Convergence Analysis of \( \ell_0 \)-RLS Adaptive Filter

B. K. Das\(^1\), Member, IEEE, S. Mukhopadhyay\(^2\), Member, IEEE, and M. Chakraborty\(^3\), Senior Member, IEEE

Copyright (c) 2017 BIJIT KUMAR DAS. Personal use of this material is permitted.

Abstract

This paper presents first and second order convergence analysis of the sparsity aware \( \ell_0 \)-RLS adaptive filter. The theorems 1 and 2 state the steady state state value of mean and mean square deviation of the adaptive filter weight vector.

Index terms—Sparse Systems, Adaptive Filter, \( \ell_0 \)-Recursive Least Squares, Mean Square Deviation.

I. INTRODUCTION

The \( \ell_0 \)-pseudonorm which measures sparsity by counting the number of nonzero elements in a vector can not be directly used for the regularization purpose and it is often approximated by some continuous functions, e.g., the \( \ell_1 \)-norm or the absolute sum, the log-sum etc. Another approximation of \( \ell_0 \)-norm by some exponential function has also been proposed in [5] and has been used to derive a new sparsity aware adaptive filter popularly called \( \ell_0 \)-LMS. This algorithm manifests excellent behavior in terms of convergence speed and steady-state mean square deviation for proper choice of a parameter responsible for promoting sparsity. Similarly, the conventional recursive least squares (RLS) algorithm has also been modified to get advantage of the sparsity using \( \ell_1 \)-norm penalty in [9]-[7], and [3]. In [7], \( \ell_0 \)-norm regularized RLS has been proposed, and it outperforms the standard RLS and the aforementioned sparsity-aware algorithms [3], [5], [2], and [3]. In this chapter, a theoretical analysis of the \( \ell_0 \)-norm regularized recursive least squares (\( \ell_0 \)-RLS) is carried out. Inspired by the work in [6], relevant common assumptions are taken along with some new ones and their applicability is discussed. Then, we also propose a combination scheme to improve the performance of the \( \ell_0 \)-RLS algorithm in a scenario where signal-to-noise (SNR) varies over time. The proposed approach combines the output of \( M \) differently parameterized \( \ell_0 \)-RLS adaptive filters, where the combiner coefficients are adapted to extract the best out of the overall combination for different levels of SNR. Finally, a new sparsity promoting non-convex function is proposed, and a sparse RLS adaptive algorithm based on this is derived.

II. BRIEF REVIEW OF THE \( \ell_0 \)-RLS ALGORITHM

In [7], an RLS based sparse adaptive filter has been proposed where the cost function uses certain differentiable approximation of the \( \ell_0 \) norm. In particular, \( \|w(n)\|_0 \) as

\[
\|w(n)\|_0 \approx \sum_{i=0}^{N-1} (1 - \exp(-\alpha|w_i(n)|)),
\]

where \( \alpha \) is a parameter to be chosen carefully so that if \( w_i(n) \neq 0 \), but \( |w_i(n)| \) is small; \( \alpha|w_i(n)| \approx 0 \). This way, if \( w_i(n) \) is zero, or \( |w_i(n)| \) is small, the corresponding factor \( (1 - \exp(-\alpha|w_i(n)|)) \approx 0 \), and if \( |w_i(n)| \) is large, \( (1 - \exp(-\alpha|w_i(n)|)) \approx 1 \), which make \( \sum_{i=0}^{N-1} (1 - \exp(-\alpha|w_i(n)|)) \) a good approximation of \( \|w(n)\|_0 \).

Using the above, the RLS cost function at index \( n \) is given by

\[
\xi(n) = \sum_{m=0}^{n} \lambda^{n-m}e^2(m) + \gamma\|w(n)\|_0,
\]

where, as before, \( e(m) = d(m) - y(m) \), \( d = 0, 1, \ldots, N \), \( y(m) = x^T(m)w(m) \), \( d(m) = x^T(m)w_0 + v(m) \), \( x(m) = [x(m), x(m-1), \ldots, x(m-N+1)]^T \), \( v(m) \) is an additive observation noise independent of any input \( x(l) \neq m,l \), and the parameter \( \gamma \) is the Lagrange multiplier that controls the balance between estimation error and sparsity promoting penalty. Also, the function \( \|w(n)\|_0 \) is to be replaced by the abovementioned approximation.

