

Copyright © 2011 IET. Reprinted with permission.

## Modified RLS algorithm with enhanced tracking capability for MIMO channel estimation

Reza Arablouei and Kutluyıl Doğançay

### Abstract

A new RLS adaptive filtering algorithm is developed by enforcing a constraint on the a posteriori error of the conventional exponentially-weighted RLS algorithm and modifying its coefficient update equation. The motivation behind the new algorithm is to attain improved tracking capability when the target system features fast variations in time. The proposed algorithm requires only slightly more computations than the conventional RLS algorithm and is specifically suitable for estimation and tracking of fast-fading MIMO channels. Simulations demonstrate that a MIMO channel estimator utilising the proposed algorithm outperforms the ones using the conventional RLS algorithm, the set-membership-based BEACON algorithm and the extended RLS algorithm without requiring prior knowledge of the Doppler shift.

### 1. Introduction

The RLS adaptive filtering algorithm can be considered as a special case of the Kalman filter with a constant multiple of the identity matrix as its state transition matrix and a zero vector as its state process noise [1, 2]. Therefore, the RLS algorithm is an optimal estimator for linear time-invariant systems. It is also able to track slow variations in the target system by exerting exponential weighting and a forgetting factor to limit the size of the input data window. However, for fast varying systems, either a very small forgetting factor is required that can lead to instability and high sensitivity to measurement noise or the algorithm loses track of the system changes.

Several methods to tackle this innate problem of the conventional RLS algorithm have been proposed, e.g. the variable-forgetting-factor, self-perturbing and extended RLS algorithms. Some other approaches assume an autoregressive (AR) or a polynomial process model for the time-varying target system. Although all these algorithms offer enhancements to some extent and exhibit performance improvement compared to the conventional RLS algorithm in certain applications, most of them have fundamental limitations and drawbacks such as high complexity, requirement of prior knowledge about rate of the variations and/or background noise power and being only suited to track jumped changes in the target system (see ,e.g., [3, 4] and references therein).

## 2. Proposed algorithm

Consider a linear system modelled by

$$d(n) = \boldsymbol{\omega}^*(n)\mathbf{x}(n) + v(n)$$

where  $\mathbf{x}(n) \in \mathbb{C}^L$  is the input,  $\boldsymbol{\omega}(n) \in \mathbb{C}^L$  is the unknown system that we wish to estimate,  $d(n) \in \mathbb{C}$  is the noisy system output (reference signal),  $v(n) \in \mathbb{C}$  is the system output noise and  $L$  is the filter order with  $\mathbb{C}$  denoting the set of complex numbers and superscript  $*$  the Hermitian (complex-conjugate transpose) operator. The exponentially-weighted RLS algorithm adjusts its coefficients,  $\mathbf{w}(n)$ , in an attempt to minimize the exponentially-weighted sum of the squared a posteriori errors via the following recursions [1]:

$$e(n) = d(n) - \mathbf{w}^*(n-1)\mathbf{x}(n)$$

$$\mathbf{q}(n) = \mathbf{P}(n-1)\mathbf{x}(n)$$

$$g(n) = \frac{1}{\lambda + \mathbf{x}^*(n)\mathbf{q}(n)}$$

$$\mathbf{k}(n) = g(n)\mathbf{q}(n)$$

$$\mathbf{P}(n) = \lambda^{-1}(\mathbf{P}(n-1) - \mathbf{k}(n)\mathbf{q}^*(n))$$

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mathbf{k}(n)e^*(n).$$

When the objective of the adaptive filter is to estimate and track a fast-varying target system, the magnitude of the a posteriori filter output error at each time instant,  $\epsilon(n) = d(n) - \mathbf{w}^*(n)\mathbf{x}(n)$ , determines the quality of estimation and tracking. Thus, it is desirable to reduce the magnitude of  $\epsilon(n)$  at each iteration as much as practicable. To achieve this goal, we develop a new modified RLS algorithm by manipulating the coefficient update equation of the conventional RLS algorithm. In view of the conventional RLS algorithm's poor performance in fast-varying environments, we propose to adjust the RLS filter coefficients by an auxiliary mechanism devised to

minimize the instantaneous a posteriori error in addition to the correction prescribed by the Kalman gain,  $\mathbf{k}(n)$ . In what follows, we describe the algorithm derivation.

In the RLS algorithm, the a priori and the a posteriori errors are related according to

$$\epsilon^*(n) = \frac{1}{1 + \lambda^{-1}\mathbf{x}^*(n)\mathbf{P}(n-1)\mathbf{x}(n)} e^*(n). \quad (1)$$

To improve the tracking performance, we propose to modify (1) to

$$\epsilon^*(n) = \frac{\alpha}{1 + \lambda^{-1}\mathbf{x}^*(n)\mathbf{P}(n-1)\mathbf{x}(n)} e^*(n) \quad (2)$$

where  $0 \leq \alpha \leq 1$  is a relaxation parameter providing control over desired reduction in the a posteriori error at each iteration. Subtracting both sides of (2) from  $e^*(n)$  and using the definitions of  $\epsilon(n)$  and  $e(n)$ , we obtain

$$\mathbf{x}^*(n)(\mathbf{w}(n) - \mathbf{w}(n-1)) = \frac{1 - \alpha + \lambda^{-1}\mathbf{x}^*(n)\mathbf{P}(n-1)\mathbf{x}(n)}{1 + \lambda^{-1}\mathbf{x}^*(n)\mathbf{P}(n-1)\mathbf{x}(n)} e^*(n). \quad (3)$$

Rewriting (3) as

$$\begin{aligned} & \mathbf{x}^*(n)(\mathbf{w}(n) - \mathbf{w}(n-1)) \\ &= \mathbf{x}^*(n) \frac{\frac{\mathbf{x}(n)}{\|\mathbf{x}(n)\|^2} (1 - \alpha) + \lambda^{-1}\mathbf{P}(n-1)\mathbf{x}(n)}{1 + \lambda^{-1}\mathbf{x}^*(n)\mathbf{P}(n-1)\mathbf{x}(n)} e^*(n), \end{aligned} \quad (4)$$

where  $\|\cdot\|$  stands for Euclidean norm, we see that a non-unique solution for  $\mathbf{w}(n) - \mathbf{w}(n-1)$  is given by

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \frac{\frac{\mathbf{x}(n)}{\|\mathbf{x}(n)\|^2} (1 - \alpha) + \lambda^{-1}\mathbf{P}(n-1)\mathbf{x}(n)}{1 + \lambda^{-1}\mathbf{x}^*(n)\mathbf{P}(n-1)\mathbf{x}(n)} e^*(n). \quad (5)$$

It is easy to show that (5) can be written as

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \left( \frac{\mathbf{x}(n)}{\|\mathbf{x}(n)\|^2} (1 - \alpha)\lambda g(n) + \mathbf{k}(n) \right) e^*(n) \quad (6)$$

which gives a modified coefficient update rule that satisfies the constraint of (2). Setting  $\alpha = 0$  results in  $\epsilon(n) = 0$  (suitable for fast-varying systems) and  $\alpha = 1$  reduces (6) to the update equation of the conventional RLS algorithm (suitable for time-invariant systems). Hence, the parameter  $\alpha$  should be duly tuned according to the relative speed of the time-variations. Generally, for a wide range of fast variations, we can set  $\alpha = 0$ .

The conventional RLS algorithm requires  $3L^2 + 4L$  multiplications and  $2L^2 + 2L$  additions per iteration. The use of the new coefficient update rule of (6) increases the complexity only by  $2L + 2$  multiplications and  $2L - 1$  additions per iteration.

### 3. Simulations

To verify the enhanced tracking capability of the proposed algorithm, its application to estimation and tracking of an extremely fast fading  $4 \times 4$  wireless MIMO channel is presented here as an example. All sub-channels independently undergo Rayleigh fading according to the Jakes model [5] with a normalised Doppler frequency of  $f_D T_s = 0.02$ . The considered MIMO channel estimators were composed of 4 adaptive filters with common input each having  $L = 4$  taps  $[\mathbf{w}_i(n), i = 1, \dots, 4]$ . The input (transmitted signal) was uncoded QPSK and organised in packets each having 160 symbol vectors. A forgetting factor of  $\lambda = 0.9$  was used and results were obtained by ensemble-averaging over  $10^3$  independent runs. Figs. 1 and 2 compare the channel estimation performance of the proposed algorithm, the RLS algorithm, the extended RLS (ERLS) algorithm [1, p. 514] and the set-membership-based BEACON algorithm [6]. The ERLS algorithm assumes a first-order auto-regressive model for time evolution of the channel coefficients, i.e.,

$$\boldsymbol{\omega}_i(n) = \sigma \boldsymbol{\omega}_i(n-1) + \sqrt{1 - \sigma^2} \mathbf{v}_i(n), \quad i = 1, \dots, 4,$$

with

$$\mathbf{v}_i(n) \sim \mathcal{N}(\mathbf{0} \in \mathbb{C}^L, \mathbf{I} \in \mathbb{C}^{L \times L}),$$

and utilises the prior knowledge of the Doppler shift to estimate  $\sigma = J_0(2\pi f_D T_s)$ ,  $J_0$  being the zeroth-order Bessel function of the first kind [1, p. 327]. For the BEACON algorithm, the error-norm bound is adjusted to achieve the best performance. The ERLS and BEACON algorithms were chosen for comparison because they have been built on sound and strong theoretical background and their computational complexities are comparable to the proposed algorithm. The ERLS algorithm has particularly been developed for tracking linear time-variant systems in light of the relationship between the RLS algorithm and the Kalman filter [2]. BEACON is an optimally-weighted RLS algorithm based on the context of set-membership filtering and is known to have good tracking properties. Fig. 1 compares the normalised misalignment, i.e.,

$$\frac{\sum_{i=1}^4 \|\mathbf{w}_i(n) - \boldsymbol{\omega}_i(n)\|^2}{\sum_{i=1}^4 \|\boldsymbol{\omega}_i(n)\|^2},$$

of different channel estimators when energy per bit to noise power spectral density ratio ( $E_b/N_0$ ) is 15 dB. Fig. 2 plots the associated curves of steady-state normalized misalignment, averaged over 60 iterations at steady state, versus  $E_b/N_0$ . Finally, Fig. 3

compares bit error rate (BER) performance of the MMSE-VBLAST MIMO-detector [7] operating in the above-mentioned MIMO communication system and using the described channel estimators. First 32 symbol vectors were used for training. Afterwards, the estimators were switched to the decision-directed mode.

#### 4. Conclusion

A new algorithm with improved tracking capability was developed by imposing a constraint on the a posteriori error of the conventional RLS algorithm. Simulations confirm that when the proposed algorithm is applied to the adaptive estimation and tracking of fast-varying MIMO channels, it is superior to the conventional RLS, BEACON and the extended RLS algorithms without either incurring a significant increase in complexity or requiring prior knowledge of the exact Doppler shift.

#### References

- [1] A. H. Sayed, *Adaptive Filters*, Hoboken, NJ: Wiley, 2008.
- [2] S. Haykin, A. H. Sayed, J. R. Zeidler, P. Yee, and P. C. Wei, "Adaptive tracking of linear time-variant systems by extended RLS algorithms," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3141–3150, 2005.
- [3] T. K. Akino, "Optimum-weighted RLS channel estimation for rapid fading MIMO channels," *IEEE Trans. Wireless Comm.*, vol. 11, no. 7, pp. 4248–4260, 2008.
- [4] S.-H. Leung and C. F. So, "Gradient-based variable forgetting factor RLS algorithm in time-varying environments," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3141–3150, 2005.
- [5] W. C. Jakes, *Microwave Mobile Communications*, New York: Wiley, 1974.
- [6] T. Wang, R. C. De Lamare, and P. D. Mitchell, "Low-complexity set-membership channel estimation for cooperative wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 6, pp. 2594–2607, 2011.
- [7] G. J. Foschini, G. D. Golden, R. A. Valenzuela and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 11, pp. 1841–1852, Nov. 1999.

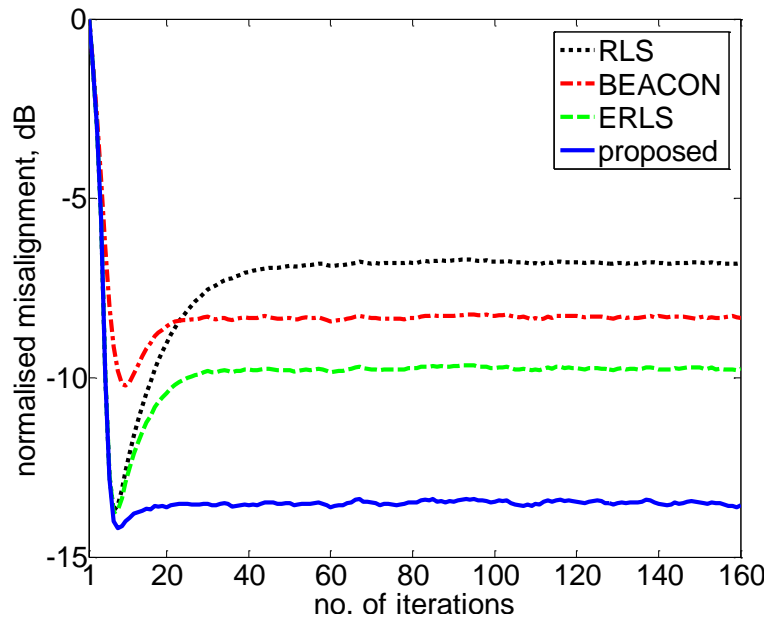


Fig. 1, Normalized misalignment of different algorithms when SNR = 15 dB.

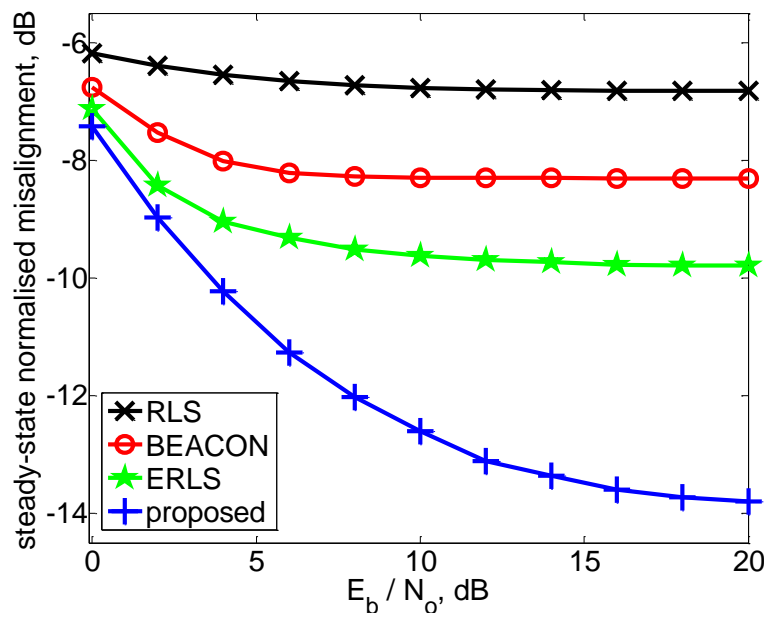


Fig. 2, Steady-state normalized misalignment versus SNR for different algorithms.

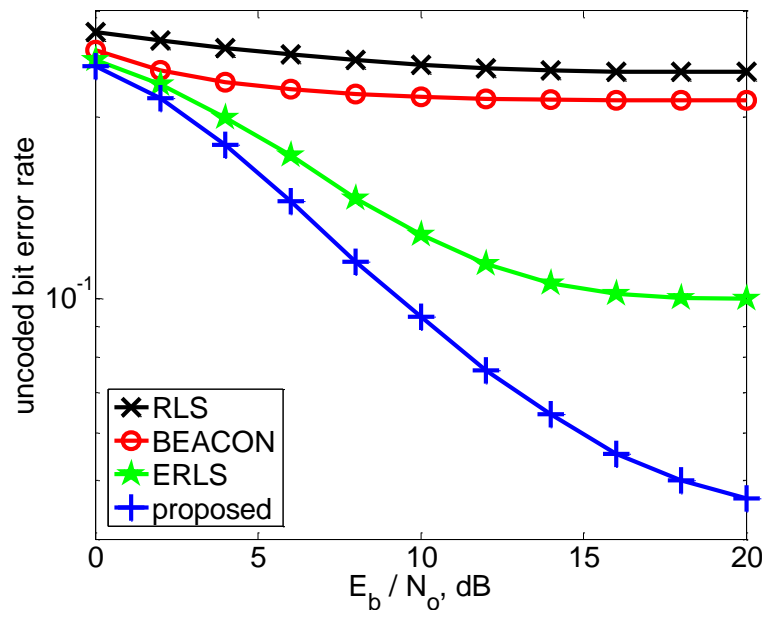


Fig. 3, BER performance of the MMSE-VBLAST MIMO-detector using different channel estimators

# Paper D

Originally published as

R. Arablouei and K. Doğançay, “Low-complexity adaptive decision-feedback equalization of MIMO channels,” *Signal Processing*, vol. 92, pp. 1515-1524, 2012.

Copyright © 2012 Elsevier. Reprinted with permission.

## Low-Complexity Adaptive Decision-Feedback Equalization of MIMO Channels

Reza Arablouei and Kutluyıl Doğançay

### Abstract

A new adaptive MIMO channel equalizer is proposed based on adaptive generalized decision-feedback equalization and ordered-successive interference cancellation. The proposed equalizer comprises equal-length subequalizers, enabling any adaptive filtering algorithm to be employed for coefficient updates. A recently proposed computationally-efficient recursive least-squares algorithm based on dichotomous coordinate descents is utilized to solve the normal equations associated with the adaptation of the new equalizer. Convergence of the proposed algorithm is examined analytically and simulations show that the proposed equalizer is superior to the previously proposed adaptive MIMO channel equalizers by providing both enhanced bit error rate performance and reduced computational complexity. Furthermore, the proposed algorithm exhibits stable numerical behavior and can deliver a trade-off between performance and complexity.

Keywords: MIMO systems; adaptive generalized decision-feedback equalization; ordered-successive interference cancelation; RLS-DCD algorithm; V-BLAST.

### 1. Introduction

Multiple-input multiple-output (MIMO) communication is a promising technology to achieve higher capacity and performance in the rapidly developing modern wireless telecommunication systems such as WiFi, WiMAX, and LTE [1]. However, the tremendous performance enhancements associated with the MIMO systems come at

the expense of drastically more complex signal processing at the receiver. Therefore, the key to the successful utilization of the MIMO technology is the availability of highly integrated and affordable mobile terminals. This brings about the need to develop new efficient receivers for future communication systems. The most challenging part of a MIMO receiver in terms of complexity is the MIMO channel equalizer/detector. Its task is to separate the spatially multiplexed data streams at the receiver. The MIMO channel equalization is performed via a great variety of methods. Among them, vertical Bell Labs layered space-time (V-BLAST) architecture [2] is the most prominent one. V-BLAST has originally been designed to deal with narrowband (flat-fading) MIMO channels and efficiently cancels inter-(sub)channel interference (ICI) to increase reliability of symbol detection.

It is proved that the receiver processing of the V-BLAST architecture may be viewed as a generalized decision-feedback equalizer (GDFE) applied to a MIMO channel [3]. In view of this fact, an adaptive MIMO channel equalizer based on GDFE has been proposed in [4] that can be considered as an adaptive implementation of the detection method introduced by V-BLAST. This equalizer, depicted in Fig. 1, can be viewed as a concatenation of two parts: a conventional linear MIMO equalizer and an ordered-successive decision-feedback interference canceller (OSIC). The former is a bank of feedforward filters (FFF) that resembles the matched filter of a CDMA multiuser detection receiver [5] and corresponds to the nulling vectors of the V-BLAST detector while the latter successively suppresses interference of the detected substreams from the undetected ones by means of cross-layer decision feedback. Feeding back the decisions enables finer equalization of the received signals, i.e. more accurate inter-substream interference cancelation. This interference cancelation is carried out in an ordered fashion to minimize the possibility of detection error propagation. Because of the special structure of this receiver, subequalizer filters of different layers (which equalize and detect different substreams) and their corresponding input vectors have unequal lengths. More notably, the filter length of the subequalizers can vary in time as the detection order of their corresponding substreams can change at different time instants. This property imposes restrictions upon adaptive filtering algorithms to be used for the adaptation of the equalizer parameters. More specifically, fixed-order adaptive filtering algorithms cannot be employed for this purpose since they are not equipped with any mechanism to handle variable filter orders [6]. In [4], the authors have formulated a set of least squares (LS) problems and derived an elegant set of time-update and order-update recursions to solve the LS problems in an adaptive manner. In their algorithm, substreams of different layers are detected successively in an ordered fashion akin to V-BLAST [2]. Only filter coefficients of the first subequalizer (detecting first substream in accordance with the detection ordering) are updated using the recursive least-squares (RLS) algorithm and the coefficients of the other subequalizers are calculated using the coefficients of the first subequalizer, input signal vector and already-

detected symbols of other layers. This coefficient update process is carried out in conjunction with determining the optimal order of detection based on ranking the least square errors. As a downside, this algorithm comprises several recursive calculations at each iteration. Exacerbated by the inherent vulnerability of the conventional RLS algorithm to numerical instability, it exhibits stability problems when long unsupervised packets of data are intended to be sent. In [7], a Cholesky-factorization-based algorithm is proposed that is mathematically equivalent to the algorithm of [4] but is less prone to numerical instability thanks to its underlying square-root implementation. It is also computationally more efficient than the algorithm of [4].

In this paper, we propose a new adaptive equalizer with ordered-successive decision-feedback interference cancelation for flat-fading MIMO channels. In the new equalizer, all subequalizers have an identical filter length. Consequently, it is possible to use any adaptive filtering algorithm to update the equalizer filter coefficients. To update the filter coefficients of the new equalizer and to determine its optimal detection ordering, we develop a new low-complexity and hardware-efficient algorithm based on the recently-proposed *dichotomous coordinate descent-recursive least-squares* (RLS-DCD) algorithm [8], in which the dichotomous coordinate descent (DCD) iterations are used to solve the associated normal equations. The proposed algorithm requires no division or square-root operation and does not suffer from numerical instability problems of the algorithm of [4]. We also show that, neither of the algorithms of [4] and [7] are amenable to the appreciable complexity reduction afforded by the RLS-DCD algorithm.

In section 2, we describe the structure and algorithm of the proposed equalizer for a flat-fading MIMO channel. In section 3, we analyze the computational complexity of the new algorithm and compare it with other existing related algorithms. Convergence properties of the new algorithm are studied in section 4. Simulation results are provided in section 5 and conclusions are drawn in section 6.

## 2. System model and algorithm description

Let us consider a MIMO communication system with  $M$  transmitters and  $N$  receivers, which operates over a flat-fading and rich-scattering wireless channel. The MIMO channel is modeled by

$$\mathbf{r}(n) = \mathbf{H}(n)\mathbf{s}(n) + \mathbf{u}(n) \quad (1)$$

where

$$\mathbf{s}(n) = [s_1(n), s_2(n), \dots, s_M(n)]^T$$

is the  $M \times 1$  vector of simultaneously-transmitted symbols by  $M$  transmitting antennas,

$$\mathbf{r}(n) = [r_1(n), r_2(n), \dots, r_N(n)]^T$$

is the  $N \times 1$  vector of received signals,  $\mathbf{H}(n)$  is the  $N \times M$  channel matrix at time index  $n$ ,  $\mathbf{u}(n)$  represents additive noise and superscript  $T$  denotes matrix transpose.

## 2.1. Equalization and detection

The proposed MIMO channel equalizer is depicted in Fig. 2. It is similar to the equalizer of [4] (Fig. 1) except for one major difference; viz., all subequalizers associated with different layers have the same filter order, i.e. the same number of filter taps. In the proposed equalizer, similar to the equalizer of [4] and other V-BLAST-like detectors, substreams of different layers are equalized and detected successively and in an ordered fashion. Therefore, the new equalizer is constituted of two parts: FFF and OSIC. The FFF is a bank of  $M$  finite impulse response (FIR) filters each having  $N$  taps with corresponding coefficients denoted by  $N \times 1$  vectors  $\mathbf{f}_j(n)$ ,  $j = 1, 2, \dots, M$ . The received signal vector,  $\mathbf{r}(n)$ , passes through this filter bank and produces the FFF outputs as

$$\bar{d}_j(n) = \mathbf{f}_j^*(n)\mathbf{r}(n), \quad j = 1, 2, \dots, M \quad (2)$$

where the superscript  $*$  denotes complex-conjugate transpose. We call these signals, i.e.  $\bar{d}_j(n)$ , *soft detections* since they can be considered as the transmitted symbols corrupted by noise and residual inter-substream interference when the FFF coefficients have converged to their optimal values. These soft detections then go through the OSIC. The OSIC, as shown in Fig. 2, comprises  $M$  hard decision devices and a bank of  $M$  FIR filters each having  $M - 1$  taps with corresponding coefficients denoted by  $(M - 1) \times 1$  vectors  $\mathbf{c}_j(n)$ ,  $j = 1, 2, \dots, M$ . Hence, in the proposed equalizer, there is an equal number of coefficients at each decision-feedback interference canceller associated to each substream. The motivation behind the design of this structure was to enable the equalizer to employ fixed-order adaptive filtering algorithms for adaptation of its coefficients. Considering the fact that the procedure of interference cancelation and detection of the substreams is supposed to be carried out sequentially and in an ordered manner, in the new equalizer's OSIC, the soft detections of the undetected substreams are used as filter input. After detection (interference cancelation and passing the decision device), the soft detections are replaced by their corresponding hard decisions. Therefore, assuming that the optimal detection ordering,

$$\mathbf{o}(n) = \{o_1(n), o_2(n), \dots, o_M(n)\},$$

which is a rearranged version of the original substream indexing set  $\{1, 2, \dots, M\}$ , has been determined in the previous iteration, input vector of the filter  $\mathbf{c}_{o_i}(n)$  is

$$\mathbf{d}_{o_i}(n) = [\hat{d}_1(n), \dots, \hat{d}_{o_i-1}(n), \hat{d}_{o_i+1}(n), \dots, \hat{d}_M(n)]^T \quad (3)$$

where

$$\hat{d}_j(n) = \begin{cases} d_j(n) & \text{if } j\text{th substream is detected before } o_i\text{th substream} \\ \bar{d}_j(n) & \text{if } j\text{th substream is detected after } o_i\text{th substream} \end{cases}, \quad (4)$$

$$j = 1, \dots, o_i - 1, o_i + 1, \dots, M$$

and  $d_j(n)$  is the detected symbol of the  $j$ th substream. In other words, the OSIC input vector  $\mathbf{d}_{o_i}(n)$  for the  $o_i$ th substream is formed by detected symbols of the substreams which come before the  $o_i$ th one in the detection ordering and soft detections of the substreams which come after it. The output of the equalizer is the detected symbols calculated by

$$d_{o_i}(n) = \begin{cases} \text{decision}(\tilde{d}_{o_i}(n)) & \text{in decision-directed mode} \\ s_{o_i}(n) & \text{in training mode} \end{cases}, \quad i = 1, 2, \dots, M \quad (5)$$

where

$$\tilde{d}_{o_i}(n) = \bar{d}_{o_i}(n) + \mathbf{c}_{o_i}^*(n)\mathbf{d}_{o_i}(n) \quad (6)$$

and  $\text{decision}(\cdot)$  is the hard decision function. The use of index  $o_i$  in (5) implies that the equalization and detection process is carried out sequentially and in accordance with  $o(n)$ .

## 2.2. Updating the coefficients

We organize the filter coefficients of the FFF and the OSIC together in the following coefficient vectors

$$\mathbf{w}_j(n) = \begin{bmatrix} \mathbf{f}_j(n) \\ \mathbf{c}_j(n) \end{bmatrix}, \quad j = 1, 2, \dots, M. \quad (7)$$

Filter input vectors associated with (7) can also be defined as

$$\mathbf{y}_j(n) = \begin{bmatrix} \mathbf{r}(n) \\ \mathbf{d}_j(n) \end{bmatrix}, \quad j = 1, 2, \dots, M. \quad (8)$$

Recalling the equivalence of the V-BLAST architecture and the MIMO-GDFE [3], [4], optimal detection ordering under the minimum mean square error (MMSE) criterion can be found by sorting the following least-squares errors (LSEs):

$$\mathcal{E}_j(n) = \sum_{l=1}^n \lambda^{n-l} |d_j(l) - \mathbf{w}_j^*(n) \mathbf{y}_j(l)|^2, \quad j = 1, 2, \dots, M \quad (9)$$

where  $\lambda$  is a forgetting factor satisfying  $0 < \lambda \leq 1$  and the equalizer coefficients  $\mathbf{w}_j(n)$ , minimizing  $\mathcal{E}_j(n)$ , are the solutions of the following normal equations:

$$\mathbf{\Phi}_j(n) \mathbf{w}_j(n) = \mathbf{z}_j(n), \quad j = 1, 2, \dots, M. \quad (10)$$

Here,  $\mathbf{\Phi}_j(n)$  is the exponentially-weighted input autocorrelation matrix of the  $j$ th subequalizer

$$\mathbf{\Phi}_j(n) = \sum_{l=1}^n \lambda^{n-l} \mathbf{y}_j(l) \mathbf{y}_j^*(l) \quad (11)$$

and  $\mathbf{z}_j(n)$  is the exponentially-weighted cross-correlation vector between input and desired signal of the  $j$ th subequalizer

$$\mathbf{z}_j(n) = \sum_{l=1}^n \lambda^{n-l} \mathbf{y}_j(l) d_j^*(l). \quad (12)$$

We can rewrite (11) and (12) as

$$\mathbf{\Phi}_j(n) = \lambda \mathbf{\Phi}_j(n-1) + \mathbf{y}_j(n) \mathbf{y}_j^*(n) \quad (13)$$

and

$$\mathbf{z}_j(n) = \lambda \mathbf{z}_j(n-1) + \mathbf{y}_j(n) d_j^*(n). \quad (14)$$

From (3), (4), and (8), we know that the filter input vector of the  $o_i$ th subequalizer,  $\mathbf{y}_{o_i}(n)$ , differs from the filter input vector of the  $o_{i-1}$ th subequalizer,  $\mathbf{y}_{o_{i-1}}(n)$ , in only one entry. We can exploit this redundancy to reduce the number of operations required to update  $\mathbf{\Phi}_j(n)$  matrices for all subequalizers at each time instant. Therefore, we first update  $\mathbf{\Phi}_{o_1}(n)$  using (13) and then update the other autocorrelation matrices successively while avoiding update of their common entries. More specifically, to update  $\mathbf{\Phi}_{o_i}(n)$  for  $i = 2, \dots, M$ , we update only one vector:

$$\mathbf{\Phi}_{o_i}^{(\kappa)}(n) = \lambda \mathbf{\Phi}_{o_i}^{(\kappa)}(n-1) + \mathbf{y}_{o_i}(n) d_{o_{i-1}}^*(n), \quad i = 2, \dots, M \quad (15)$$

where  $\mathbf{\Phi}_{o_i}^{(\kappa)}(n)$  is the  $\kappa$ th column of  $\mathbf{\Phi}_{o_i}(n)$  and

$$\kappa = \begin{cases} N + o_{i-1} & \text{if } o_i > o_{i-1} \\ N + o_{i-1} - 1 & \text{if } o_i < o_{i-1} \end{cases}, \quad i = 2, \dots, M. \quad (16)$$

In order to update  $\Phi_{o_i}(n)$ , we enlarge  $\Phi_{o_{i-1}}(n)$  by inserting  $\Phi_{o_i}^{(\kappa)}(n)$  as the  $(N + o_{i-1})$ th column and  $\Phi_{o_i}^{(\kappa)*}(n)$  as the  $(N + o_{i-1})$ th row. Then, we shrink the enlarged matrix by dropping the  $(N + o_i)$ th column and row. To reduce the computational burden further, similar to [8], we choose the forgetting factor as  $\lambda = 1 - 2^{-s}$  where  $s$  is a positive integer. Consequently, multiplications by  $\lambda$  can be replaced by bit-shifts and additions. In addition, knowing that  $\Phi_{o_1}(n)$  is symmetric, we only compute the upper triangular part of the matrix.

In order to solve the normal equations of (10) in a computationally efficient way, we employ the exponentially-weighted RLS algorithm proposed in [8], which utilizes the dichotomous coordinate descent (DCD) iterations [9]. This algorithm, called RLS-DCD, uses no division or square-root operation and requires much less number of multiplications compared to the conventional (matrix-inversion-lemma-based) or square-root (QR-decomposition-based) RLS algorithms. The DCD algorithm that iteratively solves the associated normal equations in the RLS-DCD algorithm, falls into the class of shift-and-add algorithms and is a multiplicationless method dominated by additions. Nevertheless, it yields an approximate solution and its accuracy depends on the number of exercised iterations. However, it is shown in [8] that performance of the RLS-DCD algorithm can be made arbitrarily close to that of the conventional RLS algorithm by increasing the number of iterations and the resolution of the step size.

Let us define the residual vectors as

$$\boldsymbol{\rho}_j(n) = \mathbf{z}_j(n) - \Phi_j(n)\mathbf{w}_j(n), \quad j \in o(n) \quad (17)$$

and the auxiliary vectors as

$$\boldsymbol{\beta}_j(n) = \lambda\boldsymbol{\rho}_j(n-1) - \mathbf{y}_j(n)e_j^*(n), \quad j \in o(n) \quad (18)$$

where  $e_j(n)$  is the *a priori* filter output error of the  $j$ th subequalizer

$$\begin{aligned} e_j(n) &= d_j(n) - \mathbf{w}_j^*(n-1)\mathbf{y}_j(n) \\ &= d_j(n) - \tilde{d}_j(n). \end{aligned} \quad (19)$$

As shown in [8], the solution of

$$\Phi_j(n)\Delta\mathbf{w}_j(n) = \boldsymbol{\beta}_j(n), \quad j \in o(n) \quad (20)$$

using the DCD iterations provides both coefficient update vectors,  $\Delta\mathbf{w}_j(n)$ , and residual vectors,  $\boldsymbol{\rho}_j(n)$ . Having calculated  $\Delta\mathbf{w}_j(n)$ , the coefficient vectors are updated via

$$\mathbf{w}_j(n) = \mathbf{w}_j(n-1) + \Delta\mathbf{w}_j(n), \quad j \in o(n). \quad (21)$$

### 2.3. Ordering update

Similar to the equalizers of [4] and [7], the proposed equalizer corresponds to the V-BLAST with the MMSE criterion [3] when the optimal detection order is found by sorting the LSEs in (9) and the equalizer filter coefficients are given as the solutions of the normal equations in (10). In fact, the proposed equalizer can be viewed as an adaptive implementation of the V-BLAST detector. Therefore, as shown in [3] and [4], sorting the LSEs in (9) corresponds to sorting the signal-to-noise ratios in the V-BLAST. Hence, ordering of the layers for detection at time instant  $n + 1$  is found by

$$o(n + 1) = \text{sort}\{\mathcal{E}_j(n), j = 1, 2, \dots, M\} \quad (22)$$

where the function  $\text{sort}\{a\}$  sorts the elements of the set  $a$  in ascending order and returns indices of the sorted set. We can compute the LSEs recursively via

$$\mathcal{E}_j(n) = \lambda \mathcal{E}_j(n - 1) + |\epsilon_j(n)|^2, \quad j = 1, 2, \dots, M \quad (23)$$

where  $\epsilon_j(n)$  is the *a posteriori* filter output error of the  $j$ th subequalizer

$$\epsilon_j(n) = d_j(n) - \mathbf{w}_j^*(n) \mathbf{y}_j(n), \quad j = 1, 2, \dots, M. \quad (24)$$

However, we may consider the *a priori* error as a tentative value of the *a posteriori* error before updating the coefficients [18]. Hence, in order to reduce complexity of the detection ordering, we use LSEs based on the *a priori* errors rather than the *a posteriori* errors, i.e.  $\hat{\mathcal{E}}_j(n)$  instead of  $\mathcal{E}_j(n)$  in (22) with

$$\hat{\mathcal{E}}_j(n) = \lambda \hat{\mathcal{E}}_j(n - 1) + |e_j(n)|^2, \quad j = 1, 2, \dots, M. \quad (25)$$

Extensive simulation experiments confirmed that this change does not affect the equalization performance significantly.

The proposed algorithm is summarized in Table 1 including the number of required complex arithmetic operations by each step at each iteration assuming  $M = N$ . The DCD algorithm with a leading element for solving a complex-valued linear system of equations [10] is also presented in Table 2. In the algorithm of Table 2, the variable  $\varsigma$  indicates which component is being processed, i.e.  $\varsigma = 1$  for real and  $\varsigma = \sqrt{-1}$  for imaginary component. Moreover,  $\rho_j^p(n)$  and  $\Delta w_j^p(n)$  are the  $p$ th elements of the vectors  $\boldsymbol{\rho}_j(n)$  and  $\Delta \mathbf{w}_j(n)$  respectively,  $\Phi_j^{p,p}(n)$  is the  $(p, p)$ th entry of the matrix  $\boldsymbol{\Phi}_j(n)$  and  $\text{sgn}(\cdot)$  stands for the signum function.

The DCD algorithm utilizes three user-defined parameters, namely  $N_u$ ,  $M_b$ , and  $H$ , that control its accuracy and complexity. In fact, the first two establish a trade-off between complexity and performance for the RLS-DCD algorithm. The integer parameter  $N_u$  represents the number of iterative updates performed at each run of the algorithm. In

other words,  $N_u$  determines the maximum number of filter coefficients that can be updated at each time instant. Hence, in general, adaptive filtering based on the DCD algorithm implements a form of *selective partial updates* [11]. It is known that by selective partial updating, one can trade performance for complexity [12]. In the DCD algorithm, the step-size  $\alpha$  can accept one of  $M_b$  predefined values corresponding to representation of the elements of the vector  $\Delta\mathbf{w}_j(n)$  as fixed-point words with  $M_b$  bits within an amplitude range of  $[-H, H]$ .

### 3. Computational complexity

Considering the case of equal number of transmitter and receiver antennas,  $M = N$ , the proposed algorithm requires  $8N^2 - 5N + 1$  complex multiplications,  $(16 + 4N_u)N^2 + (M_b - 3N_u - 25)N/2 + 3$  complex additions, and no division or square-root operations. The computational complexity of the proposed algorithm, the V-BLAST algorithm, and the algorithms of [4] and [7]<sup>2</sup> is presented in Table 3. For the V-BLAST algorithm, we consider the fast V-BLAST algorithm proposed in [13] with channel estimation and tracking using the RLS algorithm. It is seen that the computational complexity of the proposed algorithm is  $O(N^2)$  while the complexity of other algorithms is  $O(N^3)$ . We should note that neither the algorithm of [7] nor the order-update recursions of the algorithm of [4] can utilize the RLS-DCD algorithm to benefit from its computational efficiency. This is because of their unequal subequalizer filter lengths. However, for the time-update part of the algorithm of [4] and the channel tracking for the V-BLAST algorithm, the RLS-DCD algorithm can be employed. Replacing the conventional RLS recursions in these algorithms with the RLS-DCD recursions reduces the number of required complex multiplications by  $(5N^2 - 3N)/4$  and increases the number of required complex additions by  $(2N_u + 3/2)N^2 + (N_u + M_b + 1)N/2$ .

The number of required complex multiplications by different algorithms versus the number of transmitter/receiver antennas is shown in Fig. 3. The percentage of the saved multiplications by the new algorithm with respect to the algorithms of [4] and [7] is also shown in Fig. 3. As an example for fixed-point implementation, using the unit-gate area model of [14], a 16-bit carry-lookahead adder requires 204 gates [15] while a 16-bit array multiplier requires 2,336 gates [16]. Using these numbers, the total number of required gates by different algorithms is shown in Fig. 4 for different numbers of transmitter/receiver antennas and considering  $M_b = 12$  and

$$N_u = \begin{cases} N & \text{if } N < 4 \\ 4 & \text{if } N \geq 4 \end{cases}$$

---

<sup>2</sup> Complexity of the algorithms of [4] and [7] is from [7].

This figure also shows the percentage of the saved gates by the new algorithm with respect to the algorithms of [4] and [7]. For simplicity, we assume that a division or square-root operation has the same complexity as a multiplication operation.

The presented comparisons demonstrate that using the new algorithm, a significant saving in complexity is possible, in particular, when the number of transmitter/receiver antennas is relatively large. Moreover, the new algorithm does not require any division or square-root operation, while these operations can add to the complications of implementing the algorithms of [4] and [7] on hardware. Another important observation is that utilizing the RLS-DCD algorithm in the time-update recursions of the algorithm of [4] and channel estimation for the V-BLAST algorithm does not yield a substantial complexity reduction. This verifies that the computational efficiency of the new algorithm is mainly attributable to the proposed equalizer's special structure, which enables it to carry out the coefficient updates and especially the detection ordering with less effort compared to the algorithms of [4] and [7]. Capability of incorporating any adaptive filtering algorithm is another advantage of the proposed equalizer, which contributes to its complexity reduction and capacity of trading off performance for complexity.

## 4. Convergence analysis

### 4.1. Theory

We examine convergence of the proposed algorithm during the training mode for a time-invariant channel using the conventional RLS algorithm for coefficient updates. Since it is too complicated to prove the convergence directly, we adopt an indirect approach based on the observation that for a time-invariant channel the optimal detection order is fixed at all time instants. For convenience and without loss of generality, we assume that the optimal ordering is  $\mathbf{o}^o = \{1, 2, \dots, M\}$ . For the analysis, we assume that

- The entries of the noise vector,  $\mathbf{u}(n)$ , are independent and identically distributed (i.i.d.) complex Gaussian with zero mean and variance of  $\sigma_u^2$ .
- The transmitted symbol vector  $\mathbf{s}(n)$  satisfies  $E[\mathbf{s}(n)\mathbf{s}^*(n)] = P_t \mathbf{I}_M$ , where  $\mathbf{I}_M$  is the  $M \times M$  identity matrix and  $P_t$  is the transmitted power of each layer.

The optimal weight vector for the  $i$ th subequalizer,  $\mathbf{w}_i = [\mathbf{f}_i^T, \mathbf{c}_i^T]^T$ , is constant over a fixed channel and calculated by solving the following linear system of equations:

$$\mathbf{R}_i \mathbf{w}_i = \mathbf{p}_i \quad (26)$$

where  $\mathbf{R}_i$  is the input autocorrelation matrix for the  $i$ th subequalizer given by

$$\begin{aligned}
\mathbf{R}_i &= E[\mathbf{y}(n)\mathbf{y}^*(n)] \\
&= E \left[ \begin{bmatrix} \mathbf{r}(n) \\ d_1(n) \\ \vdots \\ d_{i-1}(n) \\ \mathbf{f}_{i+1}^* \mathbf{r}(n) \\ \vdots \\ \mathbf{f}_M^* \mathbf{r}(n) \end{bmatrix} \begin{bmatrix} \mathbf{r}(n) \\ d_1(n) \\ \vdots \\ d_{i-1}(n) \\ \mathbf{f}_{i+1}^* \mathbf{r}(n) \\ \vdots \\ \mathbf{f}_M^* \mathbf{r}(n) \end{bmatrix}^* \right] \\
&= P_t \begin{bmatrix} \mathbf{A} & \mathbf{H}_i & \mathbf{A}\mathbf{F}_i \\ \mathbf{H}_i^* & \mathbf{I}_{i-1} & \mathbf{H}_i^* \mathbf{F}_i \\ \mathbf{F}_i^* \mathbf{A} & \mathbf{F}_i^* \mathbf{H}_i & \mathbf{F}_i^* \mathbf{A}\mathbf{F}_i \end{bmatrix}
\end{aligned} \tag{27}$$

and  $\mathbf{p}_i$  is the cross-correlation vector between the input and the desired signal of the  $i$ th subequalizer

$$\mathbf{p}_i = P_t \begin{bmatrix} \mathbf{I}_N \\ \mathbf{0}_{i-1} \\ \mathbf{F}_i^* \end{bmatrix} \mathbf{h}_i. \tag{28}$$

Here,

$$\mathbf{A} = \mathbf{H}\mathbf{H}^* + \frac{\sigma_u^2}{P_t} \mathbf{I}_N, \tag{29}$$

$$\mathbf{F}_i = [\mathbf{f}_{i+1}, \mathbf{f}_{i+2}, \dots, \mathbf{f}_M], \tag{30}$$

and

$$\mathbf{H}_i = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_{i-1}] \tag{31}$$

where  $\mathbf{h}_i$  is the  $i$ th column of the channel matrix  $\mathbf{H}$ .

Since the soft detections are inner product of the received signal vector and the FFF coefficients, due to the use of them in the OSIC part of the proposed equalizer, the system of (26) is under-determined for  $i \neq M$ . This can make the autocorrelation matrix  $\mathbf{R}_i$  rank-deficient and the solution for  $\mathbf{w}_i$  non-unique. Therefore, to find the minimum-Euclidean-norm solution for  $\mathbf{w}_i$ , we use the regularized inverse matrix [17]

$$\mathbf{R}_i^\dagger = (\mathbf{R}_i^* \mathbf{R}_i + \gamma \mathbf{I}_{N+M-1})^{-1} \mathbf{R}_i^* \tag{32}$$

and compute the optimal filter coefficients via

$$\mathbf{w}_i = \mathbf{R}_i^\dagger \mathbf{p}_i. \tag{33}$$

The regularization parameter  $\gamma$  in (32) is a small positive number. Note that in the proposed algorithm, initialization of the autocorrelation matrixes to  $\Phi_i(0) = \delta \mathbf{I}_{N+M-1}$  acts as regularization.

The estimation error of the  $i$ th ideal subequalizer is defined by

$$e_i^o(n) = d_i(n) - \mathbf{w}_i^* \mathbf{y}_i(n). \quad (34)$$

Similar to [4], we assume that  $e_i^o(n)$  is white with a zero mean and variance of  $\sigma_i^2$ , which is computed as

$$\begin{aligned} \sigma_i^2 &= E[|d_i(n) - \mathbf{w}_i^* \mathbf{y}_i(n)|^2] \\ &= P_t - \mathbf{p}_i^* \mathbf{w}_i. \end{aligned} \quad (35)$$

Adopting the analysis of [18], we can show that for  $\lambda = 1$ , mean-squared deviation (MSD), which is defined as mean-squared norm of the weight-error vector, can be written as

$$\begin{aligned} E[\mathbf{v}_i^*(n) \mathbf{v}_i(n)] &= \sigma_i^2 \text{tr} \left[ E[\mathbf{\Phi}_i^{-1}(n)] \right] \\ &\cong \sigma_i^2 \frac{\text{tr}[\mathbf{R}_i^\dagger]}{n - N - M}, \quad n > N + M \end{aligned} \quad (36)$$

where  $\mathbf{v}_i(n) = \mathbf{w}_i(n) - \mathbf{w}_i$  is the weight-error vector and  $\text{tr}[\cdot]$  denotes trace of a matrix. Mean-squared error (MSE) is also calculated via

$$\begin{aligned} E[|e_i(n)|^2] &= E[|d_i(n) - \mathbf{w}_i^*(n-1) \mathbf{y}_i(n)|^2] \\ &\cong \sigma_i^2 + (N + M - 1) \frac{\sigma_i^2}{n - N - M}, \quad n > N + M. \end{aligned} \quad (37)$$

In the following, the theoretical results of (36) and (37) are compared with the corresponding experimental results.

## 4.2. Comparison with experiment

A MIMO communication system with four transmitter and four receiver antennas was considered. The transmitted signal vectors were spatially-orthogonal Walsh sequences of length 120 in BPSK modulation. The fixed channel matrix was composed by independent complex Gaussian entries with zero mean and unit power. The entries of the noise vector were also i.i.d. complex Gaussian with zero mean. Forgetting factor was  $\lambda = 1$  and energy per bit to noise power spectral density ratio was  $E_b/N_0 = 15$  dB. The theoretical results for MSE and MSD are compared with the empirical ones (ensemble-averaged over  $10^4$  independent runs) in Figs. 5 and 6, respectively. There is a good agreement between the theoretical and empirical results in both figures. These results corroborate that the proposed algorithm behaves as a set of parallel RLS algorithms and converges to the optimal detection ordering as well as the optimal filter coefficients.

## 5. Simulation studies

In this section, we provide simulation results to assess performance of the new algorithm. A MIMO communication system with  $M$  transmitter and  $N$  receiver antennas is considered. Sub-channels between all transmitter and receiver pairs are independent Rayleigh fading channels and vary in time based on Jakes model [19] with a normalized Doppler frequency  $f_D T_s = 6 \times 10^{-4}$  where  $f_D$  is the maximum Doppler frequency shift and  $T_s$  is the transmission symbol period. The transmitted signal is uncoded and modulated using QPSK scheme. It is grouped in packets of data each containing  $L_d$  vectors of transmitted symbols while  $L_t$  vectors are used for training. A forgetting factor of  $\lambda = 1 - 2^{-7} \simeq 0.9922$  was also used regarding the assumed normalized Doppler frequency. The results were obtained by ensemble-averaging over  $10^3$  independent runs and over all the layers.

In Fig. 7, we compare bit error rate (BER) performance of the new equalizer when using the conventional RLS algorithm and the RLS-DCD algorithm with different numbers of DCD iterations at each time instant ( $N_u$ ). We observe that for  $N_u = 4$ , the RLS-DCD algorithm performs almost the same as the conventional RLS algorithm. The results of Fig. 7 were obtained for a  $4 \times 4$  MIMO system but experiments with larger number of antennas also showed that  $N_u = 4$  is a good choice for a wide range of practical antenna numbers. Through numerous experiments, we also discovered that increasing  $M_b$  to more than 12 does not improve the performance noticeably. Therefore, in the simulations presented here, we chose  $M_b = 12$ .

Next, we explored the efficiency of the new algorithm in converging to the optimal detection ordering. Fig. 8 shows mean ordering difference (MOD) of the new algorithm, algorithm of [7], and the V-BLAST with RLS channel tracking for the experiment of Fig. 7. The ordering difference is evaluated with respect to the ordering of the V-BLAST algorithm with known channel, i.e.  $o^v(n) = \{o_1^v(n), o_2^v(n), \dots, o_M^v(n)\}$ , via

$$\text{MOD}(n) = E \left[ \sum_{i=1}^M \pi_i(n) \right]$$

where

$$\pi_i(n) = \begin{cases} 1 & \text{if } o_i(n) \neq o_i^v(n) \\ 0 & \text{if } o_i(n) = o_i^v(n) \end{cases}$$

The new algorithm converges slightly faster but to the same level of steady-state MOD as the algorithm of [7]. We should note that in view of the mathematical equivalence of the algorithms of [4] and [7], here we only simulate the algorithm of [7] since it is computationally more efficient than the algorithm of [4].

In Figs. 9 and 10, BER performance of the new algorithm is compared with the algorithm of [7], linear RLS equalizer (without decision-feedback interference cancellation), V-BLAST with known channel, and V-BLAST with and without channel tracking. The results of Fig. 9 were obtained for a  $4 \times 4$  MIMO system and the results of Fig. 10 for an  $8 \times 8$  MIMO system. In V-BLAST with channel tracking, an RLS adaptive filter is utilized to identify and track the channel. The noise power is also estimated recursively using the *a posteriori* error vector of this filter

$$\hat{\sigma}_u^2(n) = \lambda \hat{\sigma}_u^2(n-1) + \frac{(1-\lambda)}{N} \|\mathbf{r}(n) - \hat{\mathbf{H}}(n)\mathbf{d}(n)\|^2$$

where  $\hat{\mathbf{H}}(n)$  is the estimated channel matrix. In the V-BLAST without channel tracking, the channel and the noise power are estimated during the training mode and kept fixed during the decision-directed mode. From Figs. 9 and 10, we observe that there is an error floor at high values of  $E_b/N_0$  for each equalizer that in fact determines the performance at high  $E_b/N_0$  values. The emergence of the error floor can be attributed to the well-known error propagation phenomenon inevitably caused by occasional erroneous decisions and their destructive effect on the subsequent decisions.

The BER performance of the new algorithm is compared with the other algorithms in Fig. 11 for a fixed  $E_b/N_0 = 15$  dB and different values of the normalized Doppler shift. The forgetting factor was duly adjusted according to the corresponding Doppler shift.

Finally, in order to evaluate numerical stability of the new algorithm, a set of simulations was carried out for a very long sequence of transmitted symbol vectors, i.e.  $L_d = 10^4$ . No instability problem was observed during the experiments.

## 6. Conclusion

In order to efficiently equalize time-varying MIMO channels, a new adaptive MIMO decision-feedback equalizer was developed. The new equalizer performs ordered-successive interference cancelation by feeding the decisions of already-detected substreams to the subequalizers of other layers. The proposed receiver structure assumes identical filter orders for the subequalizers of all layers. This results in a more tractable equalization problem than that of the previously-proposed variable-order ones and enables the use of any adaptive filtering algorithm for the corresponding filter coefficient updates. For the decision-feedback interference cancelation, undetected hard decision of any substream is substituted by its soft detection. The coefficients of the proposed equalizer are updated using an approach based on the RLS-DCD algorithm. The new algorithm enjoys superior BER performance as well as appreciably reduced complexity in comparison with the previous ones. It can provide a trade-off between complexity and performance and

has a stable numerical behavior. The new algorithm can also be easily extended to equalize frequency-selective fading MIMO channels.

