A FAMILY OF LEAST-SQUARES MAGNITUDE PHASE ALGORITHMS

Scott C. Douglas[1,2]

[1]Southern Methodist University
[2]LGT Corporation
Dallas, Texas 75275 USA
douglas@lyle.smu.edu

Danilo P. Mandic

Imperial College London
Department of Electrical Engineering
London, SW7 2BT United Kingdom
d.mandic@imperial.ac.uk

ABSTRACT

This paper presents a family of least-squares algorithms for adaptive signal processing of complex-valued signals. The algorithms employ a composite cost function that allows magnitude and phase errors to be weighted differently in the parameter estimation depending on their importance, providing an opportunity for enhanced estimation performance over standard least-squares methods. We also describe a procedure for automatically adjusting this weighting based on the estimation errors themselves. Simulations show the excellent behavior of the algorithms in time-varying signal conditions.

Index Terms— Adaptive algorithm, adaptive equalizers, adaptive signal processing, adaptive systems, antenna arrays

1. INTRODUCTION

The least-squares principle is perhaps the most-used estimation method in existence, being the basis of linear regression. The exponentially-weighted recursive least-squares algorithm is well-known and a fundamental part of the Kalman filter for state-space estimation and tracking. When computational complexity is not an issue, least-squares approaches are commonly used in linear estimation tasks.

This paper considers the complex extension of the linear least-squares estimation task, in which a sequence of complex-valued input signal vectors $\mathbf{x}_k = [x_{1,k} \cdots x_{L,k}]^T$ is used to model a desired response signal d_k via the relation $y_{n,k} = \mathbf{w}_n^T \mathbf{x}_k$, where $\mathbf{w}_n = [w_{1,n} \cdots w_{L,n}]^T$ is the weight vector of a linear adaptive system with complex coefficients at iteration n. There are numerous practical situations where the linear model for $y_{n,k}$ is not adequate for estimating d_k due to some disturbance of the desired signal. One case in communications is when the desired response signal undergoes a complex phase shift due to Doppler effects that is not reflected in the input signal vector [1]. Another case is when the component of the desired response signal that is related to the input signal vector has an unknown and fast-varying amplitude. Both of these scenarios cause a degradation of estimation performance and limit the use of least-squares methods in these tasks.

In this paper, we present a family of linear least-squares estimation procedures for situations in which the uncertainty of the magnitude and/or phase errors of the signal estimates is unknown or time-varying. The algorithm family can be viewed as the least-squares extension of the recently-proposed least-mean magnitude phase (LSMP) algorithm for adaptive filtering [2]. A key result of this paper is a procedure for *automatically selecting* the weighting of the magnitude and phase costs within the composite cost used to calculate the parameter estimates. Simulations indicate that the procedures can outperform standard least-squares approaches.

2. THE LEAST-MEAN MAGNITUDE PHASE ALGORITHM REVISITED

To derive least-squares magnitude phase (LSMP) algorithms, it is useful to examine the criteria that led to the derivation of the gradient-based least-mean magnitude phase (LMMP) algorithm. For the moment, neglect time indices of signals.

As was shown in [2], the standard LMS instantaneous cost $\widehat{J}_{LMS}(d,y) = |d-y|^2$ can be decomposed as

$$\begin{aligned} \widehat{J}_{\mathrm{LMS}}(d,y) &= J_{\mathrm{m}}(d,y) + J_{\mathrm{p}}(d,y) && (1)\\ \widehat{J}_{\mathrm{m}}(d,y) &= (|d|-|y|)^2 && (2)\\ \widehat{J}_{\mathrm{p}}(d,y) &= 2|d||y|(1-\cos(\angle d - \angle y)). && (3) \end{aligned}$$