The RLS adaptive algorithm which minimizes \( \xi(n) \) and is proposed in [7] can be obtained as:

\[
w(n) = w(n-1) + k(n)e(n) + \beta P(n)g(w(n-1)),
\]

The authors are with the department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, INDIA (email: \(^1\) bijitbijit@gmail.com, \(^2\) samratphysics@gmail.com, \(^3\) mrityun@ece.iitkgp.ernet.in).
where
\[
\beta = \gamma (1 - \lambda),
\]
\[
\epsilon(n) = d(n) - w^T(n - 1)x(n) \quad \text{: a priori error}
\]
\[
k(n) = \frac{1}{\lambda + x^T(n)x(n)},
\]
\[
P(n) = \frac{1}{\lambda} [P(n - 1) - k(n)x^T(n)P(n - 1)],
\]
and the initialization is done as \( P(-1) = \delta I \) (\( \delta \) being a very small positive number).

The function \( g(w(n) - 1) \) is given by \(-\nabla^2 \{\|w(n - 1)\|_0\} \), where \( \|w(n - 1)\|_0 \) is to be replaced by the aforementioned approximation of the \( l_0 \) norm. This means \( g(w_j(n - 1)) \) is given by \(-\frac{d}{d_{w}(n-1)}[1 - \exp(-\alpha|w_j(n-1)|)] \), \( j = 0, 1, \cdots, N - 1 \).

In [7], an approximation of \( g(t) \) for scalar variable \( t' \) is used which is as follows:
For \(|t| \) large enough, \( 1 - \exp(-\alpha|t|) \approx 1 \), and thus, \( g(t) = 0 \). On the other hand, for \(|t| \) small s. t. \( \alpha|t| \leq 1 \), \( 1 - \exp(-\alpha|t|) \approx \alpha|t| - \frac{\alpha^2 t^2}{2} \), and thus \( g(t) \approx \alpha t^2 - \alpha \sgn(t) \).

Formally, \( g(t) \) is then defined as,
\[
g(t) = \begin{cases} 
\alpha^2 t - \alpha \sgn(t), & |t| \leq 1/\alpha, \\
0, & \text{elsewhere}.
\end{cases}
\]
(Note that the third term in the RHS of \( \alpha \) resulting from \( g(w(n - 1)) \) is the zero-point attraction term and the range \( (-1/\alpha, 1/\alpha) \) is called the attraction range. Lastly, the notations \( C_0, C_S \), and \( C_L \) are used to indicate sets of indices for zero valued tap (i.e. exactly zero), small valued tap (i.e. \( |w_{0,i}| \leq \alpha \)), and large valued tap (i.e. \( |w_{0,i}| > \alpha \)) of \( w_0 \) respectively.

### III. Convergence Analysis

For both first and second order convergence analysis of the \( l_0 \)-RLS algorithm [7], we adopt the following assumptions (as also made in [6]):

1. The data sequence \( x(n) \) is a white sequence with zero mean and variance \( \sigma_x^2 \) and is independent of the additive noise sequence \( v(n) \) which is also assumed to be a zero mean sequence.
2. The incoming sequence of vectors \( x(n) \) and the filter weight vector \( w(n) \) are independent. As we already know, this is the Independence Assumption [10], [11]), and is widely adopted in adaptive filtering literature for simplified analysis of the adaptive filtering algorithms.
3. As in the popular adaptive filtering literature [10], [11], the forgetting factor \( \lambda \) is chosen sufficiently close to 1 so that for large \( n \), \( P(n) \approx (1 - \lambda)R^{-1} \) where \( R \) is the autocorrelation matrix of the incoming data sequence, i.e., \( R = E[x(n)x^T(n)] \).
4. At the steady state, the \( k \)th tap of the adaptive filter, \( w_k(n), \forall k \in C_0 \) is Gaussian distributed.
5. It is assumed that, for large \( n \), \( w_k(n) \) are of the same sign as that of \( w_{0,k} \) for \( k \in C_L \cup C_S \).
6. It is also assumed that, for large \( n \), \( w_k(n) \) lies outside attraction range for \( k \in C_L \) and inside attraction range for \( k \in C_S \cup C_0 \), almost surely.