## References

- [1] H. Jafarkhani, *Space-Time Coding: Theory and Practice*, Cambridge, U.K.: Cambridge University Press, 2005.
- [2] G. J. Foschini, G. D. Golden, R. A. Valenzuela and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 11, pp. 1841–1852, Nov. 1999.
- [3] G. Ginis and J. M. Cioffi, "On the relation between V-BLAST and the GDFE," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 364–366, Sep. 2001.
- [4] J. Choi, H. Yu, and Y. H. Lee, "Adaptive MIMO decision feed-back equalization for receivers with time-varying channels," *IEEE Trans. on Signal Process.*, vol. 53, no. 11, pp. 4295–4303, 2005.
- [5] S. Verdu, *Multuser Detection*, Cambridge, U.K.: Cambridge Univ. Press, 1998.
- [6] A. H. Sayed, *Adaptive Filters*, Hoboken, NJ: Wiley, 2008.
- [7] A. A. Rontogiannis, V. Kekatos, and K. Berberidis, "A square-root adaptive V-BLAST algorithm for fast time-varying MIMO channels," *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 265–268, 2006.
- [8] Y. Zakharov, G. White, and J. Liu, "Low complexity RLS algorithms using dichotomous coordinate descent iterations," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3150–3161, Jul. 2008.
- [9] Y. V. Zakharov and T. C. Tozer, "Multiplication-free iterative algorithm for LS problem," *Electron. Lett.*, vol. 40, no. 9, pp. 567–569, Apr. 2004.
- [10] J. Liu, Y. V. Zakharov, and B. Weaver, "Architecture and FPGA design of dichotomous coordinate descent algorithms," *IEEE Trans. Circuits Syst. I*, vol. 56, no. 11, pp. 2425–2438, Nov. 2009.
- [11] K. Doğançay and O. Tanrikulu, "Adaptive filtering algorithms with selective partial updates," *IEEE Trans. Circuits Syst. II*, vol. 48, no. 8, pp. 762–769, Aug. 2001.
- [12] K. Doğançay, *Partial-Update Adaptive Signal Processing: Design, Analysis and Implementation*, Academic Press, Oxford, UK, 2008.

- [13] J. Benesty, Y. Huang, and J. Chen, "A fast recursive algorithm for optimum sequential signal detection in a BLAST system," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1722–1730, Jul. 2003.
- [14] A. Tyagi, "A reduced-area scheme for carry-select adders," *IEEE Trans. Comput.*, vol. 42, no. 10, pp. 1163–1170, Oct. 1993.
- [15] R. Zimmermann, *Binary Adder Architectures for Cell-Based VLSI and their Synthesis*, PhD dissertation, Swiss Federal Institute of Technology, Zurich, 1997.
- [16] E. E. Swartzlander, Jr. and H. H. Saleh, "Floating-point implementation of complex multiplication," in *Proc. 43th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, USA, 2009, pp. 926–929.
- [17] G. H. Golub and C. F. Van Loan, *Matrix Computations*, third ed., Baltimore, MD: Johns Hopkins Univ. Press, 1996.
- [18] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2002.
- [19] W. C. Jakes, *Microwave Mobile Communications*, New York: Wiley, 1974.

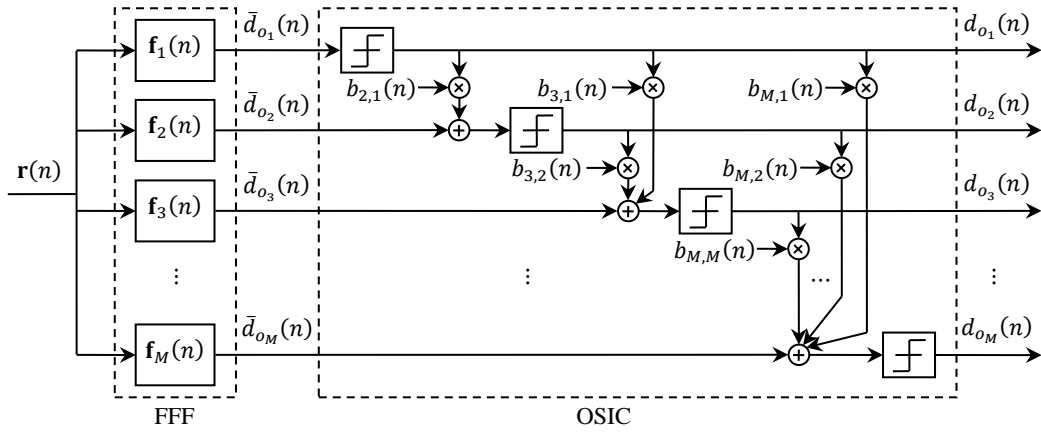


Fig. 1, Block diagram of the equalizer of [4].

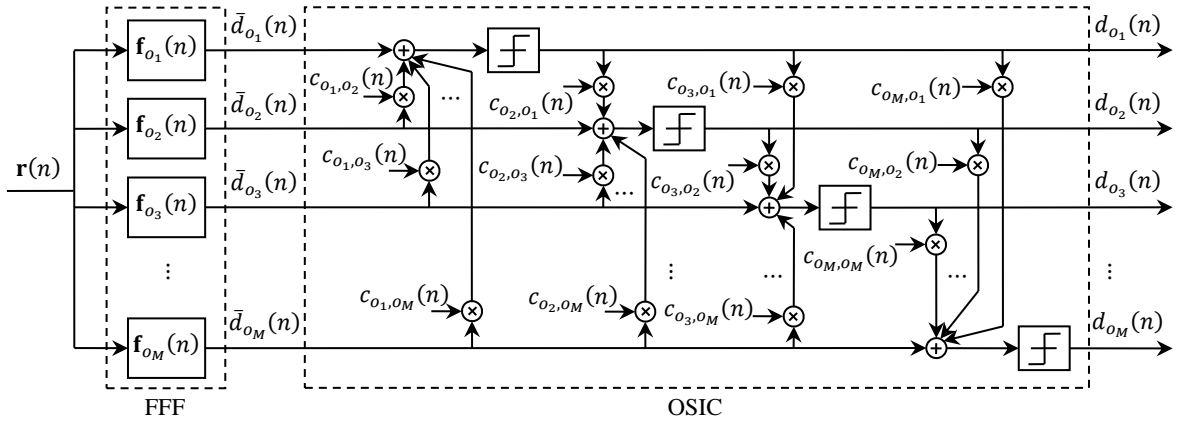


Fig. 2, Block diagram of the proposed equalizer.

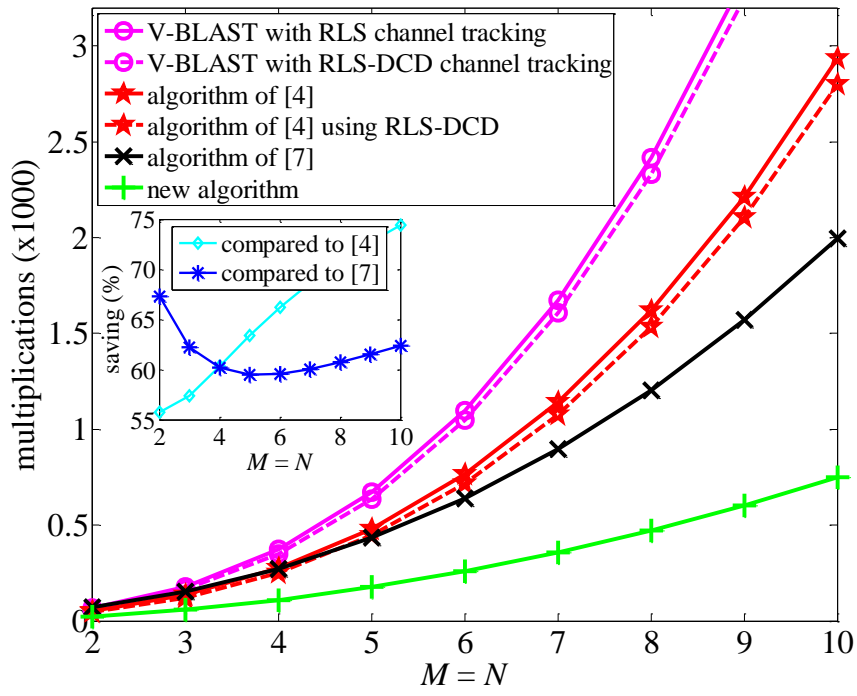


Fig. 3, Complexity comparison of different algorithms in terms of number of required multiplications.

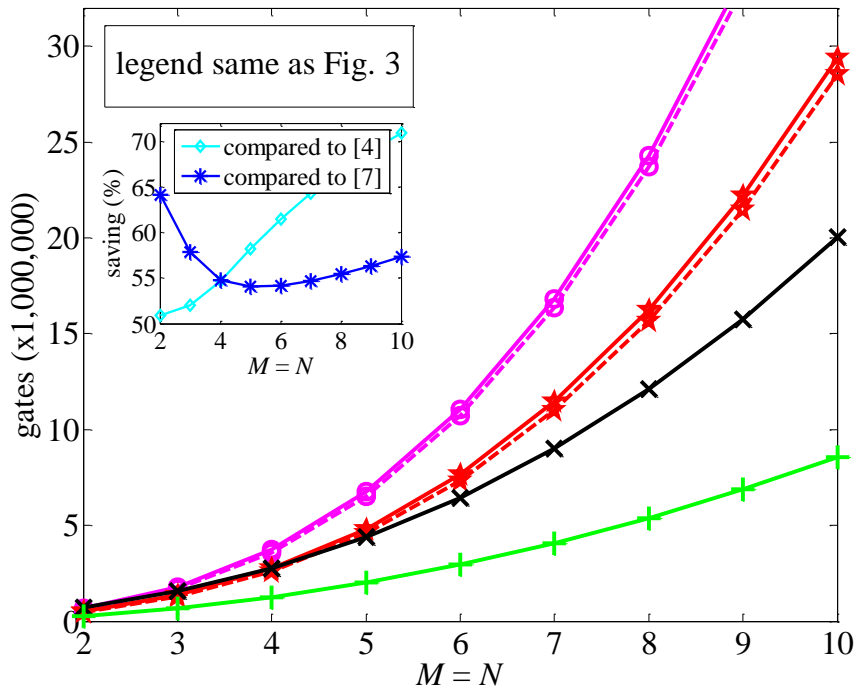


Fig. 4, Complexity comparison of different algorithms in terms of number of required gates.

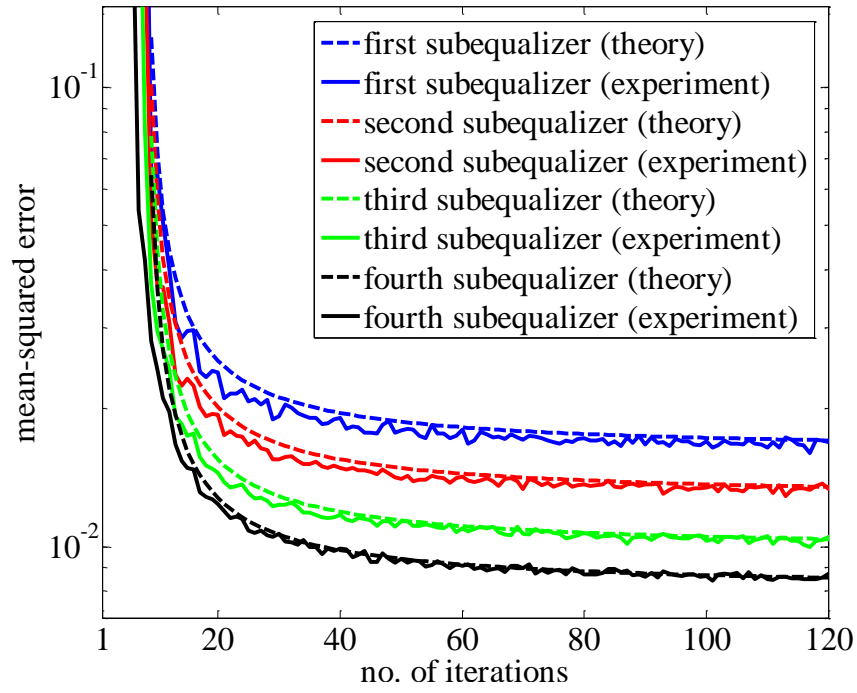


Fig. 5, Theoretical and experimental mean-squared error of the proposed equalizer for  $M = 4, N = 4, \lambda = 1$  and  $E_b/N_0 = 15$  dB.

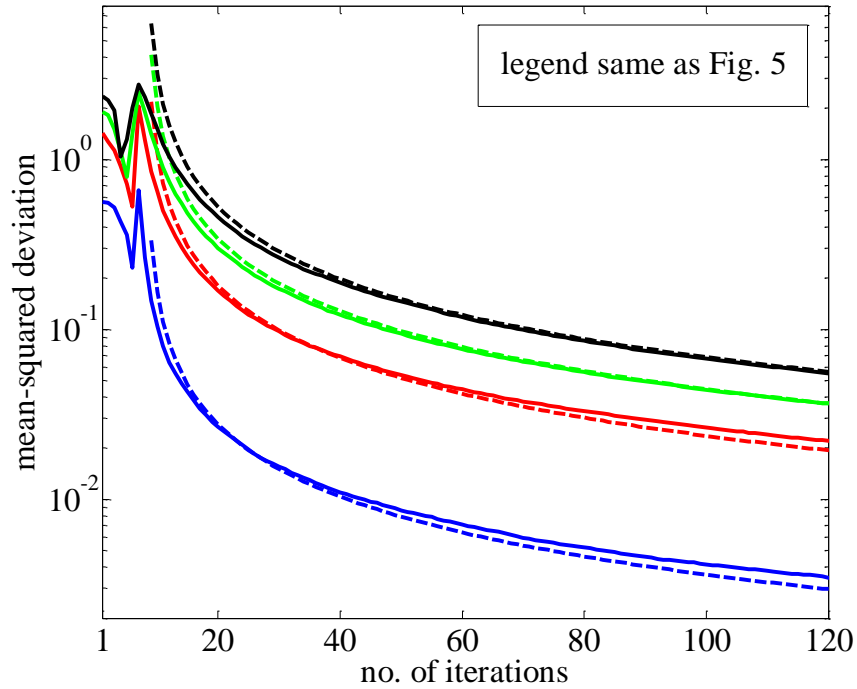


Fig. 6, Theoretical and experimental mean-squared deviation of the proposed equalizer for  $M = 4, N = 4, \lambda = 1$  and  $E_b/N_0 = 15$  dB.

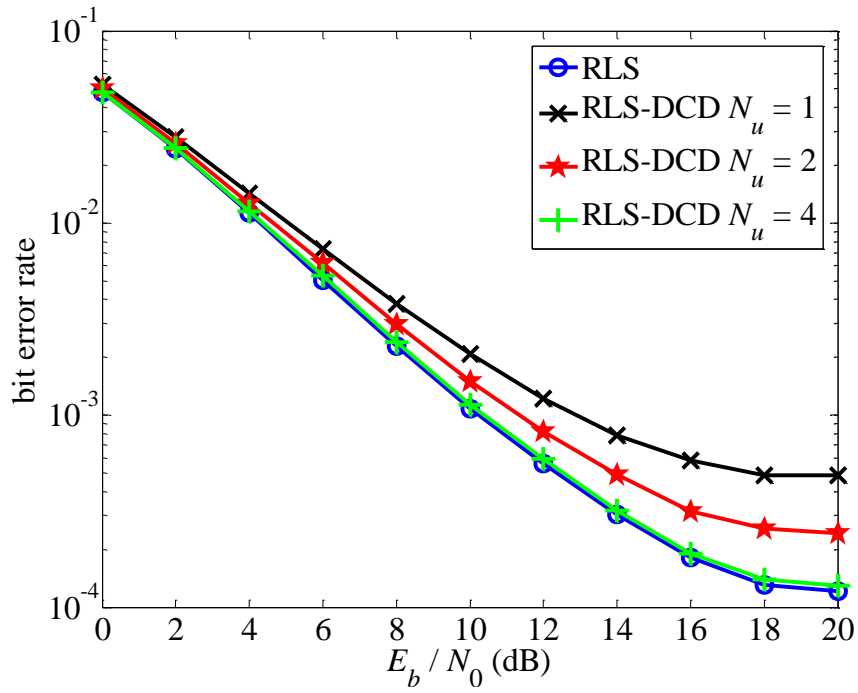


Fig. 7, Bit error rate performance of the new equalizer using conventional RLS and RLS-DCD algorithms for  $M = 4$ ,  $N = 4$ ,  $L_d = 175$ ,  $L_t = 35$ ,  $\lambda = 0.9922$ ,  $M_b = 12$ ,  $H = 2$ ,  $f_D T_s = 6 \times 10^{-4}$ .

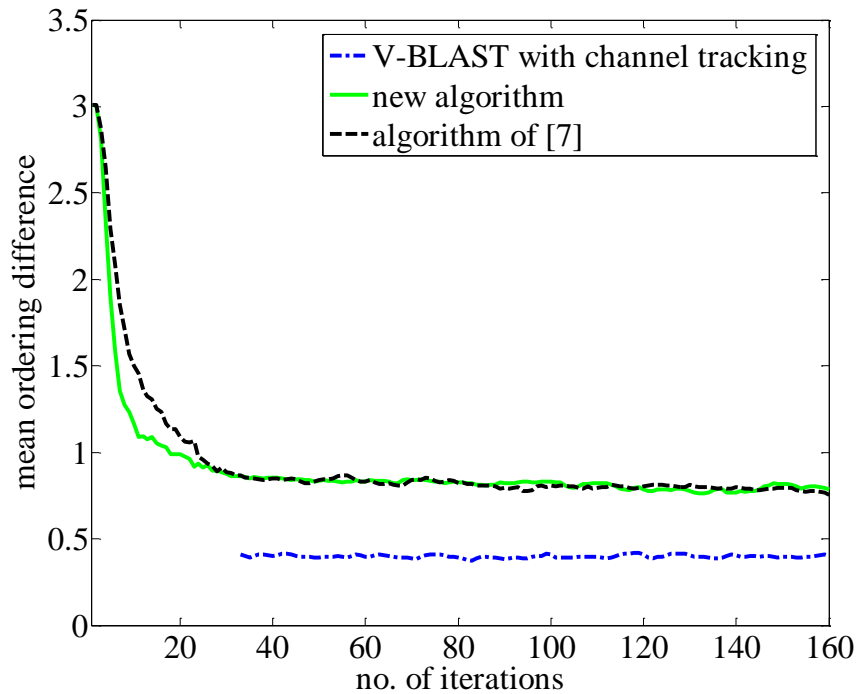


Fig. 8, Mean ordering error of different algorithms for  $M = 4$ ,  $N = 4$ ,  $L_d = 175$ ,  $L_t = 35$ ,  $\lambda = 0.9922$ ,  $M_b = 12$ ,  $N_u = 4$ ,  $H = 2$ ,  $f_D T_s = 6 \times 10^{-4}$ ,  $E_b/N_0 = 12$  dB.

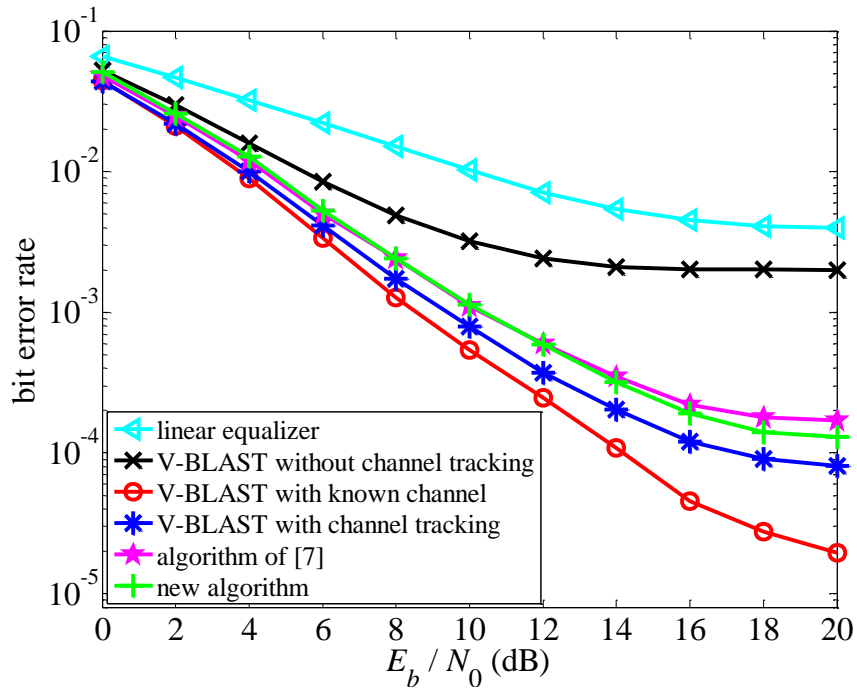


Fig. 9, Bit error rate performance of different algorithms for  $M = 4, N = 4, L_d = 175, L_t = 35, \lambda = 0.9922, M_b = 12, N_u = 4, H = 2, f_D T_s = 6 \times 10^{-4}$ .

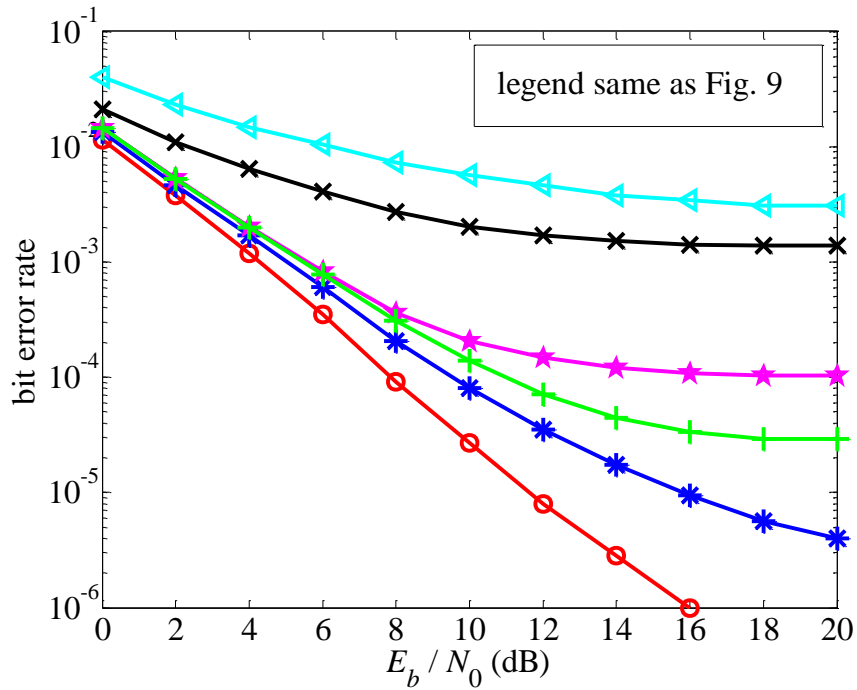


Fig. 10, Bit error rate performance of different algorithms for  $M = 8, N = 8, L_d = 260, L_t = 52, \lambda = 0.9922, M_b = 12, N_u = 4, H = 2, f_D T_s = 6 \times 10^{-4}$ .

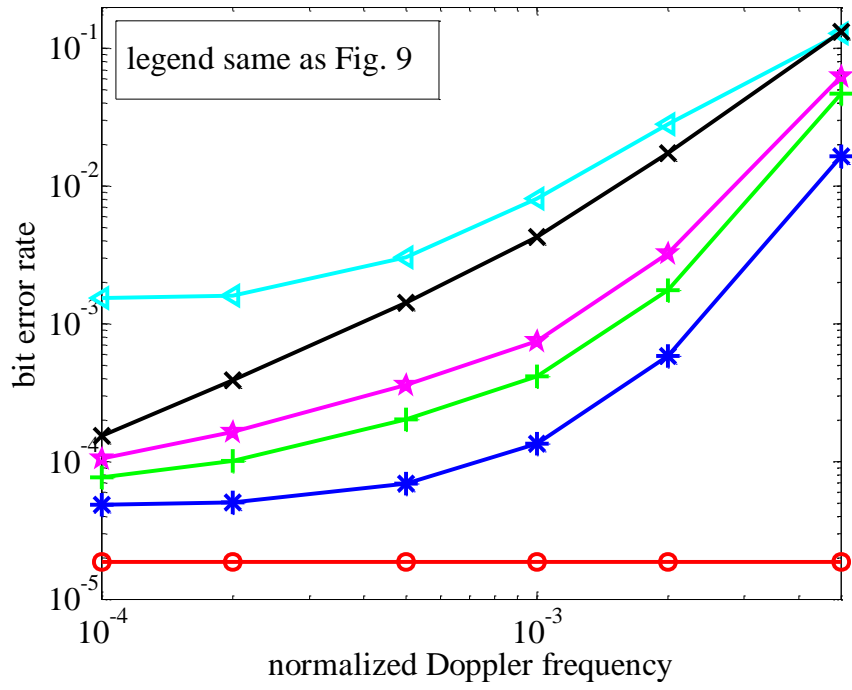


Fig. 11, Bit error rate performance of different algorithms for different normalized Doppler frequencies and  $M = 5$ ,  $N = 5$ ,  $L_d = 190$ ,  $L_t = 38$ ,  $\lambda = 0.9995$  to  $0.9688$ ,  $M_b = 12$ ,  $N_u = 4$ ,  $H = 2$ ,  $E_b/N_0 = 15$  dB.

Table 1, The proposed algorithm and required complex arithmetic operations by each step at each iteration when  $M = N$ .

Initialization		
$o(1) = \{1, 2, \dots, M\}$		
for $j = 1, 2, \dots, M$		
$\mathbf{w}_j(0) = \mathbf{0}$ (where $\mathbf{0}$ is the zero vector)		
$\boldsymbol{\rho}_j(0) = \mathbf{0}$		
$\boldsymbol{\Phi}_j(0) = \delta \mathbf{I}_{N+M-1}$ (where $\delta$ is a small positive number)		
$\hat{\mathcal{E}}_j(0) = 0$		required complex operations
At time instant $n = 1, 2, \dots, L_d$		
	$\times$	$+$
for $j = 1, 2, \dots, M$		
$\bar{d}_j(n) = \mathbf{f}_j^*(n) \mathbf{r}(n)$	$N^2$	$N^2 - N$
for $i = 1, 2, \dots, M$		
find $\mathbf{d}_{o_i}(n)$ using (3) and (4), then form $\mathbf{y}_{o_i}(n)$ as in (8)		
$\tilde{d}_{o_i}(n) = \bar{d}_{o_i}(n) + \mathbf{c}_{o_i}^*(n) \mathbf{d}_{o_i}(n)$	$N^2 - N$	$N^2 - N$
$d_{o_i}(n) = \begin{cases} \text{decision}(\tilde{d}_{o_i}(n)) & \text{if } n \leq L_t \\ s_{o_i}(n) & \text{if } n > L_t \end{cases}$		
$e_{o_i}(n) = d_{o_i}(n) - \tilde{d}_{o_i}(n)$	$0$	$N$
$\hat{\mathcal{E}}_{o_i}(n) = \lambda \hat{\mathcal{E}}_{o_i}(n-1) +  e_{o_i}(n) ^2$	$N$	$N$
$\boldsymbol{\beta}_{o_i}(n) = \lambda \boldsymbol{\rho}_{o_i}(n-1) - \mathbf{y}_{o_i}(n) e_{o_i}^*(n)$	$2N^2 - N$	$4N^2 - 2N$
update $\boldsymbol{\Phi}_{o_i}(n)$ using (13) if $i = 1$ and using (15) otherwise	$4N^2 - 4N + 1$	$8N^2 - \frac{19}{2}N + 3$
solve $\boldsymbol{\Phi}_{o_i}(n) \Delta \mathbf{w}_{o_i}(n) = \boldsymbol{\beta}_{o_i}(n)$ and obtain $\Delta \mathbf{w}_{o_i}(n)$ and $\boldsymbol{\rho}_{o_i}(n)$	$0$	$4N^2 N_u - \frac{3}{2} N N_u + \frac{1}{2} N M_b$
$\mathbf{w}_{o_i}(n) = \mathbf{w}_{o_i}(n-1) + \Delta \mathbf{w}_{o_i}(n)$	$0$	$2N^2 - N$
$o(n+1) = \text{sort}\{\hat{\mathcal{E}}_j(n), j = 1, 2, \dots, M\}$		

Table 2, The DCD algorithm with a leading element for a complex-valued linear system of equations [10].

---

Initialization

$$m = 1, \alpha = \frac{H}{2}$$

for  $j = 1, 2, \dots, M$

$$\Delta \mathbf{w}_j(n) = \mathbf{0}$$

$$\boldsymbol{\rho}_j(n) = \boldsymbol{\beta}_j(n)$$

For  $k = 1, 2, \dots, N_u$

$$[p, \zeta] = \arg \max_{l=1, \dots, N+M-1} \left\{ \left| \Re(\rho_j^l(n)) \right|, \left| \Im(\rho_j^l(n)) \right| \right\}$$

if  $\zeta = 1$ , then  $\rho_{tmp} = \Re(\rho_j^p(n))$ , else  $\rho_{tmp} = \Im(\rho_j^p(n))$

while  $|\rho_{tmp}| \leq \frac{\alpha}{2} \Phi_j^{p,p}(n)$  and  $m \leq M_b$

$$m = m + 1, \alpha = \frac{\alpha}{2}$$

if  $m > M_b$

algorithm stops

$$\Delta w_j^p(n) = \Delta w_j^p(n) + \text{sgn}(\rho_{tmp}) \zeta \alpha$$

$$\boldsymbol{\rho}_j(n) = \boldsymbol{\rho}_j(n) - \text{sgn}(\rho_{tmp}) \zeta \alpha \boldsymbol{\Phi}_j^{(p)}(n)$$


---

Table 3, Number of required complex and real arithmetic operations by different algorithms at each iteration when  $M = N$ .

	Complex		real		
	$\times$	$+$	$\langle \rangle$	$\sqrt{\quad}$	$/$
V-BLAST with channel tracking	$\frac{11}{3}N^3 + \frac{33}{4}N^2 + \frac{19}{12}N$	$3N^3 + 5N^2 - \frac{3}{2}N$	$\frac{1}{2}N^2 - \frac{1}{2}N$	–	–
Algorithm of [4]	$2N^3 + \frac{19}{2}N^2$	$\frac{4}{3}N^3 + 4N^2$	$\frac{1}{2}N^2 - \frac{1}{2}N$	–	$\frac{1}{2}N^2 - \frac{1}{2}N$
Algorithm of [7]	$\frac{2}{3}N^3 + \frac{25}{2}N^2$	$\frac{2}{3}N^3 + \frac{11}{2}N^2$	$\frac{1}{2}N^2 - \frac{1}{2}N$	$2N$	$3N$
New algorithm	$8N^2 - 5N$	$(16 + 4N_u)N^2 + \frac{1}{2}(M_b - 3N_u - 25)N$	$N \log_2 N$	–	–

---

# Paper E

Originally published as

R. Arablouei, K. Doğançay, and Sylvie Perreau, "MIMO-DFE coupled with V-BLAST for adaptive equalization of wideband MIMO channels," in *Proceedings of 19<sup>th</sup> European Signal Processing Conference*, Barcelona, Spain, Sep. 2011, pp. 639-643.

Copyright © 2011 EURASIP. Reprinted with permission.

## MIMO-DFE coupled with V-BLAST for adaptive equalization of wideband MIMO channels

Reza Arablouei, Kutluyil Doğançay, and Sylvie Perreau

### Abstract

A new adaptive equalizer is proposed by combining a MIMO-DFE and a V-BLAST detector for frequency-selective fading MIMO channels. The MIMO-DFE cancels ISI while the V-BLAST detector cancels ICI and detects the transmitted symbols. The wideband channel concatenated with the MIMO-DFE is considered as a virtual flat-fading channel. Estimation and tracking of this virtual channel, which is used by the V-BLAST detector, is realized by means of an adaptive filter. Simulation results show that the new equalizer outperforms a previously-proposed adaptive wideband MIMO channel equalizer while requiring a lower computational complexity.

### 1. Introduction

In recent decades, there has been a growing interest in multiple-input multiple-output (MIMO) communication systems. Recent information-theoretical findings corroborate that employing multiple transmit/receive antennas can increase capacity of the wireless communication systems dramatically, growing linearly with the minimum of the number of antennas used at the transmitter and the receiver [1]. To exploit this tremendous spectral efficiency, complicated receiver structures are required. Among them, vertical Bell Labs layered space-time (V-BLAST) architecture [2] is the most famous one, which has been designed to deal with flat-fading channels. It efficiently cancels inter-(sub)channel interference (ICI) to increase reliability of symbol detection. However, increased transmission rates require shorter symbol periods and thus

intersymbol interference (ISI) arises. Therefore, an equalizer for a wideband wireless communication system operating in a frequency-selective MIMO channel should be able to counter both ICI and ISI. One way to eliminate ISI is to employ MIMO orthogonal frequency-division multiplexing (OFDM) systems [3]. Using MIMO-OFDM, a frequency-selective fading channel is effectively converted to several flat-fading channels. Despite this advantage, OFDM has some drawbacks such as implementation difficulties due to high peak-to-average power ratio, identifiability of spectral nulls, and sensitivity to carrier synchronization [4]. Although, the industry has adopted MIMO-OFDM in many recent standards, the abovementioned complications make single-carrier MIMO communication systems still attractive for certain applications like uplink in 3G LTE where single-carrier frequency-division multiple-access (SC-FDMA) is used. SC-FDMA is used in view of the fact that its peak-to-average power ratio is small and more constant power enables high RF power amplifier efficiency in the mobile handsets, an important factor for battery-powered equipment [5].

Most MIMO channel equalization methods reported in the literature require perfect knowledge of the channel. They obtain this knowledge through a separate channel estimation process in the receiver. The process is performed using the information acquired during periodic training sessions in which a training sequence known to the receiver is transmitted [6], [7]. It is proven that in order to achieve the best spectral efficiency training sequences should optimally take up as much as half of the whole transmitted data [6]. Channel-estimation-based equalizers usually assume that the channel is static during transmission of a burst (packet, frame, or block). However, this assumption does not hold when long bursts of data are transmitted through time-varying channels. As an example, it is shown in [10] that for a typical slow-fading channel and an interval of 8000 symbols, channel taps may change significantly, i.e. 75% amplitude variation and  $7\pi/8$  phase rotation. Therefore, the use of adaptive receiver structures, which can adapt to channel variations without requiring excessively frequent channel estimations, is imperative. Adaptive equalizers do not need any explicit channel estimation and are inherently capable of tracking channel variations. In addition, they usually require less training and impose less computational complexity compared to other equalization schemes.

There are a limited number of papers in the literature on adaptive equalization of the frequency-selective fading MIMO channels. In [8], an adaptive MIMO-DFE based on the recursive least-squares (RLS) algorithm is proposed which equalizes the received signals by canceling ISI of all the layers. In [9], an adaptive equalizer for wideband fading MIMO channels is developed which in fact enhances the equalizer of [8] by canceling interference of the detected symbols of other layers, i.e. ICI, from ISI-canceled received signals. In [9], the ICI cancelation is carried out successively and in an ordered fashion resembling ordered-successive interference cancelation (OSIC) of V-BLAST [2].

In this paper, we propose a new equalizer that couples the MIMO-DFE of [8] with a V-BLAST detector to suppress both ISI and ICI. We show that the new equalizer is superior to the equalizer of [9] in terms of bit-error-rate (BER) performance whilst enjoying an appreciably reduced computational burden. Exploiting the low-complexity implementation of V-BLAST in [11] further contributes to mitigating the complexity.

Section 2 describes the signal and system model and section 3 explains the proposed equalizer. Section 4 analyzes complexity of the new equalizer and compares it with those of other equalizers. Section 5 provides simulation results and section 6 concludes the paper.

## 2. Signal and system model

Let us consider a MIMO communication system with  $M$  transmitters and  $N$  receivers. The system operates over a frequency-selective and time-varying wireless channel and can be described via a discrete-time complex baseband model as

$$\mathbf{r}(n) = \sum_{l=0}^{L-1} \mathbf{H}(n, l) \mathbf{s}(n-l) + \mathbf{u}(n) \quad (1)$$

where

$$\mathbf{s}(n) = \frac{1}{\sqrt{M}} [s_1(n), s_2(n), \dots, s_M(n)]^T$$

is the  $M \times 1$  vector of symbols simultaneously transmitted by  $M$  transmitting antennas,

$$\mathbf{r}(n) = [r_1(n), r_2(n), \dots, r_N(n)]^T$$

is the  $N \times 1$  vector of received signals,  $\mathbf{H}(n, l)$  is the  $N \times M$  channel impulse response coefficient matrix with lag  $l$  at time index  $n$ ,  $\mathbf{u}(n)$  represents  $N \times 1$  vector of additive white Gaussian noise and  $(\cdot)^T$  denotes matrix transposition. The channel is independent of the noise and its time dispersion,  $L$ , is assumed the same for all subchannels associated with all transmitter-receiver antenna pairs.

## 3. Algorithm description

The proposed MIMO channel equalizer is depicted in Fig. 1. It combines an adaptive MIMO-DFE with a V-BLAST detector. The MIMO-DFE suppresses intersymbol interference (ISI) and the V-BLAST detector counters inter-subchannel (inter-substream) interference (ICI).

The basic idea of the proposed equalizer was inspired by the work of [12] where it is shown that globally optimum joint ISI/ICI suppression can be achieved by carrying out ISI and ICI cancelation in separate concatenated stages. Nonetheless, the nature of the proposed equalizer is totally different from the equalizer of [12], which assumes perfect

knowledge of the frequency-selective channel,  $\mathbf{H}(n, l)_{l=0, \dots, L-1}$ , from the channel estimator in a prior stage. In contrast, the adaptive MIMO-DFE of the proposed equalizer does not require any explicit information about the channel impulse response. On the other hand, a V-BLAST detector normally requires an estimated channel matrix to cancel ICI from the received signals and to detect the transmitted symbols in an ordered-successive manner. Therefore, assuming acceptable ISI suppression by the MIMO-DFE, we consider the combination of the original frequency-selective channel and the MIMO-DFE as a *virtual flat-fading channel*,  $\mathcal{H}(n)$ . Hence, the system model of Fig. 1 can be simplified to the model of Fig. 2 with shortened channel. This virtual channel can be easily identified and tracked by an adaptive filter as shown in Fig. 3. The covariance matrix of the equivalent noise for the shortened system,  $\mathbf{u}(n)$ , can also be estimated recursively as we will show later. This noise is colored since the white noise of the original system,  $\mathbf{u}(n)$ , passes through the MIMO-DFE.

The algorithm of the proposed equalizer is described in the following subsections assuming the use of the RLS algorithm for updating the coefficients of the adaptive MIMO-DFE and identifying and tracking the virtual channel. Applying any other adaptive filtering algorithm in the proposed equalizer is straightforward.

### 3.1. Equalization and detection

Equalization of the received signals and detecting the transmitted symbols are performed in two steps:

– *ISI cancelation by the adaptive MIMO-DFE*

Intersymbol interference cancelation is carried out via

$$\hat{\mathbf{d}}(n - \delta) = \mathbf{W}^*(n)\mathbf{x}(n) \quad (2)$$

where

$$\hat{\mathbf{d}}(n - \delta) = [\hat{d}_1(n - \delta), \hat{d}_2(n - \delta), \dots, \hat{d}_M(n - \delta)]^T$$

is an  $M \times 1$  vector of ISI-suppressed outputs of the MIMO-DFE with a decision delay of  $\delta$ ,  $\mathbf{W}(n)$  is the  $K \times M$  matrix of MIMO-DFE filter coefficients and  $(\cdot)^*$  stands for complex-conjugate transposition. In addition,  $\mathbf{x}(n)$  is the input regressor vector of size  $K$  defined as

$$\mathbf{x}(n) = [\mathbf{r}_1^T(n), \mathbf{r}_2^T(n), \dots, \mathbf{r}_N^T(n), \mathbf{d}_1^T(n - \delta - 1), \mathbf{d}_2^T(n - \delta - 1), \dots, \mathbf{d}_M^T(n - \delta - 1)]^T$$

where

$$\mathbf{r}_i(n) = [r_i(n), r_i(n - 1), \dots, r_i(n - L_f + 1)]^T, \quad i = 1, \dots, N$$

and

$$\mathbf{d}_i(n - \delta - 1) = [d_i(n - \delta - 1), d_i(n - \delta - 2), \dots, d_i(n - \delta - L_b)]^T, \quad i = 1, \dots, M.$$

Here,  $d_i(n)$  is the detected symbol of the  $i$ th substream at time index  $n$  while  $L_f$  and  $L_b$  denote feedforward and feedback filter lengths for the MIMO-DFE, respectively. Consequently, the length of  $\mathbf{x}(n)$  is

$$K = N \times L_f + M \times L_b.$$

It is obvious that the past decisions,  $d_i(n - \delta - j)_{j=1, \dots, L_b}^{i=1, \dots, M}$ , are replaced by the known-for-the-receiver transmitted symbols,  $s_i(n - \delta - j)_{j=1, \dots, L_b}^{i=1, \dots, M}$ , during training.

– *ICI cancelation and symbol detection by the V-BLAST detector*

The ISI-canceled signals,  $\hat{\mathbf{d}}(n - \delta)$ , together with the estimations of the virtual flat-fading channel and the noise/interference covariance matrix from the previous iteration  $-\hat{\mathcal{H}}(n - \delta - 1)$  and  $\mathbf{Q}_u(n - 1)$  respectively (see subsection 3.2)– are fed to the V-BLAST detector to render the ICI cancelation and detect the transmitted symbols. The detected symbols are arranged in an  $M \times 1$  vector as

$$\mathbf{d}(n - \delta) = [d_1(n - \delta), d_2(n - \delta), \dots, d_M(n - \delta)]^T.$$

The V-BLAST detector is implemented based on minimum mean-square error (MMSE) criterion where the MMSE filter,  $\mathbf{G}(n)$ , uses  $\mathbf{Q}_u(n - 1)$  for regularization and is expressed as

$$\mathbf{G}(n) = \left( \hat{\mathcal{H}}^*(n - \delta - 1) \hat{\mathcal{H}}(n - \delta - 1) + \mathbf{Q}_u(n - 1) \right)^{-1} \hat{\mathcal{H}}^*(n - \delta - 1). \quad (3)$$

### 3.2. Identification and tracking of the virtual flat-fading channel

The virtual non-frequency-selective channel,  $\mathcal{H}(n - \delta)$ , can be identified and tracked using the RLS algorithm in the adaptive filtering scenario of Fig. 3 via the following set of equations

$$\mathbf{e}(n) = \hat{\mathbf{d}}(n - \delta) - \hat{\mathcal{H}}(n - \delta - 1) \mathbf{d}(n - \delta) \quad (4a)$$

$$\mathbf{q}(n) = \mathbf{P}(n - 1) \mathbf{d}(n - \delta) \quad (4b)$$

$$\mathbf{k}(n) = \frac{\mathbf{q}(n)}{\lambda + \mathbf{d}^*(n - \delta) \mathbf{q}(n)} \quad (4c)$$

$$\mathbf{P}(n) = \lambda^{-1} (\mathbf{P}(n - 1) - \mathbf{k}(n) \mathbf{q}^*(n)) \quad (4d)$$

$$\hat{\mathcal{H}}(n - \delta) = \hat{\mathcal{H}}(n - \delta - 1) + \mathbf{k}(n) \mathbf{e}^*(n) \quad (4e)$$

where  $\lambda$  is the forgetting factor and  $\hat{\mathcal{H}}(n - \delta)$  is the estimate of  $\mathcal{H}(n - \delta)$ . The channel identification is carried out with a delay equal to decision delay  $\delta$ . The noise covariance matrix of the shortened system can also be estimated recursively via

$$\mathbf{Q}_u(n) = \lambda \mathbf{Q}_u(n-1) + (1-\lambda) \mathbf{e}(n) \mathbf{e}^*(n) \quad (5)$$

where

$$\mathbf{e}(n) = \mathcal{H}(n-\delta) \mathbf{d}(n-\delta) - \hat{\mathbf{d}}(n-\delta) \quad (6)$$

Note that  $\mathbf{d}(n-\delta)$  is replaced by the vector of known transmitted symbols,  $\mathbf{s}(n-\delta)$ , during training.

### 3.3. Updating filter coefficients of the adaptive MIMO-DFE

Finally, filter tap weights of the adaptive MIMO-DFE,  $\mathbf{W}(n)$ , can be updated using the RLS algorithm:

$$\mathbf{q}(n) = \mathcal{P}(n-1) \mathbf{x}(n) \quad (7a)$$

$$\mathbf{k}(n) = \frac{\mathbf{q}(n)}{\lambda + \mathbf{x}^*(n) \mathbf{q}(n)} \quad (7b)$$

$$\mathcal{P}(n) = \lambda^{-1} (\mathcal{P}(n-1) - \mathbf{k}(n) \mathbf{q}^*(n)) \quad (7c)$$

$$\mathbf{W}(n) = \mathbf{W}(n-1) + \mathbf{e}^*(n) \otimes \mathbf{k}(n) \quad (7d)$$

where  $\otimes$  stands for Kronecker product.

It should be noted that in order to prevent the adaptive algorithms from stalling, the virtual channel identification process should be initiated with an adequate delay after initiation of the MIMO-DFE coefficient update process in the training.

## 4. Computational complexity

The proposed equalizer using the RLS algorithm requires  $3K^2 + 2KM + K$  multiplications,  $2K^2 + 2KM - M$  additions and  $K$  divisions at each iteration for equalization and coefficient update of the MIMO-DFE,  $\frac{7}{6}M^3 + \frac{7}{2}M^2 - \frac{5}{3}M + 1$  multiplications and  $\frac{7}{6}M^3 + \frac{3}{2}M^2 - \frac{13}{6}M$  additions for the V-BLAST detector (using the algorithm of [11]) and  $8M^2 + M$  multiplications,  $5M^2$  additions and  $M$  divisions for identifying and tracking the virtual channel and estimating the noise covariance matrix. Therefore, the new equalizer requires a total of  $3K^2 + 2KM + K + \frac{7}{6}M^3 + \frac{23}{2}M^2 - \frac{2}{3}M + 1$  multiplications,  $2K^2 + 2KM + \frac{7}{6}M^3 + \frac{13}{2}M^2 - \frac{19}{6}M$  additions and  $K + M$  divisions at each iteration when using the RLS algorithm. It also requires  $3KM + \frac{7}{6}M^3 + \frac{21}{2}M^2 - \frac{5}{3}M + 3$  multiplications,  $2KM + \frac{7}{6}M^3 + \frac{11}{2}M^2 - \frac{19}{6}M + 2$  additions and 2 divisions when using the NLMS algorithm. The computational complexity of the proposed algorithm and the algorithms of [8] and [9] are compared in Table 1. Assuming six floating-point operations (FLOPs) for each multiplication or division and two floating-point operations for each addition or subtraction [13], the total number of

required FLOPs for different algorithms is illustrated in Figs. 4 and 5 for different numbers of transmitter/receiver antennas and different temporal spans for feedforward and feedback filters of the MIMO-DFE. In both Figs. 4 and 5,  $M = N$  while in Fig. 4,  $L_f = L_b = 2$  and in Fig. 5,  $L_f = L_b = 6$ . It is seen that the longer the MIMO-DFE filters are, the more the computational saving by the new equalizer is.

## 5. Simulations

In order to verify the performance of the new equalizer, simulation results are provided here considering a  $4 \times 4$  MIMO wireless communication system. Subchannels between all transmitter and receiver antenna pairs are modelled independently according to the JTC indoor residential channel model A [14]. This channel model comprises three taps and each tap independently undergoes Rayleigh fading with a normalized Doppler frequency of  $T_s f_D = 1 \times 10^{-5}$  according to Jakes model [15]. The temporal spans of  $L_f = 6$  and  $L_b = 3$  are considered for the MIMO-DFE. Following the standard practice for MIMO-DFE design, a fixed decision delay of  $\delta = L_f - 1$  for all the substreams is chosen [16]. Transmitted signals are uncoded QPSK and a forgetting factor of  $\lambda = 0.99$  is used. The V-BLAST algorithm is implemented according to the MMSE criterion. The noise vectors for different time indexes are independent and identically distributed (i.i.d.) complex circular Gaussian random vectors with a zero mean vector. In Fig. 6, learning curves of the new equalizer using RLS algorithm and the equalizers of [8] and [9] are compared. The curves have been obtained by averaging over 1000 independent runs and over all substreams. Signal-to-noise ratio has been set to  $\text{SNR} = 14$  dB. First 100 transmitted symbol vectors are used for training; after that, algorithms switch to the decision-directed mode. Fig. 7 compares bit-error-rate performance of the new equalizer using the RLS and RLS-NLMS algorithms and the equalizers of [8] and [9]. The new equalizer with RLS-NLMS algorithm utilizes the RLS algorithm during training and switches to the NLMS algorithm in decision-directed mode.

It is evident from Figs. 6 and 7 that the new equalizer outperforms the previously-proposed adaptive MIMO equalizers. In addition, the new equalizer using RLS-NLMS algorithm outperforms the conventional MIMO-DFE using the RLS algorithm (equalizer of [8]) where utilizing the NLMS algorithm in decision-directed mode makes it less computationally demanding compared to the equalizer of [8].

## 6. Conclusion

A new equalizer for frequency-selective and time-varying MIMO channels was proposed by coupling an adaptive MIMO-DFE and a V-BLAST detector. The frequency-selective MIMO channel concatenated with the MIMO-DFE is considered as a virtual flat-fading channel and this virtual channel is identified and tracked adaptively. The new equalizer achieves superior performance with less computational requirements

compared to the existing adaptive wideband MIMO channel equalizers. Since the proposed equalizer can straightforwardly utilize any adaptive filtering algorithm, it has the potential to provide a sensible trade-off between performance and complexity. Furthermore, it can exploit the benefits of partial updating [17].

## References

- [1] H. Jafarkhani, *Space-Time Coding: Theory and Practice*, Cambridge, U.K.: Cambridge University Press, 2005.
- [2] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 11, pp. 1841–1852, Nov. 1999.
- [3] G. Stuber, J. Barry, S. Mclaughlin, Y. Li, M. Ingram, and T. Pratt, "Broadband MIMO-OFDM wireless communications," *Proc. IEEE*, vol. 92, pp. 271–294, Feb. 2004.
- [4] S. Haykin and M. Moher, *Modern Wireless Communications*, Englewood Cliffs, NJ: Prentice-Hall, 2003.
- [5] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS Long Term Evolution: From Theory to Practice*, Wiley, 2009.
- [6] T. L. Marzetta, "BLAST training: Estimating channel characteristics for high capacity space-time wireless," in *Proc. 37th Annual Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Sept. 1999.
- [7] S. J. Grant and J. K. Cavers, "Multiuser channel estimation for detection of cochannel signals," *IEEE Trans. Commun.*, vol. 49, pp. 1845–1855, Oct. 2001.
- [8] A. Maleki-Tehrani, B. Hassibi, and J. M. Cioffi, "Adaptive equalization of multiple-input multiple-output (MIMO) channels," in *Proc. IEEE Int. Conf. Communications*, New Orleans, LA, Jun. 2000, pp. 1670–1674.
- [9] A. A. Rontogiannis, V. Kekatos, and K. Berberidis, "An adaptive decision feedback equalizer for time-varying frequency selective MIMO channels," in *Proc. Signal Process. Advances Wireless Commun.*, Cannes, France, July 2006, pp. 1–5.
- [10] A. S. Lalos, V. Kekatos, and K. Berberidis, "Adaptive conjugate gradient DFEs for wideband MIMO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2406–2412, June 2009.
- [11] Y. Shang and X.-G. Xia, "On fast recursive algorithms for V-BLAST with optimal ordered SIC detection," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 2860–2865, June 2009.

- [12] J. He, G. Gu, and Z. Wu, "MMSE interference suppression in MIMO frequency selective and time-varying fading channels," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3638–3651, Aug. 2008.
- [13] J. Benesty, Y. Huang, and J. Chen, "A fast recursive algorithm for optimum sequential signal detection in a BLAST system," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1722–1730, Jul. 2003.
- [14] "Technical report on RF channel characterization and system deployment modeling," JTC (Air) Standards Contribution, Tech. Rep. JTC(AIR)/94.09.23-065R6, Sep. 1994.
- [15] W. C. Jakes, *Microwave Mobile Communications*, New York: Wiley, 1974.
- [16] N. Al-Dhahir and A. H. Sayed, "The finite length multi-input multi-output MMSE-DFE," *IEEE Trans. Signal Process.*, vol. 48, no. 10, pp.2921–2936, Oct. 2000.
- [17] K. Doğançay, *Partial-Update Adaptive Signal Processing: Design, Analysis and Implementation*, Academic Press, Oxford, UK, 2008.

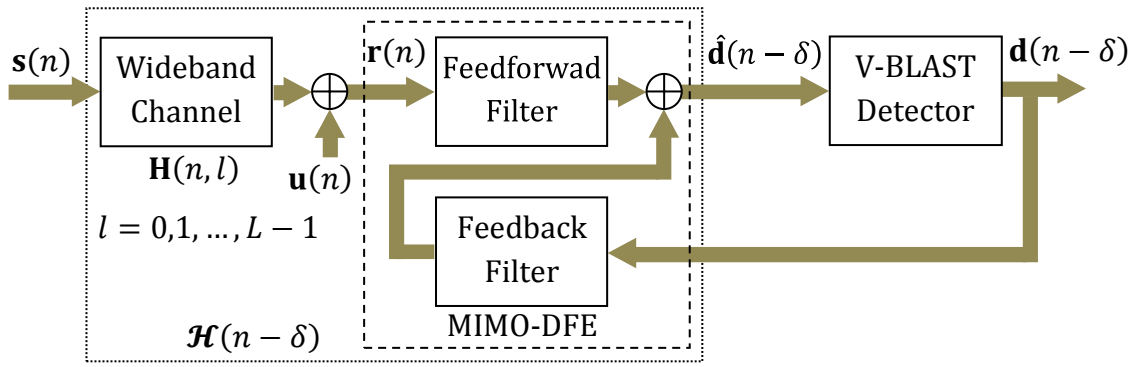


Fig. 1, System model and the proposed equalizer.