We refer to $\widehat{J}_m(d,y)$ and $\widehat{J}_p(d,y)$ as the magnitude cost and the phase cost, respectively. Let α be any real value. Then,

$$\begin{aligned} \widehat{J}_{\mathrm{LMS}}(d,y) &= \alpha \widehat{J}_{\mathrm{LMS}}(d,y) + \\ &\quad (1-\alpha)\left[\widehat{J}_{\mathrm{m}}(d,y) + \widehat{J}_{\mathrm{p}}(d,y)\right]. \quad (4) \end{aligned}$$

Because the magnitude and phase costs are a linear (additive) decomposition of the original LMS cost, the LMMP algorithm can be derived as a procedure that attempts to minimize the linear combination of the three costs $\widehat{J}_{\mathrm{LMS}}(d,y)$, $\widehat{J}_{\mathrm{m}}(d,y)$, and $\widehat{J}_{\mathrm{p}}(d,y)$, where each cost gets a different weighting in the algorithm. Similar ideas have been used before in derivations of related algorithms [3]. Define the least-mean magnitude phase instantaneous cost function as

$$\widehat{J}_{\mathrm{LMMP}}(d,y) = \widehat{J}_{\mathrm{LMS}}(d,y) + \beta_{\mathrm{m}} \widehat{J}_{\mathrm{m}}(d,y) + \beta_{\mathrm{p}} \widehat{J}_{\mathrm{p}}(d,y). \quad (5)$$

The stochastic gradient algorithm obtained from this cost function is exactly the LMMP algorithm with the values

$$\mu_{\mathrm{m}} = \mu\left(\frac{1}{2}+\beta_{\mathrm{m}}\right) \quad \text{and} \quad \mu_{\mathrm{p}} = \mu\left(\frac{1}{2}+\beta_{\mathrm{p}}\right) \tag{6}$$

where $\mu > 0$, $\beta_{\mathrm{m}} > 0$, and $\beta_{\mathrm{p}} > 0$. This relation does not guarantee unique values for μ, β_{m}, and β_{p} given μ_{m} and μ_{p}. Even so, if the behavior of the LMMP algorithm is to be translated to a least-squares context, we can consider linear combinations of magnitude-only and phase-only costs with the original quadratic cost from which the algorithm is based.

3. THE RECURSIVE LEAST-SQUARES MAGNITUDE PHASE ALGORITHM FAMILY

Consider the exponentially-weighted composite least-squares cost given by

$$\begin{aligned} &J_{\mathrm{LSMP}}(\mathbf{w}_n) \\ &= \sum_{k=1}^{n}\lambda^{n-k}\Big[\widehat{J}_{\mathrm{LMS}}(d_k,y_{n,k})+\beta_{\mathrm{m}}\widehat{J}_{\mathrm{m}}(d_k,y_{n,k}) \\ &\qquad +\beta_{\mathrm{p}}\widehat{J}_{\mathrm{p}}(d_k,y_{n,k})\Big] \qquad (7)\\ &= \sum_{k=1}^{n}\lambda^{n-k}\Big[\left|d_k-\mathbf{w}_n^T\mathbf{x}_k\right|^2+\beta_{\mathrm{m}}\left||d_k|-|\mathbf{w}_n^T\mathbf{x}_k|\right|^2 \\ &\qquad +\beta_{\mathrm{p}}2|d_k||\mathbf{w}_n^T\mathbf{x}_k|(1-\cos(\angle d_k-\angle\mathbf{w}_n^T\mathbf{x}_k)\Big]. \qquad (8)\end{aligned}$$

The connection between (7) and (5) is clear. When $\beta_{\mathrm{m}} = \beta_{\mathrm{p}}$, the standard quadratic least-squares cost is obtained. When $\beta_{\mathrm{m}} \neq \beta_{\mathrm{p}}$, the cost is non-quadratic, and an iterative solution is required.