#### A. Convergence in Mean

The first order convergence behavior of the \( l_0 \)-RLS algorithm is described by the following theorem:

**Theorem 1** Under the assumptions [1] - [6] above, the deviation coefficients \( \tilde{w}_k(n) := w_k(n) - w_{0,k}, k = 0, 1, \cdots, N - 1 \), asymptotically converge in mean to the following:
\[
\lim_{n \to \infty} E[\tilde{w}_k(n)] = \begin{cases} 
\frac{\alpha \beta}{\sigma^2 - \sigma_x^2}, & k \in C_0, \\
\frac{\beta}{\sigma^2 - \sigma_x^2} g(w_{0,k}), & k \in C_S, \\
0, & k \in C_L.
\end{cases}
\] (5)

**Proof.** From [3], we have,
\[
\tilde{w}(n) = w(n) - w_0 = \tilde{w}(n - 1) + \epsilon(n)k(n) + \beta P(n)g(w(n - 1)).
\] (6)

Replacing \( \epsilon(n) \) by \( d(n) - x^T(n)w(n - 1) = d(n) - x^T(n)w_0 + \tilde{w}(n - 1), \) and \( d(n) \) by \( x^T(n)w_0 + v(n) \), and noting from (1.2)\( I - k(n)x^T(n) = \lambda P(n)P^{-1}(n - 1) \), we can write
\[
\tilde{w}(n) = \lambda P(n)P^{-1}(n - 1)\tilde{w}(n - 1) + k(n)v(n) + \beta P(n)g(w(n - 1)).
\] (7)

Substituting \( k(n) = P(n)x(n) \) in (7) and using the assumption [3] for large \( n \), we obtain,
\[
\tilde{w}(n) = \lambda \tilde{w}(n - 1) + (1 - \lambda)R^{-1}x(n)v(n) + \beta(1 - \lambda)R^{-1}g(w(n - 1)).
\] (8)
Using assumption \([1]\) the above can be simplified further as,

\[
\dot{w}(n) = \lambda \dot{w}(n-1) + \frac{(1-\lambda)}{\sigma_x^2} x(n)v(n) + \frac{\beta(1-\lambda)}{\sigma_x^2} g(w(n-1)).
\] (9)

Using expectation operator on both sides of the above equation, and from the orthogonality of \(x(n)\) and \(v(n)\) (assumption \([1]\), we obtain, for large \(n\),

\[
E[\dot{w}(n)] = \lambda E[\dot{w}(n-1)] + \frac{\beta(1-\lambda)}{\sigma_x^2} E[g(w(n-1))].
\] (10)

To evaluate \(E[g(w(n-1))\), we analyze its \(k^{th}\) component \(g(w_k(n-1))\) for the three cases: \(k \in C_L\), \(k \in C_S\), and \(k \in C_0\) separately.

First consider the case for \(k \in C_L\). From the definition of the function \(g(\cdot)\) as given by \([4]\) and the assumption \([6]\) it follows directly that, for large \(n\), \(\forall k \in C_L\): \(g(w_k(n-1)) = 0\) almost surely, and thus,

\[
E[g(w_k(n-1))] = 0.
\] (11)

Since \(0 < \lambda < 1\), from (10), it follows that \(\lim_{n \to \infty} E[\dot{w}_k(n)] = 0\), for \(k \in C_L\).

Next, for evaluating \(E[g(w_k(n-1))\) for \(k \in C_S\), we invoke the assumptions \([5]\) and \([6]\). From the definition in (4), and following the approach in \([6]\), it is easy to see that for large \(n\) and \(k \in C_S\), the following is satisfied almost surely:

\[
g(w_k(n-1)) = \alpha^2 w_k(n-1) - \text{sgn}(w_k(n-1)) = \alpha^2 w_k(n-1) - \text{sgn}(w_{0,k}) = \alpha^2 \dot{w}_k(n-1) + \alpha^2 w_{0,k} - \text{sgn}(w_{0,k}) = \alpha^2 \dot{w}_k(n-1) + g(w_{0,k}).
\] (12)

Substituting in (10), and simplifying, we then obtain,

\[
\lim_{n \to \infty} E[\dot{w}_k(n)] = \frac{\beta}{\sigma_x^2 - \beta \alpha^2} g(w_{0,k}), \quad \text{for} \quad k \in C_S.
\] (13)

Finally, we consider the case \(k \in C_0\).