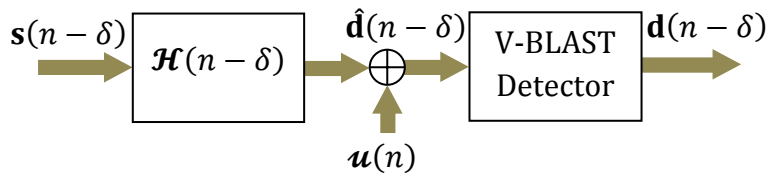


Fig. 2, Shortened system model.

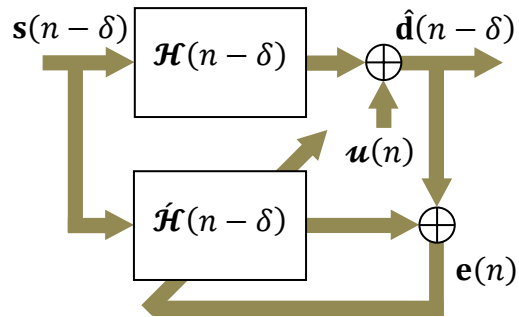


Fig. 3, Identifying and tracking  $\mathcal{H}(n - \delta)$  using an adaptive filter,  $\hat{\mathcal{H}}(n - \delta)$ , with a delay equal to decision delay  $\delta$ .

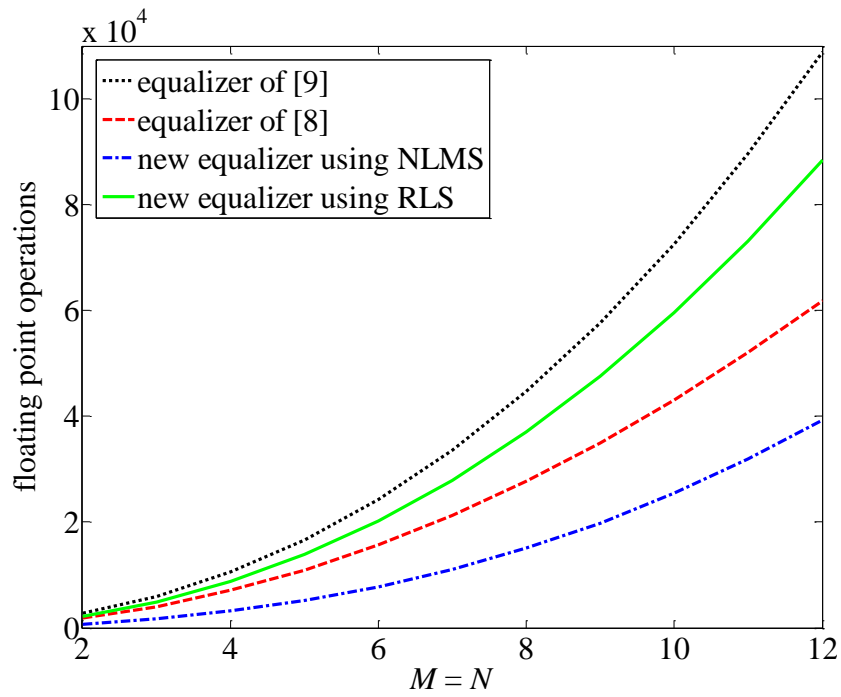


Fig. 4, Complexity comparison in terms of required floating point operations (FLOPS) for  $L_f = L_b = 2$ .

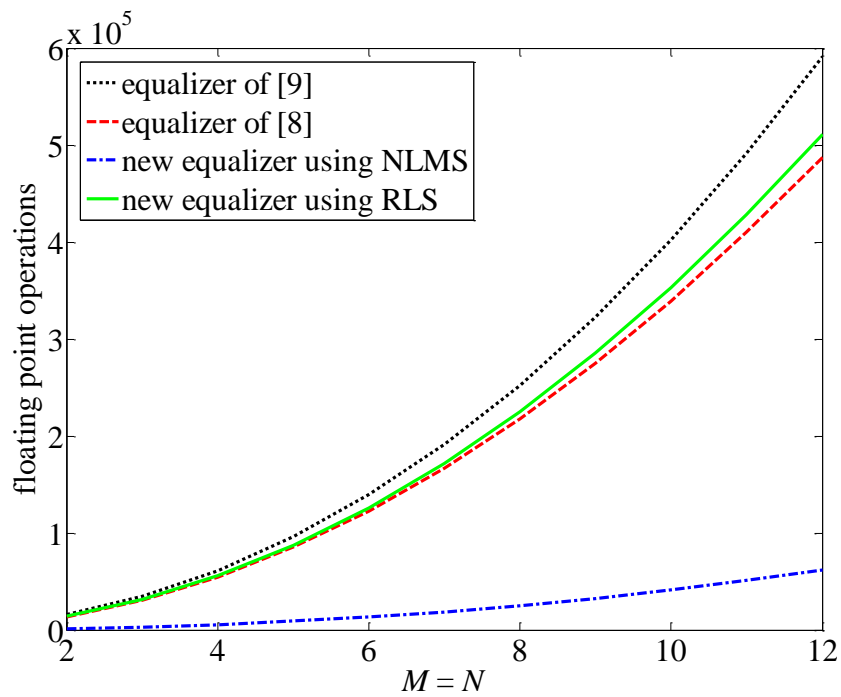


Fig. 5, Complexity comparison in terms of required floating point operations (FLOPS) for  $L_f = L_b = 6$ .

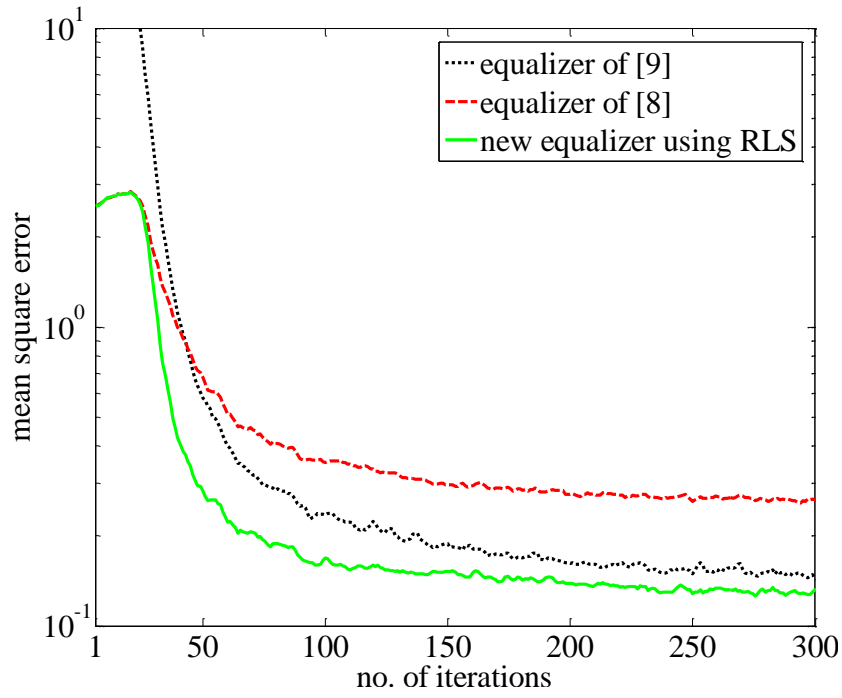


Fig. 6, Learning curves of different equalizers for a  $4 \times 4$  MIMO system with JTC indoor residential channel model A,  $T_s f_D = 1 \times 10^{-5}$ ,  $L_f = 6$ ,  $L_b = 3$ ,  $\delta = 5$ ,  $\lambda = 0.99$ , SNR = 14 dB, uncoded QPSK, and 100 symbol vectors used for training.

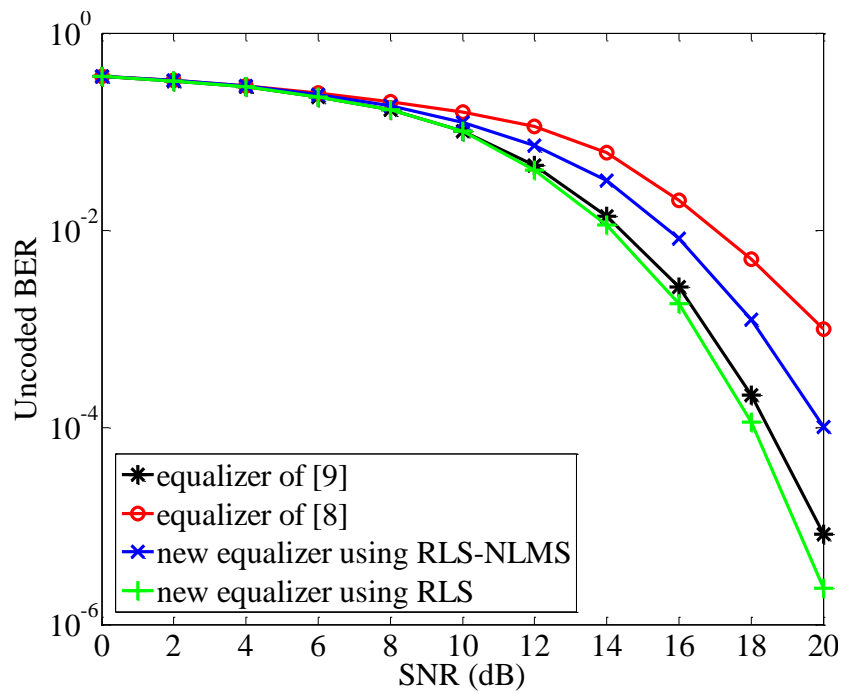


Fig. 7, BER performance of different equalizers for a  $4 \times 4$  MIMO system with JTC indoor residential channel model A,  $T_s f_D = 1 \times 10^{-5}$ ,  $L_f = 6$ ,  $L_b = 3$ ,  $\delta = 5$ ,  $\lambda = 0.99$ , uncoded QPSK, and 100 symbol vectors used for training.

Table 1, Required number of arithmetic operations for different equalizers at each iteration.

	$\times$	$+$	$/$	$\sqrt{\quad}$
Proposed equalizer using RLS	$3K^2 + 2KM + K + \frac{7}{6}M^3 + \frac{23}{2}M^2$	$2K^2 + 2KM + \frac{7}{6}M^3 + \frac{13}{2}M^2$	$K + M$	–
Proposed equalizer using NLMS	$3KM + \frac{7}{6}M^3 + \frac{21}{2}M^2$	$2KM + \frac{7}{6}M^3 + \frac{11}{2}M^2$	2	–
Equalizer of [9]	$3K^2 + \frac{1}{2}KM^2 + \frac{13}{2}KM + 5K + \frac{1}{6}M^3 + 3M^2$	$K^2 + \frac{1}{2}KM^2 + \frac{7}{2}KM + \frac{1}{6}M^3 + 2M^2$	$K + M$	$K + M$
Equalizer of [8]	$3K^2 + 2KM + 2K + M^2$	$2K^2 + 2KM + M^2$	$K$	–

# Paper F

Originally published as

R. Arablouei, K. Doğançay, and Sylvie Perreau, "Partial-update adaptive decision-feedback equalization," in *Proceedings of 19<sup>th</sup> European Signal Processing Conference*, Barcelona, Spain, Sep. 2011, pp. 2205-2209.

Copyright © 2011 EURASIP. Reprinted with permission.

## Partial-update adaptive decision-feedback equalization

Reza Arablouei, Kutluyil Doğançay, and Sylvie Perreau

### Abstract

In this paper, we consider partial updating of decision-feedback equalizers (DFEs). Application of data-dependent partial-update techniques to DFE is complicated by the fact that the feedback filter regressor has entries with discrete amplitudes, making magnitude-based sorting impossible. We present a novel selective-partial-update algorithm for DFEs that treats feedforward and feedback filters separately and utilizes past decision-directed errors for feedback filter coefficient selection. In addition, a new partial-update scheme is proposed that switches between selective partial updates and periodic partial updates to make effective use of training periods in telecommunication applications. Simulation results are presented to corroborate the effectiveness of the new partial-update DFEs in dealing with realistic frequency-selective and time-varying channels at reduced complexity.

### 1. Introduction

Ever-growing transmission rates of communication systems have led to an increased interest in decision-feedback equalizers (DFEs). The main advantage of DFE over linear equalizers is its capability to cancel inter-symbol interference (ISI) with reduced noise enhancement, resulting in considerably lower symbol error rate (SER) than that of a linear equalizer [1]. The underlying principle of DFE is to eliminate future ISI contributions of detected symbols [2].

In practice, the receiver is not initially aware of the dynamics of the channel and the channel can vary in time. A common practical way to enable tracking time variations of the channel by an equalizer is to implement it as an adaptive filter [3].

Adaptive DFE is well known for its relatively simple structure. However, in scenarios where the channel is highly dispersive or multiple antennas are utilized, DFE's complexity may become prohibitive. Fractionally-spaced equalization can make it even more expensive. Given that the equalizer/detector block is the computationally most demanding part of a communication system, it is important to have a reliable and low-complexity equalization/detection method with a sensible tradeoff between performance and complexity. Partial updating offers an attractive solution here. Although partial-update techniques have proven their capability in several linear adaptive filtering applications e.g. audio/network echo cancellation, their application to DFE has not yet been explored. As distinct from *linear* finite impulse response (FIR) adaptive filters, for which the existing partial-update techniques have been originally derived, DFE is a *nonlinear* adaptive filter. DFE comprises two FIR adaptive filters, viz. feedforward and feedback filters. These filters have input signals with different statistics and covariance matrices, implying that for partial updating, the feedforward and feedback filters must be treated separately rather than as one coefficient vector. Hence, some caveats should be taken into account when designing a partial-coefficient-update technique for DFE, especially for data-dependent partial updating which uses input regressor for coefficient selection.

The paper is organized as follows. In Section 2, we review the NLMS algorithm for DFE. In Section 3 we derive a generic partial-update NLMS-DFE using an instantaneous approximation of Newton's method. We develop a new selective-partial-update NLMS-DFE based on the principle of minimum disturbance in Section 4. We introduce a new combined selective-periodic partial-update NLMS-DFE in Section 5 with the objective of making effective use of limited training data. Computational complexity of all the algorithms considered is analysed in Section 6. We provide simulation results in Section 7 and draw conclusions in Section 8.

## 2. NLMS-DFE

The coefficients of DFE shown in Fig. 1 can be updated according to the NLMS algorithm [3] via

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{\epsilon + \|\mathbf{y}(n)\|^2} e^*(n) \mathbf{y}(n) \quad (1)$$

where

$$\mathbf{w}(n) = \begin{bmatrix} \mathbf{f}(n) \\ \mathbf{b}(n) \end{bmatrix},$$

$$\mathbf{y}(n) = \begin{bmatrix} \mathbf{x}(n) \\ \hat{\mathbf{d}}(n - \delta - 1) \end{bmatrix},$$

$$\tilde{\mathbf{d}}(n - \delta) = \mathbf{w}^*(n)\mathbf{y}(n),$$

and

$$e(n) = \hat{\mathbf{d}}(n - \delta) - \tilde{\mathbf{d}}(n - \delta).$$

Here,  $\mathbf{f}(n)$  and

$$\mathbf{x}(n) = [x(n), x(n - 1), \dots, x(n - L_f + 1)]^T$$

are  $L_f \times 1$  vectors denoting feedforward filter (FFF) coefficients and received signal regressor vector (FFF input) while  $\mathbf{b}(n)$  and

$$\hat{\mathbf{d}}(n - \delta - 1) = [\hat{d}(n - \delta - 1), \hat{d}(n - \delta - 2), \dots, \hat{d}(n - \delta - L_b)]^T$$

are  $L_b \times 1$  vectors representing feedback filter (FBF) coefficients and equalizer output regressor vector (FBF input), respectively. Moreover,  $\tilde{\mathbf{d}}(n - \delta)$  and  $\hat{\mathbf{d}}(n - \delta)$  are hard decision device input and output at time index  $n$ , whereas  $L_f$ ,  $L_b$ ,  $\mu$ ,  $\epsilon$ ,  $\delta$ ,  $\|\cdot\|^2$ ,  $(\cdot)^T$ , and  $(\cdot)^*$  stand for FFF and FBF temporal spans, adaptation step size, regularization parameter, decision delay, Euclidean norm, transposition, and complex-conjugate transposition, respectively. It should be noted that in the training mode,  $\hat{\mathbf{d}}(n - \delta)$  is replaced by the known-for-the-receiver transmitted symbol,  $s(n - \delta)$ .

### 3. Partial-update NLMS-DFE

To derive a generic partial-update NLMS algorithm for DFE, we apply an instantaneous approximation to Newton's method [3]. Newton's method is a fast-converging iterative estimation method and its regularized version is given by

$$\mathbf{w}(n + 1) = \mathbf{w}(n) + \mu(\epsilon\mathbf{I} + \mathbf{R})^{-1}(\mathbf{p} - \mathbf{R}\mathbf{w}(n)) \quad (2)$$

where

$$\mathbf{R} = E[\mathbf{y}(n)\mathbf{y}^*(n)]$$

is the input autocorrelation matrix,

$$\mathbf{p} = E[\mathbf{y}(n)\hat{\mathbf{d}}^*(n - \delta)]$$

is the cross-correlation vector between the input of the equalizer filters and the desired equalizer output and  $\mathbf{I}$  is an  $(L_f + L_b) \times (L_f + L_b)$  identity matrix. Partial updating of the adaptive filter coefficients replaces  $\mathbf{R}$  with

$$\mathbf{R}_M = E[\mathbf{I}_M(n)\mathbf{y}(n)\mathbf{y}^*(n)]$$

and  $\mathbf{p}$  with

$$\mathbf{p}_M = E[\mathbf{I}_M(n)\mathbf{y}(n)\hat{d}^*(n - \delta)]$$

where  $\mathbf{I}_M(n)$  is an  $(L_f + L_b) \times (L_f + L_b)$  diagonal matrix with zeros and/or ones as diagonal entries selecting  $M$  coefficients out of  $L_f + L_b$  for update at each iteration and  $E[\cdot]$  is the expectation operator. Consequently, partial-update version of (2) becomes

$$\mathbf{w}(n + 1) = \mathbf{w}(n) + \mu(\epsilon\mathbf{I} + \mathbf{R}_M)^{-1}(\mathbf{p}_M - \mathbf{R}_M\mathbf{w}(n)). \quad (3)$$

The instantaneous approximation of (3) is obtained by simply stripping the expectation operator off the correlation matrix and cross-correlation vector:

$$\mathbf{w}(n + 1) = \mathbf{w}(n) + \mu(\epsilon\mathbf{I} + \mathbf{I}_M(n)\mathbf{y}(n)\mathbf{y}^*(n))^{-1}e^*(n)\mathbf{I}_M(n)\mathbf{y}(n). \quad (4)$$

Applying the matrix inversion lemma [4], we get

$$\begin{aligned} & (\epsilon\mathbf{I} + \mathbf{I}_M(n)\mathbf{y}(n)\mathbf{y}^*(n))^{-1}\mathbf{I}_M(n)\mathbf{y}(n) \\ &= \frac{1}{\epsilon}\mathbf{I}_M(n)\mathbf{y}(n)\left(1 - \frac{\mathbf{y}^*(n)\mathbf{I}_M(n)\mathbf{y}(n)}{\epsilon + \mathbf{y}^*(n)\mathbf{I}_M(n)\mathbf{y}(n)}\right) \\ &= \frac{\mathbf{I}_M(n)\mathbf{y}(n)}{\epsilon + \mathbf{y}^*(n)\mathbf{I}_M(n)\mathbf{y}(n)}. \end{aligned} \quad (5)$$

Substituting (5) into (4) yields the coefficient update equation for a partial-update NLMS-DFE:

$$\mathbf{w}(n + 1) = \mathbf{w}(n) + \frac{\mu}{\epsilon + \|\mathbf{I}_M(n)\mathbf{y}(n)\|^2}e^*(n)\mathbf{I}_M(n)\mathbf{y}(n). \quad (6)$$

Several schemes have been introduced to control the computational complexity of adaptive filters by way of partial coefficient updates. Data-independent approaches are based on the primary idea of selecting coefficient subsets for update in a round-robin fashion (sequential partial updates [5]) or updating all the coefficients in a period greater than basic time index (periodic partial updates [6]). These partial-update adaptive filters may become unstable when the input signal is cyclostationary or periodic [7]. The stochastic partial updates method [8] offers a solution to such stability problems, where subsets of coefficients are randomly selected for update at each iteration.

For sequential or stochastic partial updates,  $M$  diagonal elements of  $\mathbf{I}_M(n)$  are set to 1 and the rest are set to 0 at each iteration in a sequential or stochastic fashion. For periodic partial updates,  $\mathbf{I}_M(n)$  is set to identity matrix once at each update period and is set to a zero matrix at other time instances.

Data-dependent partial-update methods use the regressor vector entries to select equalizer coefficients to be updated. Since the regressor vectors  $\mathbf{x}(n)$  and  $\hat{\mathbf{d}}(n - \delta - 1)$

do not have the same statistics, this may result in  $\mathbf{I}_M(n)$  favouring the FFF or FBF with larger input signal variance, leading to undesirable performance degradation as a result of both filters not having equal probability of receiving updates. In order to give a fair chance to all the coefficients (in FFF and FBF) for being updated, we define

$$M = M_f + M_b$$

where  $M_f$  and  $M_b$  are the number of coefficients to be updated at each time instant from FFF and FBF, respectively. Accordingly, the matrix  $\mathbf{I}_M(n)$  is redefined as

$$\mathbf{I}_M(n) = \begin{bmatrix} \mathbf{I}_{M_f}(n) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{M_b}(n) \end{bmatrix} \quad (7)$$

where  $\mathbf{I}_{M_f}(n)$  and  $\mathbf{I}_{M_b}(n)$  are  $L_f \times L_f$  and  $L_b \times L_b$  diagonal coefficient selection matrices to select  $M_f$  out of  $L_f$  and  $M_b$  out of  $L_b$  coefficients from FFF and FBF, respectively.

#### 4. Selective-partial-update NLMS-DFE

Data-independent approaches reduce the convergence rate, often proportional to the size of coefficient subsets in sequential or stochastic partial updates and the update frequency in periodic partial updates. On the other hand, data-dependent partial-update techniques offer better convergence performance, where  $M$ -max updates [9] and selective partial updates [10] are the most prominent ones.

In selective-partial-update NLMS algorithm [10], the coefficients to be updated at each time instant are determined according to the principle of minimum disturbance [3] by solving

$$\min_{\mathbf{I}_M(n)} \|\mathbf{w}(n+1) - \mathbf{w}(n)\|^2 \quad (8)$$

which, using (6), can be rewritten as

$$\min_{\mathbf{I}_M(n)} \left\| \frac{\mu}{\epsilon + \|\mathbf{I}_M(n)\mathbf{y}(n)\|^2} e^*(n)\mathbf{I}_M(n)\mathbf{y}(n) \right\|^2. \quad (9)$$

Excluding the trivial solution of zero inputs, (9) is simplified to

$$\max_{\mathbf{I}_M(n)} \|\mathbf{I}_M(n)\mathbf{y}(n)\|^2 \quad (10)$$

which means coefficients corresponding to the elements of the input regressor vector with the largest magnitudes should be selected for update at each time index [7], [10]. Using this criterion, the coefficient selection matrices become

$$\mathbf{I}_{M_f}(n) = \begin{bmatrix} i_1(n) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & i_{L_f}(n) \end{bmatrix} \quad (11)$$

$$i_k(n) = \begin{cases} 1 & \text{if } |x(n-k+1)| \in \{M_f \text{ maxima of } |x(n-l+1)|_{1 \leq l \leq L_f}\} \\ 0 & \text{otherwise} \end{cases}$$

and

$$\mathbf{I}_{M_b}(n) = \begin{bmatrix} j_1(n) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & j_{L_b}(n) \end{bmatrix} \quad (12)$$

$$j_k(n) = \begin{cases} 1 & \text{if } |\hat{d}(n-\delta-k)| \in \{M_b \text{ maxima of } |\hat{d}(n-\delta-l)|_{1 \leq l \leq L_b}\} \\ 0 & \text{otherwise} \end{cases}$$

However, the selection criterion of (12) suggests ranking the elements of  $\hat{\mathbf{d}}(n-\delta-1)$  according to their magnitudes. Clearly, the magnitudes of these elements, which are hard-decided symbols drawn from a finite-size constellation set, do not carry any information about the significance of updating their corresponding taps in the FBF. Specifically, if the employed modulation is M-ary PSK (e.g., BPSK or QPSK), all decided symbols are of the same magnitude. If one has to make a choice as to which  $M_b$  out of  $L_b$  coefficients to be updated in the FBF, the significance of the individual coefficients needs to be made explicit. The hard decisions  $\hat{d}(n-\delta-i)_{i=1,\dots,L_b}$  do not have this information. However, for each  $\hat{d}(n-\delta-i)_{1 \leq i \leq L_b}$  the equalizer output error  $e(n-i)_{1 \leq i \leq L_b}$  is readily available. The selective-partial-update method requires selection of coefficients with the largest contribution to error reduction, which is also shared by the  $M$ -max technique for LMS and NLMS [9]. This amounts to identifying coefficients with the largest regressor entries. Therefore for the FBF, the identification of coefficients must be based on the equalization errors that the FBF tap inputs have produced. In other words, past decisions with higher detection errors have larger share of the residual ISI of the FBF output. Therefore, the FBF taps corresponding to these symbols are more in need of correction and updating them provides faster convergence for partial updating. This approach leads to the following coefficient selection criterion for the FBF, which guarantees that the best  $M_b$  coefficients are updated in the selective-partial-update sense:

$$\max_{\mathbf{I}_{M_b}(n)} \|\mathbf{I}_{M_b}(n)\mathbf{e}(n-1)\|^2 \quad (13)$$

where

$$\mathbf{e}(n-1) = [e(n-1), e(n-2), \dots, e(n-L_b)]^T.$$

Hence, the selective-partial-update technique is customized for decision-feedback equalization by replacing the selection criterion of (12) with

$$j_k(n) = \begin{cases} 1 & \text{if } |e(n-k)| \in \{M_b \text{ maxima of } |e(n-l)|_{1 \leq l \leq L_b}\} \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

## 5. Combined selective-periodic partial-update NLMS-DFE

Periodic partial updating [7] is basically a downsampled version of the full-update filter exhibiting the same convergence as full-update, only slowed down by the update period. Thus, it should achieve the same steady-state mean-square error (MSE) as the full-update filter if the same step-size is used in both. In addition, it is the cheapest partial-update technique since, unlike the other techniques, it does not require calculation of the adaptation error and normalization of the step size at each iteration and performs these operations only once in each update period. On the other hand, the method of selective partial updates is the fastest converging partial-update technique. In order to exploit advantages of both these techniques, we propose a new combined partial-update NLMS-DFE, which employs selective partial updates of Section 4 during the training mode to ensure fast convergence to the steady-state MSE and switches to periodic partial updating at the end of training using the decision-directed mode. We call this new scheme *selective-periodic partial updates* (SPPU). It imposes less computational complexity in average compared to SPU and makes best use of available training data in partial-update context.

## 6. Complexity comparison

The number of required arithmetic operations and comparisons at each iteration for equalizers employing full-update and different partial-update NLMS algorithms is presented in Table 1 where  $L = L_f + L_b$  and  $p$  stands for update period in periodic partial-update and SPPU algorithms. Assuming six floating-point operations (FLOPs) for each multiplication or division and two floating-point operations for each addition, subtraction or comparison [11], the total number of required FLOPs at each iteration by different algorithms and percentage of their saved FLOPs compared to the full-update algorithm are also provided in Table 2 when  $L_f = L_b$ ,  $L = 30$ ,  $p = 2, 4$  and  $M = L/p$ . In both tables for SPPU-NLMS-DFE, training takes up one-fifth of the transmission rate.

## 7. Simulations

In order to compare performance of the proposed equalizers with the full-update and data-independent partial-update NLMS-DFEs, simulation results are provided in this section. Two different channels were utilized for this purpose, a static channel with a fractionally-spaced impulse response and a time-varying channel with a baud-spaced impulse response. In all the simulations, 5000 QPSK symbols were transmitted in each run while the first 1000 symbols were used for training. Results were obtained by

averaging over 1000 independent Monte Carlo runs. The adaptation step sizes were adjusted such that all the algorithms attained the same steady-state MSE.

### 7.1. Static channel

First part of the simulations was carried out using a terrestrial microwave channel impulse response (SPIB channel no. 3 from [12]), which is fractionally spaced ( $T/2$ ) and comprises 300 taps. The corresponding equalizers have  $L_f = 300$  (fractionally-spaced) and  $L_b = 60$  coefficients and a decision delay of  $\delta = 25$ . Except in Fig. 4, partial-update equalizers update half of the coefficients at each time instant ( $M_f = 150$  and  $M_b = 30$ ). For periodic partial updates, it means half in average ( $p = 2$ ).

Fig. 2 shows MSE curves of the full-update and different partial-update NLMS-DFEs at channel output signal to noise ratio (SNR) of 12 dB. It is clear that SPU and SPPU algorithms yield faster initial convergence than the other partial-update techniques. Fig. 3 depicts SER of different algorithms for different SNRs. It is observed that SPU and SPPU algorithms result in SER performances very close to the one of the full-update algorithm while they only update half of the filter coefficients at each iteration. SER-SNR curves of the proposed partial-update equalizers and the full-update NLMS-DFE are compared in Fig. 4 when different fractions of the coefficients are updated at each iteration. It clearly shows that how performance can be traded for complexity by using the new partial-update equalizers.

### 7.2. Time-varying channel

In this part, the 3GPP typical urban channel model [13] was considered. This channel comprises 20 taps and varies in time according to Jakes model [14] with a normalized Doppler frequency of  $12.5 \times 10^{-6}$ . The equalizers have  $L_f = 40$  and  $L_b = 20$  coefficients and a decision delay of  $\delta = 40$ . For partial updating, half of the coefficients are updated at each iteration ( $M_f = 20$ ,  $M_b = 10$ , and  $p = 2$ ).

Fig. 5 compares MSE curves of the full-update and different partial-update equalizers at SNR of 15 dB. SER performance of the equalizers for different SNRs is compared in Fig. 6. It is seen in Figs. 5 and 6 that the proposed equalizers achieve superior performance over other partial-update equalizers.

## 8. Conclusion

Employing partial coefficient updates in DFE was examined and a new selective-partial-update algorithm for NLMS-DFE was developed. The developed algorithm treats the feedforward and feedback adaptive filters separately and resolves the sorting ambiguity for the feedback filter by taking into account the errors associated with the past decisions. A new combined partial-update algorithm for NLMS-DFE was also proposed to take advantage of both selective-partial-update and periodic-partial-update

techniques. Computer simulations demonstrate the capability of the proposed equalizers to achieve a good tradeoff between performance and complexity by means of partial updating. It is shown that in practical equalization scenarios assuming both static and dynamic channels, the proposed partial-update techniques offer appreciable complexity savings at the expense of slight performance degradation.

## References

- [1] R. A. Casas, P. B. Schniter, J. Balakrishnan, C. R. Johnson, Jr., and Cornell Univ. BERG, *DFE Tutorial*, July 1998.
- [2] C. A. Belfiore and J. H. Park, "Decision feedback equalization," *Proc. IEEE*, vol. 67, no. 8, pp. 1143–1156, Aug. 1979.
- [3] A. H. Sayed, *Adaptive Filters*, Hoboken, NJ: Wiley, 2008.
- [4] G. H. Golub and C. F. Van Loan, *Matrix Computations*, third ed., Baltimore, MD: Johns Hopkins Univ. Press, 1996.
- [5] S. C. Douglas, "Adaptive filters employing partial updates," *IEEE Trans. on Circuits and Systems II*, vol. 44, no. 3, pp. 209–216, Mar. 1997.
- [6] J. R. Treichler, C. R. Johnson, Jr., and M. G. Larimore, *Theory and Design of Adaptive Filters*, Wiley, NY, 1987.
- [7] K. Doğançay, *Partial-Update Adaptive Signal Processing: Design, Analysis and Implementation*, Academic Press, Oxford, UK, 2008.
- [8] M. Godavarti, and A. O. Hero III, "Partial update LMS algorithms," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2382–2399, July 2005.
- [9] T. Aboulnasr, and K. Mayyas, "Complexity reduction of the NLMS algorithm via selective coefficient update," *IEEE Trans. Signal Process.*, vol. 47, no. 5, pp. 1421–1424, May 1999.
- [10] K. Doğançay and O. Tanrıku, "Adaptive filtering algorithms with selective partial updates," *IEEE Trans. Circuits Syst. II*, vol. 48, no. 8, pp. 762–769, Aug. 2001.
- [11] J. Benesty, Y. Huang, and J. Chen, "A fast recursive algorithm for optimum sequential signal detection in a BLAST system," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1722–1730, Jul. 2003.
- [12] Rice University Signal Processing Information Base (SPIB), <http://spib.rice.edu/spib/microwave.html>.

[13] 3GPP TR 25.943 V6.0.0, 3rd Generation Partnership Project; Technical Specification Group Radio Access Networks; Deployment aspects (Release 6), Dec. 2004.

[14] W. C. Jakes, *Microwave Mobile Communications*, New York: Wiley, 1974.

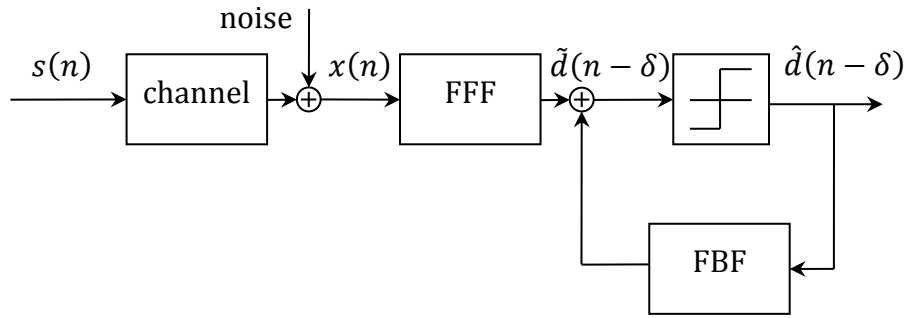


Fig. 1, Block diagram of a decision-feedback equalizer.

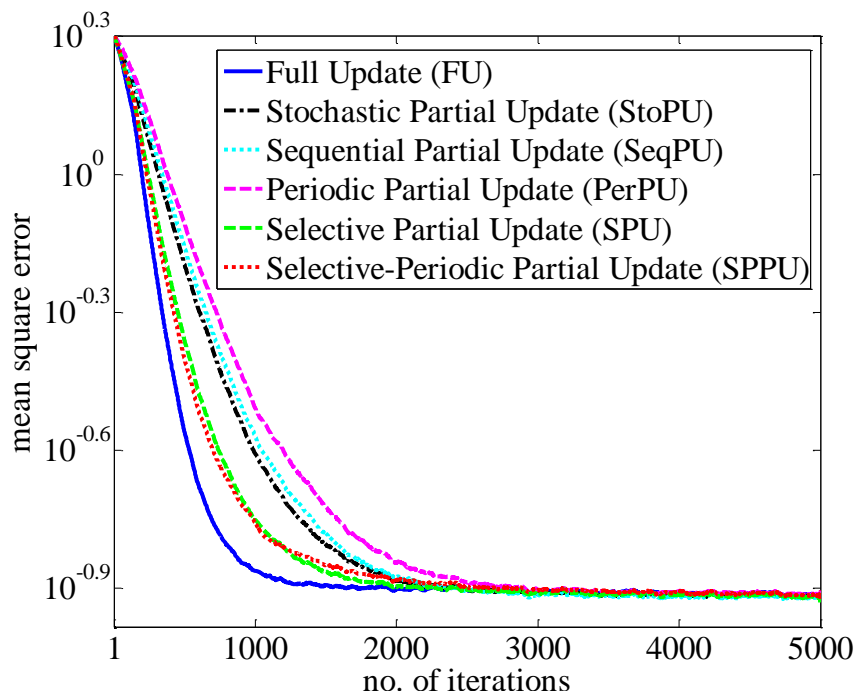


Fig. 2, MSE curves of full-update and different partial-update NLMS-DFEs for a static channel.

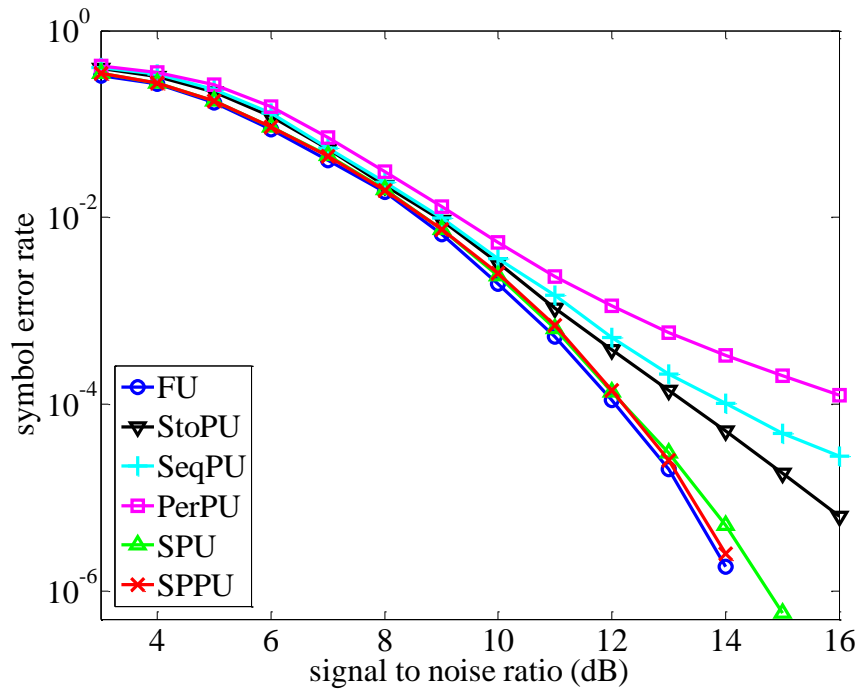


Fig. 3, SER-SNR curves of full-update and different partial-update NLMS-DFEs for a static channel.

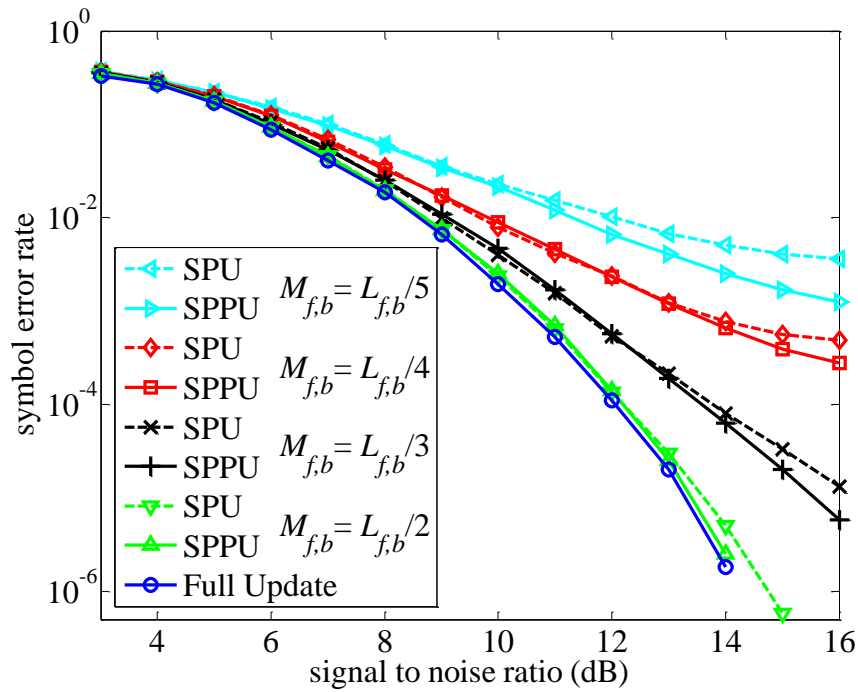


Fig. 4, SER-SNR curves of the new algorithms with different numbers of updated coefficients compared to the full-update NLMS-DFE.

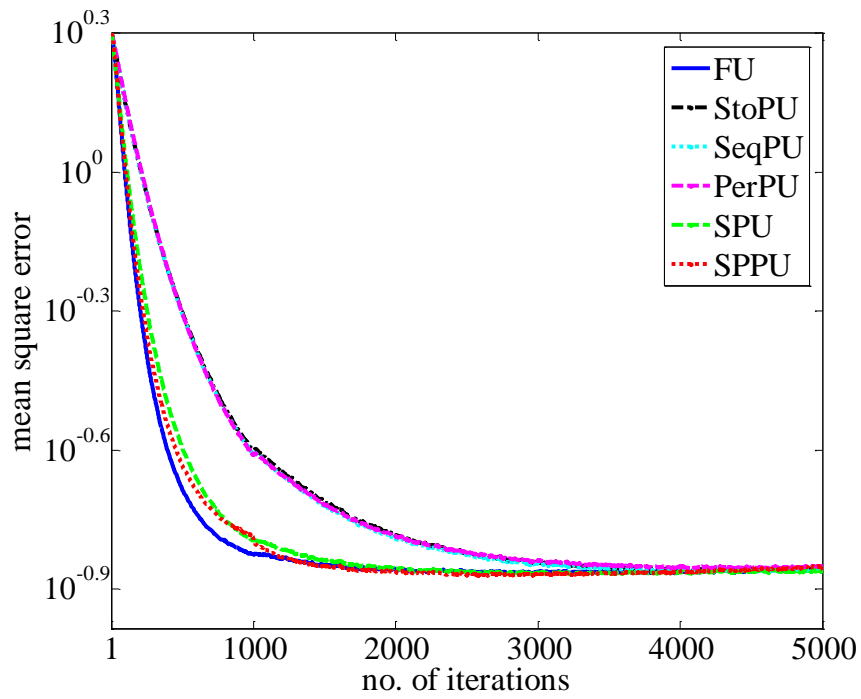


Fig. 5, MSE curves of full-update and different partial-update NLMS-DFEs for a time-varying channel.

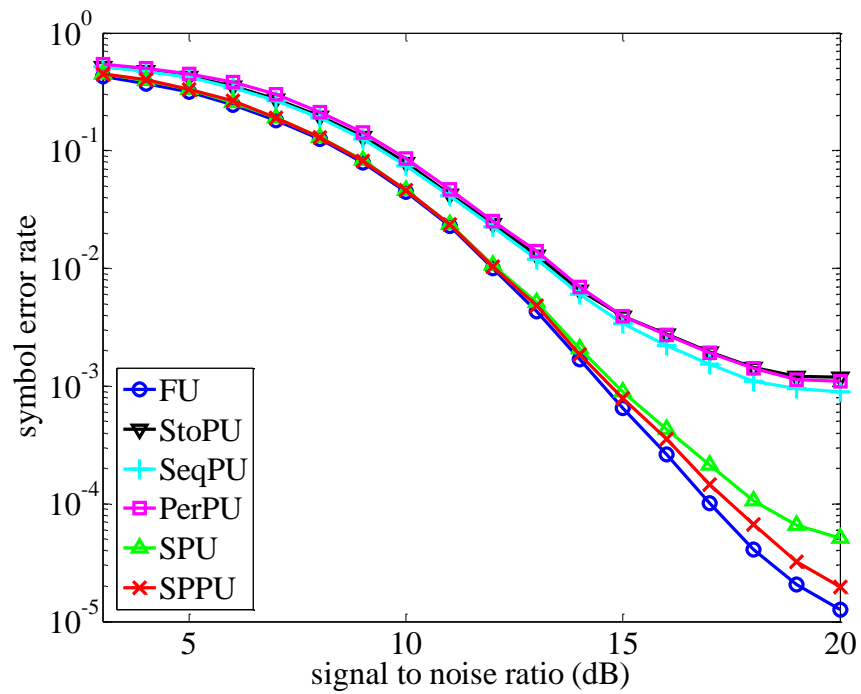


Fig. 6, SER-SNR curves of full-update and different partial-update NLMS-DFEs for a time-varying channel.

Table 1, Required number of arithmetic operations and comparisons at each iteration by different equalizers.

	$\times$	$+$	$/$	$\langle \rangle$
Full-update	$2L + 5$	$2L + 4$	1	–
Seq(Sto)PU	$L + M + 5$	$L + M + 4$	1	–
PerPU	$\left(1 + \frac{1}{p}\right)L + 4 + \frac{1}{p}$	$\left(1 + \frac{1}{p}\right)L + 3 + \frac{1}{p}$	$1/p$	–
SPU	$L + M + 5$	$L + M + 4$	1	$2(\log_2 L_f + \log_2 L_b) + 4$
SPPU	$\left(1 + \frac{4}{5p}\right)L + \frac{M}{5} + \frac{21}{5} + \frac{4}{5p}$	$\left(1 + \frac{4}{5p}\right)L + \frac{M}{5} + \frac{16}{5} + \frac{4}{5p}$	$\frac{1}{5} + \frac{4}{5p}$	$\frac{2}{5}(\log_2 L_f + \log_2 L_b) + \frac{4}{5}$

Table 2, Number of required FLOPs at each iteration and provided savings by different equalizers when  $L_f = L_b, L = 30, p = 2, 4$  and  $M = L/p$ .

	FLOPs		Saving (%)	
	$p = 2$	$p = 4$	$p = 2$	$p = 4$
Full-update NLMS-DFE	524	524	–	–
SeqPU or StoPU NLMS-DFE	404	344	22.9	34.4
PerPU NLMS-DFE	397	333	24.2	36.4
SPU-NLMS-DFE	443	383	15.4	26.9
SPPU-NLMS-DFE	406	343	22.5	34.5

# Paper G

Originally published as

R. Arablouei and K. Doğançay, “Modified quasi-OBE algorithm with improved numerical properties,” *Signal Processing*, vol. 93, pp. 797-803, 2013.

Copyright © 2013 Elsevier. Reprinted with permission.

## Modified quasi-OBE algorithm with improved numerical properties

Reza Arablouei and Kutluyıl Doğançay

### Abstract

The quasi-OBE (QOBE) algorithm is a set-membership adaptive filtering algorithm based on the principles of optimal bounding ellipsoid (OBE) processing. This algorithm can provide enhanced convergence and tracking performance as well as reduced average computational complexity in comparison with the more traditional adaptive filtering algorithms such as the recursive least-squares (RLS) algorithm. In this paper, we show that the QOBE algorithm is prone to numerical instability due to the unbounded growth/decay of its internal variables. To tackle this problem, we develop a new set-membership adaptive filtering algorithm by transforming QOBE’s internal variables into a new set of internal variables. The new algorithm, called modified quasi-OBE (MQOBE), can be viewed as an exponentially-weighted RLS algorithm with a time-varying forgetting factor, which is optimized at each iteration by imposing a bounded-magnitude constraint on the *a posteriori* filter output error. The proposed algorithm delivers the same convergence and tracking performance as the QOBE algorithm but with enhanced numerical properties. We demonstrate the improved numerical behavior of the proposed algorithm by simulation examples for a MIMO channel estimation problem.

Keywords: set-membership adaptive filtering, the quasi-OBE (BEACON) algorithm, numerical stability

### 1. Introduction

Set-membership (SM) filtering algorithms are set-theoretic estimation methods that, unlike the more traditional methods, e.g., minimum-mean-squared-error (MMSE) or

least-squares (LS) filters, estimate sets of feasible solutions rather than single-point solutions. The SM approaches are of particular interest in signal processing applications because they feature two major advantages over their more traditional counterparts. First, they exhibit superior adaptation and tracking properties. Second, they can effectively make use of innovation in the data and improve computational efficiency by establishing a data-discerning update strategy for the parameter estimates. More specifically, unlike the more traditional estimation schemes that implement a continuous update process regardless of the usefulness of the data, the SM algorithms assess the potential of the new data to improve the quality of the estimation and weigh the data accordingly. This intelligent update strategy results in discarding the data with unhelpful information content and obviating the expense of updating when the data is redundant. A more detailed background on the SM filtering paradigm can be found in [1-15] and the references therein.

An SM filtering algorithm is typically formulated as a set estimation problem and seeks solutions for a case that a certain constraining assumption is made about the filter output error. A usual assumption is a bounded magnitude for the filter output error. Several techniques have been proposed to estimate the target set of solutions, called membership set, under the bounded-error constraint. The most prominent ones are the optimal bounding ellipsoid (OBE) algorithms that approximate the membership set by tightly outer-bounding it with ellipsoids in the parameter space and optimize the size of the ellipsoids in some meaningful sense. Different optimality criteria have led to different OBE algorithms. The first OBE algorithm was introduced in [1]. A thorough review of numerous further works developing the other members of the OBE family can be found in [9, 14].

Among all the OBE algorithms, the quasi-OBE (QOBE) algorithm<sup>3</sup> [7, 14] is particularly attractive since it shares many of the desired features of the various OBE algorithms. Furthermore, it incorporates simple but efficient innovation check and optimal weight calculation processes, which make it computationally more efficient than other OBE algorithms.

In this paper, we show that the internal variables of the QOBE algorithm grow or decay unboundedly as the iterations progress. From a practical standpoint, this drawback limits the applicability of the algorithm to scenarios where the adaptation is performed for a sufficiently small number of iterations or the probability of update is low enough to prevent overflow or underflow of the internal variables. To overcome this problem, we introduce a change of the internal variables in the QOBE algorithm and develop a new SM adaptive filtering algorithm. The new algorithm can be seen as an exponentially-weighted recursive least-squares (EWRLS) algorithm

---

<sup>3</sup> This algorithm was originally published as the *Bounding Ellipsoidal Adaptive CONstrained least-squares* (BEACON) algorithm [7].

with a time-varying forgetting factor that is optimized within the framework of the SM filtering. In this sense, it differs from most of the OBE algorithms, which can be viewed as weighted recursive least-squares (WRLS) algorithms with time-varying weights. A discussion on the differences between the WRLS and EWRLS algorithms can be found in [16]. Simulation results demonstrate that the numerical behavior of the proposed algorithm is appreciably improved in comparison with the QOBE algorithm while both algorithms have the same complexity and convergence performance.

## 2. Set-membership adaptive filtering

Let us consider the affine-in-parameter model

$$d_n = \boldsymbol{\omega}^* \mathbf{x}_n + v_n$$

where  $d_n \in \mathbb{C}$  is the reference signal at time index  $n \in \mathbb{N}$ ,  $\boldsymbol{\omega} \in \mathbb{C}^L$  is the column vector of the underlying system parameters to be identified,  $\mathbf{x}_n \in \mathbb{C}^L$  is the input vector,  $L$  is the system order,  $v_n \in \mathbb{C}$  accounts for measurement noise, and superscript  $*$  denotes complex-conjugate transpose.

If we take  $\mathbf{w} \in \mathbb{C}^L$  as an estimate of  $\boldsymbol{\omega}$ , constraining the magnitude of the estimation error to be smaller than a pre-determined threshold,  $\gamma \in \mathbb{R}^+$ , yields a specification on  $\mathbf{w}$ . Consequently, there will be a set of feasible solutions for  $\mathbf{w}$  rather than a single estimate. The set of all vectors satisfying the error constraint for all possible input-desired output pairs in the model space of interest,  $\mathcal{S}$ , is called the *feasibility set* and is defined as

$$\Theta = \bigcap_{(\mathbf{x}, d) \in \mathcal{S}} \{\mathbf{w} \in \mathbb{C}^L: |d - \mathbf{w}^* \mathbf{x}| \leq \gamma\}.$$

Even with perfect knowledge of  $\mathcal{S}$ , which is not usually available, direct calculation of  $\Theta$  is formidable and computationally prohibitive. Hence, the adaptive SM filtering algorithms seek solutions that belong to a *membership set*,  $\Psi_n$ , which is a superset of  $\Theta$  and is devised to be the minimal set estimate for  $\Theta$  at time instant  $n$ . The membership set is defined as

$$\Psi_n = \bigcap_{i=1}^n \Phi_i$$

where  $\Phi_n$  is the *constraint set*, which contains all the possible estimates that are consistent with the data observed at time instant  $n$ . The set  $\Phi_n$  is also called the *observation-induced set* and is defined as

$$\Phi_n = \{\mathbf{w} \in \mathbb{C}^L: |d_n - \mathbf{w}^* \mathbf{x}_n| \leq \gamma\}.$$

The membership set  $\Psi_n$  is an  $L$ -dimensional convex polytope and still not easy to compute. Therefore, the OBE algorithms estimate a sequence of ellipsoids that tightly outer-bound  $\Psi_n$  instead. Different optimization approaches and optimality criteria regarding the bounding ellipsoids have given rise to the emergence of several OBE algorithms [9].