Using the derivatives in [2], the following necessary conditions on $\mathbf{w}_n$ for minimization of $J_{\mathrm{LSMP}}(\mathbf{w}_n)$ are obtained:

$$\begin{aligned}\sum_{k=1}^{n}\lambda^{n-k}\Bigg[&(d_k-y_{n,k})\mathbf{x}_k^*+\beta_{\mathrm{m}}\left(|d_k|\frac{y_{n,k}}{|y_{n,k}|}-y_{n,k}\right)\mathbf{x}_k^* \\ &+\beta_{\mathrm{p}}\left(d_k-\frac{|d_k|}{|y_{n,k}|}y_{n,k}\right)\mathbf{x}_k^*\Bigg] = \mathbf{0}. \end{aligned} \tag{9}$$

Unfortunately, because $J_{\mathrm{LSMP}}(\mathbf{w}_n)$ is not quadratic in $\mathbf{w}_n$, (9) is nonlinear in $\mathbf{w}_n$. To linearize this relationship, we use a concept used in other approximate solutions to non-quadratic parameter estimation problems involving exponential windows, particularly Yang's projection approximation subspace tracking (PAST) algorithm [4]. This concept replaces certain terms depending on $y_{n,k}$ in (9) that depend on $\mathbf{w}_n$ with values $y_{k-1,k}$ estimated from past parameter vectors $\mathbf{w}_k$, $k < n$. The substituted values result in a linearized relationship

$$\begin{aligned}\sum_{k=1}^{n}\lambda^{n-k}\Bigg[&(d_k-y_{n,k})\mathbf{x}_k^*+\beta_{\mathrm{m}}\left(|d_k|\frac{y_{k-1,k}}{|y_{k-1,k}|}-y_{n,k}\right)\mathbf{x}_k^* \\ &+\beta_{\mathrm{p}}\left(d_k-\frac{|d_k|}{|y_{k-1,k}|}y_{n,k}\right)\mathbf{x}_k^*\Bigg] = \mathbf{0}. \end{aligned} \tag{10}$$

Finally, we recognize that $y_{k-1,k}$ might approach zero to cause a near divide-by-zero in the term premultiplied by β_{p} in some cases. To mitigate these situations, we adjust the constraint relation as

$$\begin{aligned}\sum_{k=1}^{n}\lambda^{n-k}\Bigg[&(d_k-y_{n,k})\mathbf{x}_k^*+\beta_{\mathrm{m}}\left(|d_k|\frac{y_{k-1,k}}{|y_{k-1,k}|}-y_{n,k}\right)\mathbf{x}_k^* \\ &+\beta_{\mathrm{p}}f(y_{k-1,k})\left(d_k-\frac{|d_k|}{|y_{k-1,k}|}y_{n,k}\right)\mathbf{x}_k^*\Bigg] = \mathbf{0}, \end{aligned} \tag{11}$$

where

$$f(y) = \begin{cases} 1 & \text{if } |y| > \delta \\ 0 & \text{if } |y| \leq \delta \end{cases} \tag{12}$$

The value of δ is clearly related to the average magnitude of $|d_k|$, although its exact value does not appear to be critical. In all numerical evaluations in this paper, we chose $\delta = 0.1$.

Eq. (11) has a straightforward recursive solution. Let

$$\widehat{d}_k = [1+\beta_{\mathrm{p}}f(y_{k-1,k})]\,d_k+\beta_{\mathrm{m}}|d_k|\frac{y_{k-1,k}}{|y_{k-1,k}|} \tag{13}$$

$$\gamma_k = 1+\beta_{\mathrm{m}}+\beta_{\mathrm{p}}f(y_{k-1,k})\frac{|d_k|}{|y_{k-1,k}|} \tag{14}$$

