MINIMUM-DISTURBANCE DESCRIPTION FOR THE DEVELOPMENT OF ADAPTATION ALGORITHMS AND A NEW LEAKAGE LEAST SQUARES ALGORITHM

Fabiano T. Castoldi and Marcello L. R. de Campos

Programa de Engenharia Elétrica, COPPE/UFRJ, P.O.Box 68504, Rio de Janeiro, RJ, 21.945-970, Brazil.
ftoldi@gmail.com; campos@lps.ufrj.br

ABSTRACT

Usual methods for the development of adaptive filters are based on a stochastic approximation of the gradient vector and Hessian matrix, or on a deterministic minimization of quadratic *a posteriori* output errors. Gradient-based algorithms are usually placed in the first group, whereas least squares (LS) based algorithms are placed in the second group. These are just how algorithms are usually presented and analyzed and alternative descriptions exit. This paper proposes to shed new light onto known adaptation algorithms by means of a minimum-disturbance approach to the cost function together with constraints added to improve their robustness. The resulting algorithms are able to perform extremely well in many demanding applications.

Index Terms— Optimization methods, adaptive filters, adaptive signal processing, minimum-disturbance description

1. INTRODUCTION

Many adaptation algorithms have been proposed in the past forty years, offering trade-offs in convergence speed, robustness, and computational complexity [1]–[3]. Lack of robustness may be due to accumulation of quantization errors or due to loss of positive definiteness of the Hessian matrix caused by nonpersistently exciting input signals [4], [5].

The recursive least squares (RLS) algorithm presents good convergence speed, but its robustness is not guaranteed unless we opt for a QR-decomposition implementation, or for some regularization scheme. Robust RLS algorithm implementations with reduced computational complexity are usually based on QR decompositions, which are complex to implement and maintain [1]. There are other algorithms that have been developed based on known convex optimization methods, like the quasi-Newton (QN) [4] and interior point least squares (IPLS) algorithms [6]. These algorithms offer increased robustness at a cost of extra computational complexity, for they do not admit $O(N)$ implementations.

In this paper we present a different approach to the derivation of conventional adaptation algorithms by describing their deterministic cost function as a quadratic norm of the coefficient update minimum-disturbance, subjected or not to equality constraints. As a consequence of this approach we were able to develop new algorithms that are fast and more robust alternatives to conventional algorithms in many application scenarios.

This paper is organized as follows. In Section 2 we present the QN algorithm using its deterministic representation and using the minimum-disturbance description method. The method is also used in the following sections to derive other known algorithms as well as to propose new algorithms based on the LS cost function. A convergence performance comparision is presented in Section 7 and concluding remarks are given in Section 8.

Thanks to CAPES and CNPq for supported this work.

2. QN ALGORITHM AND THE MINIMUM-DISTURBANCE APPROACH

The QN algorithm proposed in [4] is a robust algorithm that was developed using a stochastic approach and is based on the rank-one quasi-Newton update of the Hessian matrix, i.e., satisfying the QN hereditary condition [7]. However, this algorithm can also be shown to minimize the following deterministic cost function:

$$\xi_{\mathrm{QN},n} = \sum_{i=1}^{n} (1-\mu_i) \left(\frac{d_i - \mu_i \mathbf{x}_i^T \mathbf{w}_{i-1}}{1-\mu_i} - \mathbf{x}_i^T \mathbf{w}_n \right)^2 + \frac{1}{2} \mathbf{w}_n^T \mathbf{R}_{\mathrm{QN},0} \mathbf{w}_n, \quad (1)$$

where μ_i is the step-size, d_i is the reference signal, $\mathbf{x}_i$ and $\mathbf{w}_n$ are the input data vector and the coefficient vector, both with equal length N, and $\mathbf{R}_{\mathrm{QN},0}$ is the correlation matrix at instant 0. Although interesting from an algebraic point of view, this cost function provides limited insight into the algorithm behavior. This algorithm also takes into account a constraint in the Hessian matrix, which is responsible for its good numerical properties. Minimizing (1) and considering the constraint

$$\mathbf{x}_n^T \mathbf{R}_{\mathrm{QN},n}^{-1} \mathbf{x}_n = \frac{1}{2}, \quad (2)$$

the following equations for the adaptation algorithm are obtained:

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \frac{e_n}{\tau_n} \mathbf{t}_n \quad (3)$$

$$\mathbf{R}_{\mathrm{QN},n}^{-1} = \mathbf{R}_{\mathrm{QN},n-1}^{-1} - \frac{1-\mu_n}{\tau_n} \mathbf{t}_n \mathbf{t}_n^T \quad (4)$$

$$\mu_n = \frac{1}{2\tau_n} \quad (5)$$

where e_n is the *a priori* output error given by $e_n = d_n - \mathbf{x}_n^T \mathbf{w}_{n-1}$, and τ_n and $\mathbf{t}_n$ are given by:

$$\tau_n = \mathbf{x}_n^T \mathbf{R}_{\mathrm{QN},n-1}^{-1} \mathbf{x}_n \quad (6)$$

$$\mathbf{t}_n = \mathbf{R}_{\mathrm{QN},n-1}^{-1} \mathbf{x}_n. \quad (7)$$

2.1. Minimum-Disturbance Description

An alternative cost function that yields the same algorithm described by (3)–(7) can be represented by a minimum-disturbance approach. The idea is to use some quadratic norm of the coefficient disturbance from iteration $n-1$ to n, but which also takes into account the zero instantaneous *a posteriori* output error, which is one of the features of this algorithm. The QN algorithm can be obtained as a solution to the following convex minimization problem:

$$\xi_{\mathrm{QN},n} = \|\mathbf{w}_n - \mathbf{w}_{n-1}\|^2_{\mathbf{R}_{\mathrm{QN},n-1}} \text{ s.t. } \begin{cases} d_n - \mathbf{x}_n^T \mathbf{w}_n = 0 \\ \mathbf{x}_n^T \mathbf{R}_{\mathrm{QN},n}^{-1} \mathbf{x}_n = 1/2 \end{cases} \quad (8)$$

where the weighted quadratic norm is defined as $\|\mathbf{x}\|^2_{\mathbf{A}} = \mathbf{x}^T \mathbf{A} \mathbf{x}$.

This minimum-disturbance description has been used before to derive and analyze adaptation algorithms, see, e.g., [3, 5]. However, in [5] this approach was used to derive and evaluate a least perturbation characteristic of the normalized least mean squares (NLMS) algorithm with step-size $\mu = 1$, and in [3] this approach was generalized. Our approach and objectives in this paper are slightly different: it is to bring the derivation of different adaptation algorithms under the same framework and to provide a convex cost function that is easy to work with and which gives some insight into the algorithm behavior.

Using Lagrange multipliers [8] to minimize the cost function described in (8), we can derive this algorithm with respect to $\mathbf{w}_n$, yielding

$$\xi_{\mathrm{QN},n} = \|\mathbf{w}_n - \mathbf{w}_{n-1}\|^2_{\mathbf{R}_{\mathrm{QN},n-1}} + \alpha \left[d_n - \mathbf{x}_n^T \mathbf{w}_n \right]$$

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \frac{\alpha \mathbf{R}_{\mathrm{QN},n-1}^{-1} \mathbf{x}_n}{2}. \quad (9)$$

Substituting this result in the zero instantaneous *a posteriori* error constraint to find α, yields

$$d_n - \mathbf{x}_n^T \left(\mathbf{w}_{n-1} + \frac{\alpha \mathbf{R}_{\mathrm{QN},n-1}^{-1} \mathbf{x}_n}{2} \right) = 0$$

$$2e_n = \alpha \mathbf{x}_n^T \mathbf{R}_{\mathrm{QN},n-1}^{-1} \mathbf{x}_n$$

$$\alpha = \frac{2e_n}{\tau_n}$$

where τ_n is given by (6). Now, using this value of α in (9) we obtain the same recursion formula for the update of the filter coefficients vector, (3).

The inverse correlation matrix $\mathbf{R}_{\mathrm{QN},n-1}^{-1}$ is computed so that the algorithm fits in the quasi-Newton category [7], which can be accomplished in the same way as in [4], resulting (4). Substituting (4) into (2), pre- and post-multiplied by $\mathbf{x}_n^T$ and $\mathbf{x}_n$, respectively, we obtain the value of μ_n as in (5).

As can be seen, (8) helps us understand and to explain the QN algorithm much better than (1). The idea of a new approach to interpret adaptation algorithms using the minimum-disturbance description can be extended to other adaptation algorithms.