### 3. The QOBE Algorithm

#### 3.1. Algorithm derivation

The QOBE algorithm is a quasi-OBE algorithm with an unusual optimization criterion (in the realm of OBE processing) that endows it with attractive features like enhanced tracking ability and reduced computational complexity. Although QOBE is built upon the OBE premises, it can be regarded as a WRLS algorithm with a time-varying weighting factor,  $\ell_n$ , where the input autocorrelation matrix,  $\tilde{\mathbf{R}}_n$ , and the filter coefficients are updated via

$$\tilde{\mathbf{R}}_n = \tilde{\mathbf{R}}_{n-1} + \ell_n \mathbf{x}_n \mathbf{x}_n^* \quad (1)$$

and

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \ell_n \tilde{\mathbf{P}}_n \mathbf{x}_n e_n^* \quad (2)$$

with the *a priori* estimation error being defined as

$$e_n = d_n - \mathbf{w}_{n-1}^* \mathbf{x}_n.$$

In practice,  $\tilde{\mathbf{P}}_n = \tilde{\mathbf{R}}_n^{-1}$  is updated rather than  $\tilde{\mathbf{R}}_n$  by applying the matrix inversion lemma to (1), yielding

$$\tilde{\mathbf{P}}_n = \tilde{\mathbf{P}}_{n-1} - \frac{\ell_n \tilde{\mathbf{P}}_{n-1} \mathbf{x}_n \mathbf{x}_n^* \tilde{\mathbf{P}}_{n-1}}{1 + \ell_n \mathbf{x}_n^* \tilde{\mathbf{P}}_{n-1} \mathbf{x}_n}. \quad (3)$$

At time instant  $n$ , if  $|e_n| \leq \gamma$ , it is interpreted that  $\mathbf{w}_{n-1}$  is inside the constraint set  $\Phi_n$ , so there is no need for any update, i.e.,  $\mathbf{w}_n = \mathbf{w}_{n-1}$  and  $\tilde{\mathbf{P}}_n = \tilde{\mathbf{P}}_{n-1}$ . Conversely,  $|e_n| > \gamma$  means that  $\mathbf{w}_{n-1}$  is outside  $\Phi_n$ ; thus, it needs to be updated to a new vector,  $\mathbf{w}_n$ , that lies inside  $\Phi_n$ . In this case, an update is carried out via (3) and (2) while the optimum value of the weighting factor,  $\ell_n$ , is found by satisfying the bounded-error-magnitude constraint

$$|d_n - \mathbf{w}_n^* \mathbf{x}_n| = \gamma, \quad (4)$$

which ensures that  $\mathbf{w}_n$  is a member of  $\Phi_n$  and consequently  $\Psi_n$ .

Multiplying both sides of (3) by  $\mathbf{x}_n^*$  from the left and by  $\mathbf{x}_n$  from the right, we find

$$\mathbf{x}_n^* \tilde{\mathbf{P}}_n \mathbf{x}_n = \frac{\tilde{g}_n}{1 + \ell_n \tilde{g}_n}. \quad (5)$$

where

$$\tilde{g}_n = \mathbf{x}_n^* \tilde{\mathbf{P}}_{n-1} \mathbf{x}_n. \quad (6)$$

Multiplying the complex-conjugate transpose of both sides of (2) by  $\mathbf{x}_n$  from the right and subtracting the resulting equation from  $d_n$  together with substitution of (5), we obtain

$$d_n - \mathbf{w}_n^* \mathbf{x}_n = \frac{1}{1 + \ell_n \tilde{g}_n} e_n. \quad (7)$$

Substituting (7) into (4) results in

$$\ell_n = \frac{1}{\tilde{g}_n} \left( \frac{|e_n|}{\gamma} - 1 \right). \quad (8)$$

This completes the derivation of the QOBE algorithm, which is summarized in Table 1.

### 3.2. Numerical behavior

With each update of the QOBE algorithm, the trace of  $\tilde{\mathbf{R}}_n$  increases (the trace of  $\tilde{\mathbf{P}}_n$  decreases). Growth of  $\tilde{\mathbf{R}}_{n-1}$  (decay of  $\tilde{\mathbf{P}}_{n-1}$ ) decreases  $\tilde{g}_n$  and so increases  $\ell_n$  while a larger  $\ell_n$  can in turn accelerate the growth of  $\tilde{\mathbf{R}}_n$  (decay of  $\tilde{\mathbf{P}}_n$ ). This positive feedback can eventually lead to a blowup of  $\tilde{\mathbf{R}}_n$  and  $\ell_n$  and vanishing of  $\tilde{\mathbf{P}}_n$  as the adaptation progresses. Theoretically, with a persistently exciting input, the filter coefficients of the QOBE algorithm will converge to their true values, no matter how large or small the internal variables become [14]. However, in practical finite-precision implementations, the algorithm may suffer from instability and ensuing underflow/overflow of its internal variables,  $\tilde{\mathbf{P}}_n$  and  $\ell_n$ . To be specific,  $\tilde{\mathbf{P}}_n$  may ultimately become a zero matrix and consequently adaptation may cease. In the aftermath, the filter coefficients may diverge from the true values in the presence of the background noise and/or if the target system is not stationary. In the following, in order to substantiate our intuitions, we investigate the limiting behavior of QOBE's internal variables.

At the updating time instants, we have  $|e_n| > \gamma$  or  $|e_n|/\gamma - 1 > 0$ . Besides, assuming a persistently exciting input signal,  $\tilde{\mathbf{P}}_{n-1}$  will be positive-definite; hence,  $\tilde{g}_n > 0$ . Therefore,

$$\ell_n > 0. \quad (9)$$

Calculating the trace of both sides of (1) gives

$$\text{tr}\{\tilde{\mathbf{R}}_n\} = \text{tr}\{\tilde{\mathbf{R}}_{n-1}\} + \ell_n \|\mathbf{x}_n\|^2 \quad (10)$$

where  $\text{tr}\{\cdot\}$  denotes the matrix trace. Let us make the following assumption.

A1: In addition to being persistently exciting, the input signal is stationary and has finite mean and variance. Hence,  $0 < \|\mathbf{x}_n\| < \infty$  where  $\|\cdot\|$  denotes the Euclidean norm.

Considering (9) and A1, it is obvious from (10) that

$$\text{tr}\{\tilde{\mathbf{R}}_n\} > \text{tr}\{\tilde{\mathbf{R}}_{n-1}\}.$$

This means  $\text{tr}\{\tilde{\mathbf{R}}_n\}$  is monotone increasing. Hence, since the probability of update is normally non-zero at the steady state (e.g., with Gaussian noise), we can infer that

$$\lim_{n \rightarrow \infty} \text{tr}\{\tilde{\mathbf{R}}_n\} = \infty. \quad (11)$$

In consideration of A1, we can conclude from (11) that

$$\lim_{n \rightarrow \infty} \text{tr}\{\tilde{\mathbf{P}}_n\} = 0. \quad (12)$$

Rewriting (6) as

$$\tilde{g}_n = \text{tr}\{\mathbf{x}_n \mathbf{x}_n^* \tilde{\mathbf{P}}_{n-1}\}$$

and invoking Coope's inequality, i.e.,  $\text{tr}\{\mathbf{A}\mathbf{B}\} \leq \text{tr}\{\mathbf{A}\}\text{tr}\{\mathbf{B}\}$  for positive-semidefinite matrices of the same order  $\mathbf{A}$  and  $\mathbf{B}$  [17], we have

$$\tilde{g}_n \leq \|\mathbf{x}_n\|^2 \text{tr}\{\tilde{\mathbf{P}}_{n-1}\}. \quad (13)$$

In view of A1 and (12), (13) implies that  $\lim_{n \rightarrow \infty} \tilde{g}_n = 0$  and consequently

$$\lim_{n \rightarrow \infty} \ell_n = \infty. \quad (14)$$

The limiting results of (11) and (14) are sufficient to prove that the internal variables of the QOBE algorithm are unstable in nature.

It is worth mentioning that periodic resetting of the inverse autocorrelation matrix,  $\tilde{\mathbf{P}}_n$ , is a possible way to avoid the abovementioned numerical problems. However, due to frequent removal of past information from  $\tilde{\mathbf{P}}_n$ , such a solution would be far from optimal [18].

## 4. The MQOBE Algorithm

### 4.1. Algorithm derivation

Let us define

$$\mathbf{R}_n = \ell_n^{-1} \tilde{\mathbf{R}}_n, \quad (15)$$

$$\mathbf{P}_n = \mathbf{R}_n^{-1}, \quad (16)$$

and

$$\lambda_n = \frac{\ell_{n-1}}{\ell_n}. \quad (17)$$

Multiplying both sides of (1) by  $\ell_n^{-1}$  and using (15) and (16) in (2), we find

$$\mathbf{R}_n = \lambda_n \mathbf{R}_{n-1} + \mathbf{x}_n \mathbf{x}_n^* \quad (18)$$

and

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \mathbf{P}_n \mathbf{x}_n e_n^*. \quad (19)$$

Applying the matrix inversion lemma to (18) yields

$$\mathbf{P}_n = \lambda_n^{-1} \left( \mathbf{P}_{n-1} - \frac{\mathbf{P}_{n-1} \mathbf{x}_n \mathbf{x}_n^* \mathbf{P}_{n-1}}{\lambda_n + \mathbf{x}_n^* \mathbf{P}_{n-1} \mathbf{x}_n} \right). \quad (20)$$

Substituting (8) into (17) gives

$$\lambda_n = \ell_{n-1} \frac{\tilde{g}_n}{\frac{|e_n|}{\gamma} - 1}. \quad (21)$$

By rewriting (6) as

$$\tilde{g}_n = \mathbf{x}_n^* \ell_{n-1}^{-1} \mathbf{P}_{n-1} \mathbf{x}_n$$

and substituting it into (21), we get

$$\lambda_n = \frac{g_n}{\frac{|e_n|}{\gamma} - 1}. \quad (22)$$

where

$$g_n = \mathbf{x}_n^* \mathbf{P}_{n-1} \mathbf{x}_n. \quad (23)$$

As with the QOBE algorithm, at time instant  $n$ , an update occurs only if  $|e_n| > \gamma$ ; otherwise,  $\mathbf{w}_n = \mathbf{w}_{n-1}$  and  $\mathbf{P}_n = \mathbf{P}_{n-1}$ .

We call the resultant algorithm *modified quasi-OBE* (MQOBE) and summarize it in Table 2. The new algorithm is in fact an EWRLS algorithm with a time-varying forgetting factor that is optimized to satisfy the set-membership-induced error bound in (4).

If we multiply both sides of (20) by  $\mathbf{x}_n^*$  from the left and by  $\mathbf{x}_n$  from the right, we get

$$\mathbf{x}_n^* \mathbf{P}_n \mathbf{x}_n = \frac{g_n}{\lambda_n + g_n}. \quad (24)$$

By multiplying the complex-conjugate transpose of both sides of (19) by  $\mathbf{x}_n$  from the right and subtracting the resulting equation from  $d_n$  together with using (24), we can verify that

$$d_n - \mathbf{w}_n^* \mathbf{x}_n = \frac{1}{1 + \lambda_n^{-1} g_n} e_n. \quad (25)$$

Therefore, we can alternatively attain (22) by substituting (25) into (4) and solving it with respect to  $\lambda_n$ .

#### 4.2. Numerical behavior

The QOBE and MQOBE algorithms compute the same filter coefficients but in different ways because of having different internal variables, namely  $\tilde{\mathbf{P}}_n$  and  $\ell_n$  in QOBE versus  $\mathbf{P}_n$  and  $\lambda_n$  in MQOBE. In QOBE,  $\ell_n$  is a weighting factor, whereas in MQOBE,  $\lambda_n$  acts as a forgetting factor and serves to stabilize the internal variables by means of a negative feedback.

The recursion of (18) represents a first-order autoregressive process with a variable forgetting factor ( $\lambda_n$ ). Under A1, it is known that such a process is surely stable if the forgetting factor is always smaller than one, whereas it is surely unstable if the forgetting factor is always greater than one [19]. However, the latter can never be the case since  $\lambda_n$  being always greater than one results in

$$\lim_{n \rightarrow \infty} \text{tr}\{\mathbf{R}_n\} = \infty$$

or

$$\lim_{n \rightarrow \infty} \text{tr}\{\mathbf{P}_n\} = 0$$

which in turn leads to

$$\lim_{n \rightarrow \infty} \lambda_n = 0,$$

contradicting the initial assumption. In fact, it can be deduced from (14) and (17) that for most of the time instants we have  $\lambda_n < 1$ . More specifically, since  $\ell_n$  grows in time (though not monotonically), at most of the time instants  $\ell_n > \ell_{n-1}$ , which means  $\lambda_n < 1$ .

In order to study the numerical behavior of MQOBE, one should carefully examine the inter-variable relations in the algorithm, especially the algorithm's innate negative feedback mechanism that plays a key role in maintaining the stability of the internal variables. As an example, assuming A1 and convergence of the algorithm in the mean-

squared error sense, one can verify that decrease of  $\mathbf{P}_{n-1}$  (increase of  $\mathbf{R}_{n-1}$ ) decreases  $g_n$  and hence decreases  $\lambda_n$ , whereas a smaller  $\lambda_n$  leads to a greater  $\mathbf{P}_n$  (smaller  $\mathbf{R}_n$ ); on the other hand, increase of  $\mathbf{P}_{n-1}$  (decrease of  $\mathbf{R}_{n-1}$ ) increases  $g_n$  and hence increases  $\lambda_n$ , whereas a larger  $\lambda_n$  leads to a smaller  $\mathbf{P}_n$  (greater  $\mathbf{R}_n$ ).

To demonstrate the mechanism of the abovementioned negative feedback and its stabilizing effect, we consider the special case of uncorrelated input signal by adopting the following assumption.

A2: The elements of the input signal vector are stationary correlation-ergodic and uncorrelated with zero mean and variance  $\sigma_x^2$ ; hence,  $E[\mathbf{x}_n \mathbf{x}_n^*] = \sigma_x^2 \mathbf{I}_L$  where  $\mathbf{I}_L$  is the  $L \times L$  identity matrix.

From (18), we have

$$\mathbf{R}_n = \prod_{i=1}^n \lambda_i \mathbf{R}_0 + \sum_{i=1}^{n-1} \left( \prod_{j=i+1}^n \lambda_j \right) \mathbf{x}_i \mathbf{x}_i^* + \mathbf{x}_n \mathbf{x}_n^*$$

where  $\mathbf{R}_0 = \delta \mathbf{I}_L$  is the initial covariance matrix with  $\delta$  set to a small positive number. For the sake of simplicity, let us assume that an update occurs at every iteration. Since the input signal is correlation-ergodic, as  $n \rightarrow \infty$ ,  $\mathbf{R}_n$  can be approximated by

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{R}_n &\approx \lim_{n \rightarrow \infty} E[\mathbf{R}_n] \\ &\approx \lim_{n \rightarrow \infty} E \left[ \prod_{i=1}^n \lambda_i \mathbf{R}_0 + \sum_{i=1}^{n-1} \left( \prod_{j=i+1}^n \lambda_j \right) \mathbf{x}_i \mathbf{x}_i^* + \mathbf{x}_n \mathbf{x}_n^* \right] \\ &\approx \underbrace{\sigma_x^2 \left( \lim_{n \rightarrow \infty} \sum_{i=1}^{n-1} E \left[ \prod_{j=i+1}^n \lambda_j \right] + 1 \right)}_{r_\infty} \mathbf{I}_L. \end{aligned} \quad (26)$$

where

$$\prod_{i=1}^{\infty} \lambda_i \approx 0$$

by virtue of the fact that  $\lambda_n$  cannot remain larger than one for many successive update iterations (see Fig. 3 for an illustration of this). Consequently,  $r_\infty$  is assured to be finite. Note that  $r_\infty$  cannot possibly diverge to infinity as this would result in  $\lambda_\infty \rightarrow 0$ , thereby preventing any divergence in the first place. Equation (26) suggests that, at the steady state and under A2,  $\mathbf{R}_n$  can be approximated as a multiple of the identity matrix.

## 5. Simulations

We compare the performance of the MQOBE algorithm with the QOBE algorithm and the conventional RLS algorithm in an application of flat-fading MIMO channel estimation studied in [20]. For this purpose, a MIMO communication system with four transmitter and four receiver antennas is considered. The sub-channels between all the transmitter and receiver pairs are independent Rayleigh-fading and vary in time based on Jakes model [21] with a normalized Doppler frequency of  $f_D T_s = 0.01$  where  $f_D$  is the maximum Doppler frequency shift and  $T_s$  is the transmission symbol period. A sudden random change in the channel taps is also introduced halfway through the simulations.

Four finite impulse response (FIR) filters each having four taps constitute the MIMO channel estimator. Similar to [20], in the MQOBE and QOBE algorithms, the Euclidean norm of the error vector composed by the output errors of all the filters is used for the considered MIMO case in place of the absolute of the scalar error in the SISO case. The transmitted symbols are modulated using QPSK scheme and grouped into packets of data each containing 80 symbol vectors. These vectors are the common input to the filters of the MIMO channel estimator. For the RLS algorithm, a fixed forgetting factor of 0.9 is used regarding the assumed normalized Doppler frequency. For the MQOBE and QOBE algorithms, the error threshold is set to  $\gamma = 0.1$ . The energy per bit to noise power spectral density ratio is also  $E_b/N_0 = 12$  dB.

In Fig. 1, we compare the estimation performance of the algorithms in terms of mean squared error (MSE), ensemble-averaged over  $10^4$  independent trials. Fig. 2 shows the trace of the inverse input autocorrelation matrix versus time for different algorithms. Fig. 3 also shows time evolution of the optimal weighting factor of the QOBE algorithm and the optimal forgetting factor of the MQOBE algorithm together with the fixed forgetting factor of the RLS algorithm. It should be noted that, in this experiment, both QOBE and MQOBE updated in average at 80 percent of the iterations. Figs. 4-6 correspond to Figs. 1-3, respectively, when the simulations are performed for 3000 iterations without any sudden change in the channel taps.

We observe that the MQOBE and QOBE algorithms have the same estimation performance as long as QOBE's internal variables are in the realizable range, which is  $4.94 \times 10^{-324}$  to  $1.79 \times 10^{308}$  for the double-precision floating-point format. However, when its internal variables hit the numerical limits, the QOBE algorithm stalls. The simulation results emphasize that the QOBE algorithm is prone to numerical instability, in particular, when the update frequency is high. It is also evident that QOBE's internal variables,  $\ell_n$  and  $\tilde{\mathbf{P}}_n$ , exponentially grow/decay in time. This makes their dynamic range extremely wide and consequently QOBE's practicable run-time very limited. In Figs. 2 and 3, QOBE's optimal weight grows from about 1 to more than  $10^{15}$  and the trace of its inverse autocorrelation matrix drops from about

10 to less than  $10^{-15}$  in only 80 iterations. On the other hand, the internal variables of MQOBE fluctuate around their steady-state values and are of much smaller dynamic range rendering MQOBE more suitable than QOBE for practical applications.

## 6. Conclusion

It was shown that the set-membership-based QOBE algorithm, despite all its merits, suffers from numerical instability of its internal variables. To resolve this drawback, an appropriate transformation of the internal variables was introduced leading to a new set-membership adaptive filtering algorithm, called MQOBE. Unlike most of the OBE algorithms that are known as set-membership weighted RLS algorithms, the proposed algorithm is a set-membership exponentially-weighted RLS algorithm with a time-varying forgetting factor, which is optimized within the context of the set-membership filtering. The proposed algorithm exhibits a good convergence and tracking performance, identical to QOBE, while providing a dramatically improved numerical behavior.

## References

- [1] E. Fogel and Y. F. Huang, "On the value of information in system identification – bounded noise case," *Automatica*, vol. 18, no. 2, pp. 229–238, 1982.
- [2] S. Dasgupta and Y.-F. Huang, "Asymptotically convergent modified recursive least-squares with data-dependent updating and forgetting factor for systems with bounded noise," *IEEE Trans. Inf. Theory*, vol. 33, no. 3, pp. 383–392, 1987.
- [3] J. R. Deller, Jr., M. Nayeri, and S. F. Odeh, "Least-square identification with error bounds for real-time signal processing and control," *Proc. IEEE*, vol. 81, pp. 813–849, June 1993.
- [4] J. P. Norton, Ed., Special issues on "Bounded-error methods in system identification," *Int. J. Automat. Contr. Signal Process.*, vol. 8, Jan./Feb. 1994; vol. 9, Jan./Feb. 1995.
- [5] S. Gollamudi, S. Kapoor, S. Nagaraj, and Y.-F. Huang, "Set-membership adaptive equalization and an updatator-shared implementation for multiple channel communications systems," *IEEE Trans. Signal Process.*, vol. 46, no. 9, pp. 2372–2385, 1998.
- [6] S. Gollamudi, S. Nagaraj, S. Kapoor, and Y.-F. Huang, "Set-membership filtering and a set-membership normalized LMS algorithm with an adaptive step size," *IEEE Signal Process. Lett.*, vol. 5, no. 5, pp. 111–114, 1998.

- [7] S. Nagaraj, S. Gollamudi, S. Kapoor, and Y. F. Huang, "BEACON: An adaptive set-membership filtering technique with sparse updates," *IEEE Trans. Signal Process.*, vol. 47, no. 11, pp. 2928–2941, 1999.
- [8] S. Werner and P. S. R. Diniz, "Set-membership affine projection algorithm," *IEEE Signal Process. Lett.*, vol. 8, pp. 231–235, August 2001.
- [9] J. R. Deller, Jr. and Y. F. Huang, "Set-membership identification and filtering for signal processing applications," *Circuits Systems Signal Process.*, vol. 21, no. 1, pp. 69–82, 2002.
- [10] P. S. R. Diniz and S. Werner, "Set-membership binormalized data-reusing LMS algorithms," *IEEE Trans. Signal Process.*, vol. 51, pp. 124–134, January 2003.
- [11] S. Werner, M. L. R. de Campos, and P. S. R. Diniz, "Partial-update NLMS algorithms with data-selective updating," *IEEE Trans. Signal Process.*, vol. 52, pp. 938–949, April 2004.
- [12] D. Joachim and J. R. Deller, Jr., "Multiweight optimization in optimal bounding ellipsoid algorithms," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 679–690, 2006.
- [13] S. Werner, J. A. Apolinário, Jr., and P. S. R. Diniz, "Set-membership proportionate affine projection algorithms," *EURASIP J. Audio, Speech, Music Process.*, vol. 2007, Article ID 34242, 10 pages, 2007.
- [14] J. R. Deller, Jr., S. Gollamudi, S. Nagaraj, D. Joachim, and Y. F. Huang, "Convergence analysis of the quasi-OBE algorithm and related performance issues," *Int. J. Adapt. Control Signal Process.*, vol. 21, pp. 499–527, 2007.
- [15] K. Doğançay, *Partial-Update Adaptive Signal Processing: Design, Analysis and Implementation*, Oxford: Academic Press, 2008.
- [16] F. T. Castoldi and M. L. R. de Campos, "Minimum-disturbance description for the development of adaptation algorithms and a new leakage least squares algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Taipei, Taiwan, April. 2009, pp. 3129–3132.
- [17] I. D. Coope, "On matrix trace inequalities and related topics for products of Hermitian matrix," *J. Math. Anal. Appl.*, vol. 188, pp. 999–1001, 1994.
- [18] J. T. Milek, *Stabilized Adaptive Forgetting in Recursive Parameter Estimation*. Zurich: vdf Hochschulverlag AG, 1995.
- [19] D. F. Nicholls and B. G. Quinn, *Random Coefficient Autoregressive Models: An Introduction*, New York: Springer-Verlag, 1982.

- [20] T. Wang, R. C. De Lamare, and P. D. Mitchell, "Low-complexity set-membership channel estimation for cooperative wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 6, pp. 2594–2607, 2011.
- [21] W. C. Jakes, *Microwave Mobile Communications*, New York: Wiley, 1974.

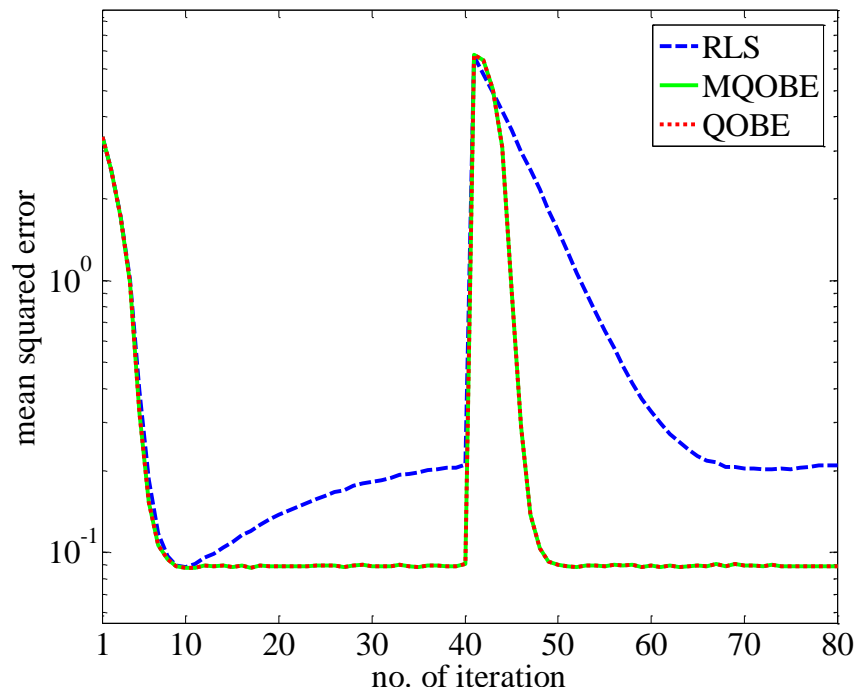


Fig. 1, Mean squared error performance of different algorithms.

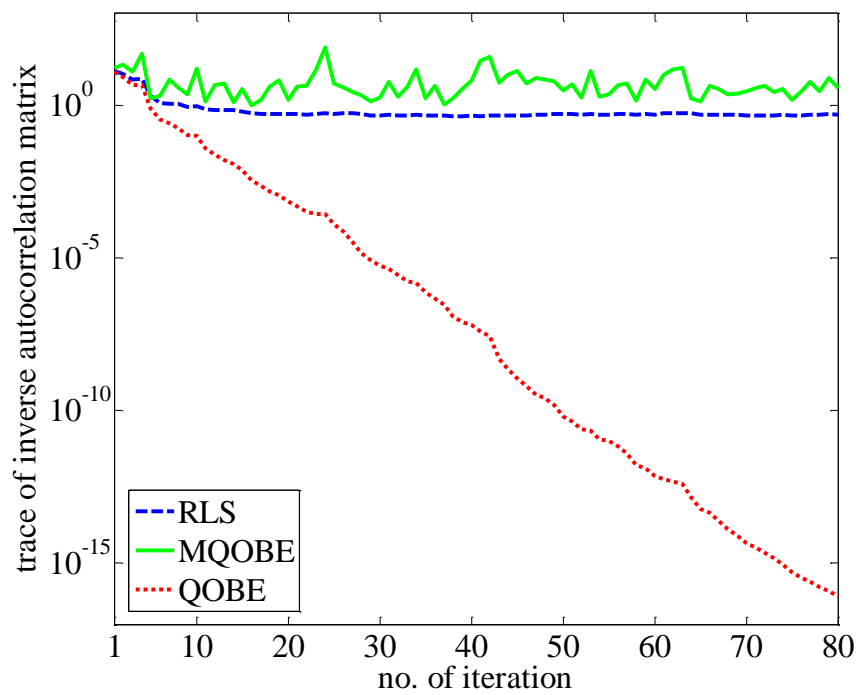


Fig. 2, Trace of the inverse autocorrelation matrix versus time for different algorithms.

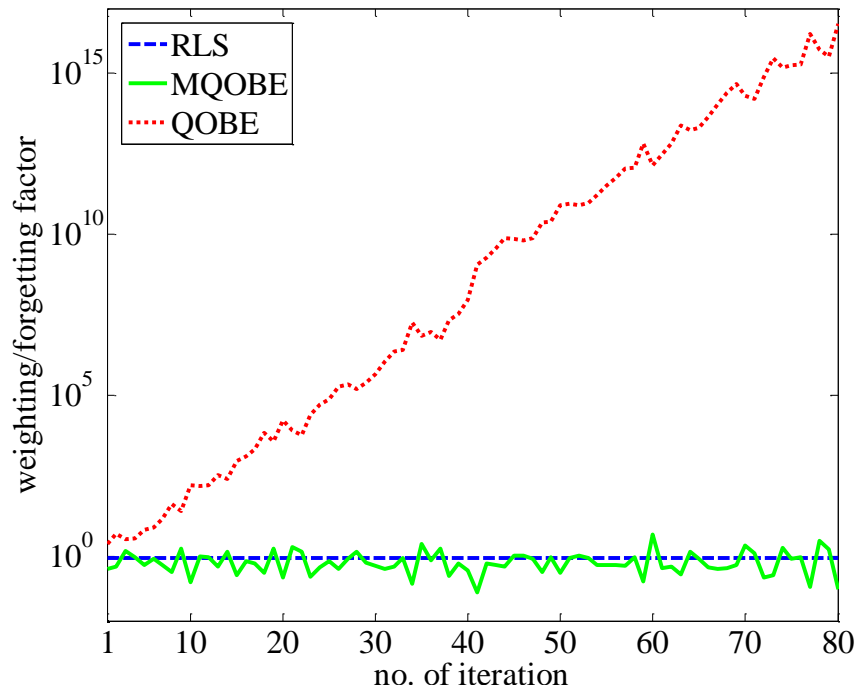


Fig. 3, Time evolution of the optimal weight of the QOBE algorithm and the optimal forgetting factor of the MQOBE algorithm together with the fixed forgetting factor of the RLS algorithm.

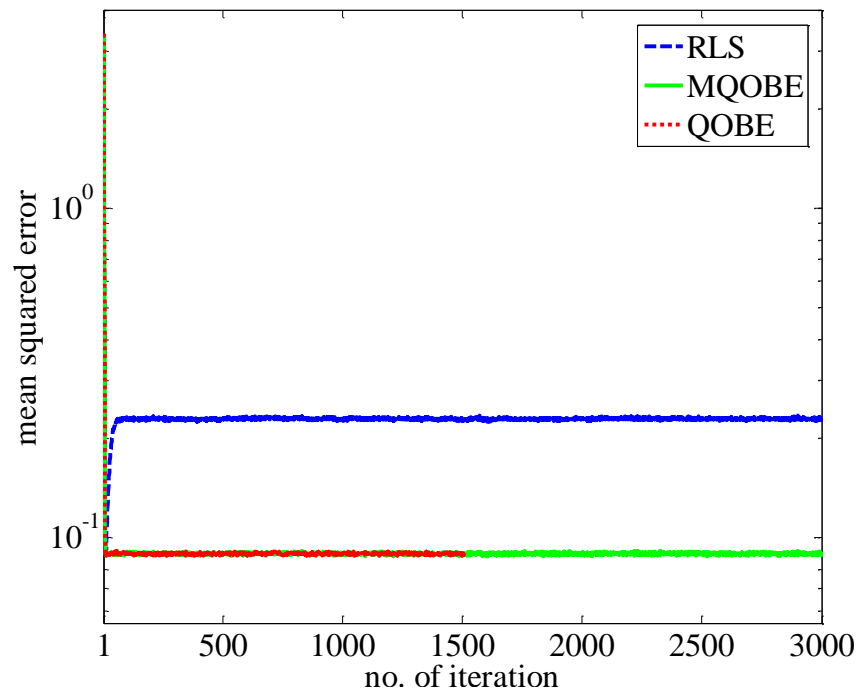


Fig. 4, Mean squared error performance of different algorithms.

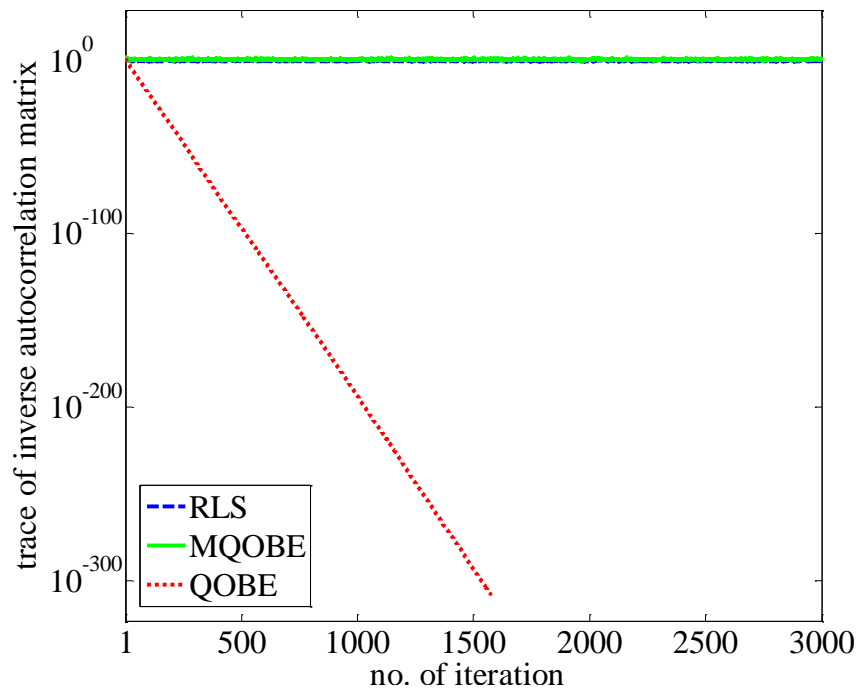


Fig. 5, Trace of the inverse autocorrelation matrix versus time for different algorithms.

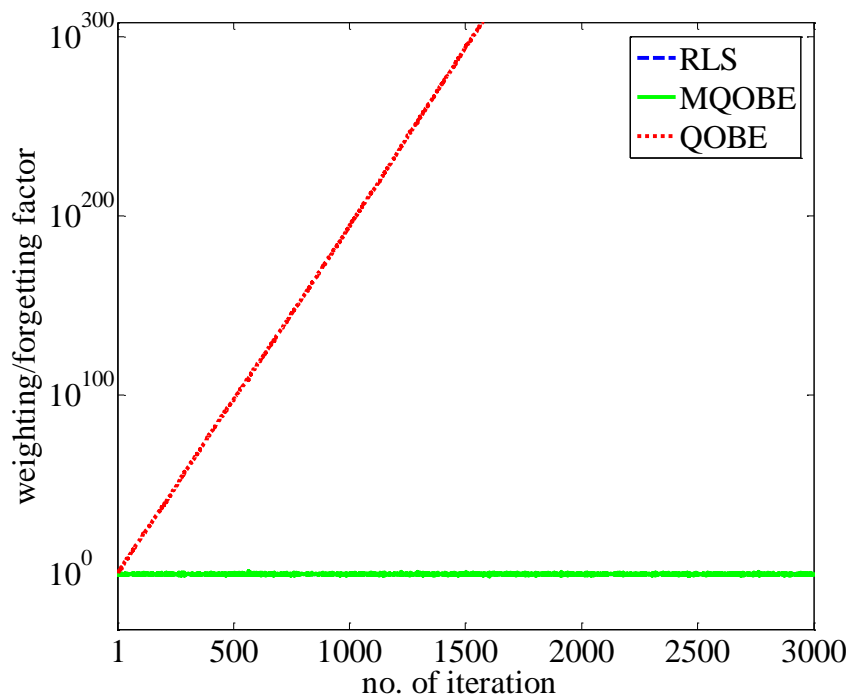


Fig. 6, Time evolution of the optimal weight of the QOBE algorithm and the optimal forgetting factor of the MQOBE algorithm together with the fixed forgetting factor of the RLS algorithm.

Table 1, The QOBE algorithm.

---

Initialization

$\tilde{\mathbf{P}}_0 = \delta^{-1} \mathbf{I}_L$  where  $\delta$  is a small positive number

$$\mathbf{w}_0 = \mathbf{0}$$

At iteration  $n = 1, 2, \dots$

$$e_n = d_n - \mathbf{w}_{n-1}^* \mathbf{x}_n$$

if  $|e_n| > \gamma$

$$\tilde{g}_n = \mathbf{x}_n^* \tilde{\mathbf{P}}_{n-1} \mathbf{x}_n$$

$$\ell_n = \frac{1}{\tilde{g}_n} \left( \frac{|e_n|}{\gamma} - 1 \right)$$

$$\tilde{\mathbf{P}}_n = \tilde{\mathbf{P}}_{n-1} - \frac{\ell_n \tilde{\mathbf{P}}_{n-1} \mathbf{x}_n \mathbf{x}_n^* \tilde{\mathbf{P}}_{n-1}}{1 + \ell_n \mathbf{x}_n^* \tilde{\mathbf{P}}_{n-1} \mathbf{x}_n}$$

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \ell_n \tilde{\mathbf{P}}_n \mathbf{x}_n e_n^*$$

otherwise

$$\tilde{\mathbf{P}}_n = \tilde{\mathbf{P}}_{n-1}$$

$$\mathbf{w}_n = \mathbf{w}_{n-1}$$


---

Table 2, The MQOBE algorithm.

---

Initialization

$$\mathbf{P}_0 = \delta^{-1} \mathbf{I}_L$$

$$\mathbf{w}_0 = \mathbf{0}$$

At iteration  $n = 1, 2, \dots$

$$e_n = d_n - \mathbf{w}_{n-1}^* \mathbf{x}_n$$

if  $|e_n| > \gamma$

$$g_n = \mathbf{x}_n^* \mathbf{P}_{n-1} \mathbf{x}_n$$

$$\lambda_n = \frac{g_n}{\frac{|e_n|}{\gamma} - 1}$$

$$\mathbf{P}_n = \lambda_n^{-1} \left( \mathbf{P}_{n-1} - \frac{\mathbf{P}_{n-1} \mathbf{x}_n \mathbf{x}_n^* \mathbf{P}_{n-1}}{\lambda_n + \mathbf{x}_n^* \mathbf{P}_{n-1} \mathbf{x}_n} \right)$$

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \mathbf{P}_n \mathbf{x}_n e_n^*$$

otherwise

$$\mathbf{P}_n = \mathbf{P}_{n-1}$$

$$\mathbf{w}_n = \mathbf{w}_{n-1}$$


---

# Paper H

Originally published as

R. Arablouei and K. Doğançay, “Low-complexity implementation of the quasi-OBE algorithm,” *Electronics Letters*, vol. 48, no. 11, pp. 621-623, 2012.

Copyright © 2012 IET. Reprinted with permission.

## Low-complexity implementation of the quasi-OBE algorithm

Reza Arablouei and Kutluyıl Doğançay

### Abstract

The quasi-OBE (QOBE) algorithm is a set-membership adaptive filtering algorithm based on the principles of optimal bounding ellipsoid (OBE) processing. This algorithm can provide enhanced convergence and tracking performance as well as reduced average computational complexity in comparison with the more traditional adaptive filtering algorithms. In this letter, we implement the QOBE algorithm using the dichotomous coordinate descent (DCD) iterations. This significantly reduces QOBE’s per-update computational complexity, especially the number of required multiplications, and delivers a trade-off between performance and complexity.

### 1. Introduction

The quasi-OBE (QOBE) algorithm [1, 2] is a set-membership adaptive filtering algorithm built upon the optimal bounding ellipsoid (OBE) concept [3]. It shares many of the desired features of the various OBE algorithms. Moreover, it incorporates simple but efficient innovation check and optimal weight calculation processes, which make it computationally more efficient than other OBE algorithms. The QOBE algorithm is of particular interest in signal processing applications because it features two major advantages over more classical adaptive filtering algorithms. First, it exhibits superior adaptation and tracking properties. Second, it can effectively make use of innovation in the data and improve computational efficiency by establishing a data-discerning update strategy for the parameter estimates. However, its computational complexity is  $\mathcal{O}(L^2)$  with  $L$  being the system order, even though there is a shift structure in the input data.

To reduce the complexity of the QOBE algorithm, we rearrange it in such a way that it requires solution of two linear systems of equations (LSOEs) instead of inverting a matrix or employing the matrix inversion lemma. We utilise the dichotomous coordinate descent (DCD) algorithm [4] to solve these LSOEs. The DCD algorithm is a multiplication-free method dominated by additions and falls into the class of shift-and-add algorithms. Nevertheless, it yields an approximate solution and its accuracy depends on the number of exercised iterations. However, it has been shown that the approximate solution can be made arbitrarily close to the true solution by increasing the number of iterations and the resolution of the step size [5]. Using the DCD algorithm, in addition to an appreciable reduction in the required computations, especially multiplications, a trade-off between performance and complexity of the QOBE algorithm is achieved while maintaining its desirable data-discerning update property.

## 2. The DCD-QOBE algorithm

Consider the affine-in-parameter model

$$d(n) = \mathbf{w}_o^* \mathbf{x}(n) + v(n) \quad (1)$$

where  $d(n) \in \mathbb{C}$  is the reference signal,  $\mathbf{w}_o \in \mathbb{C}^L$  is the column vector of the unknown system parameters,  $\mathbf{x}(n) \in \mathbb{C}^L$  is the input vector,  $v(n) \in \mathbb{C}$  is the background noise,  $\mathbb{C}$  is the set of complex numbers, and superscript  $*$  denotes complex-conjugate transposition. The original QOBE algorithm (also known as BEACON) is formulated as [2]

$$e(n) = d(n) - \mathbf{w}^*(n-1) \mathbf{x}(n) \quad (2)$$

$$\lambda(n) = \begin{cases} 0 & \text{if } |e(n)| \leq \gamma \\ \frac{1}{\mathbf{x}^*(n) \mathbf{R}^{-1}(n-1) \mathbf{x}(n)} \left( \frac{|e(n)|}{\gamma} - 1 \right) & \text{if } |e(n)| > \gamma \end{cases} \quad (3)$$

$$\mathbf{R}(n) = \mathbf{R}(n-1) + \lambda(n) \mathbf{x}(n) \mathbf{x}^*(n) \quad (4)$$

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \lambda(n) \mathbf{R}^{-1}(n) \mathbf{x}(n) e^*(n) \quad (5)$$

where  $e(n)$  is the filter output error,  $\mathbf{w}(n)$  is the vector of filter weights,  $\lambda(n)$  is the time-varying weighting factor, and  $\gamma$  is a predefined threshold. In this formulation, both the input autocorrelation matrix,  $\mathbf{R}(n)$ , and its inverse,  $\mathbf{R}^{-1}(n)$ , are used. Consequently, it requires the inversion of  $\mathbf{R}(n)$  at each iteration. One way to avoid this matrix inversion is to update  $\mathbf{R}^{-1}(n)$  instead of  $\mathbf{R}(n)$  using the matrix inversion lemma, i.e.,

$$\mathbf{R}^{-1}(n) = \mathbf{R}^{-1}(n-1) + \frac{\lambda(n) \mathbf{R}^{-1}(n-1) \mathbf{x}(n) \mathbf{x}^*(n) \mathbf{R}^{-1}(n-1)}{1 + \lambda(n) \mathbf{x}^*(n) \mathbf{R}^{-1}(n-1) \mathbf{x}(n)}. \quad (6)$$

Computational complexity of the QOBE algorithm is  $\mathcal{O}(L^2)$  and it does not exploit any possible structure, e.g., shift structure, in the input data. To reduce its peak (per-update) complexity, we propose to reformulate the QOBE algorithm with two LSOEs and solve the LSOEs utilising the low-complexity DCD iterations. Since QOBE can be viewed as a weighted recursive least-squares (WRLS) algorithm with a variable weighting factor, which is optimised according to the bounded-error-magnitude criterion [3], we can replace (5) with

$$\mathbf{z}(n) = \mathbf{z}(n-1) + \lambda(n)\mathbf{x}(n)d^*(n), \quad (7)$$

$$\mathbf{w}(n) = \mathbf{R}^{-1}(n)\mathbf{z}(n) \quad (8)$$

where  $\mathbf{z}(n)$  is the cross-correlation vector between the input vector and the reference signal. Instead of inverting  $\mathbf{R}(n)$  in (8), we can consider  $\hat{\mathbf{w}}(n)$  as an approximate solution of the following LSOE

$$\mathbf{R}(n)\mathbf{w}(n) = \mathbf{z}(n). \quad (9)$$

As shown in [5], it is more efficient to solve the following LSOE instead of (9)

$$\mathbf{R}(n)\Delta\mathbf{w}(n) = \boldsymbol{\rho}(n) \quad (10)$$

where  $\Delta\mathbf{w}(n) = \mathbf{w}(n) - \hat{\mathbf{w}}(n-1)$  and obtain an approximate solution  $\Delta\hat{\mathbf{w}}(n) = \hat{\mathbf{w}}(n) - \hat{\mathbf{w}}(n-1)$  from which we can find  $\hat{\mathbf{w}}(n)$ . Similar to the approach taken in [5], it is easy to show that

$$\boldsymbol{\rho}(n) = \mathbf{r}(n-1) + \lambda(n)\mathbf{x}(n)e^*(n) \quad (11)$$

where the residual vector  $\mathbf{r}(n-1)$  is defined as

$$\mathbf{r}(n-1) = \mathbf{z}(n-1) - \mathbf{R}(n-1)\hat{\mathbf{w}}(n-1). \quad (12)$$

Moreover, defining  $\mathbf{q}(n) = \mathbf{R}^{-1}(n-1)\mathbf{x}(n)$ , we find  $\hat{\mathbf{q}}(n)$  as an approximate solution of the following LSOE

$$\mathbf{R}(n)\mathbf{q}(n) = \mathbf{x}(n) \quad (13)$$

and calculate the weighting factor as

$$\lambda(n) = \frac{1}{\mathbf{x}^*(n)\hat{\mathbf{q}}(n)} \left( \frac{|e(n)|}{\gamma} - 1 \right).$$

To solve the LSOEs of (10) and (13), we employ the DCD algorithm with a leading element. Solving (10) in this way yields both  $\Delta\hat{\mathbf{w}}(n)$  and  $\mathbf{r}(n)$  [5]. We call the proposed algorithm DCD-QOBE.

For shift-structured input data, updating the input autocorrelation matrix is significantly simplified. In this case, the lower-right  $(L-1) \times (L-1)$  block of  $\mathbf{R}(n)$

can be obtained by copying the upper-left  $(L - 1) \times (L - 1)$  block of  $\mathbf{R}(n - 1)$ . The only part of the matrix  $\mathbf{R}(n)$  that should be directly updated is the first row and first column. Due to the symmetry of  $\mathbf{R}(n)$ , it is sufficient to update only the first column via

$$\mathbf{R}^{(1)}(n) = \mathbf{R}^{(1)}(n - 1) + \lambda(n)\mathbf{x}(n)x^*(n) \quad (14)$$

where  $\mathbf{R}^{(1)}(n)$  is the first column of  $\mathbf{R}(n)$  and  $x(n)$  is the input signal.

The DCD algorithm utilises three user-defined parameters, namely  $N_u$ ,  $M_b$ , and  $H$ , that control its accuracy and complexity [4, 5]. In fact, the first two establish the trade-off between complexity and performance. The integer parameter  $N_u$  represents the number of iterative updates performed at each run of the algorithm. In other words,  $N_u$  determines the maximum number of filter coefficients that can be updated at each time instant. Hence, in general, adaptive filtering based on the DCD algorithm implements a form of *selective partial updates*. It is known that by selective partial updating, one can trade performance for complexity [6]. In the DCD algorithm, the step-size  $\alpha$  can accept one of  $M_b$  predefined values corresponding to representation of the elements of the solution vector as fixed-point words with  $M_b$  bits within an amplitude range of  $[-H, H]$ .

### 3. Computational Complexity

For each update, the QOBE algorithm requires  $3L^2 + 4L + 2$  complex multiplications and  $3L^2 + L + 1$  complex additions while the DCD-QOBE algorithm requires  $4L + 4$  complex multiplications and  $(4L + 1)N_u + 5L + M_b$  complex additions. Assuming six and two floating-point operations (FLOPs) for each complex multiplication and addition, respectively, in Fig. 1, we compare computational complexity of QOBE and DCD-QOBE for different filter lengths when  $M_b = 12$  and  $N_u = L/4, L/2$ , and  $L$ . We also show the percentage of the saved FLOPs by DCD-QOBE in comparison with QOBE for each  $N_u$  in the subfigure inside Fig. 1.

### 4. Simulations

In Fig. 2, we compare the normalised misalignment, i.e.,

$$\frac{\|\mathbf{w}_o - \mathbf{w}(n)\|^2}{\|\mathbf{w}_o\|^2},$$

of QOBE and DCD-QOBE with different number of DCD iterations — $N_u = 2, 4, 8$ , and  $16$  for solving both (10) and (13)— in a typical system identification scenario. The results are averaged over  $10^4$  independent trials. Both the unknown system and the adaptive filter are considered to be of order  $L = 16$ . The filter weights of the unknown system are generated randomly and its output is corrupted by an additive

white complex Gaussian noise with a zero mean to achieve an approximate signal-to-noise ratio of 10 dB. The error-magnitude threshold is set to  $\gamma = \sqrt{3}\sigma_v$ , where  $\sigma_v^2$  is the power of the noise. The input signal is zero-mean complex Gaussian and the input signal vectors are shift-structured. The other DCD parameters are  $M_b = 12$  and  $H = 1$  for solving both (10) and (13). It should be noted that in this experiment, all the algorithms updated in average at about 13 percent of the simulation time.

## 5. Conclusion

Implementation of the set-membership-based QOBE algorithm using the DCD iterations was proposed. The new algorithm, DCD-QOBE, has a significantly lower peak computational complexity than the QOBE algorithm while having almost the same update frequency as QOBE for a wide range of exercised iterations in its embedded DCD algorithms. In particular, DCD-QOBE requires considerably less number of multiplications compared to QOBE. This makes it much less complicated than QOBE to implement on hardware. Simulations show that DCD-QOBE's performance can be made arbitrarily close to that of QOBE by choosing appropriate values for its DCD parameters. The DCD-QOBE algorithm, along with its inherent set-theoretic nature, practices a form of selective partial updating and establishes a trade-off between performance and complexity.

## References

- [1] J. R. Deller, Jr., S. Gollamudi, S. Nagaraj, D. Joachim, and Y. F. Huang, "Convergence analysis of the quasi-OBE algorithm and related performance issues," *Int. J. Adapt. Control Signal Process.*, vol. 21, pp. 499–527, 2007.
- [2] S. Nagaraj, S. Gollamudi, S. Kapoor, and Y. F. Huang, "BEACON: An adaptive set-membership filtering technique with sparse updates," *IEEE Trans. Signal Process.*, vol. 47, no. 11, pp. 2928–2941, 1999.
- [3] J. R. Deller, Jr. and Y. F. Huang, "Set-membership identification and filtering for signal processing applications," *Circuits Systems Signal Process.*, vol. 21, no. 1, pp. 69–82, 2002.
- [4] Y. V. Zakharov and T. C. Tozer, "Multiplication-free iterative algorithm for LS problem," *Electron. Lett.*, vol. 40, no. 9, pp. 567–569, Apr. 2004.
- [5] Y. Zakharov, G. White, and J. Liu, "Low-complexity RLS algorithms using dichotomous coordinate descent iterations," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3150–3161, Jul. 2008.
- [6] K. Doğançay, *Partial-Update Adaptive Signal Processing: Design, Analysis and Implementation*, Academic Press, Oxford, UK, 2008.

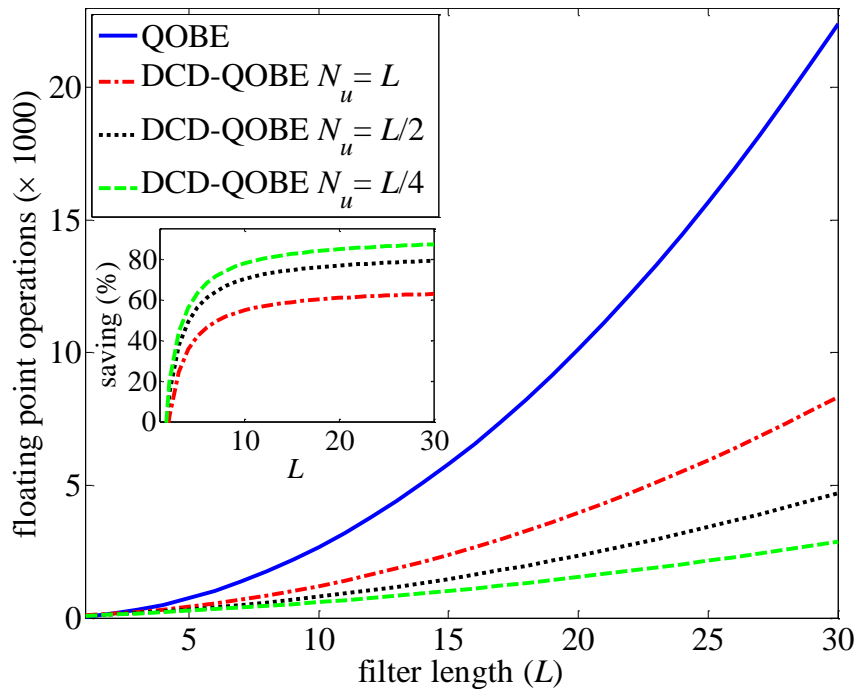


Fig. 1, Computational complexity of QOBE and DCD-QOBE.