Then, the solution for $\mathbf{w}_n$ can be propagated as

$$\mathbf{w}_n = \mathbf{R}_n^{-1}\mathbf{p}_n \tag{15}$$

$$\mathbf{R}_n = \lambda\mathbf{R}_{n-1}+\gamma_n\mathbf{x}_n^*\mathbf{x}_n^T \tag{16}$$

$$\mathbf{p}_n = \lambda\mathbf{p}_{n-1}+\widehat{d}_n\mathbf{x}_n^* \tag{17}$$

Eqs. (13)–(17) define the recursive least-squares magnitude phase (RLSMP) algorithm family. Any specific implementation of RLSMP is mathematically-equivalent to these relations with a particular method of propagating their solution recursively. For example, the well-known $\mathcal{O}(L^2)$ solution to the above relations can be derived using the Woodbury (matrix inversion) lemma, resulting in the relations

$$\mathbf{R}_n^{-1} = \frac{1}{\lambda}\left(\mathbf{R}_{n-1}^{-1}-\frac{\mathbf{R}_{n-1}^{-1}\mathbf{x}_n^*\mathbf{x}_n^T\mathbf{R}_{n-1}^{-1}}{\dfrac{\lambda}{\gamma_n}+\mathbf{x}_n^T\mathbf{R}_{n-1}^{-1}\mathbf{x}_n^*}\right) \tag{18}$$

$$\mathbf{w}_n = \mathbf{w}_{n-1}+(d_n-\gamma_n y_{n-1,n})\,\mathbf{R}_n^{-1}\mathbf{x}_n^*. \tag{19}$$

Other forms of the algorithm involving QR or Householder-based square-root implementations are also easily derived due to the fact that $\gamma_k > 0$ but will not be described here.

4. ALGORITHM ADAPTATION FOR IMPROVED PERFORMANCE

In most applications of the gradient-based LMMP algorithm [2], either the magnitude cost or the phase cost is emphasized in order to improve the algorithm's performance over that of the LMS algorithm. This emphasis has been chosen through the selection of the step sizes μ_{m} and μ_{p} at the onset of the procedure based on *a priori* or side information.

In the LSMP algorithm, we can choose $\beta_{\rm m}$ and $\beta_{\rm p}$ to emphasize either magnitude or phase information, respectively, in the algorithm based on available side information. Due to the nice way $\beta_{\rm m}$ and $\beta_{\rm p}$ appear in the update relations, however, it is reasonable to consider time-varying values of $\beta_{{\rm m},k}$ and $\beta_{{\rm p},k}$, which changes the constraint relation to be

$$\sum_{k=1}^{n} \lambda^{n-k} \left[(d_k - y_{n,k})\mathbf{x}_k^* + \beta_{{\rm m},k} \left(|d_k| \frac{y_{k-1,k}}{|y_{k-1,k}|} - y_{n,k} \right) \mathbf{x}_k^* + \beta_{{\rm p},k} f(y_{k-1,k}) \left(d_k - \frac{|d_k|}{|y_{k-1,k}|} y_{n,k} \right) \mathbf{x}_k^* \right] = \mathbf{0} \quad (20)$$

To maintain the number of parameters at a manageable level, we set

$$\beta_{{\rm m},k} = \beta_k \quad (21)$$
$$\beta_{{\rm p},k} = \beta_{\max} - \beta_k \quad (22)$$

and adjust the single parameter β_k to obtain the desired behavior. In absence of any good information for choosing $\beta_{\max}$, we set $\beta_{\max} = 1$ and $\beta_0 = 0.5$.

The advantage of the LSMP cost over the standard least-squares cost is obtained when one of either the magnitude error $e_{{\rm m},k} = |d_k|{\rm sgn}(y_{k-1,k}) - y_{k-1,k}$ or the phase error $e_{{\rm p},k} = d_k - |d_k|{\rm sgn}(y_{k-1,k})$ provides better information than the other. By emphasizing the right one of these error signals over the other in the cost, a more-accurate estimate of the parameter vector is likely to be obtained. In the absence of any prior information on d_n and $\mathbf{x}_n$, the relative magnitudes of $|e_{{\rm m},n}|$ and $|e_{{\rm p},n}|$ can serve as the metric for this choice at time sample n. Consider a simple strategy for adjusting the value of β_k:

$$\beta_n = \begin{cases} (1-\mu)\beta_{n-1} + \mu\beta_{\max} & \text{if } |e_{m,n}| < |e_{p,n}| \\ (1-\mu)\beta_{n-1} & \text{if } |e_{m,n}| > |e_{p,n}| \end{cases} \quad (23)$$

where

$$e_{m,n} = |d_n| \frac{y_{n-1,n}}{|y_{n-1,n}|} - y_{n-1,n} \quad (24)$$
$$e_{p,n} = d_n - |d_n| \frac{y_{n-1,n}}{|y_{n-1,n}|} \quad (25)$$