Using the definition of \(g(\cdot)\) in (4), and recalling the fact that, in this case, \(\dot{w}_k(n) = w_k(n)\), it is easy to see that for large \(n\), the following is satisfied almost surely:

\[
g(w_k(n-1)) = \alpha^2 w_k(n-1) - \text{sgn}(w_k(n-1)),
\]

\[
= \alpha^2 \dot{w}_k(n-1) - \text{sgn}(w_k(n-1)), \quad \text{for} \quad k \in C_0.
\] (14)

Substituting in (10) and simplifying, we can obtain

\[
\lim_{n \to \infty} E[\dot{w}_k(n)] = \frac{\alpha \beta}{\beta \alpha^2 - \sigma_x^2} \lim_{n \to \infty} E[\text{sgn}(w_k(n))], \quad k \in C_0.
\] (15)

IV. Steady State Mean Square Performance of the \(l_0\)-RLS

**Theorem 2** Under assumptions \([16]\) above, the steady state mean square deviation of the \(l_0\)-RLS adaptive filter is given by

\[
D(\infty) := \lim_{n \to \infty} E[||\dot{w}(n)||^2] = D_L(\infty) + D_S(\infty) + D_0(\infty),
\] (16)

where

\[
D_L(\infty) = \lim_{n \to \infty} \sum_{k \in C_L} E[\dot{w}_k^2(n)] = \frac{|C_L|(1-\lambda)\sigma_n^2}{(1+\lambda)\sigma_x^2},
\] (17)

\[
D_S(\infty) = \lim_{n \to \infty} \sum_{k \in C_S} E[\dot{w}_k^2(n)] = \frac{|C_S|(1-\lambda)\sigma_n^2}{(1+\lambda)\sigma_x^2} + \frac{\beta' G_s}{(1-\lambda^2)},
\] (18)

where \(\lambda' = \sqrt{\lambda^2 + \frac{2\beta(1-\lambda)}{\sigma_x^2} \alpha^2 + \frac{\beta(1-\lambda)}{\sigma_x^2} \alpha^2} \), \(\beta' = \frac{2\alpha^2 \lambda(1-\lambda)}{\sigma_x^2} \alpha^2 + \frac{\beta^2(1-\lambda)]}{\sigma_x^2} \alpha^2 + \frac{2\beta^2(1-\lambda)\alpha^2}{\sigma_x^2(1-\lambda)},\) and \(G_s = \sum_{i \in C_S} g^2(w_{0,i})\).
and,

\[ D_0(\infty) = \lim_{n \to \infty} \sum_{k \in C_0} E[\tilde{w}_k^2(n)] = -|C_0|(b_\omega + c_\omega), \]  

(19)

where \( \omega = (-b_\omega + \sqrt{b_\omega^2 - 4c_\omega})/2, \)

\( b_\omega = \frac{4\alpha \beta (1 - \lambda)}{\sqrt{2\pi(1 - \lambda^2)\sigma_z^2}} (\lambda + \frac{\alpha^2 \beta (1 - \lambda)}{\sigma_z^2}), \)

\( c_\omega = -\frac{(1 - \lambda)^2}{1 - \lambda^2} \left( \frac{\alpha^2 \beta^2 (1 - \lambda)}{\sigma_z^2} + \frac{\sigma_x^2}{\sigma_z^2} \right). \)

Proof. We begin by investigating the evolution of the autocorrelation matrix of the filter weight deviation vector. From (9), discarding the terms involving \( b \omega \)

\[ \omega = \frac{(1 - \lambda)^2}{1 - \lambda^2} \left( \frac{\alpha^2 \beta^2 (1 - \lambda)}{\sigma_z^2} + \frac{\sigma_x^2}{\sigma_z^2} \right). \]

This means that for large \( n \) and, next we consider \( S \) from (12), one can write,

\[ E[\tilde{w}(n)\tilde{w}^T(n)] = M_1 + (M_2 + M_3) + M_4, \]

(20)

where

\[ M_1 = \lambda^2 E[\tilde{w}(n-1)\tilde{w}^T(n-1)], \]

(21)

\[ M_2 = \frac{\beta \lambda (1 - \lambda)}{\sigma_z^2} E[\tilde{w}(n-1)g(w^T(n-1))], \]

(22)

\[ M_3 = \frac{\beta^2 (1 - \lambda)^2}{\sigma_z^2} E[g(w(n-1))g(w^T(n-1))], \]

(23)

\[ M_4 = \frac{(1 - \lambda)^2 \sigma_x^2}{\sigma_z^2} E[x(n)x^T(n)]. \]

(24)

Taking the \( k^{th} \) diagonal element of the weight deviation autocorrelation matrix, we obtain the corresponding evolution equation:

\[ E[\tilde{w}_k^2(n)] = \lambda^2 E[\tilde{w}_k^2(n-1)] + \frac{2\beta \lambda (1 - \lambda)}{\sigma_z^2} E[\tilde{w}_k(n-1)] \]

\[ g(w_k(n-1))] + \frac{\beta^2 (1 - \lambda)^2}{\sigma_x^2} E[g^2(w_k(n-1))] \]

\[ + \frac{(1 - \lambda)^2 \sigma_x^2}{\sigma_z^2}. \]