3. NLMS ALGORITHM

A minimum-disturbance description for the NLMS algorithm is given by [5]

$$\xi_{\mathrm{NLMS},n} = \|\mathbf{w}_n - \mathbf{w}_{n-1}\|^2 \qquad \text{s.t.} \quad d_n = \mathbf{x}_n^T \mathbf{w}_n. \quad (10)$$

The solution, after minimizing $\xi_{\mathrm{NLMS},n}$ with respect to $\mathbf{w}_n$, is the NLMS algorithm with step-size equal to 1:

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \frac{e_n}{\|\mathbf{x}_n\|^2} \mathbf{x}_n. \quad (11)$$

4. RLS ALGORITHM

The RLS algorithm has certainly become one of the preferred alternatives to gradient-type algorithms in applications where fast convergence is needed. The conventional form of the RLS algorithm can be derived as a recursive implementation that minimizes the weighted sum of the squares of the *a posteriori* output errors. One possible form of the RLS cost function can be written as [1]

$$\xi_{\mathrm{RLS},n} = \sum_{i=1}^{n} \mu_{i,n} \left(d_i - \mathbf{x}_i^T \mathbf{w}_n \right)^2 \quad (12)$$

where the weights $\mu_{i,n}$, $i = 1, \cdots, n$, control the importance of past information and are called forgetting factors:

$$\mu_{i,n} = \prod_{j=i+1}^{n} \mu_j = \mu_n \, \mu_{i,n-1}.$$

The vector $\mathbf{w}_n$ which minimizes $\xi_{\mathrm{RLS},n}$ can be obtained taking the derivative of $\xi_{\mathrm{RLS},n}$ with respect to $\mathbf{w}_n$ and setting it equal to zero.

The same algorithm can be interpreted using the minimum-disturbance description, where the instantaneous squared *a posteriori* output error is also taken into account:

$$\xi_{\mathrm{RLS},n} = \mu_n \|\mathbf{w}_n - \mathbf{w}_{n-1}\|^2_{\mathbf{R}_{\mathrm{RLS},n-1}} + \left(d_n - \mathbf{x}_n^T \mathbf{w}_n \right)^2. \quad (13)$$

Solving this minimization problem with respect to $\mathbf{w}_n$, we get

$$\nabla \xi_{\mathrm{RLS},n} = 2\mu_n \mathbf{R}_{\mathrm{RLS},n-1} \left(\mathbf{w}_n - \mathbf{w}_{n-1} \right) - 2 \left(d_n - \mathbf{w}_n^T \mathbf{x}_n \right) \mathbf{x}_n = 0 \quad (14)$$

$$\left(\mu_n \mathbf{R}_{\mathrm{RLS},n-1} + \mathbf{x}_n \mathbf{x}_n^T \right) \mathbf{w}_n = \mu_n \mathbf{R}_{\mathrm{RLS},n-1} \mathbf{w}_{n-1} + d_n \mathbf{x}_n. \quad (15)$$

Here we can define the equation for the correlation matrix as $\mathbf{R}_{\mathrm{RLS},n} = \mu_n \mathbf{R}_{\mathrm{RLS},n-1} + \mathbf{x}_n \mathbf{x}_n^T$. Continuing the solution to find the update equation for the coefficient vector, we obtain

$$\mathbf{R}_{\mathrm{RLS},n} \mathbf{w}_n = \left(\mu_n \mathbf{R}_{\mathrm{RLS},n-1} + \mathbf{x}_n \mathbf{x}_n^T \right) \mathbf{w}_{n-1} - \mathbf{x}_n \mathbf{x}_n^T \mathbf{w}_{n-1} + d_n \mathbf{x}_n \quad (16)$$

$$\Rightarrow \mathbf{R}_{\mathrm{RLS},n} \mathbf{w}_n = \mathbf{R}_{\mathrm{RLS},n} \mathbf{w}_{n-1} + e_n \mathbf{x}_n. \quad (17)$$

Multiplying both sides of the equation above by the inverse of the correlation matrix, $\mathbf{R}_{\mathrm{RLS},n}^{-1}$, which can be calculated using the matrix inversion lemma, we obtain the equations of the RLS algorithm:

$$\mathbf{w}_n = \mathbf{w}_{n-1} + e_n \mathbf{R}_{\mathrm{RLS},n}^{-1} \mathbf{x}_n \quad (18)$$

$$\mathbf{R}_{\mathrm{RLS},n}^{-1} = \frac{1}{\mu_n} \left(\mathbf{R}_{\mathrm{RLS},n-1}^{-1} - \frac{\mathbf{t}_n \mathbf{t}_n^T}{\mu_n + \tau_n} \right). \quad (19)$$

4.1. Leakage QN Algorithm

In [9], a new algorithm was developed with the objective of relaxing the normalization of the QN algorithm but retaining its robust convergence properties. Using the minimum-disturbance description, it is straighforward to incorporate the constraint to any cost function. The constraint of (2) was used together with the RLS cost function, resulting in the leakage quasi-Newton (LQN) algorithm. We can calculate the step-size for which this constraint is satisfied.