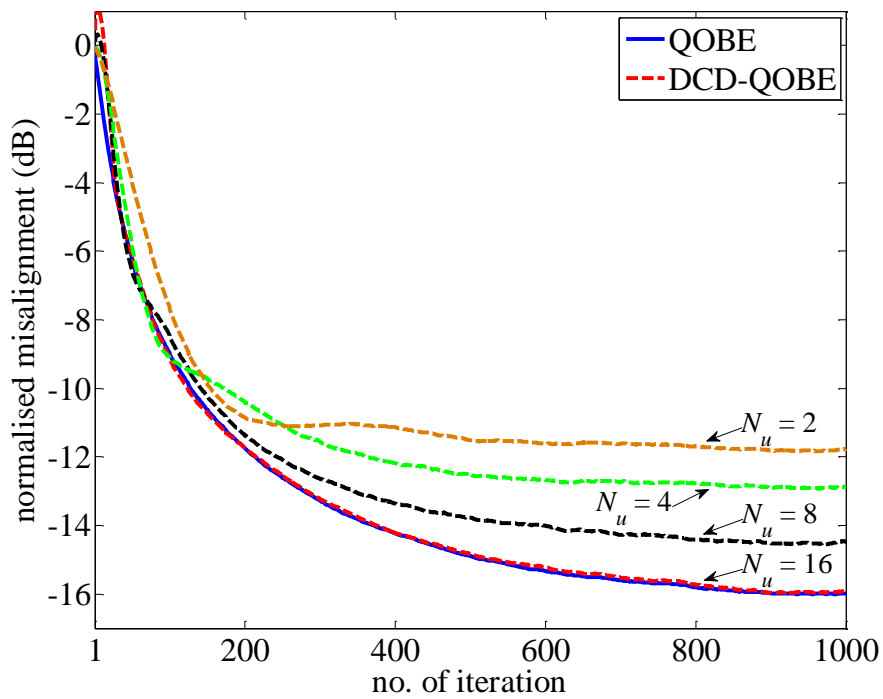


Fig. 2, Normalised misalignment of QOBE and DCD-QOBE.

# Paper I

Originally published as

R. Arablouei and K. Doğançay, "Steady-state mean squared error and tracking performance analysis of the quasi-OBE algorithm," *Signal Processing*, vol. 93, pp. 100-108, 2013.

Copyright © 2013 Elsevier. Reprinted with permission.

## Steady-state mean squared error and tracking performance analysis of the quasi-OBE algorithm

Reza Arablouei and Kutluyıl Doğançay

### Abstract

The quasi-OBE (QOBE) algorithm is a set-membership adaptive filtering algorithm based on the principles of optimal bounding ellipsoid (OBE) processing. This algorithm can provide enhanced convergence and tracking performance as well as reduced average computational complexity in comparison with the more traditional adaptive filtering algorithms such as the recursive least-squares (RLS) algorithm. In this paper, we analyse the steady-state mean squared error (MSE) and tracking performance of the QOBE algorithm. For this purpose, we derive energy conservation relation of the QOBE algorithm. The analysis leads to a nonlinear equation whose solution gives the steady-state MSE of the QOBE algorithm in both stationary and nonstationary environments. We prove that there is always a unique solution for this equation. The results predicted by the analysis show good agreement with the simulation experiments.

**Keywords:** set-membership adaptive filtering; the quasi-OBE (BEACON) algorithm; steady-state mean squared error performance analysis; tracking performance analysis; energy conservation

### 1. Introduction

Set-membership adaptive filtering (SMAF) algorithms are known for their superiority over the more classical estimation techniques in three aspects. First, they provide sets

of acceptable estimates rather than single estimates. Second, they feature improved convergence and tracking properties generally due to the execution of some kind of set-theoretic optimization, e.g., optimization of the step size or the weighting sequence. Third, they enjoy a data-discerning update strategy, which enables them to check for the innovation in the new data at each time instant and determine whether an update is required or not. In other words, they perform an update only when it can improve the quality of estimation and obviate the expense of updating when there is no useful information in the incoming data. As a result, in addition to the enhanced convergence and tracking performance, they can also provide an appreciable reduction in the average computational complexity [1-3].

The optimal bounding ellipsoids (OBE) algorithms are well-established SMAF algorithms that tightly outer-bound the set of feasible solutions in the associated parameter space using ellipsoids. They optimize the size of the ellipsoids in some meaningful sense. Different optimality criteria have led to different OBE algorithms. Among them, the quasi-OBE (QOBE) algorithm<sup>4</sup> [4, 5] is particularly interesting since it shares many of the desired features of the various OBE algorithms. Furthermore, it incorporates simple but efficient innovation check and optimal weight calculation processes, which make it computationally more efficient than other OBE algorithms.

The SMAF algorithms generally have high degrees of nonlinearities, which make their performance analysis complicated, especially at the transient state. The QOBE algorithm involves both data nonlinearity [6] and error nonlinearity [7]. Few works have been reported on the theoretical performance analysis of the SMAF algorithms. Steady-state mean squared error (MSE) of the set-membership normalized least mean squares (SM-NLMS) algorithm [8] is analysed in [9, 10] and steady-state MSE of the set-membership affine projection (SM-AP) algorithm [11] is analysed in [12-14]. Analytical results on the steady-state MSE performance of the set-membership binormalized data-reusing LMS algorithm and the multichannel filtered-x set-membership affine projection algorithm are given in [15] and [16], respectively. The work of [4] is also dedicated to prove that, with persistent excitation [17], the QOBE algorithm exhibits central-estimator convergence while its associated hyperellipsoidal membership set cannot converge to a point set. To the best of our knowledge, no steady-state MSE or tracking performance analysis of the QOBE algorithm has been reported in the literature to date.

In this paper, we present a steady-state MSE and tracking performance analysis of the QOBE algorithm. The classic approach is to analyse the transient behavior of the adaptive filter and then obtain the steady-state MSE as a limiting behavior of the transient MSE [18, 19]. However, due to high nonlinearity of the QOBE algorithm, this

---

<sup>4</sup> This algorithm was originally published as the *Bounding Ellipsoidal Adaptive CONstrained least-squares* (BEACON) algorithm [5].

approach would be very difficult. To circumvent the analysis of the transient behaviour and its complications, we employ the energy conservation argument [20, 21] and initiate our steady-state analysis from the energy conservation relation of the QOBE algorithm. For the tracking performance analysis, we consider a random walk model for time variations of the unknown system in a nonstationary environment. The outcome of the analysis is a nonlinear equation whose solution yields the theoretical steady-state MSE of the QOBE algorithm in both stationary and nonstationary environments. Simulations confirm the accuracy of the theoretical predictions for the steady-state MSE in addition to justifying the assumptions made in the derivations.

In section 2, we describe the QOBE algorithm. We analyse its steady-state MSE and tracking performance in sections 3 and 4, provide some simulation results in section 5, and conclude the paper in section 6.

## 2. The QOBE algorithm

Let us consider the affine-in-parameter model

$$d_n = \boldsymbol{\omega}^T \mathbf{x}_n + \eta_n \quad (1)$$

where  $d_n \in \mathbb{R}$  is the reference signal at time index  $n \in \mathbb{N}$ ,  $\boldsymbol{\omega} \in \mathbb{R}^L$  is the column vector of the unknown system parameters,  $\mathbf{x}_n \in \mathbb{R}^L$  is the input vector,  $\eta_n \in \mathbb{R}$  accounts for background noise, and superscript  $T$  denotes matrix/vector transposition.

Conventional filtering schemes produce an estimate of  $\boldsymbol{\omega}$  by minimizing a cost function, which is usually a direct function of the estimation error. On the contrary, set-membership filtering algorithms estimate a set of feasible solutions rather than a single-point estimate by specifying a bound on the magnitude of the estimation error over the model space of interest. The model space,  $\mathcal{S}$ , comprises all the input-reference signal pairs used in the estimation. Hence, any parameter vector,  $\mathbf{w}$ , resulting in an error less than a predetermined bound,  $\gamma > 0$ , for all data pairs from  $\mathcal{S}$  is an acceptable solution. The *feasibility set* contains all these possible solutions and is defined as

$$\Theta = \bigcap_{(\mathbf{x}, d) \in \mathcal{S}} \{\mathbf{w} \in \mathbb{R}^L: |d - \mathbf{w}^T \mathbf{x}| \leq \gamma\}.$$

Direct calculation of  $\Theta$  is often unachievable since, in practice,  $\mathcal{S}$  is not fully known and may be time-varying. Therefore, SMAF algorithms aim to estimate a minimal set estimate of  $\Theta$  at each time instant. This set is called the *membership set* and is defined as

$$\Psi_n = \bigcap_{i=1}^n \Phi_i$$

where  $\Phi_n$  is the observation-induced *constraint set* at time instant  $n$ :

$$\Phi_n = \{\mathbf{w} \in \mathbb{R}^L: |d_n - \mathbf{w}^T \mathbf{x}_n| \leq \gamma\}.$$

Obviously,  $\Theta \subset \Psi_n$ . However,  $\Psi_n$  is a convex polytope in  $\mathcal{S}$  and usually too complex to compute. One solution is to outer-bound  $\Psi_n$  at each iteration by a mathematically tractable ellipsoid  $E_n$ . This is the basic idea behind the OBE algorithms. Given an initial ellipsoid

$$E_0 = \{\mathbf{w} \in \mathbb{R}^L: (\mathbf{w} - \mathbf{w}_0)^T \mathbf{P}_0^{-1} (\mathbf{w} - \mathbf{w}_0) \leq \sigma_0\}$$

with some suitably chosen initial estimates  $\mathbf{w}_0 \in \mathbb{R}^L$ ,  $\mathbf{P}_0^{-1} \in \mathbb{R}^{L \times L}$ , and  $\sigma_0 > 0$ , an OBE algorithm recursively computes  $E_n$  using the knowledge of  $E_{n-1}$  in such a way that

$$E_n \supset (\Phi_n \cap E_{n-1}) \supset \Psi_n \quad \forall n. \quad (2)$$

As  $E_n$  outer-bounds the intersection of  $\Phi_n$  and  $E_{n-1}$ ,  $\Theta$  lies within  $E_n$ . To satisfy (2),  $E_n$  is calculated via

$$E_n = \{\mathbf{w} \in \mathbb{R}^L: (\mathbf{w} - \mathbf{w}_n)^T \mathbf{P}_n^{-1} (\mathbf{w} - \mathbf{w}_n) \leq \sigma_n\}$$

where

$$\mathbf{P}_n^{-1} = \mathbf{P}_{n-1}^{-1} + \lambda_n \mathbf{x}_n \mathbf{x}_n^T, \quad (3)$$

$$e_n = d_n - \mathbf{w}_{n-1}^T \mathbf{x}_n, \quad (4)$$

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \lambda_n \mathbf{P}_n \mathbf{x}_n e_n, \quad (5)$$

$$\sigma_n = \sigma_{n-1} - \frac{\lambda_n e_n^2}{1 + \lambda_n G_n} + \lambda_n \gamma^2,$$

and

$$G_n = \mathbf{x}_n^T \mathbf{P}_{n-1} \mathbf{x}_n.$$

The centroid of  $E_n$ ,  $\mathbf{w}_n$ , can be regarded as a point estimate at time instant  $n$  and  $e_n$  is the *a priori* estimation error. The QOBE (also known as BEACON) algorithm calculates the optimal weighting factor,  $\lambda_n$ , by minimizing  $\sigma_n$  under the constraint that  $\lambda_n > 0$ . The result is a simple but efficient update-checking rule and a relatively simple expression for  $\lambda_n$ :

$$\lambda_n = \begin{cases} \frac{1}{G_n} \left( \frac{|e_n|}{\gamma} - 1 \right) & \text{if } |e_n| > \gamma. \\ 0 & \text{if } |e_n| \leq \gamma \end{cases} \quad (6)$$

In practice,  $\mathbf{P}_n$  is updated rather than  $\mathbf{P}_n^{-1}$  by applying the matrix inversion lemma [22] to (3):

$$\mathbf{P}_n = \mathbf{P}_{n-1} - \frac{\lambda_n \mathbf{P}_{n-1} \mathbf{x}_n \mathbf{x}_n^T \mathbf{P}_{n-1}}{1 + \lambda_n G_n}. \quad (7)$$

### 3. Steady-state mean squared error analysis

Multiplying both sides of (7) by  $\mathbf{x}_n$  from the right and using (6), we can show that

$$\mathbf{P}_n \mathbf{x}_n = \begin{cases} \mathbf{P}_{n-1} \mathbf{x}_n \frac{\gamma}{|e_n|} & \text{if } |e_n| > \gamma. \\ \mathbf{P}_{n-1} \mathbf{x}_n & \text{if } |e_n| \leq \gamma \end{cases}. \quad (8)$$

Substituting (6) and (8) into (5) yields

$$\mathbf{w}_n = \begin{cases} \mathbf{w}_{n-1} + \frac{\mathbf{P}_{n-1} \mathbf{x}_n}{G_n} \left( 1 - \frac{\gamma}{|e_n|} \right) e_n & \text{if } |e_n| > \gamma. \\ \mathbf{w}_{n-1} & \text{if } |e_n| \leq \gamma \end{cases}. \quad (9)$$

Defining

$$f_n = \begin{cases} e_n - \gamma & \text{if } e_n > \gamma \\ 0 & \text{if } -\gamma \leq e_n \leq \gamma, \\ e_n + \gamma & \text{if } e_n < -\gamma \end{cases} \quad (10)$$

we can rewrite (9) as

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \frac{\mathbf{P}_{n-1} \mathbf{x}_n}{G_n} f_n. \quad (11)$$

Subtracting both sides of (11) from  $\boldsymbol{\omega}$  results in

$$\mathbf{v}_n = \mathbf{v}_{n-1} - \frac{\mathbf{P}_{n-1} \mathbf{x}_n}{G_n} f_n \quad (12)$$

where

$$\mathbf{v}_n = \boldsymbol{\omega} - \mathbf{w}_n$$

is the weight-error vector. The noiseless *a priori* and *a posteriori* estimation errors are respectively defined as

$$e_{a,n} = \mathbf{v}_{n-1}^T \mathbf{x}_n \quad (13)$$

and

$$e_{p,n} = \mathbf{v}_n^T \mathbf{x}_n.$$

From (1), (4), and (13), we have

$$e_n = e_{a,n} + \eta_n. \quad (14)$$

Multiplying the transpose of both sides of (12) by  $\mathbf{x}_n$  from the right results in

$$e_{p,n} = e_{a,n} - f_n. \quad (15)$$

By substituting  $f_n = e_{a,n} - e_{p,n}$  into (12), we can verify that

$$\mathbf{v}_n + \frac{\mathbf{P}_{n-1}\mathbf{x}_n}{G_n} e_{a,n} = \mathbf{v}_{n-1} + \frac{\mathbf{P}_{n-1}\mathbf{x}_n}{G_n} e_{p,n}. \quad (16)$$

By evaluating the energies (i.e., the squared Euclidean norms) of both sides of (16), we obtain the following energy conservation relation:

$$\|\mathbf{v}_n\|^2 + \left\| \frac{\mathbf{P}_{n-1}\mathbf{x}_n}{G_n} \right\|^2 e_{a,n}^2 = \|\mathbf{v}_{n-1}\|^2 + \left\| \frac{\mathbf{P}_{n-1}\mathbf{x}_n}{G_n} \right\|^2 e_{p,n}^2. \quad (17)$$

Taking the expectations of both sides of (17), we get

$$E[\|\mathbf{v}_n\|^2] + E\left[\left\| \frac{\mathbf{P}_{n-1}\mathbf{x}_n}{G_n} \right\|^2 e_{a,n}^2\right] = E[\|\mathbf{v}_{n-1}\|^2] + E\left[\left\| \frac{\mathbf{P}_{n-1}\mathbf{x}_n}{G_n} \right\|^2 e_{p,n}^2\right]. \quad (18)$$

Here,  $\|\cdot\|$  and  $E[\cdot]$  denote the Euclidean norm and the expectation operation, respectively.

We know that after convergence, gradient of the mean square deviation approaches zero [4], i.e.,

$$\lim_{n \rightarrow \infty} E[\|\mathbf{v}_{n-1}\|^2] - E[\|\mathbf{v}_n\|^2] \rightarrow 0.$$

Therefore, at the steady state, (18) becomes

$$E\left[\left\| \frac{\mathbf{P}_{n-1}\mathbf{x}_n}{G_n} \right\|^2 e_{a,n}^2\right] = E\left[\left\| \frac{\mathbf{P}_{n-1}\mathbf{x}_n}{G_n} \right\|^2 e_{p,n}^2\right]. \quad (19)$$

In order to make the analysis tractable, let us adopt the following assumptions:

A1: At the steady state,  $\|\mathbf{P}_{n-1}\mathbf{x}_n/G_n\|^2$  is statistically independent of both  $e_{a,n}^2$  and  $e_{p,n}^2$ .

A2: At the steady state, the noiseless *a priori* estimation error,  $e_{a,n}$ , is white Gaussian with zero mean and variance of  $\sigma_a^2 = E[e_{a,n}^2]$ .

A3: The background noise,  $\eta_n$ , is white Gaussian with zero mean and variance of  $\sigma_\eta^2 = E[\eta_n^2]$ . It is also independent of the input data.

Using A1, (19) simplifies to

$$E[e_{a,n}^2] = E[e_{p,n}^2]. \quad (20)$$

Substituting (15) into (20), we find that at the steady state

$$E[f_n^2 - 2e_{a,n}f_n] = 0. \quad (21)$$

Considering (14), A2 and A3 imply that, at the steady state, the *a priori* estimation error,  $e_n$ , is also zero-mean white Gaussian. In addition, the variance of  $e_n$  at the steady-state is computed as

$$\sigma^2 = E[e_n^2] = \sigma_a^2 + \sigma_\eta^2. \quad (22)$$

Based on this fact and using (10), we can write

$$\begin{aligned} E[f_n^2] &= \int_{-\infty}^{-\gamma} (x + \gamma)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx + \int_{\gamma}^{\infty} (x - \gamma)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx \\ &= (\sigma^2 + \gamma^2) \left[ 1 - \operatorname{erf}\left(\frac{\gamma}{\sqrt{2\sigma^2}}\right) \right] - \gamma \sqrt{\frac{2\sigma^2}{\pi}} e^{-\frac{\gamma^2}{2\sigma^2}} \end{aligned} \quad (23)$$

where  $\operatorname{erf}(\cdot)$  is the error function and is defined as

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

Under A2 and A3, we can also verify that, at the steady state,  $e_{a,n}$  and  $e_n$  are jointly Gaussian and the random variable

$$\zeta = e_{a,n} - \frac{\sigma_a^2}{\sigma^2} e_n \quad (24)$$

is independent of  $e_n$  and thus  $f_n$ . Consequently, using (22) and (24), we can show that

$$\begin{aligned} E[e_{a,n}f_n] &= E[\zeta f_n] + \frac{\sigma^2 - \sigma_\eta^2}{\sigma^2} E[e_n f_n] \\ &= E[\zeta]E[f_n] + \frac{\sigma^2 - \sigma_\eta^2}{\sigma^2} E[e_n f_n] \\ &= \frac{\sigma^2 - \sigma_\eta^2}{\sigma^2} E[e_n f_n]. \end{aligned} \quad (25)$$

Besides, we have

$$\begin{aligned}
E[e_n f_n] &= \int_{-\infty}^{-\gamma} x(x + \gamma) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx + \int_{\gamma}^{\infty} x(x - \gamma) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx \\
&= \sigma^2 \left[ 1 - \operatorname{erf}\left(\frac{\gamma}{\sqrt{2\sigma^2}}\right) \right].
\end{aligned} \tag{26}$$

Hence, by substituting (26) into (25), we get

$$E[e_{a,n} f_n] = (\sigma^2 - \sigma_\eta^2) \left[ 1 - \operatorname{erf}\left(\frac{\gamma}{\sqrt{2\sigma^2}}\right) \right]. \tag{27}$$

Substitution of (23) and (27) into (21) leads to

$$(-\sigma^2 + \gamma^2 + 2\sigma_\eta^2) \left[ 1 - \operatorname{erf}\left(\frac{\gamma}{\sqrt{2\sigma^2}}\right) \right] - \gamma \sqrt{\frac{2\sigma^2}{\pi}} e^{-\frac{\gamma^2}{2\sigma^2}} = 0. \tag{28}$$

The steady-state MSE,  $\sigma^2$ , is found as a unique positive solution of (28). This equation is the same as the one given in [9] for the relaxed and regularized SM-NLMS algorithm when its relaxation and regularization parameters are respectively set to one and zero. This resemblance can be interpreted in view of the intrinsic similarity of the SM-NLMS and the QOBE algorithms, especially in the sense of steady-state behavior, and the parallel nature of the assumptions made in the analyses, i.e., A1 here and S2-S4 in [9]. Although existence and uniqueness of the solution of (28) has been proved in [9], we provide a more immediate and less complicated proof in the Appendix.

Once the steady-state MSE is found, the probability of update (PU) at the steady state can also be computed as

$$\begin{aligned}
P_u &= P(|e_n| > \gamma) \\
&= 1 - \operatorname{erf}\left(\frac{\gamma}{\sqrt{2\sigma^2}}\right).
\end{aligned} \tag{29}$$

#### 4. Tracking analysis

To analyse the tracking performance of the QOBE algorithm in a nonstationary environment, we change the model of (1) to

$$d_n = \boldsymbol{\omega}_n^T \mathbf{x}_n + \eta_n \tag{30}$$

where the unknown system is considered time-varying. To model the variation of  $\boldsymbol{\omega}_n$ , we use the well-known random walk model [23], i.e.,

$$\boldsymbol{\omega}_n = \boldsymbol{\omega}_{n-1} + \mathbf{q}_n \tag{31}$$

where  $\mathbf{q}_n \in \mathbb{R}^L$  denotes the random perturbation. By redefining the weight-error vector as

$$\mathbf{v}_n = \boldsymbol{\omega}_n - \mathbf{w}_n$$

and the noiseless *a posteriori* error as

$$e_{p,n} = (\mathbf{v}_n - \mathbf{q}_n)^T \mathbf{x}_n,$$

we can rewrite (12) as

$$\mathbf{v}_n = \mathbf{v}_{n-1} - \frac{\mathbf{P}_{n-1} \mathbf{x}_n}{G_n} f_n + \mathbf{q}_n. \quad (32)$$

Multiplying the transpose of both sides of (32) by  $\mathbf{x}_n$  from the right, we can verify that (15) still holds and (16) changes to

$$\mathbf{v}_n - \mathbf{q}_n + \frac{\mathbf{P}_{n-1} \mathbf{x}_n}{G_n} e_{a,n} = \mathbf{v}_{n-1} + \frac{\mathbf{P}_{n-1} \mathbf{x}_n}{G_n} e_{p,n}. \quad (33)$$

Evaluating the energies of both sides of (33) gives the following energy conservation relation:

$$\|\mathbf{v}_n - \mathbf{q}_n\|^2 + \left\| \frac{\mathbf{P}_{n-1} \mathbf{x}_n}{G_n} \right\|^2 e_{a,n}^2 = \|\mathbf{v}_{n-1}\|^2 + \left\| \frac{\mathbf{P}_{n-1} \mathbf{x}_n}{G_n} \right\|^2 e_{p,n}^2. \quad (34)$$

Taking expectations of both sides of (34), we get

$$E[\|\mathbf{v}_n - \mathbf{q}_n\|^2] + E\left[\left\| \frac{\mathbf{P}_{n-1} \mathbf{x}_n}{G_n} \right\|^2 e_{a,n}^2\right] = E[\|\mathbf{v}_{n-1}\|^2] + E\left[\left\| \frac{\mathbf{P}_{n-1} \mathbf{x}_n}{G_n} \right\|^2 e_{p,n}^2\right]. \quad (35)$$

For tractability of the tracking analysis, let us adopt the following additional assumptions:

A4: The perturbation  $\mathbf{q}_n$  is independent and identically distributed (i.i.d.) with zero mean and covariance matrix of  $\mathbf{Q} = E[\mathbf{q}_n \mathbf{q}_n^T]$ . It is also independent of the input data and the noise.

A5: At the steady state, the inverse of the input autocorrelation matrix can be approximated as  $\mathbf{P}_n \cong p_n \mathbf{I}_L$  where  $\mathbf{I}_L \in \mathbb{R}^{L \times L}$  is the identity matrix.

A6: The stochastic process  $\|\mathbf{x}_n\|^{-2}$  is stationary and the initial input,  $\mathbf{x}_0$ , is a nonzero vector that satisfies  $E[\|\mathbf{x}_0\|^{-2}] < \infty$ .

Under A4, we have

$$\begin{aligned} E[\|\mathbf{v}_n - \mathbf{q}_n\|^2] &= E[\|\mathbf{v}_n\|^2] - 2E[\mathbf{q}_n^T \mathbf{v}_n] + E[\|\mathbf{q}_n\|^2] \\ &= E[\|\mathbf{v}_n\|^2] - 2E[\mathbf{q}_n^T (\boldsymbol{\omega}_n - \mathbf{w}_n)] + E[\|\mathbf{q}_n\|^2] \\ &= E[\|\mathbf{v}_n\|^2] - 2E[\mathbf{q}_n^T (\boldsymbol{\omega}_{n-1} + \mathbf{q}_n - \mathbf{w}_n)] + E[\|\mathbf{q}_n\|^2] \\ &= E[\|\mathbf{v}_n\|^2] - 2E[\|\mathbf{q}_n\|^2] + E[\|\mathbf{q}_n\|^2] \\ &= E[\|\mathbf{v}_n\|^2] - \text{Tr}(\mathbf{Q}) \end{aligned} \quad (36)$$

where  $\text{Tr}(\cdot)$  denotes the matrix trace.

Assuming convergence and using A1, (15), (35), and (36), we can verify that at the steady state

$$E[f_n^2 - 2e_{a,n}f_n] + \frac{\text{Tr}(\mathbf{Q})}{E\left[\left\|\frac{\mathbf{P}_{n-1}\mathbf{x}_n}{G_n}\right\|^2\right]} = 0. \quad (37)$$

In view of A5, we can make the following approximation at the steady state

$$\left\|\frac{\mathbf{P}_{n-1}\mathbf{x}_n}{G_n}\right\|^2 \cong \frac{1}{\|\mathbf{x}_n\|^2}. \quad (38)$$

Substituting (23), (27), and (38) into (37) together with using A6 gives

$$(-\sigma^2 + \gamma^2 + 2\sigma_\eta^2) \left[1 - \text{erf}\left(\frac{\gamma}{\sqrt{2\sigma^2}}\right)\right] - \gamma \sqrt{\frac{2\sigma^2}{\pi}} e^{-\gamma^2/2\sigma^2} + \frac{\text{Tr}(\mathbf{Q})}{E[\|\mathbf{x}_0\|^{-2}]} = 0. \quad (39)$$

The solution of this equation yields the steady-state MSE of the QOBE algorithm when the underlying system is nonstationary and modelled via (30) and (31). Existence and uniqueness of the solution of (39) is proved in the Appendix. It is clear that (39) is a generalized form of (28) and can also be used in a stationary environment by simply setting  $\text{Tr}(\mathbf{Q})$  to zero.

## 5. Numerical studies

### 5.1. Stationary case

We consider a system identification scenario where the unknown system has  $L = 16$  taps. The taps are generated randomly and normalized to unit energy. The corresponding adaptive filter is also assumed to have the same number of taps. The input vector is shift-structured, i.e.,

$$\mathbf{x}_n = [x_n, x_{n-1}, \dots, x_{n-L+1}]^T$$

where  $x_n$  is generated by passing a zero-mean Gaussian signal through a first-order autoregressive process with transfer function

$$\frac{1}{1 - 0.95z^{-1}}.$$

The power of the additive white Gaussian noise is  $\sigma_\eta^2 = 0.001$ .

In Figs. 1 and 2, we compare the simulated and predicted steady-state MSEs and PUs of the QOBE algorithm for different values of  $\gamma^2/\sigma_\eta^2$ . The simulated results are obtained by averaging the instantaneous values over 100 time instants at the steady state and ensemble-averaging over  $10^3$  independent trials. The theoretical results are

achieved by solving (28) and using (29). Figs. 1 and 2 exhibit a good agreement between the experiment and the theory.

We also investigate the validity of A1 by plotting

$$\rho_{a,n} = \frac{E \left[ \left\| \frac{\mathbf{P}_{n-1} \mathbf{x}_n}{G_n} \right\|^2 e_{a,n}^2 \right]}{E \left[ \left\| \frac{\mathbf{P}_{n-1} \mathbf{x}_n}{G_n} \right\|^2 \right] E[e_{a,n}^2]}$$

and

$$\rho_{p,n} = \frac{E \left[ \left\| \frac{\mathbf{P}_{n-1} \mathbf{x}_n}{G_n} \right\|^2 e_{p,n}^2 \right]}{E \left[ \left\| \frac{\mathbf{P}_{n-1} \mathbf{x}_n}{G_n} \right\|^2 \right] E[e_{p,n}^2]}$$

versus time in Fig. 3. The expectations are approximated by ensemble-averaging over  $10^4$  independent trials when  $\gamma^2/\sigma_\eta^2 = 1$ . We observe that both  $\rho_{a,n}$  and  $\rho_{p,n}$  are close to one at the steady state. This test together with the results given in Figs. 1 and 2 justifies A1 as a reasonable assumption.

## 5.2. Nonstationary case

We consider estimation and tracking of a flat-fading multiple-input single-output (MISO) wireless channel with  $N_t = 4$  transmitter antennas. Sub-channels between each transmitter antenna and the receiver fade independently according to the random walk model. Thus, the channel vector,  $\mathbf{h}_n \in \mathbb{R}^{N_t}$ , is modeled via

$$\mathbf{h}_n = \mathbf{h}_{n-1} + \alpha \mathbf{k}_n$$

where  $\alpha$  is a parameter that controls the rate of the variations and  $\mathbf{k}_n \in \mathbb{R}^{N_t}$  is a circular Gaussian random vector with zero mean and identity covariance matrix. The channel estimator is an adaptive filter with  $L = 4$  taps. The transmitted symbols (input to the channel estimator) are i.i.d. and modulated using the BPSK scheme with a modulus of  $1/\sqrt{N_t} = 1/2$ .

In Figs. 4-7, we depict the experimental and theoretical steady-state MSEs and PUs as a function of  $\gamma^2/\sigma_\eta^2$  for different values of  $\alpha$  and two values of the additive white noise power, i.e.,  $\sigma_\eta^2 = 0.01$  and  $\sigma_\eta^2 = 0.001$ . The experimental steady-state MSEs and PUs are obtained by averaging 100 steady-state values and ensemble-averaging over  $10^3$  independent runs. The theoretical steady-state MSEs and PUs are calculated by solving (39) and using (29) where the last term of the left-hand side of (39) simplifies to

$$\frac{\text{Tr}(E[\alpha \mathbf{k}_n \alpha \mathbf{k}_n^T])}{E[\|\mathbf{x}_0\|^{-2}]} = N_t \alpha^2.$$

Figs. 4-7 show that the theoretical results are in good match with the simulations.

## 6. Concluding remarks

*Assumptions:* Despite seeming unnatural, A1 is reasonably realistic and simplifies the analysis tremendously. Several similar assumptions have been made in the literature (see, e.g., [9, 18, 21, 23]). Without utilizing A2, calculation of the expectations in (21) would be arduous. This assumption is commonly made to deal with error nonlinearities (see, e.g., [7], [9]). It is justified for long adaptive filters via central limit arguments [24]. A3 is natural and regularly used while A4 is typical in the context of tracking performance analysis of the adaptive filters (see, e.g., [21], [25]). A5 is necessary to make the analysis tractable. Such assumptions are frequently used in the MSE performance analysis of the adaptive filters (see, e.g., [26]). A6 is also required for tractability of the analysis. However, assuming a stationary  $\|\mathbf{x}_n\|^{-2}$  is not as strict as assuming a stationary input. Therefore, our analysis is valid for any input signal as long as  $\|\mathbf{x}_n\|^{-2}$  is stationary.

*Input signal correlation:* In the numerical studies of the stationary case, we provided a comparison between the experimental and theoretical steady-state MSEs for a typical correlated input signal. Extensive investigations showed that the theoretical predictions are in good agreement with experiments for a wide range of correlated and uncorrelated input signals.

*Simulations of the nonstationary case:* The primary objective of the provided numerical results is to corroborate our theoretical findings. That is why the simulated MISO wireless channel varies in time according to the random walk model, which is the assumed model in the analysis. Admittedly, this is not the most appropriate way to model variations of a wireless channel. However, it is well established that a Rayleigh-fading channel can be suitably modelled by a first-order autoregressive model [27]. For slow variations, the corresponding autoregressive model can be approximated by a random walk model [23].

*Accuracy of the analysis:* Comparison with the simulation results shows that the analytical predictions for the steady-state MSE are less accurate when  $\gamma^2/\sigma_\eta^2$  is considerably small. Nonetheless, in the presented comparisons, the largest gap between the experimental and the theoretical MSEs is still only about one dB. This lesser accuracy for the low values of  $\gamma^2/\sigma_\eta^2$  can mainly be attributed to the fact that the assumptions employed in the analysis, particularly A1 and A2, become less tenable in such cases. However, it is known that the practically appropriate range of

values is  $\gamma^2/\sigma_\eta^2 \geq 1$  and a value of  $\gamma^2/\sigma_\eta^2 < 1$  cannot usually be a good choice in practice [19].

*Summary:* Steady-state MSE and tracking performance of the QOBE algorithm was analysed utilizing the energy conservation argument. The analysis resulted in a nonlinear equation that by solving it, the theoretical steady-state MSE of the QOBE algorithm for both stationary and nonstationary systems is calculated. Existence and uniqueness of the solution of this equation was also proved. Simulations showed that the theoretical predictions match the experimental ones well.

## Appendix

Introducing the change of variable

$$x = \frac{\gamma}{\sqrt{2\sigma^2}}$$

and defining

$$a = 1 + \frac{2\sigma_\eta^2}{\gamma^2}$$

and

$$b = \frac{\text{Tr}(\mathbf{Q})}{E[\|\mathbf{x}_0\|^{-2}]}$$

we can rewrite (39) as

$$ax^2 - \frac{1}{2} + \frac{b}{\gamma^2} \frac{x^2}{1 - \text{erf}(x)} = \frac{1}{\sqrt{\pi}} \frac{xe^{-x^2}}{1 - \text{erf}(x)}. \quad (\text{A1})$$

*Lemma 1:* Let

$$Y(x) = ax^2 - \frac{1}{2} + \frac{b}{\gamma^2} \frac{x^2}{1 - \text{erf}(x)} \quad \text{for } x > 0. \quad (\text{A2})$$

Then, the following holds:

- 1)  $\lim_{x \rightarrow 0} Y(x) = -1/2$ ;
- 2)  $\lim_{x \rightarrow \infty} Y(x) = \infty$ ;
- 3)  $Y(x)$  is monotone increasing.

*Proof:*

- 1) This is obvious.

2) This can be easily verified since  $a \geq 1$ ,  $b \geq 0$ ,  $1 - \text{erf}(x) > 0$  for  $x > 0$ , and  $\lim_{x \rightarrow \infty} [1 - \text{erf}(x)] = 0$ .

3) This is also easy to verify since, for  $x > 0$ ,  $ax^2 - \frac{1}{2}$  is monotone increasing and  $1 - \text{erf}(x)$  is monotone decreasing in  $(0,1]$  so

$$\frac{b}{\gamma^2} \frac{x^2}{1 - \text{erf}(x)}$$

is also monotone increasing. ■

*Lemma 2:* Let

$$Z(x) = \frac{1}{\sqrt{\pi}} \frac{xe^{-x^2}}{1 - \text{erf}(x)} \quad \text{for } x > 0. \quad (\text{A3})$$

Then, the following holds:

- 1)  $\lim_{x \rightarrow 0} Z(x) = 0$ ;
- 2)  $\lim_{x \rightarrow \infty} Z(x) = \infty$ ;
- 3)  $Z(x)$  is monotone increasing.

*Proof:*

- 1) This is obvious.
- 2) Using L'Hôpital's rule, we have

$$\begin{aligned} \lim_{x \rightarrow \infty} Z(x) &= \frac{1}{\sqrt{\pi}} \lim_{x \rightarrow \infty} \frac{\frac{d}{dx} [xe^{-x^2}]}{\frac{d}{dx} [1 - \text{erf}(x)]} \\ &= \frac{1}{\sqrt{\pi}} \lim_{x \rightarrow \infty} \frac{-(2x^2 - 1)e^{-x^2}}{-\frac{2}{\sqrt{\pi}}e^{-x^2}} \\ &= \infty. \end{aligned}$$

3) Let us define

$$\Omega(x) = \frac{e^{-x^2}}{1 - \text{erf}(x)}$$

and calculate its derivative

$$\frac{d\Omega(x)}{dx} = \frac{2/\sqrt{\pi} - 2xe^{x^2}[1 - \text{erf}(x)]}{e^{2x^2}[1 - \text{erf}(x)]^2}.$$

Invoking Gordon's inequality [28, 29], i.e.,

$$\sqrt{\pi}e^{x^2}[1 - \operatorname{erf}(x)] < \frac{1}{x} \quad \text{for } x > 0,$$

we can show that  $d\Omega(x)/dx > 0$  for  $x > 0$ . This means  $\Omega(x)$  and subsequently  $Z(x) = \frac{x}{\sqrt{\pi}}\Omega(x)$  are monotone increasing for  $x > 0$ . ■

*Lemma 3:* Given  $Y(x)$  and  $Z(x)$  defined by (A2) and (A3), for  $b = 0$

$$\lim_{x \rightarrow \infty} \frac{Y(x)}{Z(x)} = a$$

and for  $b > 0$

$$\lim_{x \rightarrow \infty} \frac{Y(x)}{Z(x)} = \infty.$$

*Proof:*

We can write

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{Y(x)}{Z(x)} &= \lim_{x \rightarrow \infty} \frac{\left(ax^2 - \frac{1}{2}\right)[1 - \operatorname{erf}(x)] + \frac{b}{\gamma^2}x^2}{\frac{1}{\sqrt{\pi}}xe^{-x^2}} \\ &= \sqrt{\pi} \left( a \lim_{x \rightarrow \infty} \frac{x[1 - \operatorname{erf}(x)]}{e^{-x^2}} - \frac{1}{2} \lim_{x \rightarrow \infty} \frac{1 - \operatorname{erf}(x)}{xe^{-x^2}} + \frac{b}{\gamma^2} \lim_{x \rightarrow \infty} xe^{x^2} \right). \end{aligned} \quad (\text{A4})$$

Using L'Hôpital's rule, we have

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{x[1 - \operatorname{erf}(x)]}{e^{-x^2}} &= \lim_{x \rightarrow \infty} \frac{\frac{d}{dx}[x - x \operatorname{erf}(x)]}{\frac{d}{dx}[e^{-x^2}]} \\ &= \lim_{x \rightarrow \infty} \frac{1 - \operatorname{erf}(x) - \frac{2}{\sqrt{\pi}}xe^{-x^2}}{-2xe^{-x^2}} \\ &= \frac{1}{\sqrt{\pi}} \end{aligned} \quad (\text{A5})$$

and

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{1 - \operatorname{erf}(x)}{xe^{-x^2}} &= \lim_{x \rightarrow \infty} \frac{\frac{d}{dx}[1 - \operatorname{erf}(x)]}{\frac{d}{dx}[xe^{-x^2}]} \\ &= \lim_{x \rightarrow \infty} \frac{-\frac{2}{\sqrt{\pi}}e^{-x^2}}{-(2x^2 - 1)e^{-x^2}} \\ &= 0. \end{aligned} \quad (\text{A6})$$

Substituting (A5) and (A6) into (A4) results in

$$\begin{aligned}\lim_{x \rightarrow \infty} \frac{Y(x)}{Z(x)} &= a + \sqrt{\pi} \frac{b}{\gamma^2} \lim_{x \rightarrow \infty} x e^{x^2} \\ &= \begin{cases} a & \text{if } b = 0 \\ \infty & \text{if } b > 0 \end{cases} \quad \blacksquare\end{aligned}$$

From Lemmas 1-3 and the fact that  $a \geq 1$ , one can deduce that the equation  $Y(x) = Z(x)$ , i.e., (A1), surely has a unique positive solution.

In Fig. 8,  $Y(x)$  and  $Z(x)$  are plotted for  $\sigma_\eta^2 = 0.01$ ,  $\gamma^2/\sigma_\eta^2 = 4$ , and different values of  $b$ . It is observed that a larger  $b$  (higher degree of nonstationarity) leads to a lower  $x = \gamma/\sqrt{2\sigma^2}$  and consequently a higher  $\sigma^2$ .

## References

- [1] J. R. Deller, Jr. and Y. F. Huang, "Set-membership identification and filtering for signal processing applications," *Circuits Systems Signal Process.*, vol. 21, no. 1, pp. 69-82, 2002.
- [2] J. R. Deller, Jr., M. Nayeri, and S. F. Odeh, "Least-square identification with error bounds for real-time signal processing and control," *Proc. IEEE*, vol. 81, pp. 813-849, Jun. 1993.
- [3] J. P. Norton, Ed., Special issues on "Bounded-error methods in system identification," *Int. J. Automat. Contr. Signal Process.*, vol. 8, Jan./Feb. 1994; vol. 9, Jan./Feb. 1995.
- [4] J. R. Deller, Jr., S. Gollamudi, S. Nagaraj, D. Joachim, and Y. F. Huang, "Convergence analysis of the quasi-OBE algorithm and related performance issues," *Int. J. Adapt. Control Signal Process.*, vol. 21, pp. 499-527, 2007.
- [5] S. Nagaraj, S. Gollamudi, S. Kapoor, and Y. F. Huang, "BEACON: An adaptive set-membership filtering technique with sparse updates," *IEEE Trans. Signal Process.*, vol. 47, no. 11, pp. 2928-2941, 1999.
- [6] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of data-normalized adaptive filters," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 639-652, Mar. 2003.
- [7] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of adaptive filters with error nonlinearities," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 653-663, Mar. 2003.

- [8] S. Gollamudi, S. Nagaraj, S. Kapoor, and Y. F. Huang, "Set-membership filtering and a set-membership normalized LMS algorithm with an adaptive step size," *IEEE Signal Process. Lett.*, vol. 5, no. 5, pp. 111–114, May 1998.
- [9] N. Takahashi and I. Yamada, "Steady-state mean-square performance analysis of a relaxed set-membership NLMS algorithm by the energy conservation argument," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3361–3372, 2009.
- [10] N. Takahashi and I. Yamada, "Steady-state performance of hyperslab projection algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Las Vegas, USA, Apr. 2008, pp. 3829–3832.
- [11] S. Werner and P. S. R. Diniz, "Set-membership affine projection algorithm," *IEEE Signal Process. Lett.*, vol. 8, no. 8, pp. 231–235, Aug. 2001.
- [12] P. S. R. Diniz, "Convergence performance of the simplified set-membership affine projection algorithm," *Circuits Systems Signal Process.*, vol. 30, pp. 439–462, 2011.
- [13] M. V. S. Lima and P. S. R. Diniz, "Steady-state analysis of the set-membership affine projection algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Dallas, USA, Mar. 2010, pp. 3802–3805.
- [14] P. S. R. Diniz, "Analysis of a set-membership affine projection algorithm in nonstationary environment," in *Proc. European Signal Process. Conf.*, Glasgow, Scotland, Aug. 2009, pp. 2623–2627.
- [15] P. S. R. Diniz, S. Werner, "Set-membership binormalized data-reusing LMS algorithms," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 124–134, Jan. 2003.
- [16] A. Carini, G. L. Sicuranza, "Analysis of a multichannel filtered-x set-membership affine projection algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Toulouse, France, May 2006, pp. 193–196.
- [17] R. R. Bitmead, "Persistence of excitation conditions and the convergence of adaptive schemes," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 2, Mar. 1984.
- [18] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2002.
- [19] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, 3rd ed. New York: Springer, 2008.
- [20] M. Rupp and A. H. Sayed, "A time-domain feedback analysis of filtered-error adaptive gradient algorithms," *IEEE Trans. Signal Process.*, vol. 44, pp. 1428–1439, Jun. 1996.

- [21] N. R. Yousef and A. H. Sayed, "A unified approach to the steady-state and tracking analyses of adaptive filters," *IEEE Trans. Signal Process.*, vol. 49, no. 2, pp. 314–324, Feb. 2001.
- [22] G. H. Golub and C. F. Van Loan, *Matrix Computations*, third ed., Baltimore, MD: Johns Hopkins Univ. Press, 1996.
- [23] A. H. Sayed, *Adaptive Filters*, Hoboken, NJ: Wiley, 2008.
- [24] D. L. Duttweiler, "Adaptive filter performance with nonlinearities in the correlation multiplier," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 30, pp. 578–586, Aug. 1982.
- [25] E. Eweda, "Comparison of RLS, LMS, and sign algorithms for tracking randomly time-varying channels," *IEEE Trans. Signal Process.*, vol. 42, pp. 2937-2944, Nov. 1994.
- [26] A. W. H. Khong and P. A. Naylor, "Selective-tap adaptive filtering with performance analysis for identification of time-varying systems," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 15, no. 5, pp. 1681–1695, Jul. 2007.
- [27] C. Komninakis, C. Fragouli, A. H. Sayed, and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1065–1076, May 2002.
- [28] R. D. Gordon, "Values of Mills' ratio of area bounding ordinate and of the normal probability integral for large values of the argument," *Ann. Math. Statistics*, vol. 12, pp. 364–366, 1941.
- [29] Á Baricz, "Mills' ratio: Monotonicity patterns and functional inequalities," *J. Math. Anal. Appl.*, vol. 340, no. 2, pp. 1362–1370, 2008.

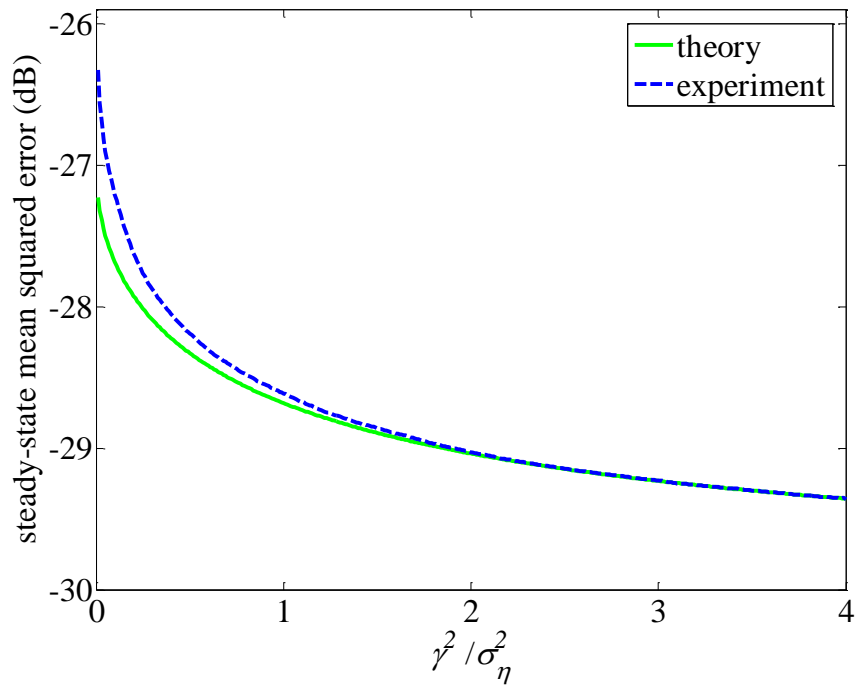


Fig. 1, Comparing experimental and theoretical steady-state MSEs of the QOBE algorithm in a stationary environment.

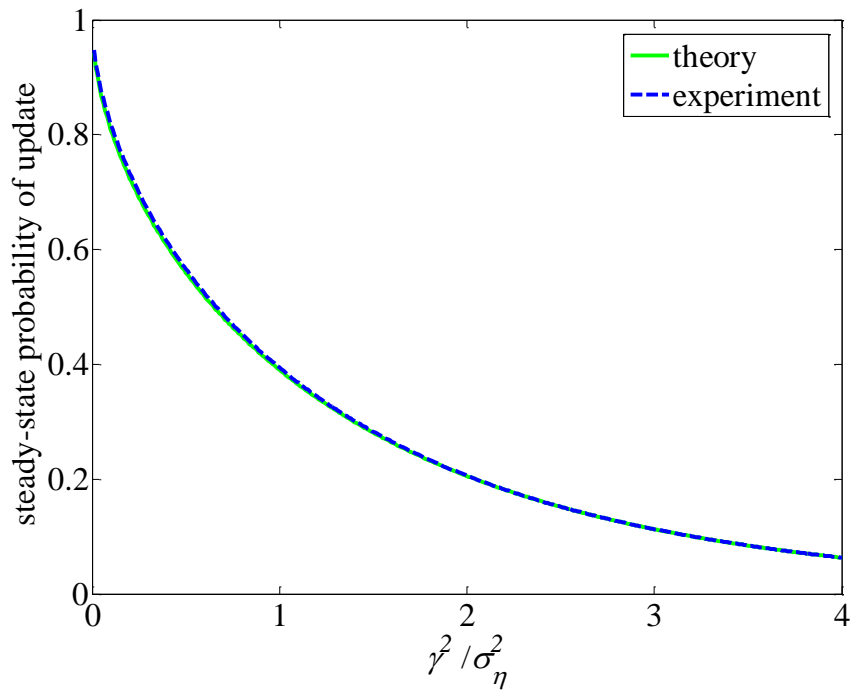


Fig. 2, Comparing experimental and theoretical steady-state PUs of the QOBE algorithm in a stationary environment.

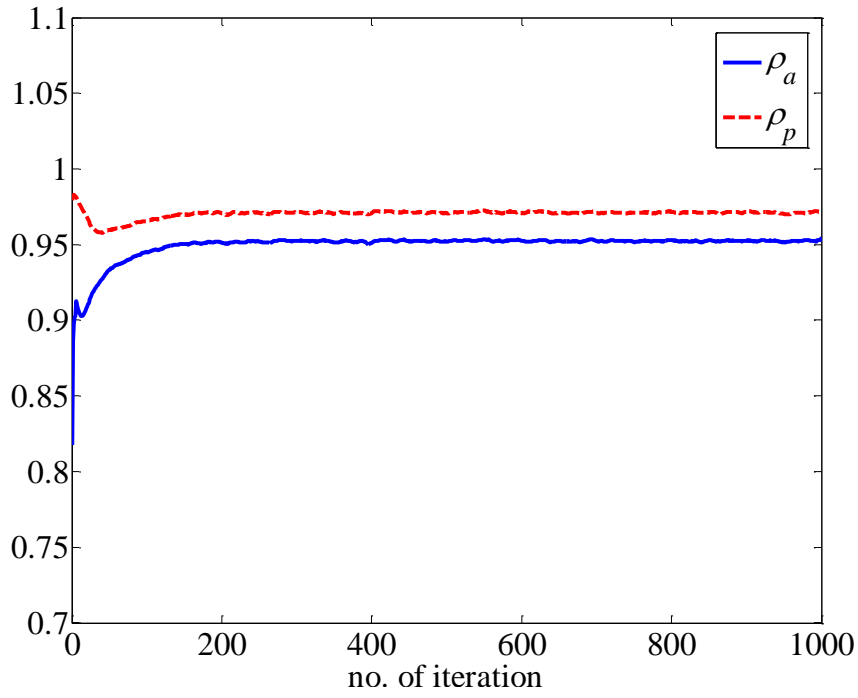


Fig. 3, Testing the validity of A1.

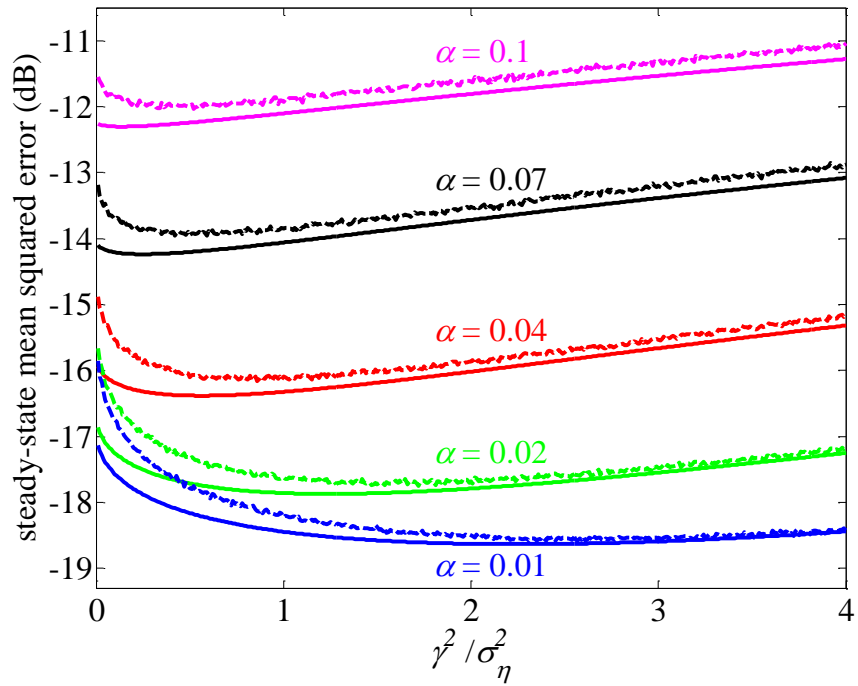


Fig. 4, Comparing experimental (dashed lines) and theoretical (solid lines) steady-state MSEs of the QOBE algorithm in nonstationary environments with  $\sigma_\eta^2 = 0.01$  and different values of  $\alpha$ .