(25)

To evaluate \( \lim_{n \to \infty} E[\tilde{w}_k^2(n)] \) recursively using (25), we need to evaluate the terms \( E[\tilde{w}_k(n-1)g(w_k(n-1))] \) and \( E[g^2(w_k(n-1))] \), for each \( k \in \{1, 2, \ldots, N\} \). First consider the case of \( k \in C_L \). From assumptions (6) and for large \( n \), as seen earlier, \( g(w_k(n)) = 0 (\forall k \in C_L) \) almost surely.

This means that for large \( n \), \( E[\tilde{w}_k(n-1)g(w_k(n-1))] = 0 \) and \( E[g^2(w_k(n-1))] = 0 \). Substituting in (25) and noting that \( 0 < \lambda < 1 \), (25) gives rise to the following stable, steady state solution:

\[ \lim_{n \to \infty} E[\tilde{w}_k^2(n)] = \frac{(1 - \lambda) \sigma_x^2}{(1 + \lambda) \sigma_z^2}. \]

(26)

From this, \( D_L(\infty) \) as given by (17) follows trivially.

Next we consider \( k \in C_S \). From (12), one can write,

\[ E[\tilde{w}_k(n-1)g(w_k(n-1))] = \alpha^2 E[\tilde{w}_k^2(n-1)] + E[\tilde{w}_k(n-1)]g(w_0,k), \]

(27)

and

\[ E[g^2(w_k(n-1))] = \alpha^4 E[\tilde{w}_k^2(n-1)] + g^2(w_0,k) + 2\alpha^2 g(w_0,k)E[\tilde{w}_k(n-1)]. \]

(28)

Substituting in (25), we have, for \( k \in C_S \),

\[ E[\tilde{w}_k^2(n)] = \lambda^2 E[\tilde{w}_k^2(n-1)] + \frac{(1 - \lambda) \sigma_x^2}{\sigma_z^2} \]

\[ + \frac{2\beta \lambda (1 - \lambda)}{\sigma_z^2} + \frac{\beta^2 (1 - \lambda)^2 \alpha^2}{\sigma_x^2} \]

\[ g(w_0,k)E[\tilde{w}_k(n-1)] \]

\[ + \frac{\beta^2 (1 - \lambda)^2}{\sigma_x^2} g^2(w_0,k), \]

(29)
where $\lambda' = \sqrt{\lambda^2 + \frac{2\beta\lambda(1-\lambda)\sigma_x^2}{\sigma_z^2} + \frac{\beta^2(1-\lambda)^2\sigma_z^4}{\sigma_x^4}}$.

For large $n$, from (5), $E[\hat{w}_k(n-1)]$ can be approximated by its steady state value $\frac{\beta}{\sigma_z^2 - \beta\alpha^2} g(w_0, k)$.

The stability of (29) will then require $\lambda'$ to be less than 1. Since $\beta = \gamma(1 - \lambda)$ where $\gamma$ is very very small $(\approx 10^{-4})$ while $\lambda \approx 1$, and $\alpha$ is typically in the range of 50 - 60, for practical signals with $\sigma_z^2 \approx 1$, it is easy to see that $\lambda'^2 \approx \lambda^2$ and since $0 < \lambda < 1$, (29) corresponds to a stable system, with

$$
\lim_{n \to \infty} E[\hat{w}_k^2(n)] = \frac{(1 - \lambda)^2\sigma_n^2}{(1 - \lambda^2)\sigma_z^2} + \frac{\beta'g^2(w_0,k)}{(1 - \lambda^2)}, \quad \forall k \in C_S,
$$

where $\beta' = \frac{2\beta^2(1-\lambda)}{\sigma_z^2 - \beta\alpha^2} + \frac{\beta^2(1-\lambda)^2\sigma_z^4}{\sigma_x^4} + 2\frac{\beta^2(1-\lambda)^2\sigma_z^4}{\sigma_x^4} - \frac{\beta^2}{\sigma_z^2} - \beta\alpha^2$.

From (30), $D_S(\infty)$ as given by (18), follows directly.