The solution for $\mathbf{w}_n$ is the same as before, except

$$\widehat{d}_n = [1 + (1-\beta_n) f(y_{n-1,n})]\, d_n + \beta_n |d_n| \frac{y_{n-1,n}}{|y_{n-1,n}|} \quad (26)$$
$$\gamma_n = 1 + \beta_n + (1-\beta_n) f(y_{n-1,n}) \frac{|d_n|}{|y_{n-1,n}|} \quad (27)$$

This version of the RLSMP algorithm has two parameters: λ and μ. The value of the exponential forgetting factor λ can be chosen according to the statistics of the data and the desired tracking behavior. We typically choose μ to be somewhat smaller than $1-\lambda$ so that the tracking of the error function occurs more slowly than the minimization of the error function through the evolution of $\mathbf{w}_n$.

5. A BLOCK-BASED LEAST-SQUARES MAGNITUDE PHASE ALGORITHM

We now describe an iterative block-based version of the LSMP algorithm required in many applications. *This algorithm has only a single adaptive parameter μ.* This procedure attempts to minimize the following criterion:

$$J_{\rm BLSMP}(\mathbf{w}_n) = \sum_{k=1}^{N} \left[\widehat{J}_{\rm LMS}(d_k, y_{n,k}) + \beta_{n,k} \widehat{J}_{\rm m}(d_k, y_{n,k}) + (1-\beta_{n,k}) \widehat{J}_{\rm p}(d_k, y_{n,k}) \right], \quad (28)$$

where the sequence $\beta_{n,k}$, $1 \le k \le N$ is automatically adjusted at each iteration to emphasize magnitude and/or phase errors in the criterion and achieve good performance. The approximate constraint equation to minimize (28) is

$$\sum_{k=1}^{N} \left[(d_k - y_{n,k})\mathbf{x}_k^* + \beta_{n,k} \left(|d_k| \frac{y_{n-1,k}}{|y_{n-1,k}|} - y_{n,k} \right) \mathbf{x}_k^* + (1-\beta_{n,k}) f(y_{n-1,k}) \left(d_k - \frac{|d_k|}{|y_{n-1,k}|} y_{n,k} \right) \mathbf{x}_k^* \right] = \mathbf{0} \quad (29)$$

Eq. (29) differs from (20) in that (a) $\lambda = 1$ and (b) the sample *a priori* estimate $y_{k-1,k}$ is replaced by the block iteration *a priori* estimate $y_{n-1,k}$. This expression is in the form of a system of linear equations in $\mathbf{w}_n$ which is solved iteratively at each n, and the updated $\mathbf{w}_n$ is used to form the new system of linear equations at iteration $n+1$. Then, the algorithm is iterated P times to obtain adequate convergence, where P is a small integer.

In this algorithm, the entire sequence $\beta_{n,k}$ is adjusted at each iteration n using a relation that is similar to (23), namely,

$$\beta_{n,k} = \begin{cases} (1-\mu)\beta_{n-1,k} + \mu\beta_{\max} & \text{if } |e_{{\rm m},n-1,k}| < |e_{{\rm p},n-1,k}| \\ (1-\mu)\beta_{n-1,k} & \text{if } |e_{{\rm m},n-1,k}| > |e_{{\rm p},n-1,k}| \end{cases} \quad (30)$$

where

$$e_{{\rm m},n,k} = |d_k|{\rm sgn}(y_{n,k}) - y_{n,k} \quad (31)$$
$$e_{{\rm p},n,k} = d_k - |d_k|{\rm sgn}(y_{n,k}). \quad (32)$$

In practice, the local estimate of $\beta_{n,k}$ can be improved somewhat by a small amount of averaging of the sequence across the k domain. In the simulations that follow, we replace $\beta_{n,k}$ by $(0.25\beta_{n,k+1} + 0.5\beta_{n,k} + 0.25\beta_{n,k-1})$ after application of (30) at each iteration to achieve this smoothing.