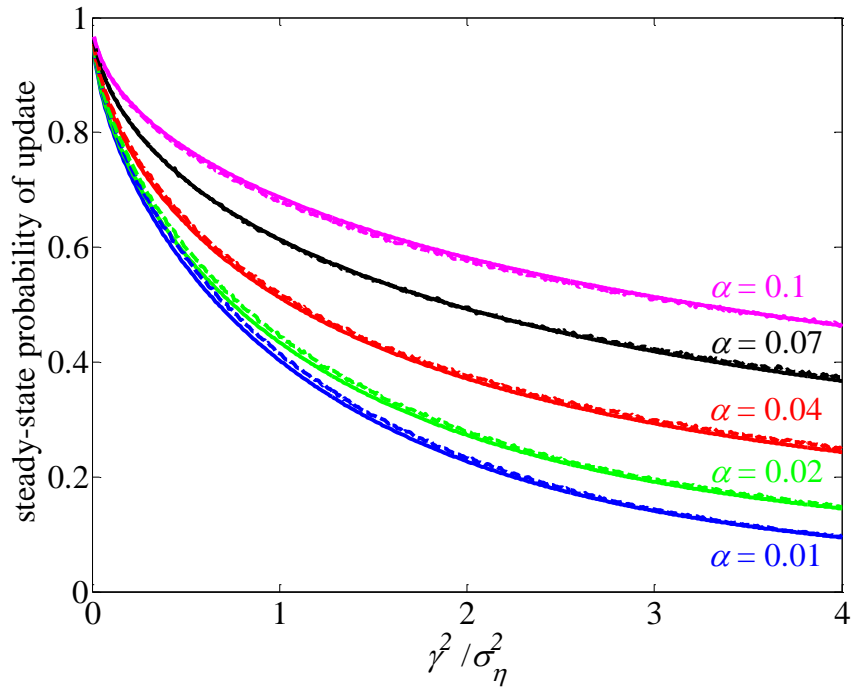


Fig. 5, Comparing experimental (dashed lines) and theoretical (solid lines) steady-state PUs of the QOBE algorithm in nonstationary environments with  $\sigma_\eta^2 = 0.01$  and different values of  $\alpha$ .

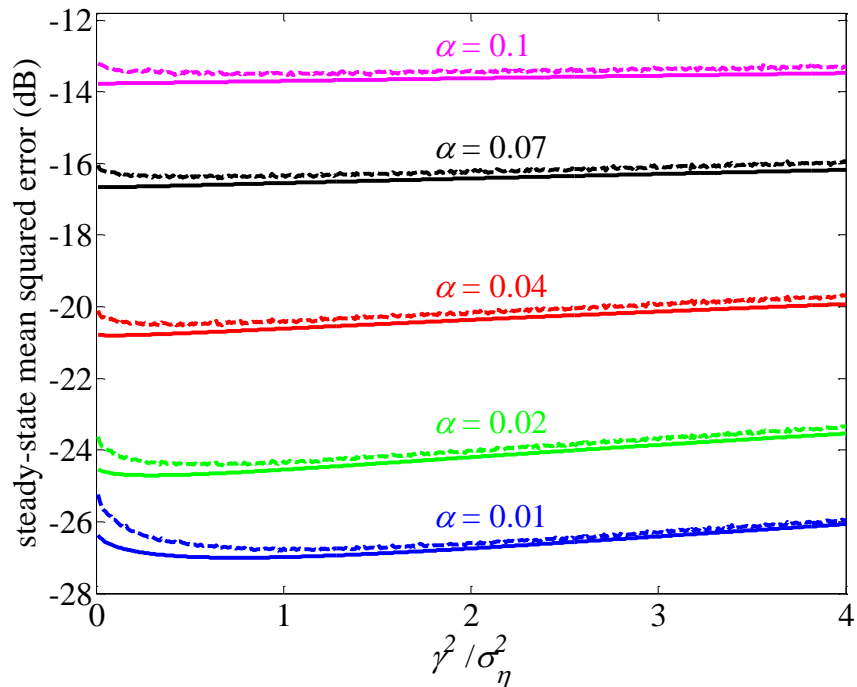


Fig. 6, Comparing experimental (dashed lines) and theoretical (solid lines) steady-state MSEs of the QOBE algorithm in nonstationary environments with  $\sigma_\eta^2 = 0.001$  and different values of  $\alpha$ .

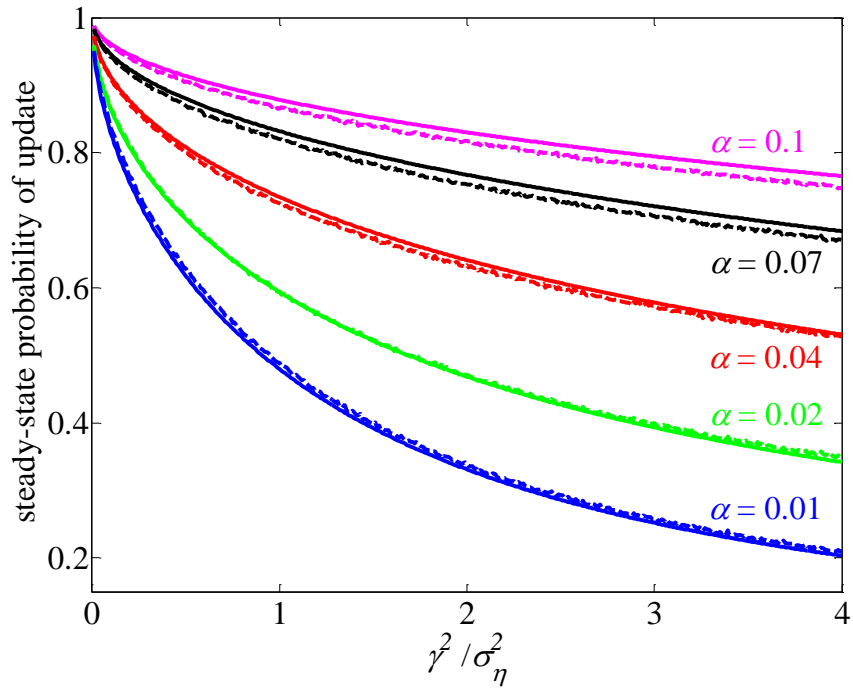


Fig. 7, Comparing experimental (dashed lines) and theoretical (solid lines) steady-state PUs of the QOBE algorithm in nonstationary environments with  $\sigma_\eta^2 = 0.001$  and different values of  $\alpha$ .

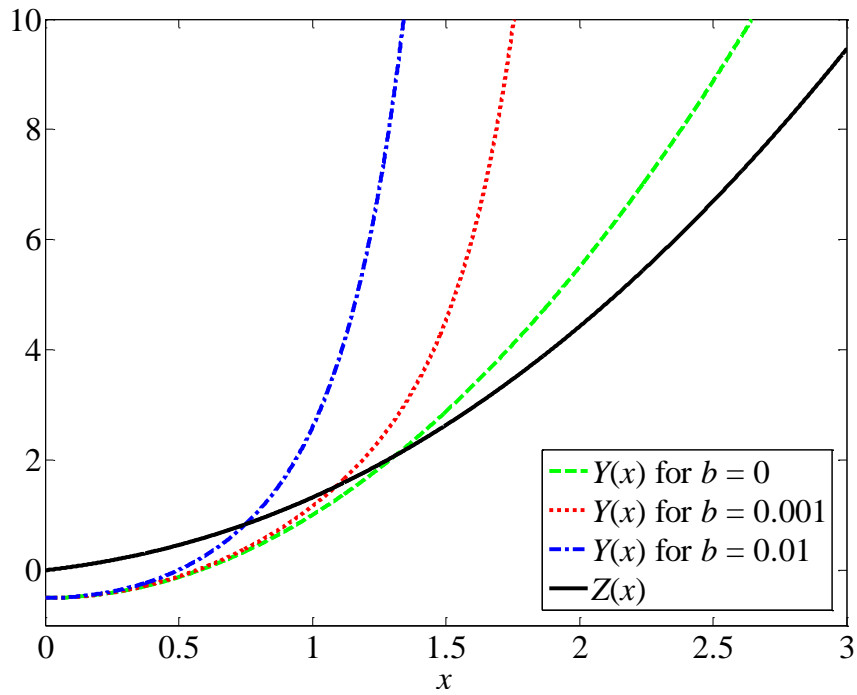


Fig. 8, The functions  $Z(x)$  and  $Y(x)$  for different values of  $b$ .

# Paper J

Originally published as

R. Arablouei and K. Doğançay, "Tracking performance analysis of the set-membership NLMS algorithm," in *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, Hollywood, USA, Dec. 2012, paper id: 200.

Copyright © 2012 APSIPA. Reprinted with permission.

## Tracking performance analysis of the set-membership NLMS adaptive filtering algorithm

Reza Arablouei and Kutluyıl Doğançay

### Abstract

In this paper, we analyze the tracking performance of the set-membership normalized least mean squares (SM-NLMS) adaptive filtering algorithm using the energy conservation argument. The analysis leads to a nonlinear equation whose solution gives the steady-state mean squared error (MSE) of the SM-NLMS algorithm in a nonstationary environment. We prove that there is always a unique positive solution for this equation. The results predicted by the analysis show good agreement with the simulation experiments.

### 1. Introduction

Set-membership (SM) adaptive filtering algorithms are set-theoretic techniques, which estimate the members of a *set* of solutions that are consistent with observations and *a priori* signal or model constraints. Most SM adaptive filtering algorithms have been developed assuming a *bounded-error* constraint where the *a priori* knowledge emerges as a magnitude bound on the estimation error. The capacity of incorporating the *a priori* information in the estimation process has endowed the SM-based approaches with several advantages over their more classical counterparts in the realm of adaptive signal processing. Superior tracking capabilities and reduced average computational complexity are possibly the most notable ones. The former is achieved because every SM adaptive filtering algorithm executes a form of set-theoretic optimization that allows faster readjustment to the changes in the

underlying system. The latter is also due to the inherent data-discerning update strategy of the SM approaches, which enables them to check for the innovation in the new data and perform an update only when it can improve the quality of the estimation. In the other words, the SM-based techniques are able to detect and discard the unhelpful data and save the expense of associated updating [1]-[5].

The set-membership normalized least mean squares (SM-NLMS) algorithm [6] is a well-known SM adaptive filtering algorithm that can be viewed as a SM equivalent of the stochastic-gradient NLMS algorithm. Formulation of the SM-NLMS algorithm is the same as that of the NLMS algorithm except having an optimized adaptive step-size instead of a fixed one. The SM-NLMS algorithm has attracted a lot of interest in adaptive signal processing mainly because of its relative simplicity, low computational complexity, robustness against noise, and numerical stability in finite precision implementations [7].

The SM adaptive filtering algorithms generally have high degrees of nonlinearities, which make their performance analysis complicated, especially at the transient state. The SM-NLMS algorithm involves both data nonlinearity [8] and error nonlinearity [9]. Few works have been published on the theoretical performance analysis of the SM adaptive filtering algorithms. Steady-state mean squared error (MSE) of the SM-NLMS algorithm in a stationary environment is analyzed in [10] and [11]. Steady-state MSE of the set-membership affine projection (SM-AP) algorithm in a stationary environment is also analyzed in [20] and [21]. Since the SM-NLMS algorithm is a special case of the SM-AP algorithm (when the projection order is equal to one), the results of [20] and [21] can apply to the SM-NLMS algorithm as well. To the best of our knowledge, the work of [22] is the only one to date that reports a tracking performance analysis of the SM-NLMS algorithm, i.e., steady-state MSE analysis in a nonstationary environment. This work originally deals with the SM-AP algorithm. It is based on an empirical formula for the probability of updating and employs some assumptions, which are in general difficult to justify.

In this paper, we analyze the tracking performance of the SM-NLMS algorithm with a more rigorous approach than the one taken in [22]. Following the customary trend in the literature, we take the steady-state MSE in a nonstationary environment as the measure of tracking performance and assume that the target system varies in time according to the random walk model. The classic approach is to analyze the transient behavior of the adaptive filter and then obtain the steady-state MSE as a limiting behavior of the transient MSE [15]. However, due to high nonlinearity of the SM-NLMS algorithm, this approach would be very difficult. To circumvent the analysis of the transient behavior and its complications, we initiate the steady-state analysis from the energy conservation relation [12]-[14] of the SM-NLMS algorithm. The outcome of the analysis is a nonlinear equation whose solution yields the theoretical

steady-state MSE of the SM-NLMS algorithm in a nonstationary environment. Simulations confirm the accuracy of the theoretical findings.

In section 2, we describe the SM-NLMS algorithm. We analyze its tracking performance in section 3, provide some simulation results in section 4, and conclude the paper in section 5.

## 2. The Set-Membership NLMS algorithm

Let us consider the affine-in-parameter model of the form

$$d(n) = \mathbf{w}_o^T(n)\mathbf{x}(n) + \eta(n) \quad (1)$$

where  $d(n) \in \mathbb{R}$  is the reference signal at time index  $n \in \mathbb{N}$ ,  $\mathbf{w}_o(n) \in \mathbb{R}^L$  is the column vector of the underlying system parameters,  $\mathbf{x}(n) \in \mathbb{R}^L$  is the input vector,  $\eta(n) \in \mathbb{R}$  accounts for background noise,  $\mathbb{N}$  and  $\mathbb{R}$  respectively denote the sets of all nonnegative integers and all real numbers, and superscript  $T$  stands for transposition. Moreover, let us assume that the underlying system varies in time according to the well-known random walk model [14], i.e.,

$$\mathbf{w}_o(n) = \mathbf{w}_o(n-1) + \mathbf{q}(n) \quad (2)$$

where  $\mathbf{q}(n) \in \mathbb{R}^L$  denotes the random perturbation.

SM adaptive filtering algorithms estimate a set of feasible solutions rather than a single-point estimate by specifying a bound on the magnitude of the estimation error over the model space of interest,  $\mathcal{S}$ . Hence, the *feasibility set* that contains all the parameter vectors resulting in an error less than a predetermined bound,  $\gamma > 0$ , is defined as

$$\Theta = \bigcap_{(\mathbf{x}, d) \in \mathcal{S}} \{\mathbf{w} \in \mathbb{R}^L: |d - \mathbf{w}^* \mathbf{x}| \leq \gamma\}.$$

In general, SM adaptive filtering algorithms aim to approximate adaptively the minimal-set estimate of  $\Theta$  that is called the *membership set* and is defined as

$$\Psi(n) = \bigcap_{i=1}^n \mathcal{H}(i)$$

where  $\mathcal{H}(n)$  is the observation-induced *constraint set* at time instant  $n$ :

$$\mathcal{H}(n) = \{\mathbf{w} \in \mathbb{R}^L: |d(n) - \mathbf{w}^* \mathbf{x}(n)| \leq \gamma\}.$$

One way is to outer-approximate  $\Psi(n)$  at each time instant by a minimal-volume spheroid described by

$$S(n) = \{\mathbf{w} \in \mathbb{R}^L: \|\mathbf{w} - \mathbf{w}(n)\|^2 \leq \sigma^2(n)\}$$

where  $\|\cdot\|$  denotes the Euclidean norm,  $\mathbf{w}(n)$  is the center of the spheroid (which is usually taken as the point estimate), and  $\sigma(n)$  is its radius. This is the basic idea behind the SM-NLMS algorithm, which, at each iteration, computes  $S(n)$  as the smallest spheroid that contains the intersection of  $S(n-1)$  and  $\mathcal{H}(n)$ . It is shown in [6] that this can be realized using the following recursions:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu(n) \frac{\mathbf{x}(n)}{\|\mathbf{x}(n)\|^2} e(n), \quad (3)$$

$$\sigma^2(n) = \sigma^2(n-1) - \mu^2(n) \frac{e^2(n)}{\|\mathbf{x}(n)\|^2}$$

where the estimation error,  $e(n)$ , and the variable step-size,  $\mu(n)$ , are given by

$$e(n) = d(n) - \mathbf{w}^T(n-1)\mathbf{x}(n) \quad (4)$$

and

$$\mu(n) = \begin{cases} 1 - \frac{\gamma}{|e(n)|} & \text{if } |e(n)| > \gamma \\ 0 & \text{if } |e(n)| \leq \gamma \end{cases}. \quad (5)$$

### 3. Tracking performance analysis

Substituting (5) into (3) yields

$$\mathbf{w}(n) = \begin{cases} \mathbf{w}(n-1) + \left(1 - \frac{\gamma}{|e(n)|}\right) \frac{\mathbf{x}(n)}{\|\mathbf{x}(n)\|^2} e(n) & \text{if } |e(n)| > \gamma \\ \mathbf{w}(n-1) & \text{if } |e(n)| \leq \gamma \end{cases}. \quad (6)$$

Defining

$$f(n) = \begin{cases} e(n) - \gamma & \text{if } e(n) > \gamma \\ 0 & \text{if } -\gamma \leq e(n) \leq \gamma \\ e(n) + \gamma & \text{if } e(n) < -\gamma \end{cases}, \quad (7)$$

we can rewrite (6) as

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \frac{\mathbf{x}(n)}{\|\mathbf{x}(n)\|^2} f(n). \quad (8)$$

Subtracting both sides of (8) from  $\mathbf{w}_o(n)$  together with using (2) results in

$$\mathbf{v}(n) = \mathbf{v}(n-1) - \frac{\mathbf{x}(n)}{\|\mathbf{x}(n)\|^2} f(n) + \mathbf{q}(n) \quad (9)$$

where  $\mathbf{v}(n) = \mathbf{w}_o(n) - \mathbf{w}(n)$  is the weight-error vector.

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

$$e_a(n) = \mathbf{v}(n-1)^T \mathbf{x}(n) \quad (10)$$

and

$$e_p(n) = (\mathbf{v}(n) - \mathbf{q}(n))^T \mathbf{x}(n).$$

From (1), (4), and (10), we have

$$e(n) = e_a(n) + \eta(n). \quad (11)$$

Multiplying the transpose of both sides of (9) by  $\mathbf{x}(n)$  from the right results in

$$e_p(n) = e_a(n) - f(n). \quad (12)$$

By substituting  $f(n) = e_a(n) - e_p(n)$  into (9), we get

$$\mathbf{v}(n) - \mathbf{q}(n) + \frac{\mathbf{x}(n)}{\|\mathbf{x}(n)\|^2} e_a(n) = \mathbf{v}(n-1) + \frac{\mathbf{x}(n)}{\|\mathbf{x}(n)\|^2} e_p(n). \quad (13)$$

We obtain the following energy conservation relation by evaluating the energies (i.e., the squared Euclidean norms) of both sides of (13):

$$\|\mathbf{v}(n) - \mathbf{q}(n)\|^2 + \frac{e_a^2(n)}{\|\mathbf{x}(n)\|^2} = \|\mathbf{v}(n-1)\|^2 + \frac{e_p^2(n)}{\|\mathbf{x}(n)\|^2}. \quad (14)$$

Taking the expectations of both sides of (14) gives

$$E[\|\mathbf{v}(n) - \mathbf{q}(n)\|^2] + E\left[\frac{e_a^2(n)}{\|\mathbf{x}(n)\|^2}\right] = E[\|\mathbf{v}(n-1)\|^2] + E\left[\frac{e_p^2(n)}{\|\mathbf{x}(n)\|^2}\right]. \quad (15)$$

In order to make the analysis tractable, let us adopt the following assumptions:

- A0: The stochastic process  $\|\mathbf{x}(n)\|^{-2}$  is strictly stationary and the initial input,  $\mathbf{x}(0)$ , is a nonzero vector that satisfies  $E[\|\mathbf{x}(0)\|^{-2}] < \infty$ .
- A1: At the steady state,  $\|\mathbf{x}(n)\|^{-2}$  is statistically independent of both  $e_a^2(n)$  and  $e_p^2(n)$ .
- A2: At the steady state, the noiseless *a priori* estimation error,  $e_a(n)$ , is white Gaussian with zero mean and variance of  $\sigma_a^2 = E[e_a^2(n)]$ .
- A3: The background noise,  $\eta(n)$ , is white Gaussian with zero mean and variance of  $\sigma_\eta^2 = E[\eta^2(n)]$ . It is also independent of the input data.
- A4: The perturbation  $\mathbf{q}(n)$  is independent and identically distributed (i.i.d.) with zero mean and covariance matrix of  $\mathbf{Q} = E[\mathbf{q}(n)\mathbf{q}^T(n)]$ . It is also independent of the input data and the noise.

Under A4, we have

$$\begin{aligned}
E[\|\mathbf{v}(n) - \mathbf{q}(n)\|^2] &= E[\|\mathbf{v}(n)\|^2] - 2E[\mathbf{q}^T(n)\mathbf{v}(n)] + E[\|\mathbf{q}(n)\|^2] \\
&= E[\|\mathbf{v}(n)\|^2] - 2E[\|\mathbf{q}(n)\|^2] + E[\|\mathbf{q}(n)\|^2] \\
&= E[\|\mathbf{v}(n)\|^2] - \text{Tr}(\mathbf{Q})
\end{aligned} \tag{16}$$

where  $\text{Tr}(\cdot)$  denotes the matrix trace.

We know that after convergence, gradient of the mean square deviation approaches zero [6], i.e.,

$$\lim_{n \rightarrow \infty} E[\|\mathbf{v}(n-1)\|^2] - E[\|\mathbf{v}(n)\|^2] \rightarrow 0. \tag{17}$$

Using (16) and (17), at the steady state, (15) becomes

$$E\left[\frac{e_a^2(n)}{\|\mathbf{x}(n)\|^2}\right] = E\left[\frac{e_p^2(n)}{\|\mathbf{x}(n)\|^2}\right] + \text{Tr}(\mathbf{Q}). \tag{18}$$

and using A0 and A1, (18) simplifies to

$$E[e_a^2(n)] = E[e_p^2(n)] + \frac{\text{Tr}(\mathbf{Q})}{E[\|\mathbf{x}(n)\|^{-2}]}. \tag{19}$$

Substituting (12) into (19), we find that at the steady state

$$E[f^2(n) - 2e_a(n)f(n)] + \frac{\text{Tr}(\mathbf{Q})}{E[\|\mathbf{x}(n)\|^{-2}]} = 0. \tag{20}$$

Considering (11), A2 and A3 imply that, at the steady state, the *a priori* estimation error,  $e(n)$ , is also zero-mean white Gaussian. In addition, the variance of  $e(n)$  at the steady-state is computed as

$$\sigma^2 = E[e^2(n)] = \sigma_a^2 + \sigma_\eta^2. \tag{21}$$

Based on this fact and using (7), we can write

$$\begin{aligned}
E[f^2(n)] &= \int_{-\infty}^{-\gamma} (x + \gamma)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx + \int_{\gamma}^{\infty} (x - \gamma)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx \\
&= (\sigma^2 + \gamma^2) \left(1 - \text{erf}\left(\frac{\gamma}{\sqrt{2\sigma^2}}\right)\right) - \gamma \sqrt{\frac{2\sigma^2}{\pi}} e^{-\frac{\gamma^2}{2\sigma^2}}.
\end{aligned} \tag{22}$$

where  $\text{erf}(\cdot)$  is the error function and is defined as

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

We can also verify that, at the steady state,  $e_a(n)$  and  $e(n)$  are jointly Gaussian and the random variable

$$\zeta = e_a(n) - \frac{\sigma_a^2}{\sigma^2} e(n). \quad (23)$$

is independent of  $e(n)$  and thus  $f(n)$ . Consequently, using (21) and (23), we can show that

$$\begin{aligned} E[e_a(n)f(n)] &= E[\zeta f(n)] + \frac{\sigma^2 - \sigma_\eta^2}{\sigma^2} E[e(n)f(n)] \\ &= E[\zeta]E[f(n)] + \frac{\sigma^2 - \sigma_\eta^2}{\sigma^2} E[e(n)f(n)] \\ &= \frac{\sigma^2 - \sigma_\eta^2}{\sigma^2} E[e(n)f(n)]. \end{aligned} \quad (24)$$

Besides, we have

$$\begin{aligned} E[e(n)f(n)] &= \int_{-\infty}^{-\gamma} x(x + \gamma) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx \\ &\quad + \int_{\gamma}^{\infty} x(x - \gamma) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \sigma^2 \left( 1 - \operatorname{erf}\left(\frac{\gamma}{\sqrt{2\sigma^2}}\right) \right). \end{aligned} \quad (25)$$

Hence, by substituting (25) into (24), we get

$$E[e_a(n)f(n)] = (\sigma^2 - \sigma_\eta^2) \left( 1 - \operatorname{erf}\left(\frac{\gamma}{\sqrt{2\sigma^2}}\right) \right). \quad (26)$$

Substitution of (22) and (26) into (20) together with using A0 leads to the following nonlinear equation:

$$(-\sigma^2 + \gamma^2 + 2\sigma_\eta^2) \left( 1 - \operatorname{erf}\left(\frac{\gamma}{\sqrt{2\sigma^2}}\right) \right) - \gamma \sqrt{\frac{2\sigma^2}{\pi}} e^{-\frac{\gamma^2}{2\sigma^2}} + \frac{\operatorname{Tr}(\mathbf{Q})}{E[\|\mathbf{x}(0)\|^{-2}]} = 0. \quad (27)$$

The unique positive solution of (27) for  $\sigma^2$  yields the steady-state MSE of the SM-NLMS algorithm when the underlying system is time-varying. In the Appendix, we prove existence and uniqueness of the solution. It is clear that (27) is the extension of the equation given in [11] (for the unrelaxed and unregularized SM-NLMS algorithm) to the case of nonstationary environment.

## 4. Simulations

We consider estimation and tracking of a flat-fading multiple-input single-output (MISO) wireless channel with  $N_t = 4$  transmitter antennas. Sub-channels between each transmitter and the receiver antennas fade independently according to the random walk model. Thus, the channel vector,  $\mathbf{h}(n) \in \mathbb{R}^{N_t}$ , is modeled via

$$\mathbf{h}(n) = \mathbf{h}(n-1) + \alpha \mathbf{k}(n)$$

where  $\alpha$  is a parameter that controls the rate of the variations and  $\mathbf{k}(n) \in \mathbb{R}^{N_t}$  is a circular Gaussian random vector with zero mean and identity covariance matrix. The channel estimator is an adaptive filter with  $L = 4$  taps. The transmitted symbols (input to the channel estimator) are i.i.d. and modulated using the BPSK scheme with a modulus of  $1/\sqrt{N_t} = 1/2$ .

In Fig. 1, we depict the experimental and theoretical steady-state MSEs as a function of  $\gamma^2/\sigma_\eta^2$  for different values of  $\alpha$  and two values of the power of the additive white Gaussian noise, i.e.,  $\sigma_\eta^2 = 0.01$  and  $0.001$ . The experimental steady-state MSEs are obtained by averaging 100 steady-state values and ensemble-averaging over  $10^3$  independent runs. The theoretical steady-state MSEs are calculated by solving (27) where the last term of its left-hand side simplifies to

$$\frac{\text{Tr}(E[\alpha \mathbf{k}(n) \alpha \mathbf{k}^T(n)])}{E[\|\mathbf{x}(0)\|^{-2}]} = N_t \alpha^2.$$

Fig. 1 testifies that our theoretical results are in excellent match with the simulations. It is also noteworthy that the theoretical predictions of [22] severely mismatch the experimental results in the examined scenario. The mismatch is more pronounced for higher values of  $\alpha$  and  $\gamma$ . To avoid confusion and maintain legibility of the figures, we do not include the results of [22] in Fig. 1.

## 5. Concluding Remarks

*Assumptions:* A0 is required for tractability of the analysis. However, assuming a stationary  $\|\mathbf{x}(n)\|^{-2}$  is not as strict as assuming a stationary input. Therefore, our analysis is valid for any input signal as long as  $\|\mathbf{x}(n)\|^{-2}$  is stationary. Despite seeming unnatural, A1 is reasonably realistic and simplifies the analysis in a great deal. Several similar assumptions have been made in the literature (see, e.g., [11], [13], [14], [15]). Without utilizing A2, calculation of the expectations in (20) would be arduous. This assumption is commonly made to deal with error nonlinearities (see, e.g., [9], [11]). It is justified for long adaptive filters via central limit arguments [16]. Lastly, A3 is natural and regularly used while A4 is typical in the context of tracking performance analysis of the adaptive filters (see, e.g., [13], [17]).

*Simulations:* The primary objective of the provided numerical results is to corroborate our theoretical findings. That is why the simulated MISO wireless channel varies in time according to the random walk model, which is the assumed model in the analysis. Admittedly, this is not the most appropriate way to model variations of a wireless channel. However, it is well established that a Rayleigh-fading channel can be suitably modeled by a first-order autoregressive model [18]. For slow

variations, the corresponding autoregressive model can be approximated by a random walk model [14].

*Summary:* Tracking performance of the SM-NLMS algorithm was analyzed utilizing the energy conservation argument. The analysis resulted in a nonlinear equation that by solving it, the theoretical steady-state MSE of the SM-NLMS algorithm in a nonstationary system is calculated. Existence and uniqueness of the solution of this equation was also proved. Simulations showed that the theoretical MSE values match the experimental ones well.

## Appendix

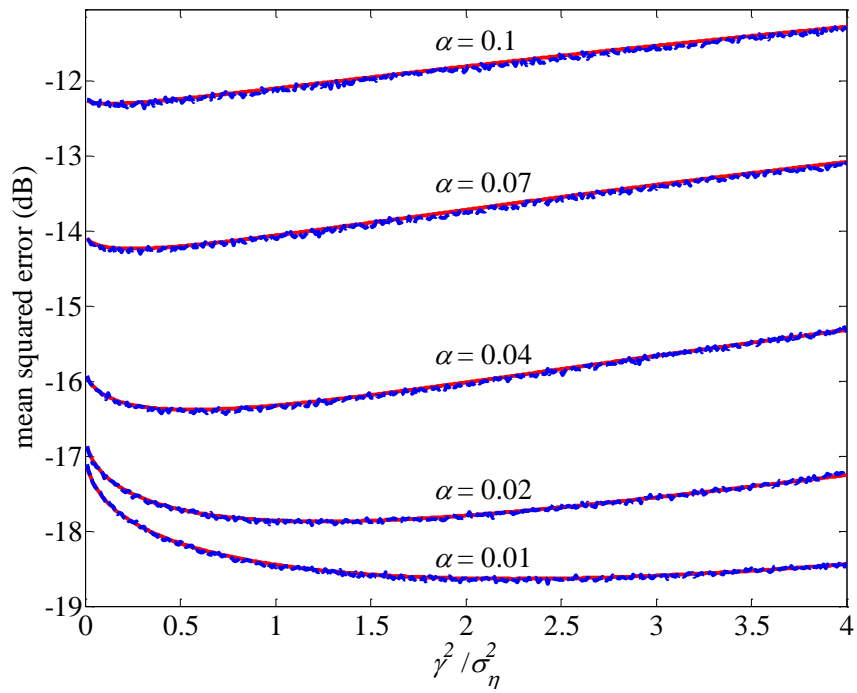
*The same as the Appendix of Paper I.*

## References

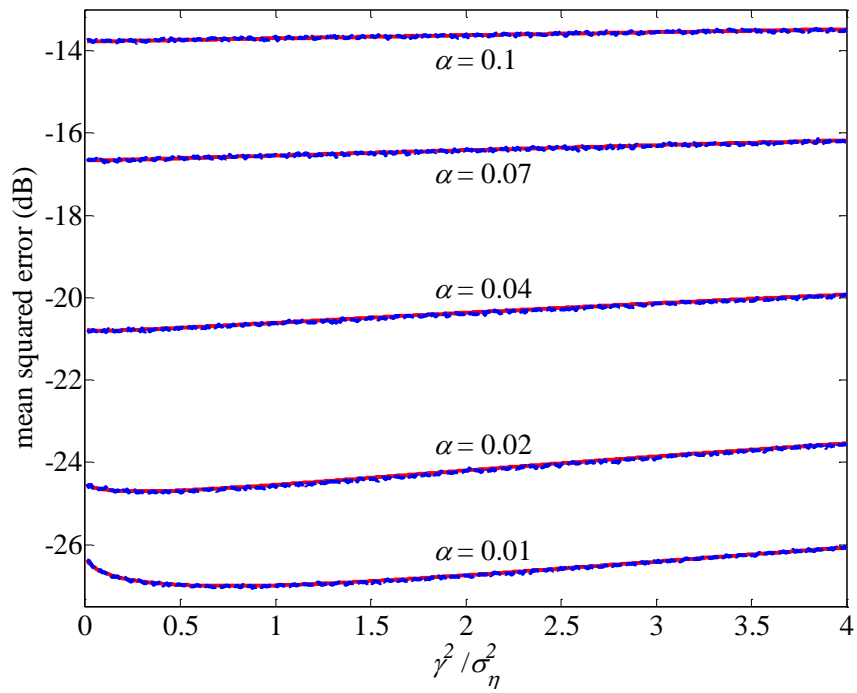
- [1] M. Milanese, J. P. Norton, H. Piet-Lahanier, and E. Walter (eds.), *Bounding Approaches to System Identification*, London: Plenum, 1996.
- [2] J. R. Deller, Jr. and Y. F. Huang, "Set-membership identification and filtering for signal processing applications," *Circuits Systems Signal Process.*, vol. 21, no. 1, pp. 69–82, 2002.
- [3] E.-W. Bai, K. M. Nagpal, and R. Tempo, "Bounded-error estimation: noise models and recursive algorithms," *Automatica*, vol. 32, pp. 1–15, 2002.
- [4] J. R. Deller, Jr., M. Nayeri, and S. F. Odeh, "Least-square identification with error bounds for real-time signal processing and control," *Proc. IEEE*, vol. 81, pp. 813–849, June 1993.
- [5] P. L. Combettes, "The foundations of set-theoretic estimation," *Proc. IEEE*, vol. 81, pp. 182–208, 1993.
- [6] S. Gollamudi, S. Nagaraj, S. Kapoor, and Y. F. Huang, "Set-membership filtering and a set-membership normalized LMS algorithm with an adaptive step size," *IEEE Signal Process. Lett.*, vol. 5, no. 5, pp. 111–114, May 1998.
- [7] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, 3rd ed. New York: Springer, 2008.
- [8] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of data-normalized adaptive filters," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 639–652, Mar. 2003.

- [9] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of adaptive filters with error nonlinearities," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 653–663, Mar. 2003.
- [10] N. Takahashi and I. Yamada, "Steady-state performance of hyperslab projection algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Las Vegas, USA, April 2008, pp. 3829–3832.
- [11] N. Takahashi and I. Yamada, "Steady-state mean-square performance analysis of a relaxed set-membership NLMS algorithm by the energy conservation argument," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3361–3372, 2009.
- [12] M. Rupp and A. H. Sayed, "A time-domain feedback analysis of filtered-error adaptive gradient algorithms," *IEEE Trans. Signal Process.*, vol. 44, pp. 1428–1439, June 1996.
- [13] N. R. Yousef and A. H. Sayed, "A unified approach to the steady-state and tracking analyses of adaptive filters," *IEEE Trans. Signal Process.*, vol. 49, no. 2, pp. 314–324, Feb. 2001.
- [14] A. H. Sayed, *Adaptive Filters*, Hoboken, NJ: Wiley, 2008.
- [15] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2002.
- [16] D. L. Duttweiler, "Adaptive filter performance with nonlinearities in the correlation multiplier," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 30, pp. 578–586, Aug. 1982.
- [17] E. Eweda, "Comparison of RLS, LMS, and sign algorithms for tracking randomly time-varying channels," *IEEE Trans. Signal Process.*, vol. 42, pp. 2937–2944, Nov. 1994.
- [18] C. Komninakis, C. Fragouli, A. H. Sayed, and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1065–1076, May 2002.
- [19] R. D. Gordon, "Values of Mills' ratio of area bounding ordinate and of the normal probability integral for large values of the argument," *Ann. Math. Statistics*, vol. 12, pp. 364–366, 1941.
- [20] M. V. S. Lima and P. S. R. Diniz, "Steady-state analysis of the set-membership affine projection algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Dallas, USA, March 2010, pp. 3802–3805.

- [21] P. S. R. Diniz, "Convergence performance of the simplified set-membership affine projection algorithm," *Circuits Systems Signal Process.*, vol. 30, pp. 439-462, 2011.
- [22] P. S. R. Diniz, "Analysis of a set-membership affine projection algorithm in nonstationary environment," in *Proc. European Signal Process. Conf.*, Glasgow, Scotland, Aug. 2009, pp. 2623-2627.



(a)



(b)

Fig. 1, Comparing experimental (dashed lines) and theoretical (solid lines) MSEs of the SM-NLMS algorithm in nonstationary environments with different values of  $\alpha$  when  $\sigma_\eta^2 = 0.01$  (a) and  $\sigma_\eta^2 = 0.001$  (b).

# Paper K

Originally published as

R. Arablouei and K. Doğançay, "Reduced-complexity constrained recursive least-squares adaptive filtering algorithm," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6687-6692, Dec. 2012.

Copyright © 2012 IEEE. Reprinted with permission.

## Reduced-complexity constrained recursive least-squares adaptive filtering algorithm

Reza Arablouei and Kutluyıl Doğançay

### Abstract

A linearly-constrained recursive least-squares adaptive filtering algorithm based on the method of weighting and the dichotomous coordinate descent (DCD) iterations is proposed. The method of weighting is employed to incorporate the linear constraints into the least-squares problem. The normal equations of the resultant unconstrained least-squares problem are then solved using the DCD iterations. The proposed algorithm has a significantly smaller computational complexity than the previously proposed constrained recursive least-squares (CRLS) algorithm while delivering convergence performance on par with CRLS. The effectiveness of the proposed algorithm is demonstrated by simulation examples.

*Index Terms*—Linearly-constrained adaptive filtering; constrained least-squares; the method of weighting; dichotomous coordinate descent algorithm.

### 1. Introduction

Linearly-constrained adaptive filtering algorithms are powerful signal processing tools with widespread use in several applications such as beamforming, blind multiuser detection in code-division multiple-access (CDMA) systems, and system identification [1]. The linear constraints usually emerge from the prior knowledge about the system, e.g., directions of arrival in array processing [2], spreading codes in blind multiuser detection [3], and linear phase in system identification [4]

The constrained least mean squares (CLMS) algorithm [5] is the first linearly-constrained adaptive filtering algorithm. It was originally introduced as an adaptive solution for the well-known problem of linearly-constrained minimum-variance (LCMV) filtering, which is commonly encountered in antenna array processing [2]. Since CLMS belongs to the family of LMS-type stochastic-gradient algorithms, it is simple, robust, and computationally efficient. However, it is also subject to the limitations of this family. Most notably, it has a slow convergence that is further aggravated when the input signal is correlated. The constrained recursive least-squares (CRLS) algorithm [6] solves the problem of slow convergence in CLMS at the cost of increased computational complexity. This algorithm is developed by directly incorporating the linear constraints into a recursive least-squares solution. The constrained affine projection (CAP) algorithm [7] features complexity and convergence rate between those of the CLMS and CRLS algorithms.

The generalized sidelobe canceler (GSC) [8] is an alternative approach to the implementation of the linearly-constrained adaptive filtering. The GSC offers the advantage of employing any unconstrained adaptive filtering algorithm by transforming the constrained optimization problem to an unconstrained one. However, at each iteration, it multiplies the input vector by a blocking matrix that is orthogonal to the constraint matrix. This generally makes the GSC computationally more expensive than the direct-form constrained adaptive filtering algorithms, e.g., CLMS, CRLS, and CAP, which incorporate the linear constraints into their adaptation process. Householder-transform constrained adaptive filtering [9] is an efficient implementation of the GSC that takes advantage of the efficiency of the Householder transformation to reduce the complexity. In [10], it is proven that the GSC using the RLS algorithm is mathematically equivalent to the CRLS algorithm. A review and comparison of different RLS-type constrained adaptive filtering algorithms can be found in [12].

In this correspondence, we develop a linearly-constrained recursive least-squares algorithm employing the method of weighting [11] and the dichotomous coordinate descent (DCD) iterations [13]. The proposed algorithm can be viewed as a reduced-complexity implementation of the CRLS algorithm. Despite the approximate nature of the method of weighting and the DCD algorithm, the performance of the new algorithm can be made arbitrary close to that of the CRLS algorithm by appropriate selection of the design parameters. The basic idea behind the proposed algorithm is to convert the constrained least-squares problem to an unconstrained one by applying the method of weighting and to solve the system of linear equations (SLE) of the normal equations corresponding to the unconstrained problem by means of the DCD algorithm. The DCD algorithm is an excellent multiplication-free tool for solving SLEs and falls into the category of shift-and-add algorithms. The number of iterations that the DCD algorithm exercises at each run and the resolution of its step-size

control the trade-off between its accuracy and complexity [14].

## 2. Algorithm description

Let us consider a linear system described by

$$d_n = \mathbf{h}^T \mathbf{x}_n + v_n$$

where  $d_n \in \mathbb{R}$  is the desired signal at time index  $n$ ,  $\mathbf{h} \in \mathbb{R}^{L \times 1}$  is the vector of unknown system parameters,  $\mathbf{x}_n \in \mathbb{R}^{L \times 1}$  is the input vector,  $v_n \in \mathbb{R}$  represents the background noise, and superscript  $T$  stands for matrix/vector transpose. Adaptive filtering algorithms recursively estimate the unknown system parameters exploiting the data available up to the current time instant. Generally, the tap weights of an adaptive filter, denoted by  $\mathbf{w}_n \in \mathbb{R}^{L \times 1}$ , are taken as the estimate.

In linearly-constrained adaptive filtering, a set of constraints is imposed upon the filter weights at each time instant as

$$\mathbf{C}^T \mathbf{w}_n = \mathbf{f}$$

where  $\mathbf{C} \in \mathbb{R}^{L \times K}$  is the constraint matrix and  $\mathbf{f} \in \mathbb{R}^{K \times 1}$  is the response vector. A constrained exponentially-weighted least-squares (LS) estimate can be found as

$$\begin{aligned} \mathbf{w}_n = \arg \min_{\mathbf{w}} & \|\mathbf{d}_n - \mathbf{X}_n^T \mathbf{w}\|^2 \\ & \text{subject to } \mathbf{C}^T \mathbf{w} = \mathbf{f} \end{aligned} \quad (1)$$

where

$$\begin{aligned} \mathbf{d}_n &= [d_n, \lambda^{1/2} d_{n-1}, \dots, \lambda^{(n-1)/2} d_1]^T, \\ \mathbf{X}_n &= [\mathbf{x}_n, \lambda^{1/2} \mathbf{x}_{n-1}, \dots, \lambda^{(n-1)/2} \mathbf{x}_1], \end{aligned}$$

$\|\cdot\|$  denotes the Euclidean norm, and  $0 \ll \lambda < 1$  is the forgetting factor. The solution of this optimization problem obtained through the method of Lagrange multipliers is [6]

$$\mathbf{w}_n = \mathbf{R}_n^{-1} \mathbf{p}_n + \mathbf{R}_n^{-1} \mathbf{C} (\mathbf{C}^T \mathbf{R}_n^{-1} \mathbf{C})^{-1} (\mathbf{f} - \mathbf{C}^T \mathbf{R}_n^{-1} \mathbf{p}_n) \quad (2)$$

where

$$\begin{aligned} \mathbf{R}_n &= \mathbf{X}_n \mathbf{X}_n^T \\ &= \sum_{i=1}^n \lambda^{n-i} \mathbf{x}_i \mathbf{x}_i^T \end{aligned}$$

is the exponentially-weighted input autocorrelation matrix and

$$\begin{aligned}\mathbf{p}_n &= \mathbf{X}_n \mathbf{d}_n \\ &= \sum_{i=1}^n \lambda^{n-i} \mathbf{x}_i d_i\end{aligned}$$

is the exponentially-weighted cross-correlation vector between the input vector and the desired signal. The constrained recursive least-squares (CRLS) algorithm [6] is a recursive calculation of (2) that avoids the matrix inversions by applying the matrix inversion lemma [15].

The expression of (2) is an exact solution for the constrained LS problem of interest, (1). However, employing the method of weighting [11], an approximate solution,  $\tilde{\mathbf{w}}_n$ , can be obtained by solving the following unconstrained optimization problem:

$$\tilde{\mathbf{w}}_n = \arg \min_{\mathbf{w}} \left\| \begin{bmatrix} \mathbf{d}_n \\ \mu \mathbf{f} \end{bmatrix} - \begin{bmatrix} \mathbf{X}_n^T \\ \mu \mathbf{C}^T \end{bmatrix} \mathbf{w} \right\|^2$$

where  $\mu \gg 1$  is a weighting parameter. Consequently,  $\tilde{\mathbf{w}}_n$  is the solution of the following normal equations [15]:

$$\left( \begin{bmatrix} \mathbf{X}_n & \mu \mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{X}_n^T \\ \mu \mathbf{C}^T \end{bmatrix} \right) \tilde{\mathbf{w}}_n = \begin{bmatrix} \mathbf{X}_n & \mu \mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{d}_n \\ \mu \mathbf{f} \end{bmatrix}$$

or

$$(\mathbf{R}_n + \mu^2 \mathbf{C} \mathbf{C}^T) \tilde{\mathbf{w}}_n = \mathbf{p}_n + \mu^2 \mathbf{C} \mathbf{f}. \quad (3)$$

In the Appendix, we show that the solution for  $\tilde{\mathbf{w}}_n$  in (3) approaches  $\mathbf{w}_n$  of (2) as  $\mu$  increases.

Defining

$$\Phi_n = \mathbf{R}_n + \mu^2 \mathbf{C} \mathbf{C}^T$$

and

$$\mathbf{z}_n = \mathbf{p}_n + \mu^2 \mathbf{C} \mathbf{f},$$

we may rewrite (3) as

$$\Phi_n \tilde{\mathbf{w}}_n = \mathbf{z}_n. \quad (4)$$

Considering the recursive properties of  $\mathbf{R}_n$  and  $\mathbf{p}_n$ , we have

$$\mathbf{R}_n = \lambda \mathbf{R}_{n-1} + \mathbf{x}_n \mathbf{x}_n^T$$

and

$$\mathbf{p}_n = \lambda \mathbf{p}_{n-1} + \mathbf{x}_n d_n.$$

Accordingly, it is easy to show that

$$\Phi_n = \lambda \Phi_{n-1} + \mathbf{x}_n \mathbf{x}_n^T + (1 - \lambda) \mu^2 \mathbf{C} \mathbf{C}^T \quad (5)$$

and

$$\mathbf{z}_n = \lambda \mathbf{z}_{n-1} + \mathbf{x}_n d_n + (1 - \lambda) \mu^2 \mathbf{C} \mathbf{f}. \quad (6)$$

The system of linear equations (SLE) of (4) can be efficiently solved utilizing the dichotomous coordinate descent (DCD) algorithm [13]. However, in order to exploit the full potential of the DCD algorithm, similar to the approach taken in [14], we solve

$$\Phi_n \Delta \tilde{\mathbf{w}}_n = \boldsymbol{\rho}_n \quad (7)$$

instead of (4). Here,

$$\Delta \tilde{\mathbf{w}}_n = \tilde{\mathbf{w}}_n - \hat{\mathbf{w}}_{n-1}$$

and

$$\boldsymbol{\rho}_n = \mathbf{z}_n - \Phi_n \hat{\mathbf{w}}_{n-1} \quad (8)$$

where  $\hat{\mathbf{w}}_{n-1}$  is the vector of filter weights calculated at the previous time instant. Substituting (5) and (6) into (8) results in

$$\boldsymbol{\rho}_n = \lambda \mathbf{z}_{n-1} + \mathbf{x}_n d_n + (1 - \lambda) \mu^2 \mathbf{C} \mathbf{f} - \lambda \Phi_{n-1} \hat{\mathbf{w}}_{n-1} - \mathbf{x}_n \mathbf{x}_n^T \hat{\mathbf{w}}_{n-1} - (1 - \lambda) \mu^2 \mathbf{C} \mathbf{C}^T \hat{\mathbf{w}}_{n-1},$$

which can be simplified to

$$\boldsymbol{\rho}_n = \lambda \mathbf{z}_{n-1} - \lambda \Phi_{n-1} \hat{\mathbf{w}}_{n-1} + \mathbf{x}_n e_n + (1 - \lambda) \mu^2 \mathbf{C} (\mathbf{f} - \mathbf{C}^T \hat{\mathbf{w}}_{n-1}) \quad (9)$$

where the *a priori* estimation error is defined as

$$e_n = d_n - \hat{\mathbf{w}}_{n-1}^T \mathbf{x}_n.$$

Since  $\hat{\mathbf{w}}_{n-1}$  is an approximate solution, we can define the residual vector at time instant  $n - 1$  as

$$\mathbf{r}_{n-1} = \mathbf{z}_{n-1} - \Phi_{n-1} \hat{\mathbf{w}}_{n-1}. \quad (10)$$

Using (10) in (9) yields

$$\boldsymbol{\rho}_n = \lambda \mathbf{r}_{n-1} + \mathbf{x}_n e_n + (1 - \lambda) \mu^2 \mathbf{C} (\mathbf{f} - \mathbf{C}^T \hat{\mathbf{w}}_{n-1}). \quad (11)$$

Solving (7) using the DCD algorithm gives both  $\Delta \hat{\mathbf{w}}_n$ , an approximate solution for  $\Delta \tilde{\mathbf{w}}_n$ , and  $\mathbf{r}_n$  at each time instant [14]. Having found  $\Delta \hat{\mathbf{w}}_n$ , we can calculate the filter weights as

$$\hat{\mathbf{w}}_n = \hat{\mathbf{w}}_{n-1} + \Delta \hat{\mathbf{w}}_n.$$

We call the proposed algorithm *reduced-complexity constrained recursive least-squares* (RC-CRLS) and summarize it in Table 1. We also present the DCD algorithm in Table 2 where  $a_n^\kappa$  denotes the  $\kappa$ th element of a vector  $\mathbf{a}_n$  while  $\Phi_n^{\kappa,\kappa}$  and  $\Phi_n^{(\kappa)}$  are the  $(\kappa, \kappa)$ th entry and the  $\kappa$ th column of the matrix  $\Phi_n$ , respectively.

The accuracy and complexity of the DCD algorithm is controlled by its user-defined parameters,  $N$ ,  $M$ , and  $H$ , which establish a trade-off between complexity and performance [13]. At each run of the algorithm, maximum  $N$  iterative updates are performed. In other words,  $N$  determines the maximum number of filter coefficients that can be updated at each time instant. Hence, in general, adaptive filtering based on the DCD algorithm implements a form of *selective partial updates* [16]. It is known that by selective partial updating, one can trade performance for complexity [17]. In the DCD algorithm, the step-size  $\alpha$  can accept one of  $M$  predefined values corresponding to representation of the elements of the solution vector as fixed-point words with  $M$  bits within an amplitude range of  $[-H, H]$ .

### 3. Computational complexity

Since  $\Phi_n$  is symmetric, we only need to compute the upper triangular part of the matrix. In addition, similar to [14], we choose the forgetting factor as  $\lambda = 1 - 2^{-s}$  where  $s$  is a positive integer. Consequently, multiplications by  $\lambda$  can be replaced by bit-shifts and additions.

For shift-structured input data, i.e., when

$$\mathbf{x}_n = [x_n, x_{n-1}, \dots, x_{n-L+1}]^T$$

where  $x_n$  is the input signal, updating  $\mathbf{R}_n$  is significantly simplified. In this case, the lower-right  $(L - 1) \times (L - 1)$  block of  $\mathbf{R}_n$  can be obtained by copying the upper-left  $(L - 1) \times (L - 1)$  block of  $\mathbf{R}_{n-1}$ . The only part of the matrix that should be directly updated is the first row and the first column. Due to the symmetry of  $\mathbf{R}_n$ , it is sufficient to update only the first column via

$$\mathbf{R}_n^{(1)} = \lambda \mathbf{R}_{n-1}^{(1)} + \mathbf{x}_n x_n.$$

In the shift-structured input case, the Kalman gain vector in CRLS can be updated via any fast RLS algorithm. The stabilized fast transversal filter (S-FTF) algorithm [18] requires  $7L + 16$  multiplications,  $6L - 1$  additions, and 4 division for this purpose.

In Table 1, we have included the complexity of each step of the proposed algorithm at each iteration for the case of non-shift-structured input. In Table 3, we also present the total number of required multiplications and additions per iteration by the CRLS and RC-CRLS algorithms for both cases of non-shift-structured and shift-structured input.

## 4. Simulations

### 4.1. Linear-phase system identification (LPSI)

In the first part of simulations, we consider a system identification problem where the filter length is an odd number and the filter weights are constrained to preserve linear phase at each iteration. Thus, the constraint matrix is  $\mathbf{C} = [\mathbf{I}_K, \mathbf{0}, -\mathbf{J}_K]^T \in \mathbb{R}^{L \times K}$  and the response vector is  $\mathbf{f} = \mathbf{0} \in \mathbb{R}^{K \times 1}$  where  $\mathbf{I}_K \in \mathbb{R}^{K \times K}$  is the identity matrix of size  $K = (L - 1)/2$  and  $\mathbf{J}_K \in \mathbb{R}^{K \times K}$  is the reversal matrix of size  $K$  (identity matrix with all rows in reversed order). The input vectors are independent and have multivariate normal distribution with zero mean vector and identity covariance matrix. The unknown system is an arbitrary finite-impulse-response filter of order  $L = 15$  with linear phase and unit energy. The background noise is white Gaussian with a power of  $\sigma_v^2 = 10^{-2}$ . We set  $\lambda = 0.99$ ,  $M = 16$ , and  $H = 1$ . We use two performance metrics: misalignment, defined as  $E[\|\mathbf{h} - \mathbf{w}_n\|^2]$ , and constraint mismatch, defined as  $E[\|\mathbf{C}^T \mathbf{w}_n - \mathbf{f}\|^2]$ , where  $E[\cdot]$  is the expectation operator and the expected values are computed by ensemble-averaging over  $10^3$  independent runs.