Lastly, we consider the case of $k \in C_0$, for which we have $\hat{w}_k(n-1) = w_k(n-1)$. Since $\beta$ is very very small and thus $\beta\alpha^2 \ll \sigma_z^2$, from (30), it is safe to assume that $\lim_{n \to \infty} E[\hat{w}_k(n)] = \lim_{n \to \infty} E[w_k(n)] \approx 0, \quad \forall k \in C_0$.

Also, as per assumption (12) $w_k(n) \forall k \in C_0$, is Gaussian distributed in the steady state. One can then apply Price’s theorem (12) and (14) to write the following:

$$
E[\hat{w}_k(n-1)g(w_k(n-1))] = \alpha^2 E[w_k^2(n-1)] - \sqrt{\frac{2}{\pi}}\alpha \sqrt{E[w_k^2(n-1)]},
$$

and

$$
E[g^2(w_k(n-1))] = \alpha^4 E[w_k^2(n-1)] + \alpha^2 - 2\sqrt{\frac{2}{\pi}}\alpha^3 \sqrt{E[w_k^2(n-1)]}.
$$

Substituting in (25), and using the notation $\omega = \lim_{n \to \infty} \sqrt{E[w_k^2(n)]}$, we then obtain,

$$
b_\omega + b_\omega \omega + c_\omega = 0, \quad \text{Eq.}(33)
$$

The roots of (33) are given by $\omega = -\frac{b_\omega \pm \sqrt{b_\omega^2 - 4c_\omega}}{2}$, $\forall k \in C_0$.

Under stability assumption, $0 < \lambda' < 1$ and thus $b_\omega > 0$, $c_\omega < 0$. Eq. (33) then has only one positive root given by

$$
\omega = -\frac{b_\omega + \sqrt{b_\omega^2 - 4c_\omega}}{2},
$$

which provides the steady state value of $\sqrt{E[w_k^2(n)]}$, $\forall k \in C_0$.

For the above choice of $\omega$, from (33), one can also write $\lim_{n \to \infty} E[\hat{w}_k^2(n)] = \lim_{n \to \infty} E[w_k^2(n)] = \omega^2 = -b_\omega \omega - c_\omega$. From this, $D_0(\infty)$ as given by (19) follows trivially.

REFERENCES

[1] E. Hansler, “The hands-free telephone problem- an annotated bibliography”, Signal Processing, vol. 27, no. 3, pp. 259-271, Jun. 1992.
[2] C. Paleologou, J. Benesty, and S. Ciochina, Sparse Adaptive Filters for Echo Cancellation, Morgan and Claypool Publishers, June 2010.
[3] Y. Chen, Y. Gu, and A. O. Hero, “Sparse LMS for System Identification”, Proc. IEEE ICASSP-2009, April 2009, Taipei, Taiwan.
[4] Y. Chen, Y. Gu, and A. O. Hero, “Regularized least-mean-square algorithms,” Arxiv preprint arXiv:1012.5066v2 [stat.ME], Dec. 2010.
[5] Y. Gu, J. Jin, and S. Mei, "ℓ₀ norm constraint LMS algorithm for Sparse System Identification", IEEE Signal Processing Letters, vol. 16, No. 9, pp. 774-777, Sept. 2009.
[6] G. Su, J. Jin, Y. Gu, and J. Wang, “Performance analysis of ℓ₀ norm constraint least mean square algorithm”, IEEE Trans. Signal Proc., vol. 60, No. 5, pp. 2223-2235, Sept. 2012.
[7] E. M. Eksioglu and A. K. Tanc, “RLS algorithm with convex regularization”, IEEE Signal Processing Letters, vol. 18, No. 8, pp. 470-473, Aug. 2011.
[8] D. Angelosante, J. A. Bazerque, and G. B. Giannakis, “Online adaptive estimation of sparse signals: where RLS meets the ℓ₁-norm”, IEEE Trans. Signal Proc., vol. 58, No. 7, pp. 3436-3447, July 2010.
[9] B. Babadi, N. Kalouptsidis, and V. Tarokh, “SPARLS: the sparse RLS algorithm”, IEEE Trans. Signal Proc., vol. 58, No. 8, pp. 4013-4025, Aug. 2010.
[10] S. Haykin, Adaptive Filter Theory, Englewood Cliffs, NJ: Prentice-Hall, 1986.
[11] A. H. Sayed, Fundamentals of Adaptive Filtering, New York, Wiley, 2003.
[12] A. Papoulis and S. U. Pillai, Probability, Random Variables and Stochastic Processes, McGraw Hills.