6. NUMERICAL EVALUATIONS

We first explore the behavior of RLSMP as compared to NLMS and standard RLS in a FIR filter system identification task, in which $\mathbf{w}_{opt}$ is of length $L = 10$ and has complex Gaussian entries and d_k is generated according to

$$d_k = a_k \mathbf{w}_{{\rm opt},k}^T \mathbf{x}_k, \quad (33)$$

where x_k is from a time series that is created from an IIR filter with system function $H(z) = \sqrt{0.51}(1 - z^{-1})/(1 - 0.7\exp(-j\pi/4)z^{-1})$ and the statistics of a_k change with time. For $1 \leq k \leq 2000$, a_k is uniformly-distributed on the interval $[0.5, 1.5]$. For $2001 \leq k \leq 6000$, a_k has unity amplitude and has a phase that is uniformly-distributed over the interval $-0.05\pi \leq \angle a_k \leq 0.05\pi$. At $k = 4000$, $\mathbf{w}_{\text{opt},k} = -0.7j\mathbf{w}_{\text{opt},k-1}$. For $6001 \leq k \leq 10000$, a_k is uniformly-distributed on the interval $[0.5, 1.5]$. At $k = 8000$, $\mathbf{w}_{\text{opt},k} = 1.5j\mathbf{w}_{\text{opt},k-1}$. This combination of nonstationarity and unknown system change considers various combinations of changes in both unknown system coefficients and unknown system statistics. At each time instant, we estimate the value of $||\mathbf{w}_n - \mathbf{w}_{\text{opt},n}||^2/||\mathbf{w}_{\text{opt},n}||^2$. One thousand simulations are averaged to produce the curves shown. In order to equalize convergence rates, we choose $\lambda = 0.981$ and $\mu = 1 - \lambda$ for the RLSMP algorithm, $\lambda = 0.99$ for the standard RLS algorithm, and $\mu = 1$ for the NLMS algorithm.

Fig. 1 shows the averaged normalized misalignment for the NLMS, standard RLS, and RLSMP algorithms. The performance improvement for RLSMP is clearly shown in terms of lower misalignment in steady-state with no loss of convergence rate as compared to RLS, and a faster convergence rate as compared to NLMS. Fig. 2 shows the value of β_n. Note how its value tracks the variations in both phase certainty and magnitude certainty quite nicely. The high degree of variability of β_n is not a problem because the cost function naturally averages the error values that matter to overall performance.

Next, we explore the behavior of the block-based LSMP algorithm in a channel equalization task. The scenario is similar to that described in [2], in which a received signal x_k is generated from an i.i.d. 16-QAM source as

$$x_k = \eta_k + e^{j2\pi k f_k}[h_0 s_k + h_1 s_{k-1} + h_2 s_{k-2}], \quad (34)$$

where $\{h_0 \; h_1 \; h_2\} = \{0.2e^{j0.1\pi} \; 1e^{j0.2\pi} \; 0.1e^{j0.3\pi}\}$ and the channel noise η_k is complex circular Guassian with variance 0.01. The parameter f_k models frequency offset effects due to physical motion, and its value varies from $f_k = 0.01$ to $f_k = 0$ to $f_k = 0.02$ over three different 2000-sample periods, such that $N = 6000$. Both a block-based LS and the iterative block LSMP algorithm are applied to x_k, with $d_k = s_{k-5}$, where $L = 11$, $\mu = 0.1$ and $P = 10$.