In Fig. 1, we compare the misalignment curves of the CRLS algorithm and the proposed algorithm with  $\mu = 50$  and different values of  $N$ . It is clear that the larger  $N$  is, the faster the RC-CRLS algorithm converges. We obtain the curves of Fig. 2 by repeating the experiment of Fig. 1 for  $10^4$  iterations and  $N = 16$  only. It is seen that, unlike the CRLS algorithm that diverges after about 3200 iterations, the proposed algorithm maintains its numerical stability throughout the simulation. In Fig. 3, we plot the constraint mismatch of the proposed algorithm as a function of  $\mu$  with  $N$  fixed at 16. As expected, with a larger  $\mu$ , the constraints are better satisfied.

### 4.2. Linearly-constrained minimum-variance (LCMV) filtering

In the second part, we consider an LCMV filtering problem. The input signal is made of three sinusoids with random phases,  $\phi_1$ ,  $\phi_2$ , and  $\phi_3$ , and normalized frequencies  $f_1 = 0.065$ ,  $f_2 = 0.195$ , and  $f_3 = 0.41$  corrupted by additive white Gaussian noise,  $\eta_n$ , with power of  $\sigma_\eta^2 = 10^{-2}$ :

$$x_n = \sin(2\pi f_1 n + \phi_1) + \sin(2\pi f_2 n + \phi_2) + \sin(2\pi f_3 n + \phi_3) + \eta_n.$$

The input vectors are shift-structured. We require the filter to have unity response at two frequencies,  $f_4 = 0.14$  and  $f_5 = 0.325$ . Hence, we impose two point constraints and two first-order derivative constraints [19] apropos  $f_4$  and  $f_5$  and form the constraint matrix as

$$\mathbf{C} = \begin{bmatrix} 1, \cos(2\pi f_4), \dots, \cos\{2\pi f_4(L-1)\} \\ 1, \cos(2\pi f_5), \dots, \cos\{2\pi f_5(L-1)\} \\ 0, \sin(2\pi f_4), \dots, \sin\{2\pi f_4(L-1)\} \\ 0, \sin(2\pi f_5), \dots, \sin\{2\pi f_5(L-1)\} \end{bmatrix}$$

and the response vector as  $\mathbf{f} = [1, 1, 0, 0]^T$ . The filters are assumed to be of length  $L = 9$ . The desired signal is zero. The optimal solution is calculated as

$$\mathbf{h} = \Psi^{-1} \mathbf{C} (\mathbf{C}^T \Psi^{-1} \mathbf{C})^{-1} \mathbf{f}$$

where  $\Psi = E[\mathbf{x}_n \mathbf{x}_n^T]$  is the covariance matrix of the input signal [4]. We also set  $\lambda = 0.99$ ,  $M = 16$ , and  $H = 0.1$ .

In Fig. 4, we compare the misalignment curves of the CRLS algorithm and the proposed algorithm with  $\mu = 30$  and different values of  $N$ . In Fig. 5, we plot the constraint mismatch of the proposed algorithm as a function of  $\mu$  with  $N$  fixed at 16. Fig. 6 depicts the frequency response of the filters after 400 iterations when  $\mu = 20$  and only a single iteration of the DCD algorithm is exercised at each time instant, i.e.,  $N = 1$ .

In Table 4, we present the number of required multiplications and additions per iteration by the CRLS algorithm and the proposed algorithm with  $N = 16$  for both the above experiments.

In addition to the experiment of Fig. 2 and in order to assess numerical stability of the proposed algorithm, we carried out several numerical experiments in different scenarios for very large numbers of iterations, e.g.,  $10^5$ . Consistent with what has been reported in [14] concerning the numerical stability of the DCD-RLS algorithm, we did not observe any instability problem in our RC-CRLS algorithm during the experiments.

## 5. Concluding remarks

We have used the method of weighting and the DCD iterations to develop a new linearly-constrained recursive least-squares algorithm, which we call RC-CRLS. The proposed algorithm requires considerably less arithmetic operations than the previously proposed CRLS algorithm. Moreover, it can be easily tuned to perform almost as well as CRLS, even though the method of weighting and the DCD algorithm are approximate techniques. This can be explained in view of the fact that, at each iteration, an LS-type adaptive filter seeks an approximate solution to an overdetermined problem based on current and previous noisy observations. Hence, if the imprecision introduced by an incorporated approximate technique, such as the method of weighting, constitutes only a small portion of the total error, it will not have a noticeable impact on the overall performance of the adaptive filter.

For the sake of simplicity, we have considered that all the signals, parameters, and variables are real-valued. However, the proposed algorithm (summarized in Table 1) can be straightforwardly extended to the complex domain by replacing the transpose with complex-conjugate transpose,  $e_n$  in (11) with its complex conjugate, and the DCD algorithm of Table 2 with the one in Table VII of [20].

We observe that a larger value of the weighting parameter,  $\mu$ , results in a lower constraint mismatch, i.e., better satisfaction of the constraints. However, it should be noted that a large  $\mu$  can make the associated SLE hard to solve necessitating a large  $N$  and requiring a large number of addition operations at each iteration. Therefore, there exists a trade-off between the convergence performance, satisfaction of the constraints, and complexity of the proposed algorithm.

## Appendix

If we were not concerned about complexity, (3) could be solved as

$$\tilde{\mathbf{w}}_n = (\mathbf{R}_n + \mu^2 \mathbf{C}\mathbf{C}^T)^{-1}(\mathbf{p}_n + \mu^2 \mathbf{C}\mathbf{f}). \quad (\text{A1})$$

Using the matrix inversion lemma [15],

$$(\mathbf{A} - \mathbf{B}\mathbf{\Delta}^{-1}\mathbf{\Gamma})^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{\Delta} - \mathbf{\Gamma}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{\Gamma}\mathbf{A}^{-1} \quad (\text{A2})$$

with  $\mathbf{A} = \mathbf{R}_n$ ,  $\mathbf{B} = -\mathbf{C}$ ,  $\mathbf{\Gamma} = \mathbf{C}^T$ , and  $\mathbf{\Delta} = \mu^{-2}\mathbf{I}$ , we have

$$(\mathbf{R}_n + \mu^2 \mathbf{C}\mathbf{C}^T)^{-1} = \mathbf{R}_n^{-1} - \mathbf{R}_n^{-1}\mathbf{C}(\mu^{-2}\mathbf{I} + \mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C})^{-1}\mathbf{C}^T\mathbf{R}_n^{-1}. \quad (\text{A3})$$

Substituting (A3) into (A1) results in

$$\begin{aligned} \tilde{\mathbf{w}}_n &= [\mathbf{R}_n^{-1} - \mathbf{R}_n^{-1}\mathbf{C}(\mu^{-2}\mathbf{I} + \mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C})^{-1}\mathbf{C}^T\mathbf{R}_n^{-1}](\mathbf{p}_n + \mu^2 \mathbf{C}\mathbf{f}) \\ &= \mathbf{R}_n^{-1}\mathbf{p}_n - \mathbf{R}_n^{-1}\mathbf{C}(\mu^{-2}\mathbf{I} + \mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C})^{-1}\mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{p}_n + \mu^2 \mathbf{R}_n^{-1}\mathbf{C}\mathbf{f} \\ &\quad - \mu^2 \mathbf{R}_n^{-1}\mathbf{C}(\mu^{-2}\mathbf{I} + \mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C})^{-1}\mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C}\mathbf{f}. \end{aligned} \quad (\text{A4})$$

Using (A2) with  $\mathbf{A} = \mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C}$ ,  $\mathbf{B} = -\mu^{-2}\mathbf{I}$ ,  $\mathbf{\Gamma} = \mathbf{I}$ , and  $\mathbf{\Delta} = \mathbf{I}$  gives

$$\begin{aligned} (\mu^{-2}\mathbf{I} + \mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C})^{-1} &= (\mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C})^{-1} \\ &\quad - \mu^{-2}(\mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C})^{-1}[\mathbf{I} + \mu^{-2}(\mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C})^{-1}]^{-1}(\mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C})^{-1}. \end{aligned} \quad (\text{A5})$$

Substituting (A5) into the fourth term on the right-hand side of (A4) yields

$$\begin{aligned} \tilde{\mathbf{w}}_n &= \mathbf{R}_n^{-1}\mathbf{p}_n - \mathbf{R}_n^{-1}\mathbf{C}(\mu^{-2}\mathbf{I} + \mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C})^{-1}\mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{p}_n + \mu^2 \mathbf{R}_n^{-1}\mathbf{C}\mathbf{f} \\ &\quad - \mu^2 \mathbf{R}_n^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C})^{-1}\mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C}\mathbf{f} + \mathbf{R}_n^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C})^{-1} \\ &\quad \times [\mathbf{I} + \mu^{-2}(\mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C})^{-1}]^{-1}(\mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C})^{-1}\mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C}\mathbf{f} \\ &= \mathbf{R}_n^{-1}\mathbf{p}_n - \mathbf{R}_n^{-1}\mathbf{C}(\mu^{-2}\mathbf{I} + \mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C})^{-1}\mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{p}_n \\ &\quad + \mathbf{R}_n^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C})^{-1}[\mathbf{I} + \mu^{-2}(\mathbf{C}^T\mathbf{R}_n^{-1}\mathbf{C})^{-1}]^{-1}\mathbf{f}. \end{aligned} \quad (\text{A6})$$

For large  $\mu$ , the following approximations hold:

$$\mu^{-2}\mathbf{I} + \mathbf{C}^T \mathbf{R}_n^{-1} \mathbf{C} \approx \mathbf{C}^T \mathbf{R}_n^{-1} \mathbf{C} \quad (\text{A7})$$

and

$$\mathbf{I} + \mu^{-2}(\mathbf{C}^T \mathbf{R}_n^{-1} \mathbf{C})^{-1} \approx \mathbf{I}. \quad (\text{A8})$$

Using (A7) and (A8), (A6) becomes

$$\tilde{\mathbf{w}}_n \approx \mathbf{R}_n^{-1} \mathbf{p}_n - \mathbf{R}_n^{-1} \mathbf{C}(\mathbf{C}^T \mathbf{R}_n^{-1} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{R}_n^{-1} \mathbf{p}_n + \mathbf{R}_n^{-1} \mathbf{C}(\mathbf{C}^T \mathbf{R}_n^{-1} \mathbf{C})^{-1} \mathbf{f}. \quad (\text{A9})$$

The right-hand side of (A9) is equal to the exact solution of (2). Therefore, we conclude that for sufficiently large  $\mu$ , we have  $\tilde{\mathbf{w}}_n \approx \mathbf{w}_n$ . Note that the accuracy of approximation improves as  $\mu$  gets larger.

## References

- [1] M. L. R. de Campos, S. Werner, and J. A. Apolinário Jr., "Constrained adaptive filters," in *Adaptive Antenna Arrays: Trends and Applications*, S. Chrandran, Ed. New York: Springer-Verlag, 2004.
- [2] H. L. Van Trees, *Detection, Estimation, and Modulation Theory Part IV: Optimum Array Processing*, New York: Wiley, 2002.
- [3] S. Verdu, *Multiuser Detection*, Cambridge, U.K.: Cambridge Uni. Press, 1998.
- [4] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementations*, 3<sup>rd</sup> ed., Boston: Springer, 2008.
- [5] O. L. Frost III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.
- [6] L. S. Resende, J. M. T. Romano, and M. G. Bellanger, "A fast least-squares algorithm for linearly constrained adaptive filtering," *IEEE Trans. Signal Process.*, vol. 44, no. 5, pp. 1168–1174, May 1996.
- [7] S. Werner, J. A. Apolinário Jr., M. L. R. de Campos, and P. S. R. Diniz, "Low-complexity constrained affine-projection algorithms," *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 4545–4555, Dec. 2005.
- [8] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas and Propag.*, vol. AP-30, no. 1, pp. 27–34, Jan. 1982.
- [9] M. L. R. de Campos, S. Werner, and J. A. Apolinário Jr., "Constrained adaptation algorithms employing Householder transformation," *IEEE Trans. Signal Process.*, vol. 50, no. 9, pp. 2187–2195, Sep. 2002.

- [10] S. Werner, J. A. Apolinário Jr., and M. L. R. de Campos, "On the equivalence of RLS implementations of LCMV and GSC processors," *IEEE Signal Process. Lett.*, vol. 10, no. 12, pp. 356–359, Dec. 2003.
- [11] C. Van Loan, "On the method of weighting for equality-constrained least-squares problems," *SIAM J. Numer. Anal.*, vol. 22, pp. 851–864, Oct. 1985.
- [12] A. A. L. Ramos, J. A. Apolinário Jr., and M. L. R. Campos "On numerical robustness of constrained RLS-like algorithms," in *Proc. XXI Simpósio Brasileiro de Telecomunicações*, Belém, Brazil, Sep. 2004.
- [13] Y. V. Zakharov and T. C. Tozer, "Multiplication-free iterative algorithm for LS problem," *Electron. Lett.*, vol. 40, no. 9, pp. 567–569, Apr. 2004.
- [14] Y. Zakharov, G. White, and J. Liu, "Low complexity RLS algorithms using dichotomous coordinate descent iterations," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3150–3161, Jul. 2008.
- [15] G. H. Golub and C. F. Van Loan, *Matrix Computations*, third ed., Baltimore, MD: Johns Hopkins Univ. Press, 1996.
- [16] K. Doğançay and O. Tanrikulu, "Adaptive filtering algorithms with selective partial updates," *IEEE Trans. Circuits Syst. II*, vol. 48, no. 8, pp. 762–769, Aug. 2001.
- [17] K. Doğançay, *Partial-Update Adaptive Signal Processing: Design, Analysis and Implementation*, Academic Press, Oxford, UK, 2008.
- [18] D. T. M. Slock and T. Kailath, "Numerically stable fast transversal filters for recursive least-squares adaptive filtering," *IEEE Trans. Signal Process.*, vol. 39, pp. 92–114, Jan. 1991.
- [19] M. H. Er and A. Cantoni, "Derivative constraints for broad-band element space antenna array processors," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, no. 6, pp. 1378–1393, Dec. 1983.
- [20] J. Liu, Y. V. Zakharov, and B. Weaver, "Architecture and FPGA design of dichotomous coordinate descent algorithms," *IEEE Trans. Circuits Syst. I*, vol. 56, no. 11, pp. 2425–2438, Nov. 2009.

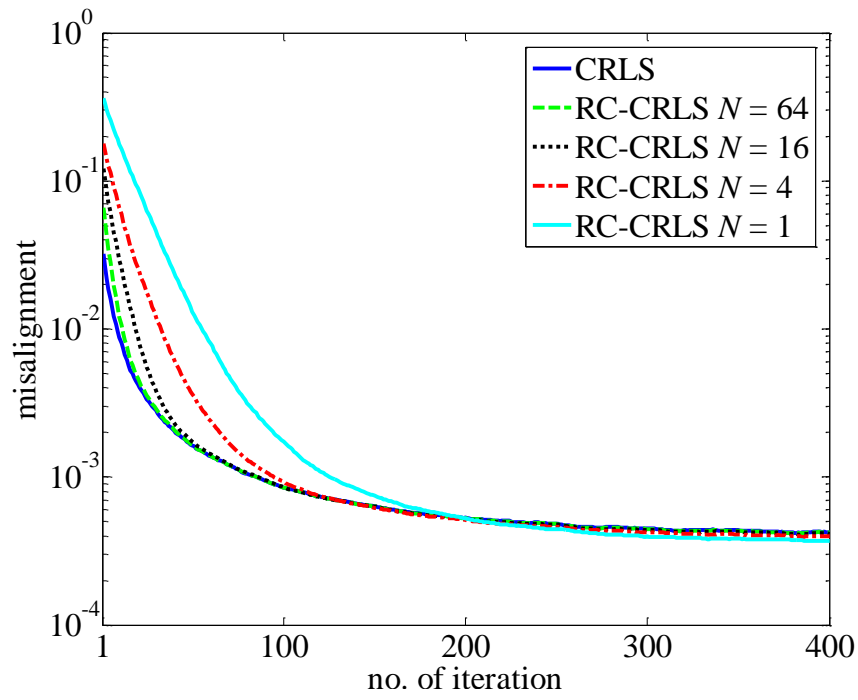


Fig. 1, Misalignment,  $E[\|\mathbf{h} - \mathbf{w}_n\|^2]$ , of the CRLS algorithm and the RC-CRLS algorithm with different values of the DCD iterations,  $N$ , in an LPSI experiment.

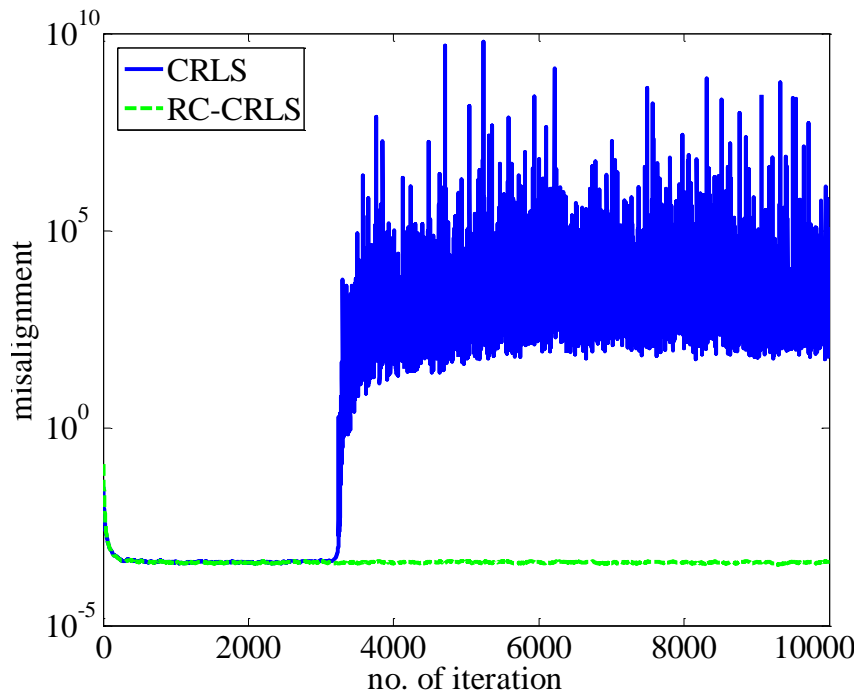


Fig. 2, Misalignment,  $E[\|\mathbf{h} - \mathbf{w}_n\|^2]$ , of the CRLS and RC-CRLS algorithms in an LPSI experiment simulated for  $10^4$  iterations.

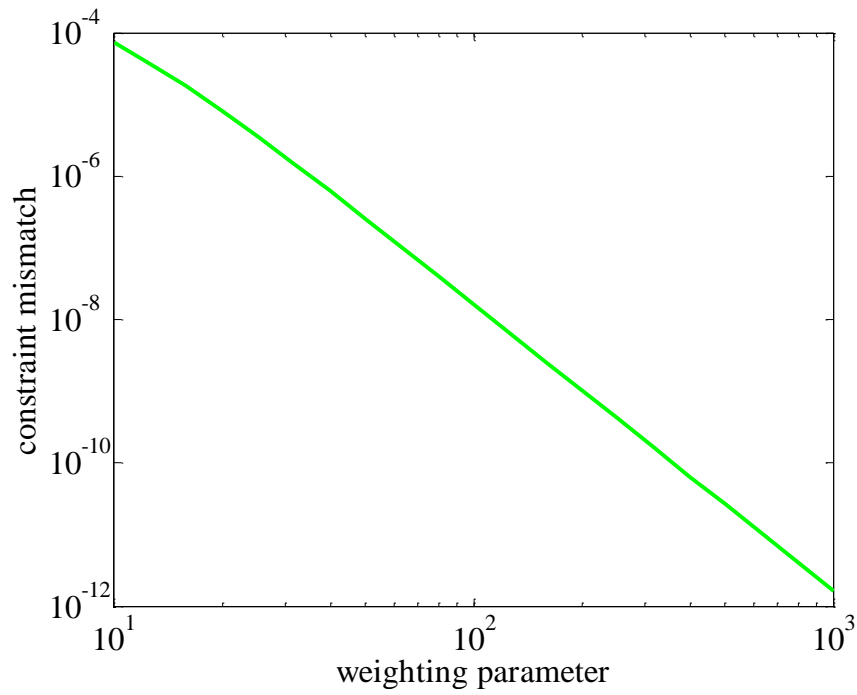


Fig. 3, Constraint mismatch,  $E[\|\mathbf{C}^T \mathbf{w}_n - \mathbf{f}\|^2]$ , of the RC-CRLS algorithm as a function of the weighting parameter,  $\mu$ , in an LPSI experiment.

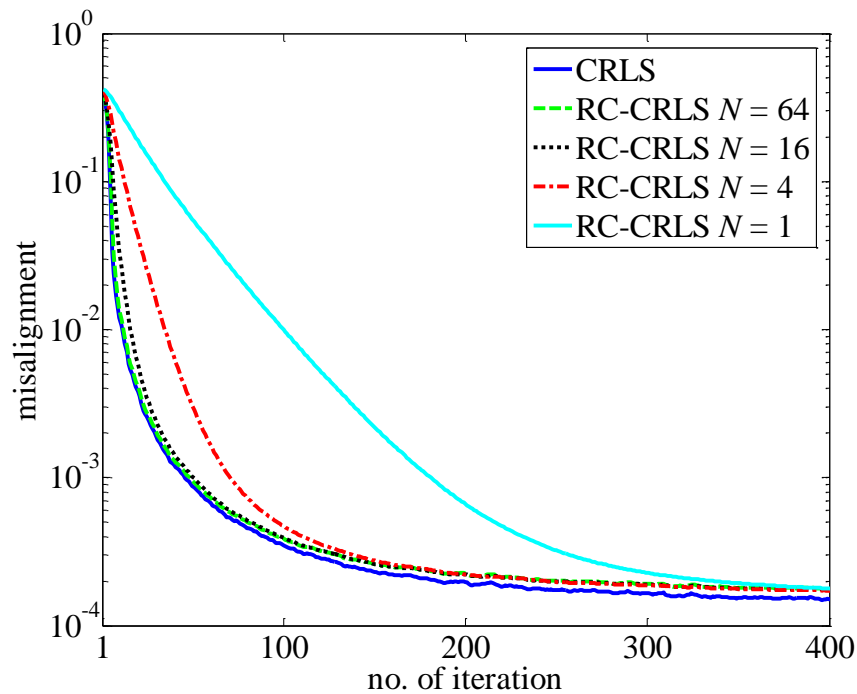


Fig. 4, Misalignment,  $E[\|\mathbf{h} - \mathbf{w}_n\|^2]$ , of the CRLS algorithm and the RC-CRLS algorithm with different values of the DCD iterations,  $N$ , in an LCMV filtering experiment.

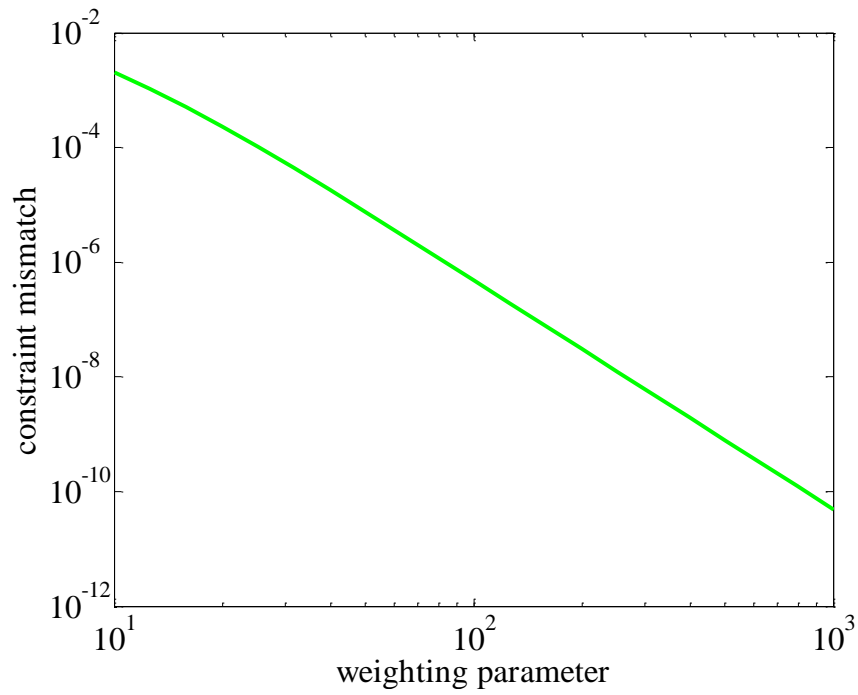


Fig. 5, Constraint mismatch,  $E[\|\mathbf{C}^T \mathbf{w}_n - \mathbf{f}\|^2]$ , of the RC-CRLS algorithm as a function of the weighting parameter,  $\mu$ , in an LCMV experiment.

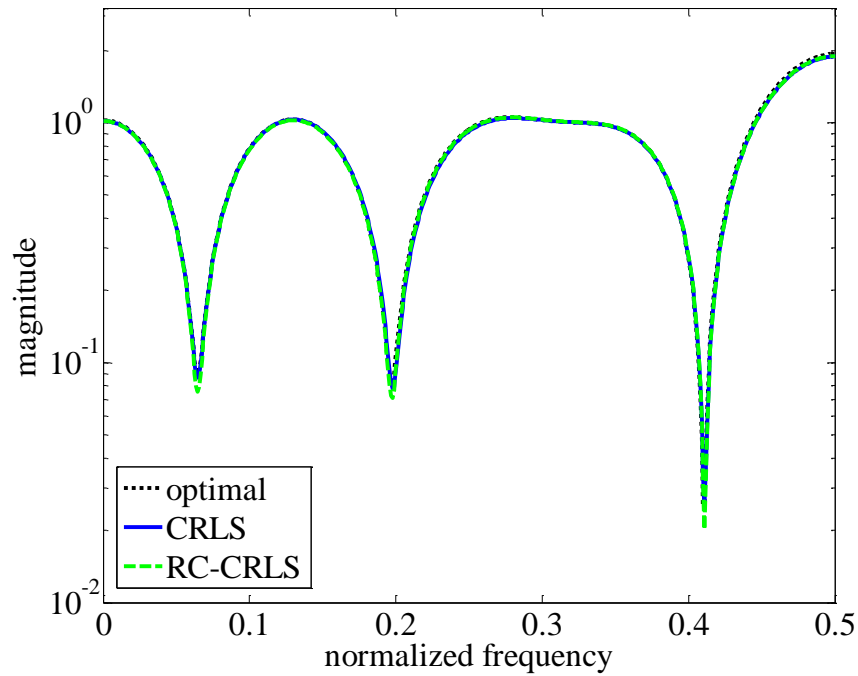


Fig. 6, Frequency response of the filters after 400 iterations.

Table 1, The RC-CRLS algorithm.

Initialization		
$\mathbf{R}_0 = \delta \mathbf{I}_L$		
$\hat{\mathbf{w}}_0 = \mathbf{C}(\mathbf{C}^T \mathbf{C})^{-1} \mathbf{f}$		
$\mathbf{r}_0 = \mathbf{0}$		
$\mathbf{T} = \mu^2 \mathbf{C} \mathbf{C}^T$		
$\mathbf{U} = (1 - \lambda) \mu^2 \mathbf{C}$	×	+
At iteration $n = 1, 2, \dots$		
$\mathbf{R}_n = \lambda \mathbf{R}_{n-1} + \mathbf{x}_n \mathbf{x}_n^T$	$\frac{1}{2} L^2 + \frac{1}{2} L$	$L^2 + L$
$\Phi_n = \mathbf{R}_n + \mathbf{T}$	0	$\frac{1}{2} L^2 + \frac{1}{2} L$
$e_n = d_n - \hat{\mathbf{w}}_{n-1}^T \mathbf{x}_n$	$L$	$L$
$\rho_n = \lambda \mathbf{r}_{n-1} + \mathbf{x}_n e_n + \mathbf{U}(\mathbf{f} - \mathbf{C}^T \hat{\mathbf{w}}_{n-1})$	$(2K + 1)L$	$(2K + 2)L$
solve $\Phi_n \Delta \tilde{\mathbf{w}}_n = \rho_n$ and obtain $\Delta \hat{\mathbf{w}}_n$ and $\mathbf{r}_n$	0	$2NL + N + M$
$\hat{\mathbf{w}}_n = \hat{\mathbf{w}}_{n-1} + \Delta \hat{\mathbf{w}}_n$	0	$L$
sum: $\frac{1}{2} L^2 + \left(2K + \frac{5}{2}\right) L$ $\frac{3}{2} L^2 + \left(2K + 2N + \frac{11}{2}\right) L + N + M$		

Table 2, The DCD algorithm solving  $\Phi_n \Delta \tilde{\mathbf{w}}_n = \rho_n$ .

Initialization
$m = 1, \alpha = \frac{H}{2}$
$\Delta \hat{\mathbf{w}}_n = \mathbf{0}$
$\mathbf{r}_n = \rho_n$
For $k = 1, 2, \dots, N$
$p = \arg \max_{i=1, \dots, L} \{ r_n^i \}$
while $ r_n^p  \leq \frac{\alpha}{2} \Phi_n^{p,p}$ and $m \leq M$
$m = m + 1, \alpha = \frac{\alpha}{2}$
if $m > M$
algorithm stops
$\Delta \hat{\mathbf{w}}_n^p = \Delta \hat{\mathbf{w}}_n^p + \text{sign}\{r_n^p\} \alpha$
$\mathbf{r}_n = \mathbf{r}_n - \text{sign}\{r_n^p\} \alpha \Phi_n^{(p)}$

Table 3, Computational complexity of the CRLS and RC-CRLS algorithms in terms of number of required multiplications and additions per iteration.

	×	+
Non-shift-structured input		
CRLS	$4L^2 + (3K^2 + 5K + 4)L + K^2 + 2K$	$3L^2 + (3K^2 + 4K + 1)L - K^2 + K$
RC-CRLS	$\frac{1}{2}L^2 + \left(2K + \frac{5}{2}\right)L$	$\frac{3}{2}L^2 + \left(2K + 2N + \frac{11}{2}\right)L + N + M$
Shift-structured input		
CRLS	$(3K^2 + 5K + 9)L + K^2 + 2K + 16$	$(3K^2 + 4K + 8)L - K^2 + K - 1$
RC-CRLS	$(2K + 3)L$	$\frac{1}{2}L^2 + \left(2K + 2N + \frac{13}{2}\right)L + N + M$

Table 4, Computational complexity of the CRLS and RC-CRLS algorithms in the simulated experiments.

	×	+
LPSI		
CRLS	3753	3273
RC-CRLS	360	1142
LCMV		
CRLS	733	635
RC-CRLS	99	491

# Paper L

Originally published as

R. Arablouei and K. Doğançay, "Linearly-constrained recursive total least-squares algorithm," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 821-824, Dec. 2012.

Copyright © 2012 IEEE. Reprinted with permission.

## Linearly-constrained recursive total least-squares algorithm

Reza Arablouei and Kutluyıl Doğançay

### Abstract

We develop a new linearly-constrained recursive total least-squares adaptive filtering algorithm by incorporating the linear constraints into the underlying total least-squares problem using an approach similar to the method of weighting and searching for the solution (filter weights) along the input vector. The proposed algorithm outperforms the previously proposed constrained recursive least-squares (CRLS) algorithm when both input and output data are observed with noise. It also has a significantly smaller computational complexity than CRLS. Simulations demonstrate the efficacy of the proposed algorithm.

*Index Terms*—Linearly-constrained adaptive filtering; recursive total least-squares.

### 1. Introduction

Linearly-constrained adaptive filtering algorithms are powerful signal processing tools with widespread use in several applications such as beamforming, blind multiuser detection in code-division multiple-access (CDMA) systems, and system identification [1]. The linear constraints usually emerge from the prior knowledge about the system, e.g., directions of arrival in array processing [2], spreading codes in blind multiuser detection [3], and linear phase in system identification [4].

Several linearly-constrained adaptive filtering algorithms have been proposed, which can be loosely divided into two categories: 1) stochastic-gradient-based algorithms such as the constrained least mean squares (CLMS) algorithm [5] and the constrained affine projection (CAP) algorithm [6]; 2) least-squares(LS)-based algorithms such as the constrained recursive least-squares (CRLS) algorithm [7] and the constrained

conjugate-gradient (CCG) algorithm [8]. The algorithms of the former category are generally simpler, more robust, and computationally more efficient than the algorithms of the latter category. However, they typically have slower convergence rate, which is further aggravated when the input signal is correlated.

The LS-based algorithms assume that the input signal is observed accurately and only the filter output is corrupted by noise. Thus, when the input signal is observed with noise, the LS-based algorithms show poor estimation performance as a result of failing to account for error in the input signal. On the other hand, algorithms based on the total least-squares (TLS) method produce improved estimation performance as TLS compensates for both input and output data errors [9]. Among the few proposed recursive TLS adaptive filtering algorithms, the recursive TLS (RTLS) algorithm of [10] and the new RTLS (N-RTLS) algorithm of [11] are the most prominent and computationally-efficient ones. The latter is in fact a reduced-complexity version of the former. However, to the best of our knowledge, no linearly-constrained recursive TLS algorithm has been proposed to date.

In this letter, we develop a linearly-constrained recursive total least-squares algorithm in Section II by converting the constrained TLS problem to an unconstrained one that incorporates the linear constraints and then solving it on a linear manifold defined by the input vector. We provide some simulation results in Section 3 and draw conclusions in Section 4.

## 2. Algorithm description

Let us consider a linear system described by

$$\tilde{\mathbf{x}}_n^T \mathbf{h} = \tilde{d}_n$$

and

$$\tilde{\mathbf{x}}_n = [\tilde{x}_n, \tilde{x}_{n-1}, \dots, \tilde{x}_{n-L+1}]^T$$

where  $\mathbf{h} \in \mathbb{R}^L$  is the column vector of unknown system parameters,  $\tilde{x}_n \in \mathbb{R}$  and  $\tilde{d}_n \in \mathbb{R}$  are respectively the input and output signals at time index  $n \in \mathbb{N}$ , and superscript  $T$  denotes matrix transposition. We observe noisy versions of the signals. To be precise, instead of  $\tilde{x}_n$  and  $\tilde{d}_n$ , we measure  $x_n = \tilde{x}_n + \xi_n$  and  $d_n = \tilde{d}_n + v_n$  where  $\xi_n \in \mathbb{R}$  and  $v_n \in \mathbb{R}$  represent the corresponding noises, which are assumed to be white Gaussian with respective powers of  $\sigma_i^2$  and  $\sigma_o^2$ .

Adaptive filtering algorithms iteratively estimate  $\mathbf{h}$  exploiting the data available up to the current time. The tap weights of the adaptive filter, denoted by  $\mathbf{w}_n \in \mathbb{R}^L$ , are taken as the estimate at iteration  $n$ . In linearly-constrained adaptive filtering, a set of constraints is imposed upon the filter weights at each time instant as

$$\mathbf{C}^T \mathbf{w}_n = \mathbf{f}$$

where  $\mathbf{C} \in \mathbb{R}^{L \times K}$  is the constraint matrix and  $\mathbf{f} \in \mathbb{R}^K$  is the response vector.

A constrained exponentially-weighted TLS solution for the adaptive filter weights can be found by solving the following optimization problem [9]:

$$\mathbf{w}_n = \arg \min_{\mathbf{w}} \frac{\left\| \mathbf{D} \begin{bmatrix} \mathbf{X}_n^T & \mathbf{d}_n \end{bmatrix} \mathbf{T} \mathbf{T}^{-1} \begin{bmatrix} \mathbf{w} \\ -1 \end{bmatrix} \right\|^2}{\left\| \mathbf{T}^{-1} \begin{bmatrix} \mathbf{w} \\ -1 \end{bmatrix} \right\|^2} \quad (1)$$

subject to  $\mathbf{C}^T \mathbf{w} = \mathbf{f}$

where

$$\begin{aligned} \mathbf{X}_n &= [\mathbf{x}_n, \mathbf{x}_{n-1}, \dots, \mathbf{x}_1], \\ \mathbf{x}_n &= [x_n, x_{n-1}, \dots, x_{n-L+1}]^T, \\ \mathbf{d}_n &= [d_n, d_{n-1}, \dots, d_1]^T, \\ \mathbf{D} &= \text{diag}\{1, \lambda^{1/2}, \dots, \lambda^{(n-1)/2}\}, \end{aligned}$$

$0 \ll \lambda < 1$  is the forgetting factor, and  $\mathbf{T}$  is a weighting matrix that is chosen as

$$\mathbf{T} = \begin{bmatrix} \mathbf{I}_L & \mathbf{0} \\ \mathbf{0} & \beta^{-1/2} \end{bmatrix}$$

with  $\mathbf{I}_L \in \mathbb{R}^{L \times L}$  being the identity matrix and  $\beta = \sigma_o^2 / \sigma_i^2$  to remove the bias from the TLS solution when  $\sigma_i^2 \neq \sigma_o^2$  [10].

An exact solution for the constrained problem of (1) is hard to obtain. Therefore, we propose an approximate solution:

$$\mathbf{w}_n = \arg \min_{\mathbf{w}} \frac{\left\| \begin{bmatrix} \mu \mathbf{I}_K & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{C}^T & \mathbf{f} \\ \mathbf{X}_n^T & \mathbf{d}_n \end{bmatrix} \mathbf{T} \mathbf{T}^{-1} \begin{bmatrix} \mathbf{w} \\ -1 \end{bmatrix} \right\|^2}{\left\| \mathbf{T}^{-1} \begin{bmatrix} \mathbf{w} \\ -1 \end{bmatrix} \right\|^2}. \quad (2)$$

where  $\mu \gg 1$  is a weighting factor that emphasizes the certainty of the constraints. Note that in (2), the constrained TLS problem of (1) is converted to an unconstrained TLS problem that incorporates the constraints. The idea behind this approach is similar to the one behind the method of weighting [12], which is applicable to the constrained LS problems.

We can rewrite (2) as

$$\mathbf{w}_n = \arg \min_{\mathbf{w}} \frac{\mathbf{w}^T (\mathbf{R}_n + \mu^2 \mathbf{C} \mathbf{C}^T) \mathbf{w} - 2 \mathbf{w}^T (\mathbf{p}_n + \mu^2 \mathbf{C} \mathbf{f}) + \delta_n + \mu^2 \mathbf{f}^T \mathbf{f}}{\|\mathbf{w}\|^2 + \beta}. \quad (3)$$

where

$$\begin{aligned}\mathbf{R}_n &= \mathbf{X}_n \mathbf{D}^2 \mathbf{X}_n^T \\ &= \sum_{i=0}^n \lambda^{n-i} \mathbf{x}_i \mathbf{x}_i^T \\ &= \lambda \mathbf{R}_{n-1} + \mathbf{x}_n \mathbf{x}_n^T,\end{aligned}$$

$$\begin{aligned}\mathbf{p}_n &= \mathbf{X}_n \mathbf{D}^2 \mathbf{d}_n \\ &= \sum_{i=0}^n \lambda^{n-i} \mathbf{x}_i d_i \\ &= \lambda \mathbf{p}_{n-1} + \mathbf{x}_n d_n,\end{aligned}$$

and

$$\begin{aligned}\delta_n &= \mathbf{d}_n^T \mathbf{D}^2 \mathbf{d}_n \\ &= \sum_{i=0}^n \lambda^{n-i} d_i^2 \\ &= \lambda \delta_{n-1} + d_n^2.\end{aligned}$$

Inspired by the line-search algorithms [13] and similar to the approach taken in [11], we consider a first-order autoregressive model for the adaptive filter weights and search them along the input vector:

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \alpha_n \mathbf{x}_n.$$

Hence, the problem boils down to finding  $\alpha_n$  such that

$$\alpha_n = \arg \min_{\alpha} J_n(\alpha) \quad (4)$$

where

$$\begin{aligned}J_n(\alpha) &= \frac{(\mathbf{w}_{n-1} + \alpha \mathbf{x}_n)^T (\mathbf{R}_n + \mu^2 \mathbf{C} \mathbf{C}^T) (\mathbf{w}_{n-1} + \alpha \mathbf{x}_n) - 2(\mathbf{w}_{n-1} + \alpha \mathbf{x}_n)^T (\mathbf{p}_n + \mu^2 \mathbf{C} \mathbf{f}) + \delta_n + \mu^2 \mathbf{f}^T \mathbf{f}}{\|\mathbf{w}_{n-1} + \alpha \mathbf{x}_n\|^2 + \beta}.\end{aligned}$$

The optimization problem of (4) is solved by equating the gradient of the cost function with respect to its argument to zero:

$$\left. \frac{\partial J_n(\alpha)}{\partial \alpha} \right|_{\alpha=\alpha_n} = 0,$$

which leads to

$$\begin{aligned} & (\mathbf{w}_{n-1}^T \mathbf{g}_n - \mathbf{x}_n^T \mathbf{p}_n + \alpha_n \mathbf{x}_n^T \mathbf{g}_n + \mu^2 \mathbf{s}_n \mathbf{r}_n + \mu^2 \alpha_n \|\mathbf{s}_n\|^2) \\ & \times (\|\mathbf{w}_{n-1} + \alpha_n \mathbf{x}_n\|^2 + \beta) - (\mathbf{w}_{n-1}^T \mathbf{x}_n + \alpha_n \|\mathbf{x}_n\|^2) \eta_n (\|\mathbf{w}_n\|^2 + \beta) = 0 \end{aligned} \quad (5)$$

where we define the auxiliary variables

$$\begin{aligned}\mathbf{s}_n &= \mathbf{x}_n^T \mathbf{C}, \\ \mathbf{r}_n &= \mathbf{C}^T \mathbf{w}_{n-1} - \mathbf{f}, \\ \mathbf{g}_n &= \mathbf{R}_n \mathbf{x}_n,\end{aligned}$$

and

$$\eta_n = \frac{\mathbf{w}_n^T (\mathbf{R}_n + \mu^2 \mathbf{C} \mathbf{C}^T) \mathbf{w}_n - 2 \mathbf{w}_n^T (\mathbf{p}_n + \mu^2 \mathbf{C} \mathbf{f}) + \delta_n + \mu^2 \mathbf{f}^T \mathbf{f}}{\|\mathbf{w}_n\|^2 + \beta}.$$

It is easy to verify that

$$\eta_n = \frac{\zeta_n + 2\alpha_n (\mathbf{w}_{n-1}^T \mathbf{g}_n - \mathbf{x}_n^T \mathbf{p}_n + \mu^2 \mathbf{s}_n \mathbf{r}_n) + \alpha_n^2 (\mathbf{x}_n^T \mathbf{g}_n + \mu^2 \|\mathbf{s}_n\|^2)}{\|\mathbf{w}_n\|^2 + \beta}$$

where

$$\zeta_n = \mathbf{w}_{n-1}^T (\mathbf{R}_n + \mu^2 \mathbf{C} \mathbf{C}^T) \mathbf{w}_{n-1} - 2 \mathbf{w}_{n-1}^T (\mathbf{p}_n + \mu^2 \mathbf{C} \mathbf{f}) + \delta_n + \mu^2 \mathbf{f}^T \mathbf{f}$$

or equivalently

$$\zeta_n = \lambda \eta_{n-1} (\|\mathbf{w}_{n-1}\|^2 + \beta) + (\mathbf{w}_{n-1}^T \mathbf{x}_n - d_n)^2 + \mu^2 (1 - \lambda) \|\mathbf{r}_n\|^2 - \mu^2 (1 - \lambda) \mathbf{f}^T \mathbf{f}.$$

Using the definition of  $\eta_n$ , (5) becomes

$$\begin{aligned}& (\mathbf{w}_{n-1}^T \mathbf{g}_n - \mathbf{x}_n^T \mathbf{p}_n + \alpha_n \mathbf{x}_n^T \mathbf{g}_n + \mu^2 \mathbf{s}_n \mathbf{r}_n + \mu^2 \alpha_n \|\mathbf{s}_n\|^2) \\ & \times (\|\mathbf{w}_{n-1}\|^2 + \beta + 2\alpha_n \mathbf{w}_{n-1}^T \mathbf{x}_n + \alpha_n^2 \|\mathbf{x}_n\|^2) - (\mathbf{w}_{n-1}^T \mathbf{x}_n + \alpha_n \|\mathbf{x}_n\|^2) \\ & \times (\zeta_n + 2\alpha_n \mathbf{w}_{n-1}^T \mathbf{g}_n - 2\alpha_n \mathbf{x}_n^T \mathbf{p}_n + 2\mu^2 \alpha_n \mathbf{s}_n \mathbf{r}_n + \alpha_n^2 \mathbf{x}_n^T \mathbf{g}_n + \mu^2 \alpha_n^2 \|\mathbf{s}_n\|^2) = 0,\end{aligned}$$

which can be written in the form of a quadratic polynomial equation:

$$a_n \alpha_n^2 + b_n \alpha_n + c_n = 0 \tag{6}$$

with

$$\begin{aligned}a_n &= \mathbf{w}_{n-1}^T \mathbf{x}_n (\mathbf{x}_n^T \mathbf{g}_n + \mu^2 \|\mathbf{s}_n\|^2) - \|\mathbf{x}_n\|^2 (\mathbf{w}_{n-1}^T \mathbf{g}_n - \mathbf{x}_n^T \mathbf{p}_n + \mu^2 \mathbf{s}_n \mathbf{r}_n), \\ b_n &= (\|\mathbf{w}_{n-1}\|^2 + \beta) (\mathbf{x}_n^T \mathbf{g}_n + \mu^2 \|\mathbf{s}_n\|^2) - \zeta_n \|\mathbf{x}_n\|^2,\end{aligned}$$

and

$$c_n = (\|\mathbf{w}_{n-1}\|^2 + \beta) (\mathbf{w}_{n-1}^T \mathbf{g}_n - \mathbf{x}_n^T \mathbf{p}_n + \mu^2 \mathbf{s}_n \mathbf{r}_n) - \zeta_n \mathbf{w}_{n-1}^T \mathbf{x}_n.$$

The smaller root of (6) in absolute value is the desired solution [10], [11]:

$$\alpha_n = \frac{-b_n + \sqrt{b_n^2 - 4a_n c_n}}{2a_n}.$$

We call the proposed algorithm *constrained recursive total least-squares* (CRTLS) and summarize it in Table 1 where  $\mathbf{g}_n$  is updated using the technique introduced in [14] to minimize the complexity. We also present the total number of required multiplications and additions per iteration by the CRTLS and CRLS algorithms in Table 2.

Note that the cost function in (2) is in fact the Rayleigh quotient of  $\mathbf{M}_n^T \mathbf{M}_n$ ,

$$\mathbf{M}_n = \begin{bmatrix} \mu \mathbf{I}_K & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{C}^T & \mathbf{f} \\ \mathbf{X}_n^T & \mathbf{d}_n \end{bmatrix} \mathbf{T},$$

which reaches its minimum value  $\lambda_{\min}$  (the smallest eigenvalue of  $\mathbf{M}_n^T \mathbf{M}_n$ ) when its argument  $\mathbf{T}^{-1} \begin{bmatrix} \mathbf{w} \\ -1 \end{bmatrix}$  is the eigenvector corresponding to  $\lambda_{\min}$  [15]. It has been proved that  $\lambda_{\min}$  is the unique stable critical value (the only local and global minimum) of the Rayleigh quotient cost function and can be reached using the line-search approach (see, e.g., [10], [11], [16]). Similar to the algorithms of [10] and [11], the proposed algorithm minimizes a Rayleigh quotient cost function by means of the line search and updates the filter weights at each iteration by taking a step along the search direction (input vector) such that the gradient of the cost function evaluated at the updated filter weights becomes orthogonal to the search (update) direction (see [10, Section VI], [11, Lemma 5.1]). Accordingly and in the light of the Theorems 3.1 and 5.1 of [11], it is straightforward to verify that convergence of the proposed algorithm to the unique solution of (2), which is the global minimum of the corresponding Rayleigh quotient cost function, is guaranteed.

### 3. Simulations

We consider an errors-in-variables system identification problem where the filter length is an odd number and the filter weights are constrained to preserve linear phase at each iteration. Thus, the constraint matrix is  $\mathbf{C} = [\mathbf{I}_K, \mathbf{0}, -\mathbf{J}_K]^T \in \mathbb{R}^{L \times K}$  and the response vector is  $\mathbf{f} = \mathbf{0} \in \mathbb{R}^K$  where  $K = (L - 1)/2$  and  $\mathbf{J}_K \in \mathbb{R}^{K \times K}$  is the reversal matrix (identity matrix with all rows in reversed order). The input signal is zero-mean unit-variance Gaussian. The unknown system is an arbitrary finite-impulse-response filter of order  $L = 15$  with linear phase and unit energy. We set  $\lambda = 0.998$  and use two performance metrics: misalignment, defined as  $E[\|\mathbf{h} - \mathbf{w}_n\|^2]$ , and constraint mismatch, defined as  $E[\|\mathbf{C}^T \mathbf{w}_n - \mathbf{f}\|^2]$ , where  $E[\cdot]$  is the expectation operator and the expected values are computed by ensemble-averaging over  $10^3$  independent runs.

In Figs. 1 and 3, we compare the misalignment curves of the CRLS algorithm and the proposed algorithm with different values of  $\mu$  when  $\sigma_i^2 = \sigma_o^2 = 0.5$  and  $\sigma_i^2 = \sigma_o^2 = 0.1$ . It is clear that the larger  $\mu$  is, the smaller the steady-state misalignment is but the slower the algorithm converges and vice versa. In Figs. 2 and 4, we plot the constraint

mismatch of the proposed algorithm as a function of  $\mu$  for the two different values of the noise powers. As expected, with larger  $\mu$ , the constraints are better satisfied.

In this experiment, the CRLS algorithm requires 2944 multiplications and 2702 additions per iteration while the CRTLS algorithm requires only 502 multiplications and 433 additions per iteration.

#### 4. Conclusion

We have developed a linearly-constrained recursive total least-squares (CRTLS) adaptive filtering algorithm by incorporating the linear constraints into the TLS problem in hand and minimizing the resultant cost function on a linear manifold defined by the input vector. The new algorithm considerably outperforms the CRLS algorithm when both input and output data are noisy. Moreover, it requires appreciably less arithmetic operations than CRLS.

We observe that a large value of the weighting parameter,  $\mu$ , results in a lower misalignment (better estimation) and a lower constraint mismatch (better satisfaction of the constraints) while it slows down the convergence. On the other hand, a small  $\mu$  results in a faster convergence but to a higher misalignment and constraint mismatch. Therefore, there exists a trade-off between estimation accuracy and satisfaction of the constraints on one hand and the convergence rate of the proposed algorithm on the other.

#### References

- [1] M. L. R. de Campos, S. Werner, and J. A. Apolinário Jr., "Constrained adaptive filters," in *Adaptive Antenna Arrays: Trends and Applications*, S. Chrandran, Ed. New York: Springer-Verlag, 2004.
- [2] H. L. Van Trees, *Detection, Estimation, and Modulation Theory Part IV: Optimum Array Processing*, New York: Wiley, 2002.
- [3] S. Verdu, *Multuser Detection*, Cambridge, U.K.: Cambridge University Press, 1998.
- [4] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementations*, 3<sup>rd</sup> ed., Boston: Springer, 2008.
- [5] O. L. Frost III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.
- [6] S. Werner, J. A. Apolinário Jr., M. L. R. de Campos, and P. S. R. Diniz, "Low-complexity constrained affine-projection algorithms," *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 4545–4555, Dec. 2005.

- [7] L. S. Resende, J. M. T. Romano, and M. G. Bellanger, "A fast least-squares algorithm for linearly constrained adaptive filtering," *IEEE Trans. Signal Process.*, vol. 44, no. 5, pp. 1168–1174, May 1996.
- [8] J. A. Apolinário Jr., M. L. R. de Campos, and C. P. Bernal, "The constrained conjugate-gradient algorithm," *Signal Process. Lett.*, vol. 7, no. 12, pp. 351–354, Dec. 2000.
- [9] G. H. Golub and C. F. Van Loan, "An analysis of the total least squares problem," *SIAM J. Numer. Anal.*, vol. 17, no. 6, pp. 883–893, Dec. 1980.
- [10] C. E. Davila, "An efficient total least squares algorithm for FIR adaptive filtering," *IEEE Trans. Signal Process.*, vol. 42, no. 2, pp. 268–280, Feb. 1994.
- [11] D.-Z. Feng, X.-D. Zhang, D.-X. Chang, and W. X. Zheng, "A fast recursive total least squares algorithm for adaptive FIR filtering," *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 2729–2737, Oct. 2004.
- [12] C. Van Loan, "On the method of weighting for equality-constrained least-squares problems," *SIAM J. Numer. Anal.*, vol. 22, no. 5, pp. 851–864, Oct. 1985.
- [13] C. E. Davila, "Line search algorithms for adaptive filtering," *IEEE Trans. Signal Process.*, vol. 41, no. 7, pp. 2490–2494, Jul. 1993.
- [14] C. E. Davila, "Efficient, high performance, subspace tracking for time-domain data," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3307–3315, Dec. 2000.
- [15] R. A. Horn and C. A. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
- [16] M. S. Bazaraa, C. M. Shetty, and J. J. Goode, "On the convergence of a class of algorithms using linearly independent search directions," *Math. Prog.*, vol. 18, pp. 89–93, 1980.