Across one thousand trials, the average intersymbol interference (ISI) for the standard block LS procedure is -18.5 dB, whereas the average ISI for the iterative block LSMP algorithm is -23.2 dB, a 4.7 dB improvement. Shown in Fig. 3 is an example $\beta_{P,k}$ sequence obtained from one of the block LSMP trials. The sharp dips in β_k correspond to time instants where the signal rotation is near a multiple of 2π such that the phase information is reliable.

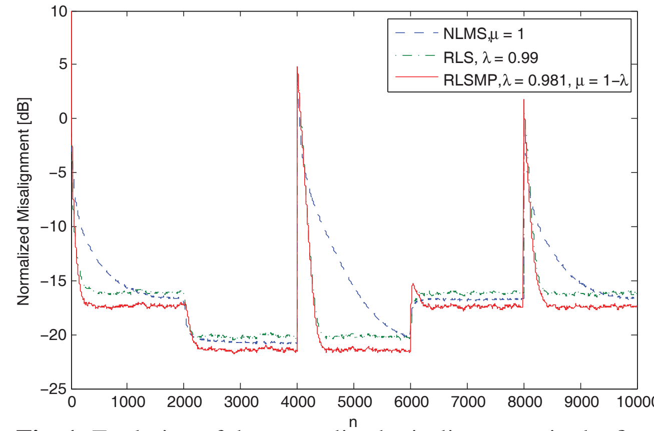

Fig. 1. Evolution of the normalized misalignments in the first example.

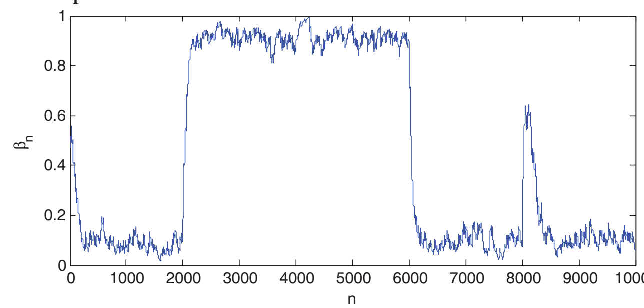

Fig. 2. Evolution of β_n for RLSMP in the first example.

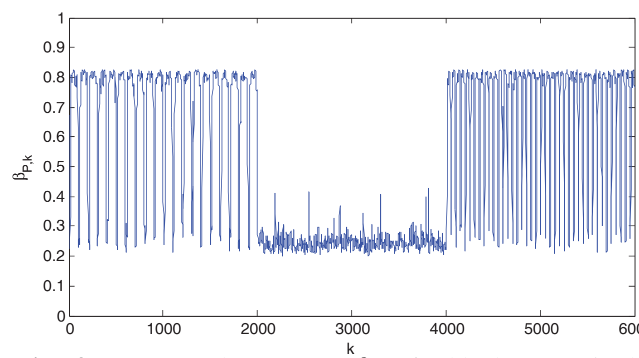

Fig. 3. An example sequence $\beta_{P,k}$ for block LSMP in the channel equalization example.

7. CONCLUSIONS

In this paper, we have described a family of iterative least-squares procedures that adjust their estimation capabilities depending on the quality of the magnitude and phase information in a complex-valued linear estimation task. The simplest algorithm versions have one or two adjustable parameters and are shown to outperform standard LS approaches when magnitude and/or phase uncertainties are present.

8. REFERENCES

[1] M. Rupp, "On the separation of channel and frequency estimation," *Conf. Record 36th IEEE Asilomar Conf. Signals, Syst., Comput.,* Pacific Grove, CA, vol. 2, pp. 1186-1190, Nov. 1998.

[2] S.C. Douglas and D.P. Mandic, "The least-mean magnitude-phase algorithm with applications to communication systems," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Prague, Czech Republic, pp. 4152–4155, May 2011.

[3] A. Tarighat and A.H. Sayed, "Least mean-phase adaptive filters with applications to communications systems," *IEEE Signal Processing Lett.,* vol. 11, no. 2, pp. 220-223, Feb. 2004.

[4] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Processing,* vol. 43, pp. 95–107, Jan. 1995.