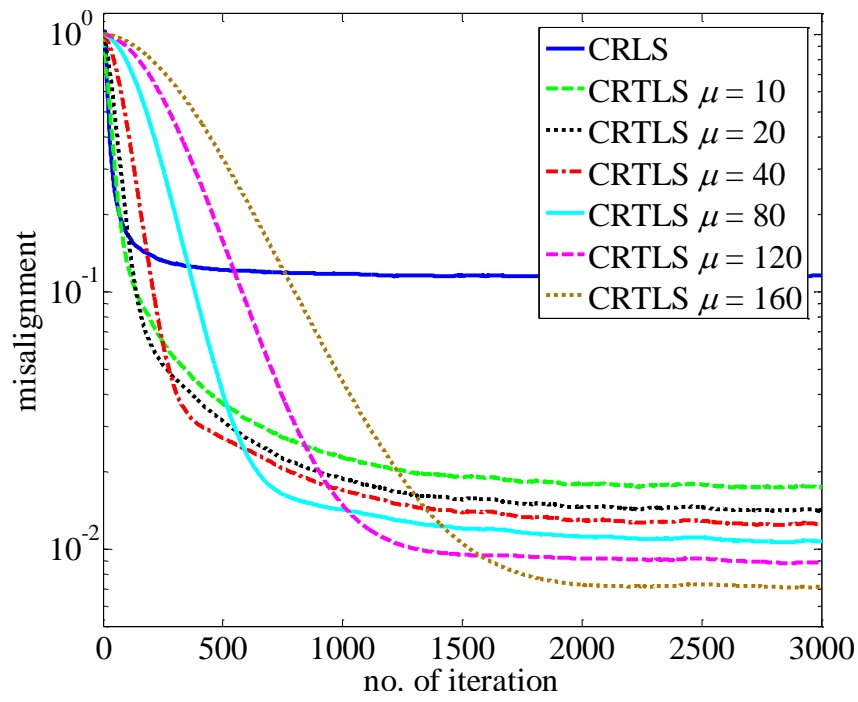


Fig. 1, Misalignment of the CRLS algorithm and the CRTLS algorithm with different values of  $\mu$  when  $\sigma_i^2 = \sigma_o^2 = 0.5$ .

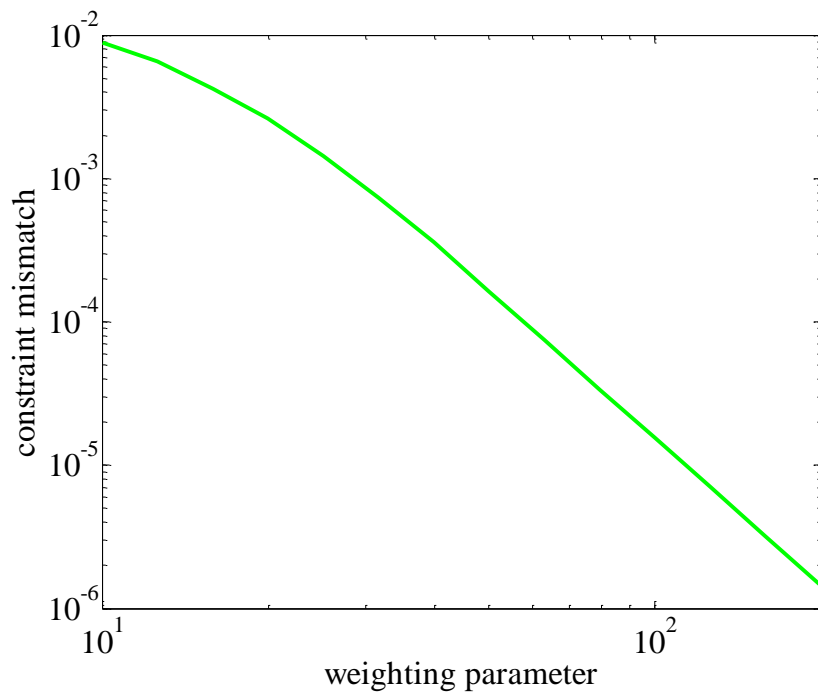


Fig. 2, Constraint mismatch of the CRTLS algorithm as a function of  $\mu$  when  $\sigma_i^2 = \sigma_o^2 = 0.5$ .

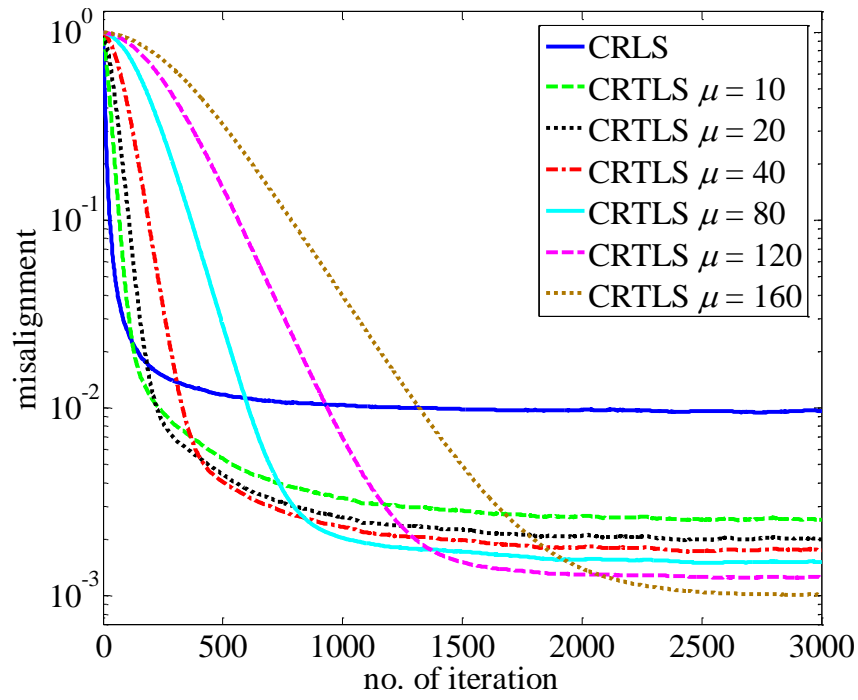


Fig. 3, Misalignment of the CRLS algorithm and the CRTLS algorithm with different values of  $\mu$  when  $\sigma_i^2 = \sigma_o^2 = 0.1$ .

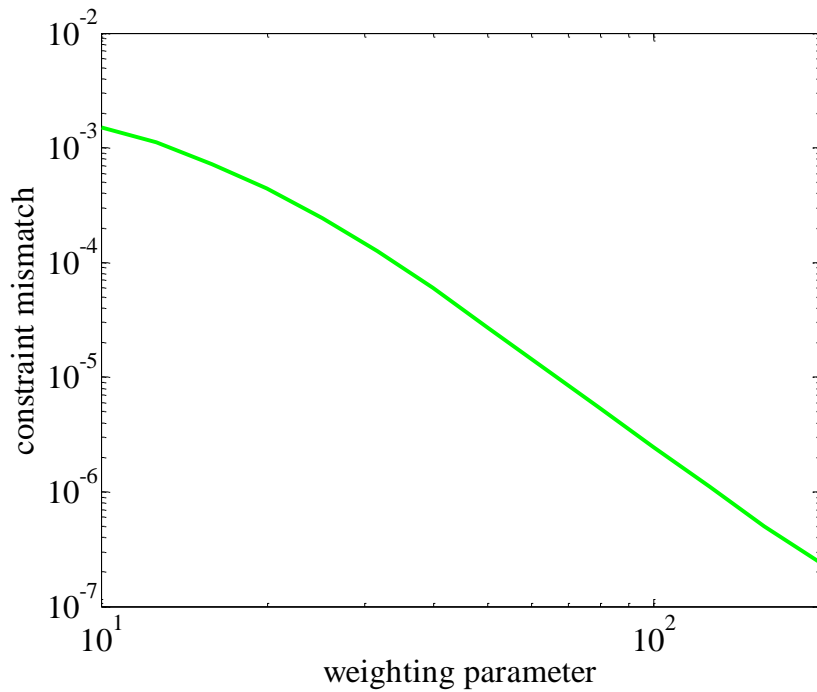


Fig. 4, Constraint mismatch of the CRTLS algorithm as a function of  $\mu$  when  $\sigma_i^2 = \sigma_o^2 = 0.1$ .

Table 1, The CRTLS algorithm.

---

Initialization

$$\boldsymbol{\rho}_0 = \mathbf{0}, \boldsymbol{q}_0 = \mathbf{0}, \epsilon_0 = 0, \mathbf{g}_0 = \mathbf{0}$$

$$\mathbf{p}_0 = \mathbf{0}, \eta_0 = 0$$

$$\mathbf{w}_0 = \mathbf{C}(\mathbf{C}^T \mathbf{C})^{-1} \mathbf{f}$$

At iteration  $n = 1, 2, \dots$

$$\boldsymbol{\rho}_n = \lambda \boldsymbol{\rho}_{n-1} + \mathbf{x}_{n-1} x_n$$

$$\boldsymbol{q}_n = \lambda \boldsymbol{q}_{n-1} + \mathbf{x}_n x_{n-L}$$

$$\epsilon_n = \lambda \epsilon_{n-1} + x_n^2$$

$$\mathbf{g}_n = \begin{bmatrix} \epsilon_n x_n + \boldsymbol{\rho}_n^T \mathbf{x}_{n-1} \\ \boldsymbol{\rho}_n x_n + \mathbf{g}_{n-1} \end{bmatrix}_{1:L-1} - \boldsymbol{q}_n x_{n-L}$$

$$\mathbf{s}_n = \mathbf{x}_n^T \mathbf{C}$$

$$\mathbf{r}_n = \mathbf{C}^T \mathbf{w}_{n-1} - \mathbf{f}$$

$$\mathbf{p}_n = \lambda \mathbf{p}_{n-1} + \mathbf{x}_n d_n$$

$$\zeta_n = \lambda \eta_{n-1} (\|\mathbf{w}_{n-1}\|^2 + \beta) + (\mathbf{w}_{n-1}^T \mathbf{x}_n - d_n)^2 + \mu^2 (1 - \lambda) (\|\mathbf{r}_n\|^2 - \mathbf{f}^T \mathbf{f})$$

$$a_n = \mathbf{w}_{n-1}^T \mathbf{x}_n (\mathbf{x}_n^T \mathbf{g}_n + \mu^2 \|\mathbf{s}_n\|^2) - \|\mathbf{x}_n\|^2 (\mathbf{w}_{n-1}^T \mathbf{g}_n - \mathbf{x}_n^T \mathbf{p}_n + \mu^2 \mathbf{s}_n \mathbf{r}_n)$$

$$b_n = (\|\mathbf{w}_{n-1}\|^2 + \beta) (\mathbf{x}_n^T \mathbf{g}_n + \mu^2 \|\mathbf{s}_n\|^2) - \zeta_n \|\mathbf{x}_n\|^2$$

$$c_n = (\|\mathbf{w}_{n-1}\|^2 + \beta) (\mathbf{w}_{n-1}^T \mathbf{g}_n - \mathbf{x}_n^T \mathbf{p}_n + \mu^2 \mathbf{s}_n \mathbf{r}_n) - \zeta_n \mathbf{w}_{n-1}^T \mathbf{x}_n$$

$$\alpha_n = (-b_n + \sqrt{b_n^2 - 4a_n c_n}) / 2a_n$$

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \alpha_n \mathbf{x}_n$$

$$\eta_n = \{\zeta_n + 2\alpha_n (\mathbf{w}_{n-1}^T \mathbf{g}_n - \mathbf{x}_n^T \mathbf{p}_n + \mu^2 \mathbf{s}_n \mathbf{r}_n) + \alpha_n^2 (\mathbf{x}_n^T \mathbf{g}_n + \mu^2 \|\mathbf{s}_n\|^2)\} / (\|\mathbf{w}_n\|^2 + \beta)$$


---

Table 2, Computational complexity of the CRLS and CRTLS algorithms in terms of number of required arithmetic operations per iteration.

---

	×	+
CRLS	$(3K^2 + 5K + 9)L + K^2 + 2K + 16$	$(3K^2 + 4K + 8)L - K^2 + K - 1$
CRTLS	$(2K + 16)L + 4K + 24$	$(2K + 13)L + 3K + 7$

---

# Paper M

Originally published as

R. Arablouei and K. Doğançay, "Linearly-constrained line-search algorithm for adaptive filtering," *Electronics Letters*, vol. 48, no. 19, pp. 1208-1209, 2012.

Copyright © 2012 IET. Reprinted with permission.

## Linearly-constrained line-search algorithm for adaptive filtering

Reza Arablouei and Kutluyıl Doğançay

### Abstract

We develop a linearly-constrained line-search adaptive filtering algorithm by incorporating the linear constraints into the least-squares problem and searching the solution (filter weights) along the Kalman gain vector. The proposed algorithm performs close to the constrained recursive least-squares (CRLS) algorithm while having a computational complexity comparable to the constrained least mean square (CLMS) algorithm. Simulations demonstrate its effectiveness.

### 1. Introduction

Linearly-constrained adaptive filtering algorithms are powerful signal processing tools with widespread use in several applications such as beamforming, blind multiuser detection in code-division multiple-access systems, and system identification [1]. The linear constraints usually emerge from the prior knowledge about the system, e.g., directions of arrival in array processing, spreading codes in blind multiuser detection, and linear phase in system identification [2].

Several linearly-constrained adaptive filtering algorithms have been proposed, which can be loosely divided into two categories:

- 1) stochastic-gradient-based algorithms such as the constrained least mean square (CLMS) algorithm [3] and the constrained affine projection (CAP) algorithm [2];
- 2) least-squares(LS)-based algorithms such as the constrained recursive least-squares (CRLS) algorithm [4] and the constrained conjugate-gradient (CCG) algorithm [2].

The algorithms of the former category are generally simpler, more robust, and computationally more efficient than the algorithms of the latter category. However, they typically have slower convergence rate that is further aggravated when the input signal is correlated.

In this letter, we develop a linearly-constrained line-search algorithm by converting the constrained LS problem to an unconstrained one, which incorporates the linear constraints, and then solving it on a linear manifold defined by the Kalman gain vector. The proposed algorithm features CRLS-like performance with CLMS-like complexity.

## 2. Algorithm description

Let us consider a linear system described by

$$d_n = \boldsymbol{\omega}^T \mathbf{x}_n + v_n$$

where  $d_n \in \mathbb{R}$  is the desired signal at time index  $n \in \mathbb{N}$ ,  $\boldsymbol{\omega} \in \mathbb{R}^L$  is the vector of unknown system parameters,  $\mathbf{x}_n = [x_n, x_{n-1}, \dots, x_{n-L+1}]^T$  is the input vector,  $x_n \in \mathbb{R}$  is the input signal,  $v_n \in \mathbb{R}$  is the background noise, and superscript  $T$  stands for matrix transposition.

Adaptive filtering algorithms iteratively estimate  $\boldsymbol{\omega}$  exploiting the data available up to the current time. The tap weights of the adaptive filter, denoted by  $\mathbf{w}_n \in \mathbb{R}^L$ , are taken as the estimate at iteration  $n$ . In linearly-constrained adaptive filtering, a set of constraints is imposed upon the filter weights at each time instant as

$$\mathbf{C}^T \mathbf{w}_n = \mathbf{f}$$

where  $\mathbf{C} \in \mathbb{R}^{L \times K}$  is the constraint matrix ( $K < L$ ) and  $\mathbf{f} \in \mathbb{R}^K$  is the response vector.

A constrained exponentially-weighted LS estimate can be found as

$$\begin{aligned} \mathbf{w}_n &= \arg \min_{\mathbf{w}} J_n(\mathbf{w}) \\ &\text{subject to } \mathbf{C}^T \mathbf{w} = \mathbf{f} \end{aligned} \quad (1)$$

where

$$\begin{aligned} J_n(\mathbf{w}) &= \|\mathbf{X}_n^T \mathbf{w} - \mathbf{d}_n\|^2, \\ \mathbf{X}_n &= [\mathbf{x}_n, \lambda^{1/2} \mathbf{x}_{n-1}, \dots, \lambda^{(n-1)/2} \mathbf{x}_1], \\ \mathbf{d}_n &= [d_n, \lambda^{1/2} d_{n-1}, \dots, \lambda^{(n-1)/2} d_1]^T, \end{aligned}$$

$\|\cdot\|$  denotes the Euclidean norm and  $0 \ll \lambda \leq 1$  is the forgetting factor. The solution of (1) obtained through the method of Lagrange multipliers is [4]

$$\mathbf{w}_n = \mathbf{R}_n^{-1} \mathbf{p}_n + \mathbf{R}_n^{-1} \mathbf{C} (\mathbf{C}^T \mathbf{R}_n^{-1} \mathbf{C})^{-1} (\mathbf{f} - \mathbf{C}^T \mathbf{R}_n^{-1} \mathbf{p}_n) \quad (2)$$

where

$$\begin{aligned} \mathbf{R}_n &= \mathbf{X}_n \mathbf{X}_n^T \\ &= \sum_{i=1}^n \lambda^{n-i} \mathbf{x}_i \mathbf{x}_i^T \end{aligned}$$

is the exponentially-weighted input autocorrelation matrix and

$$\begin{aligned} \mathbf{p}_n &= \mathbf{X}_n \mathbf{d}_n \\ &= \sum_{i=1}^n \lambda^{n-i} \mathbf{x}_i d_i \\ &= \lambda \mathbf{p}_{n-1} + \mathbf{x}_n d_n \end{aligned}$$

is the exponentially-weighted cross-correlation vector between the input vector and the desired signal. The CRLS algorithm [4] is a recursive calculation of (2) that avoids the matrix inversions by applying the matrix inversion lemma.

By relaxing the hard constraints, we can convert (1) to an unconstrained problem that incorporates the constraints:

$$\mathbf{w}_n = \arg \min_{\mathbf{w}} \tilde{J}_n(\mathbf{w})$$

where

$$\tilde{J}_n(\mathbf{w}) = \|\mathbf{X}_n^T \mathbf{w} - \mathbf{d}_n\|^2 + \mu^2 \|\mathbf{C}^T \mathbf{w} - \mathbf{f}\|^2$$

and  $\mu \gg 1$  is a weighting parameter that emphasises the certainty of the constraints. Note that this approach is the same as the method of weighting [5] and yields an approximate solution.

In the same vein as the line-search algorithms [6], we consider a first-order autoregressive model for the adaptive filter weights and search them along the Kalman gain vector,  $\mathbf{k}_n = \mathbf{R}_n^{-1} \mathbf{x}_n$ , as

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \alpha_n \mathbf{k}_n.$$

Hence, the problem turns into finding  $\alpha_n$  such that

$$\alpha_n = \arg \min_{\alpha} \tilde{J}_n(\alpha) \quad (3)$$

where

$$\begin{aligned}
\tilde{J}_n(\alpha) &= \|\mathbf{X}_n^T(\mathbf{w}_{n-1} + \alpha\mathbf{k}_n) - \mathbf{d}_n\|^2 + \mu^2\|\mathbf{C}^T(\mathbf{w}_{n-1} + \alpha\mathbf{k}_n) - \mathbf{f}\|^2 \\
&= \mathbf{w}_{n-1}^T\mathbf{R}_n\mathbf{w}_{n-1} - 2\mathbf{w}_{n-1}^T\mathbf{p}_n + \delta_n + \mu^2\|\mathbf{r}_n\|^2 \\
&\quad + 2\alpha(\mathbf{w}_{n-1}^T\mathbf{x}_n - \mathbf{k}_n^T\mathbf{p}_n + \mu^2\mathbf{r}_n^T\mathbf{s}_n) + \alpha^2(\mathbf{x}_n^T\mathbf{k}_n + \mu^2\|\mathbf{s}_n\|^2)
\end{aligned}$$

and the auxiliary variables are defined as

$$\mathbf{s}_n = \mathbf{C}^T\mathbf{k}_n,$$

$$\mathbf{r}_n = \mathbf{C}^T\mathbf{w}_{n-1} - \mathbf{f}.$$

Equating the derivative of  $\tilde{J}_n(\alpha)$  with respect to  $\alpha$  to zero gives the solution of (3), which is

$$\alpha_n = \frac{\mathbf{k}_n^T\mathbf{p}_n - \mathbf{w}_{n-1}^T\mathbf{x}_n - \mu^2\mathbf{r}_n^T\mathbf{s}_n}{\mathbf{x}_n^T\mathbf{k}_n + \mu^2\|\mathbf{s}_n\|^2}.$$

We call the proposed algorithm *constrained line-search* (CLS) and summarise it in Table 1 where, similar to CRLS,  $\mathbf{k}_n$  is updated using the stabilised fast transversal filter (S-FTF) algorithm [7]. The proposed algorithm requires  $(2K + 13)L + 2K + 18$  multiplications per iteration while the CRLS and CLMS algorithms require  $(3K^2 + 5K + 9)L + K^2 + 2K + 16$  and  $L^2 + 2L + 1$  multiplications per iteration, respectively.

### 3. Simulations

We consider a system identification problem where the filter length is an odd number and the filter weights are constrained to preserve linear phase at each iteration. Thus, the constraint matrix is  $\mathbf{C} = [\mathbf{I}_K, \mathbf{0}, -\mathbf{J}_K]^T \in \mathbb{R}^{L \times K}$  and the response vector is  $\mathbf{f} = \mathbf{0} \in \mathbb{R}^K$  where  $K = (L - 1)/2$ ,  $\mathbf{I}_K \in \mathbb{R}^{K \times K}$  is the identity matrix and  $\mathbf{J}_K \in \mathbb{R}^{K \times K}$  is the reversal matrix (identity matrix with all rows in reversed order). The input signal is zero-mean unit-variance Gaussian. The unknown system is an arbitrary finite-impulse-response filter of order  $L = 15$  with linear phase and unit energy. The background noise is white Gaussian with a power of 0.1. We set  $\lambda = 0.995$  and obtain the results by ensemble-averaging over  $10^3$  independent runs.

In Fig. 1, we plot the misalignment curves of the CRLS and CLMS algorithms as well as those of the proposed algorithm with different values of  $\mu$ . The misalignment is defined as  $\|\boldsymbol{\omega} - \mathbf{w}_n\|^2$ . It is clear that the larger  $\mu$  is, the smaller the steady-state misalignment is but the slower the algorithm converges and vice versa. In Fig. 2, we plot the constraint mismatch of the proposed algorithm as a function of  $\mu$ . The constraint mismatch is defined as  $\|\mathbf{C}^T\mathbf{w}_n - \mathbf{f}\|^2$ . As expected, with larger  $\mu$ , the constraints are better satisfied.

In this experiment, the CRLS and CLMS algorithms respectively require 2944 and 256 multiplications per iteration while the CLS algorithm requires 437 multiplications per iteration.

#### 4. Conclusion

We have developed a linearly-constrained line-search (CLS) adaptive filtering algorithm by incorporating the linear constraints into the LS problem and minimising the resultant cost function on a linear manifold. The new algorithm performs nearly as good as the CRLS algorithm with a considerably reduced computational complexity, which is comparable to the CLMS algorithm.

We observe that a large value of the weighting parameter,  $\mu$ , results in a lower misalignment (better estimation) and a lower constraint mismatch (better satisfaction of the constraints) while it slows down the convergence. On the other hand, a small  $\mu$  results in a faster convergence but to a higher misalignment and constraint mismatch. Therefore, there exists a trade-off between estimation accuracy and satisfaction of the constraints on one hand and the convergence rate of the proposed algorithm on the other.

#### References

- [1] M. L. R. de Campos, S. Werner, and J. A. Apolinário Jr., "Constrained adaptive filters," in *Adaptive Antenna Arrays: Trends and Applications*, S. Chandraan, Ed. New York: Springer-Verlag, 2004.
- [2] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementations*, 3<sup>rd</sup> ed., Boston: Springer, 2008.
- [3] O. L. Frost III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.
- [4] L. S. Resende, J. M. T. Romano, and M. G. Bellanger, "A fast least-squares algorithm for linearly constrained adaptive filtering," *IEEE Trans. Signal Process.*, vol. 44, no. 5, pp. 1168–1174, May 1996.
- [5] C. Van Loan, "On the method of weighting for equality-constrained least-squares problems," *SIAM J. Numer. Anal.*, vol. 22, no. 5, pp. 851–864, Oct. 1985.
- [6] C. E. Davila, "Line search algorithms for adaptive filtering," *IEEE Trans. Signal Process.*, vol. 41, no. 7, pp. 2490–2494, Jul. 1993.
- [7] D. T. M. Slock and T. Kailath, "Numerically stable fast transversal filters for recursive least-squares adaptive filtering," *IEEE Trans. Signal Process.*, vol. 39, pp. 92–114, Jan. 1991.

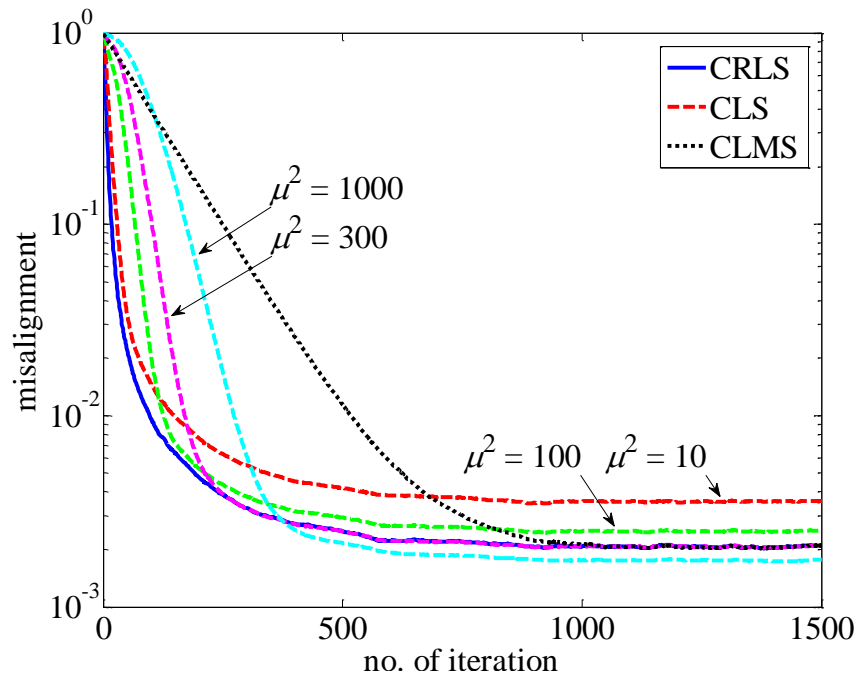


Fig. 1, Misalignment of different algorithms.

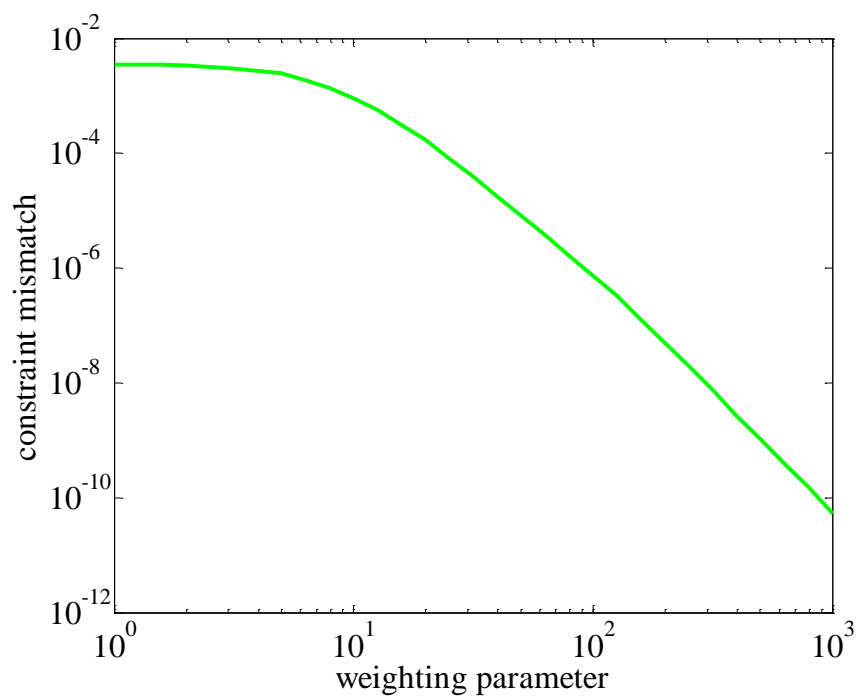


Fig. 2, Constraint mismatch of the CLS algorithm as a function of  $\mu$ .

Table 1, The CLS algorithm.

---

Initialisation

$$\mathbf{p}_0 = \mathbf{0}$$

$$\mathbf{w}_0 = \mathbf{C}(\mathbf{C}^T \mathbf{C})^{-1} \mathbf{f}$$

At iteration  $n = 1, 2, \dots$

update  $\mathbf{k}_n$

$$\mathbf{p}_n = \lambda \mathbf{p}_{n-1} + \mathbf{x}_n d_n$$

$$\mathbf{s}_n = \mathbf{x}_n^T \mathbf{C}$$

$$\mathbf{r}_n = \mathbf{C}^T \mathbf{w}_{n-1} - \mathbf{f}$$

$$\alpha_n = \frac{\mathbf{k}_n^T \mathbf{p}_n - \mathbf{w}_{n-1}^T \mathbf{x}_n - \mu^2 \mathbf{r}_n^T \mathbf{s}_n}{\mathbf{x}_n^T \mathbf{k}_n + \mu^2 \|\mathbf{s}_n\|^2}$$

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \alpha_n \mathbf{k}_n$$

---

# Epilogue

## Summary

In this thesis, several low-complexity adaptive filtering algorithms for various communications applications have been developed and analysed. In the considered applications, the adaptive filters are implemented in *direct*, *inverse*, *set-membership*, and *linearly-constrained* system identification configurations.

The main contributions of the thesis can be summarized as

➤ the new equalizers:

- 1) low-complexity adaptive MIMO decision-feedback equalizer (Paper D)
- 2) adaptive wideband MIMO channel equalizer coupling MIMO-DFE with V-BLAST (Paper E)
- 3) partial-update adaptive decision-feedback equalizer (Paper F),

➤ the new adaptive filtering algorithms:

- 1) affine projection algorithm with selective projections (Paper A)
- 2) proportionate affine projection algorithm with selective projections (Paper B)
- 3) modified RLS algorithm with enhanced tracking capability (Paper C)
- 4) modified quasi-OBE algorithm with improved numerical properties (Paper G)
- 5) low-complexity quasi-OBE algorithm (Paper H)
- 6) low-complexity constrained recursive least-squares algorithm (Paper K)
- 7) linearly-constrained recursive total least-squares algorithm (Paper L)
- 8) linearly-constrained line-search algorithm (Paper M),

➤ and the new theoretical analyses:

- 1) convergence analysis of the low-complexity adaptive MIMO decision-feedback equalizer (Paper D)
- 2) steady-state mean squared error and tracking performance analysis of the quasi-OBE algorithm (Paper I)
- 3) tracking performance analysis of the set-membership NLMS algorithm (Paper J).

The main techniques utilized to mitigate the computational complexity in the proposed algorithms are

- adoption of a *time-varying projection order* for the affine projection algorithm (Papers A and B)
- the concept of *partial updates* (Paper F)

- the *dichotomous coordinate-descent* iterations for solving the associated systems of linear equations (Papers D, H, and K)
- the *method of weighting* for linearly-constrained filters (Papers K, L, and M).

The proposed algorithms deliver good convergence and steady-state performance as well as significant reduction in computational complexity compared with the existing algorithms and techniques. They also offer useful trade-offs between complexity and performance. Finally, the predictions made by the theoretical analyses are in close match with the simulation results.

## Future work

Devising new low-complexity algorithms for the adaptive filtering configurations, which have not been considered in this thesis, is a subject of future research. Prominent examples are

- adaptive *errors-in-variables* system identification
- adaptive *widely-linear* system identification
- adaptive *distributed* system identification.

In errors-in-variables system identification [SV91] applications, usually complicated matrix operations, such as singular value decomposition or eigen-decomposition, are required. In adaptive widely-linear system identification [BP95], the parameter vector of the underlying model is augmented, i.e., the number of sought-after parameters is doubled compared with the case of strictly-linear model. Therefore, computational complexity of the system is accordingly increased. Alleviating the computational burden in these applications is certainly desirable and worthy of additional study.

Adaptive distributed system identification over networks [ST12] can be performed in several ways. Diffusion strategies [AS13] are known to accomplish the task very efficiently. However, they require frequent exchange of data between the neighbouring nodes. Devising a way to reduce the diffusion load as well as the computational complexity without compromising the performance noticeably is a subject for future research.

Theoretical performance analysis of some of the proposed algorithms, e.g., the affine projection algorithm with selective projections (Paper A), is another topic for further investigation.

Finally, actual implementation of the proposed algorithms on hardware and/or software platforms will help substantiate their effectiveness and can provide more insight into their advantages and possible shortcomings.

# Appendices

## Contributions of the co-authors to the publications

I am the primary author of all the publications included in this *thesis containing published research*. My principal supervisor, Associate Professor Kutluyıl Doğançay, is the co-author of all of the papers and my associate supervisor, Dr Sylvie Perreau, is the co-author of three of them (Papers B, E, and F). I have carried out the majority of the works and my own contribution to each paper has been at least 70%. My supervisors (mostly my principle supervisor) have contributed to the publications by

- assisting in the formation of the underpinning ideas
- assisting in planning and organizing the works
- supervising the development of the works
- assisting in the interpretation of the results
- providing critical evaluation of the works
- assessing and editing the manuscripts.



Reza Arablouei

29/Jan./2013

I confirm the above declaration to be correct.



Kutluyıl Doğançay

29/Jan./2013

# Copyright permissions

The permissions for reusing the published works in both print and electronic versions of this thesis have been duly acquired from the publishers and attached herein.

## ELSEVIER LICENSE TERMS AND CONDITIONS

Jan 23, 2013

This is a License Agreement between Reza Arablouei ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, UK
Registered Company Number	1982084
Customer name	Reza Arablouei
Customer address	ITR, W Building, UniSA, Mawson Lakes, SA 5095
License number	3067350606298
License date	Jan 13, 2013
Licensed content publisher	Elsevier
Licensed content publication	Signal Processing
Licensed content title	Affine projection algorithm with selective projections
Licensed content author	Reza Arablouei, Kutluyıl Doğançay
Licensed content date	September 2012
Licensed content volume number	92
Licensed content issue number	9
Number of pages	11
Start Page	2253
End Page	2263
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	Reduced-Complexity Adaptive Filtering Techniques for Communications Applications
Expected completion date	Jan 2013
Estimated size (number of pages)	200
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.0 USD / 0.0 GBP
Total	0.00 USD

## ELSEVIER LICENSE TERMS AND CONDITIONS

Jan 23, 2013

This is a License Agreement between Reza Arablouei ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, UK
Registered Company Number	1982084
Customer name	Reza Arablouei
Customer address	ITR, W Building, UniSA, Mawson Lakes, SA 5095
License number	3067371328362
License date	Jan 13, 2013
Licensed content publisher	Elsevier
Licensed content publication	Signal Processing
Licensed content title	Low-complexity adaptive decision-feedback equalization of MIMO channels
Licensed content author	Reza Arablouei, Kutluyıl Doğançay
Licensed content date	June 2012
Licensed content volume number	92
Licensed content issue number	6
Number of pages	10
Start Page	1515
End Page	1524
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	Reduced-Complexity Adaptive Filtering Techniques for Communications Applications
Expected completion date	Jan 2013
Estimated size (number of pages)	200
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.0 USD / 0.0 GBP
Total	0.00 USD

## ELSEVIER LICENSE TERMS AND CONDITIONS

Jan 23, 2013

This is a License Agreement between Reza Arablouei ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, UK
Registered Company Number	1982084
Customer name	Reza Arablouei
Customer address	ITR, W Building, UniSA, Mawson Lakes, SA 5095
License number	3067381330842
License date	Jan 13, 2013
Licensed content publisher	Elsevier
Licensed content publication	Signal Processing
Licensed content title	Steady-state mean squared error and tracking performance analysis of the quasi-OBE algorithm
Licensed content author	Reza Arablouei, Kutluyıl Doğançay
Licensed content date	January 2013
Licensed content volume number	93
Licensed content issue number	1
Number of pages	9
Start Page	100
End Page	108
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	Reduced-Complexity Adaptive Filtering Techniques for Communications Applications
Expected completion date	Jan 2013
Estimated size (number of pages)	200
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.0 USD / 0.0 GBP
Total	0.00 USD

## ELSEVIER LICENSE TERMS AND CONDITIONS

Jan 23, 2013

This is a License Agreement between Reza Arablouei ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, UK
Registered Company Number	1982084
Customer name	Reza Arablouei
Customer address	ITR, W Building, UniSA, Mawson Lakes, SA 5095
License number	3067380585708
License date	Jan 13, 2013
Licensed content publisher	Elsevier
Licensed content publication	Signal Processing
Licensed content title	Modified quasi-OBE algorithm with improved numerical properties
Licensed content author	Reza Arablouei, Kutluyıl Doğançay
Licensed content date	April 2013
Licensed content volume number	93
Licensed content issue number	4
Number of pages	7
Start Page	797
End Page	803
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	Reduced-Complexity Adaptive Filtering Techniques for Communications Applications
Expected completion date	Jan 2013
Estimated size (number of pages)	200
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.0 USD / 0.0 GBP
Total	0.00 USD

## INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

## GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

“Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER].” Also Lancet special credit - “Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier.”

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at [permissions@elsevier.com](mailto:permissions@elsevier.com))

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and

shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. **Warranties:** Publisher makes no representations or warranties with respect to the licensed material.

10. **Indemnity:** You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. **No Transfer of License:** This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. **No Amendment Except in Writing:** This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. **Objection to Contrary Terms:** Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. **Revocation:** Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

### **LIMITED LICENSE**

The following terms and conditions apply only to specific license types:

15. **Translation:** This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. **Website:** The following terms and conditions apply to electronic reserve and author websites:

**Electronic reserve:** If licensed material is to be posted to website, the web site is to be password-protected and made available only to bona fide students registered on a relevant course if: This license was made in connection with a course, This permission is granted for 1 year only. You may obtain a license for future website posting, All content posted to the web site must maintain the copyright information line on the bottom of each image, A hyper-text must be included to the Homepage of the journal from which you are licensing at <http://www.sciencedirect.com/science/journal/xxxxx> or the Elsevier homepage for books at <http://www.elsevier.com>, and Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

17. **Author website** for journals with the following additional clauses:

All content posted to the web site must maintain the copyright information line on the bottom of each image, and the permission granted is limited to the personal version of your paper. You are not allowed to download and post the published electronic version of your article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version. A hyper-text must be included to the Homepage of the journal from which you are licensing at <http://www.sciencedirect.com/science/journal/xxxxx> . As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier's online service ScienceDirect ([www.sciencedirect.com](http://www.sciencedirect.com)). That e-mail will include the article's Digital Object Identifier (DOI). This number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

18. **Author website** for books with the following additional clauses: Authors are permitted to place a brief summary of their work online only. A hyper-text must be included to the Elsevier homepage at <http://www.elsevier.com>. All content posted to the web site must maintain the copyright information line on the bottom of each image. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

19. **Website** (regular and for author): A hyper-text must be included to the Homepage of the journal from which you are licensing at <http://www.sciencedirect.com/science/journal/xxxxx> or for books to the Elsevier homepage at <http://www.elsevier.com>

20. **Thesis/Dissertation:** If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

**21. Other Conditions:**

v1.6

**If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK500933295.**

**Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.**

**Make Payment To:  
Copyright Clearance Center  
Dept 001  
P.O. Box 843006  
Boston, MA 02284-3006**

**For suggestions or comments regarding this order, contact RightsLink Customer Support: [customercare@copyright.com](mailto:customercare@copyright.com) or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.**

**Gratis licenses (referencing \$0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.**



**Title:** Reduced-Complexity Constrained Recursive Least-Squares Adaptive Filtering Algorithm  
**Author:** Arablouei, R.; Dogancay, K.  
**Publication:** Signal Processing, IEEE Transactions on  
**Publisher:** IEEE  
**Date:** Dec. 2012  
Copyright © 2012, IEEE

Logged in as:  
Reza Arablouei  
Account #:  
3000610896

## Thesis / Dissertation Reuse

**The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:**

*Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:*

- 1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
- 2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
- 3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

*Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:*

- 1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
- 2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
- 3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to [http://www.ieee.org/publications\\_standards/publications/rights/rights\\_link.html](http://www.ieee.org/publications_standards/publications/rights/rights_link.html) to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.



**Title:** Linearly-Constrained Recursive Total Least-Squares Algorithm  
**Author:** Arablouei, R.; Dogancay, K.  
**Publication:** IEEE Signal Processing Letters  
**Publisher:** IEEE  
**Date:** Dec. 2012  
Copyright © 2012, IEEE

Logged in as:  
Reza Arablouei  
Account #:  
3000610896

## Thesis / Dissertation Reuse

**The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:**

*Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:*

- 1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
- 2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
- 3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

*Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:*

- 1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
- 2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
- 3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to [http://www.ieee.org/publications\\_standards/publications/rights/rights\\_link.html](http://www.ieee.org/publications_standards/publications/rights/rights_link.html) to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

**RE: Enquiry from IET Web Site - Copyright Permission Request**

Ratty,Rebecca [RebeccaRatty@theiet.org]

**Sent:** Monday, 21 January 2013 10:19 PM

**To:** Arablouei, Reza - arary003

---

Dear Reza

Permission to reproduce as requested is given, provided that the source of the material including the author, title, date, and publisher is acknowledged.

A reproduction fee is not due to the IET on this occasion because you are the author.

Thank you for your enquiry. Please do not hesitate to contact me should you require any further information.

Kind regards,  
Rebecca

Rebecca Ratty  
Editorial Assistant  
Research Journals  
T: 01438 767309

FAQs  
Author Guide

Michael Faraday House, Six Hills Way, Stevenage, SG1 2AY  
Please consider the environment before printing this email

-----Original Message-----

From: webmaster@theiet.org [<mailto:webmaster@theiet.org>]  
Sent: Friday, January 18, 2013 12:52 AM  
To: Ratty,Rebecca  
Subject: Enquiry from IET Web Site - Copyright Permission Request

A visitor to the IET Web site has used your contact form to send the following enquiry.

The form is here:  
<http://www.theiet.org/resources/contacts/journalsofficer.cfm>

Name: Reza Arablouei

Subject: Copyright Permission Request

Their email address: arary003@mymail.unisa.edu.au

Comments:

Dear Sir/Ma'am,

I wish to request IET's permission for reusing the following three articles, which have been authored by me, in my PhD thesis:

R. Arablouei and K. Doğançay, "Modified RLS algorithm with enhanced tracking capability for MIMO channel estimation," *Electronics Letters*, vol. 47, no. 19, pp. 1101-1103, 2011.

R. Arablouei and K. Doğançay, "Low-complexity implementation of the quasi-OBE algorithm," *Electronics Letters*, vol. 48, no. 11, pp. 621-623, 2012.

R. Arablouei and K. Doğançay, "Linearly-constrained line-search algorithm for adaptive filtering," *Electronics Letters*, vol. 48, no. 19, pp. 1208-1209, 2012.

My thesis is titled: "Reduced-complexity adaptive filtering techniques for communications applications" and will be produced both in print and electronically.

Kind regards,  
Reza Arablouei

---

The Institution of Engineering and Technology is registered as a Charity in England and Wales (No. 211014) and Scotland (No. SC038698). The information transmitted is intended only for the person or entity to which it is addressed and may contain confidential and/or privileged material. Any review, retransmission, dissemination or other use of, or taking of any action in reliance upon, this information by persons or entities other than the intended recipient is prohibited. If you received this in error, please contact the sender and delete the material from any computer. The views expressed in this message are personal and not necessarily those of the IET unless explicitly stated.

**RE: APSIPA conference copyright permission request**

C.-C. Jay Kuo [cckuo@sipi.usc.edu]

**Sent:** Monday, 28 January 2013 2:09 PM

**To:** Arablouei, Reza - arary003

---

Dear Reza,

You are able to re-use these two APSIPA articles in your PhD thesis without any problem.

Jay

---

**From:** Arablouei, Reza - arary003

[mailto:reza.arablouei@mymail.unisa.edu.au]

**Sent:** Sunday, January 27, 2013 7:27 PM

**To:** C.-C. Jay Kuo [cckuo@sipi.usc.edu]

**Subject:** APSIPA conference copyright permission request

Dear Prof Kuo,

I wish to request APSIPA's permission for reusing the following two articles, which have been authored by me, in my PhD thesis:

R. Arablouei, K. Doğançay, and Sylvie Perreau, "Proportionate affine projection algorithm with selective projections for sparse system identification," in Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Singapore, Dec. 2010, pp. 362-366.

R. Arablouei and K. Doğançay, "Tracking performance analysis of the set-membership NLMS algorithm," in Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Hollywood, USA, Dec. 2012, paper id: 200.

My thesis is titled: "Reduced-complexity adaptive filtering techniques for communications applications" and will be produced both in print and electronically.

Kind regards,  
Reza Arablouei



# European Association for Signal Processing

## EUSIPCO-2011 Conference Copyright Agreement

It is required that Authors publishing at EUSIPCO-2011 in Barcelona, Spain provide a Transfer of Copyright to the European Association for Signal Processing ("EURASIP"), in the following just labelled "EURASIP". This empowers "EURASIP" on behalf of the Author to protect the Work and its image against unauthorised use and to properly authorise dissemination of the Work by means of printed publications, offprint, reprints, electronic files, licensed photocopies, microfilm editions, translations, document delivery, and secondary information sources such as abstracting, reviewing and indexing services, including converting the work into machine readable form and storing in electronic databases.

EDAS Paper number: 1569418705

Title of contribution ("Work"): **MIMO-DFE coupled with V-BLAST for adaptive equalization of wideband MIMO channels**

Author(s): **Reza Arablouei, Kutluyıl Doğançay, and Sylvie Perreau**

Name of Publication: 19th European Signal Processing Conference, 2011 (EUSIPCO-2011) The Author(s) hereby consents that the publisher appointed by EURASIP will publish the work.

- (1) The signatory(ies) on this form warrants all other authors/co-authors are properly credited, and generally that the author(s) has the right to make the grants made to "EURASIP" complete and unencumbered. The Author(s) also warrants that the Work is novel and has not been published elsewhere. The Author(s) furthermore warrant that the Work does not libel anyone, infringe anyone's copyright, or otherwise violate anyone's statutory or common law rights.
- (2) The Author(s) hereby transfers to "EURASIP" the copyright of the Work named above. "EURASIP" shall have the exclusive and unlimited right to publish the said Work and to translate (or authorize others to translate) it wholly or in part throughout the World during the full term of copyright including renewals and extensions and all subsidiary rights.
- (3) The Work may be reproduced by any means for educational and scientific purposes by the author or by others without fee or permission with the exception of reproduction by services that collect fee for delivery of documents. The Author(s) may use part or all of this Work or its image in any future works of his/her (their) own. In any reproduction, the original publication by "EURASIP" must be credited in the following manner: "First published in the Proceedings of the 19th European Signal Processing Conference (EUSIPCO-2011) in 2011, published by EURASIP", and such a credit notice must be placed on all copies. Any publication or other form of reproduction not meeting these requirements will be deemed to be unauthorized.
- (4) In the event of receiving any request to reprint or translate all or part of the Work, "EURASIP" shall seek to inform the author. This form is to be signed by the Author(s) or in case of a "work-made-for-hire", by the employer. If there is more than one Author, then either all must sign the Copyright Agreement, or one Author signs in consent for all, taking on full responsibility for the content of the publication.

Date:

30/5/2011

Full name(s):

Reza Arablouei

Signature(s):

A handwritten signature in blue ink that reads "Reza Arablouei".

Please return this form on or before June 6, 2011 by electronic upload of a scanned copy in EDAS.



# European Association for Signal Processing

## EUSIPCO-2011 Conference Copyright Agreement

It is required that Authors publishing at EUSIPCO-2011 in Barcelona, Spain provide a Transfer of Copyright to the European Association for Signal Processing ("EURASIP"), in the following just labelled "EURASIP". This empowers "EURASIP" on behalf of the Author to protect the Work and its image against unauthorised use and to properly authorise dissemination of the Work by means of printed publications, offprint, reprints, electronic files, licensed photocopies, microfilm editions, translations, document delivery, and secondary information sources such as abstracting, reviewing and indexing services, including converting the work into machine readable form and storing in electronic databases.

EDAS Paper number: 1569422151

Title of contribution ("Work"): **Partial-update adaptive decision-feedback equalization**

Author(s): **Reza Arablouei, Kutluyl Doğançay, and Sylvie Perreau**

Name of Publication: 19th European Signal Processing Conference, 2011 (EUSIPCO-2011) The Author(s) hereby consents that the publisher appointed by EURASIP will publish the work.

- (1) The signatory(ies) on this form warrants all other authors/co-authors are properly credited, and generally that the author(s) has the right to make the grants made to "EURASIP" complete and unencumbered. The Author(s) also warrants that the Work is novel and has not been published elsewhere. The Author(s) furthermore warrant that the Work does not libel anyone, infringe anyone's copyright, or otherwise violate anyone's statutory or common law rights.
- (2) The Author(s) hereby transfers to "EURASIP" the copyright of the Work named above. "EURASIP" shall have the exclusive and unlimited right to publish the said Work and to translate (or authorize others to translate) it wholly or in part throughout the World during the full term of copyright including renewals and extensions and all subsidiary rights.
- (3) The Work may be reproduced by any means for educational and scientific purposes by the author or by others without fee or permission with the exception of reproduction by services that collect fee for delivery of documents. The Author(s) may use part or all of this Work or its image in any future works of his/her (their) own. In any reproduction, the original publication by "EURASIP" must be credited in the following manner: "First published in the Proceedings of the 19th European Signal Processing Conference (EUSIPCO-2011) in 2011, published by EURASIP", and such a credit notice must be placed on all copies. Any publication or other form of reproduction not meeting these requirements will be deemed to be unauthorized.
- (4) In the event of receiving any request to reprint or translate all or part of the Work, "EURASIP" shall seek to inform the author. This form is to be signed by the Author(s) or in case of a "work-made-for-hire", by the employer. If there is more than one Author, then either all must sign the Copyright Agreement, or one Author signs in consent for all, taking on full responsibility for the content of the publication.

Date:

30/5/2011

Full name(s):

Reza Arablouei

Signature(s):

A handwritten signature in blue ink that reads "Reza Arablouei".

Please return this form on or before June 6, 2011 by electronic upload of a scanned copy in EDAS.



# Bibliography

- [PD08] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementations*, 3<sup>rd</sup> ed., Boston: Springer, 2008.
- [AS08] A. H. Sayed, *Adaptive Filters*, Hoboken, NJ: Wiley, 2008.
- [SH02] S. Haykin, *Adaptive Filter Theory*, 4<sup>th</sup> ed., Upper Saddle River, NJ: Prentice-Hall, 2002.
- [BW60] B. Widrow and M. E. Hoff, Jr., "Adaptive switching circuits," *IRE WESCON Convention Record*, part 4, pp. 96–104, 1960.
- [K084] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electron. Commun. Jpn.*, vol. 67-A, no. 5, pp. 19–27, 1984.
- [TG00] T. Gaensler, S. L. Gay, M. M. Sondhi, and J. Benesty, "Double-talk robust fast converging algorithms for network echo cancellation," *IEEE Trans. Speech, Audio Process.*, vol. 8, pp. 656–663, Nov. 2000.
- [GF99] G. J. Foschini, G. D. Golden, R. A. Valenzuela and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 11, pp. 1841–1852, Nov. 1999.
- [KD08] K. Doğançay, *Partial-Update Adaptive Signal Processing: Design, Analysis and Implementation*, Academic Press, Oxford, UK, 2008.
- [SN99] S. Nagaraj, S. Gollamudi, S. Kapoor, and Y. F. Huang, "BEACON: An adaptive set-membership filtering technique with sparse updates," *IEEE Trans. Signal Process.*, vol. 47, no. 11, pp. 2928–2941, 1999.
- [YZ04] Y. V. Zakharov and T. C. Tozer, "Multiplication-free iterative algorithm for LS problem," *Electron. Lett.*, vol. 40, no. 9, pp. 567–569, Apr. 2004.
- [MR96] M. Rupp and A. H. Sayed, "A time-domain feedback analysis of filtered-error adaptive gradient algorithms," *IEEE Trans. Signal Process.*, vol. 44, pp. 1428–1439, Jun. 1996.
- [CV85] C. Van Loan, "On the method of weighting for equality-constrained least-squares problems," *SIAM J. Numer. Anal.*, vol. 22, no. 5, pp. 851–864, Oct. 1985.

- [LR96] L. S. Resende, J. M. T. Romano, and M. G. Bellanger, "A fast least-squares algorithm for linearly constrained adaptive filtering," *IEEE Trans. Signal Process.*, vol. 44, no. 5, pp. 1168–1174, May 1996.
- [SV91] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*, Philadelphia, PA: SIAM, 1991.
- [OF72] O. L. Frost III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.
- [BP95] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 2030–2033, Aug. 1995.
- [ST12] S.-Y. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6217–6234, Dec. 2012.
- [AS13] A. Sayed, "Diffusion adaptation over networks," to appear in *E-Reference Signal Processing*, R. Chellapa and S. Theodoridis, Eds., Elsevier, 2013, available at <http://arxiv.org/abs/1205.4220>.