One way to do this is by directly substituting (19) in the constraint $\mathbf{x}_n^T \mathbf{R}_{\text{RLS},n}^{-1} \mathbf{x}_n = 1/2$ and solving it with respect to μ_n, which results in $\mu_n = \tau_n$.

The LQN algorithm does not have user-defined parameters to adjust, which gives a different degree of robustness, avoiding bad choice of the step-size by the user, and giving fast convergence speed and low misadjustment, lower than that of the QN algorithm [9].

5. LEAKAGE LS ALGORITHM

A different approach to the LS cost function than that presented in Section 4 is to use a simpler weight factor instead of the one used in (12). The parameter μ_n acts only in its respective quadratic *a posteriori* error, instead of upon the whole sequence of past *a posteriori* errors, as in the RLS algorithm:

$$\xi_{\text{LLS},n} = \sum_{i=1}^{n} \mu_i \left(d_i - \mathbf{x}_i^T \mathbf{w}_n \right)^2. \tag{20}$$

This cost function can be modified using the minimum-disturbance description as we did for the RLS algorithm:

$$\xi_{\text{LLS},n} = \|\mathbf{w}_n - \mathbf{w}_{n-1}\|^2_{\mathbf{R}_{\text{LLS},n-1}} + \mu_n \left(d_n - \mathbf{w}_n^T \mathbf{x}_n \right)^2. \tag{21}$$

Minimizing this cost function gives:

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \frac{\mu_n e_n}{1 + \mu_n \tau_n} \mathbf{t}_n \tag{22}$$

$$\mathbf{R}_{\text{LLS},n}^{-1} = \mathbf{R}_{\text{LLS},n-1}^{-1} - \frac{\mu_n \mathbf{t}_n \mathbf{t}_n^T}{1 + \mu_n \tau_n} \tag{23}$$

where the inverse of the correlation matrix was obtained using the matrix inversion lemma in the correlation matrix update formula, $\mathbf{R}_{\text{LLS},n} = \sum_{i=1}^{n} \mu_i \mathbf{x}_i \mathbf{x}_i^T = \mathbf{R}_{\text{LLS},n-1} + \mu_n \mathbf{x}_n \mathbf{x}_n^T$. This algorithm is similar to the BEACON algorithm [10], the difference residing in the method used to update the step-size.

One can use the method presented in [10] to calculate the step-size and to get exactly the BEACON algorithm. However, using the constraint of (2) another algorithm can be developed that does not need user-defined parameters with the step-size given by $\mu_n = (2\tau_n - 1)/\tau_n$. This equation was derived by directly substituting the (23) in the constraint, (2). This algorithm, called leakage least squares (LLS), was analyzed in [11] under persistently and non-persistently exciting input signals, and presented poor convergence performance. However, a modification in this algorithm using box constraints such that $\alpha < \mathbf{x}_n^T \mathbf{R}_{\text{LLS},n}^{-1} \mathbf{x}_n < \beta$ was also suggested in [11], resulting in a robust algorithm for any kind of input signal.

As in the previous case, the constraint does not directly modify the derivation of the algorithm's equations. Using an interior-point method, called the logarithmic barrier [8], we can solve this new minimization problem obtained by the two inequality functions:

$$\begin{aligned}\min_{\mu_n} \quad & -\left(\frac{1}{p_1}\right) \log \left(\mathbf{x}_n^T \mathbf{R}_{\text{LLS},n}^{-1} \mathbf{x}_n - \alpha \right) \\ & - \left(\frac{1}{p_2}\right) \log \left(-\mathbf{x}_n^T \mathbf{R}_{\text{LLS},n}^{-1} \mathbf{x}_n + \beta \right)\end{aligned} \tag{24}$$

where α and β are the limits of the constraints, and p_1 and p_2 control the barrier precision. As p_i increases, so does the precision.

Solving this problem one can get the following step-size update equation:

$$\mu_n = \frac{\tau_n (p_1 + p_2) - (\alpha p_1 + \beta p_2)}{\tau_n (\alpha p_1 + \beta p_2)}. \tag{25}$$

The step-size must always be kept positive. We opted to use a hard decision, setting $\mu_n = 0$ whenever (25) produced a negative value. The convergence analysis of this algorithm can follow the same considerations used for the BEACON algorithm [10].

6. CONVEX LS ALGORITHM

If we closely examine the minimum-disturbance description of the algorithms presented in the last two sections, a natural, perhaps trivial, modification of the objective function is to make it a convex combination between minimum disturbance and *a posteriori* output error, as follows:

$$\begin{aligned}\xi_{\text{CONV},n} = (1 - \mu_n) \|\mathbf{w}_n - \mathbf{w}_{n-1}\|^2_{\mathbf{R}_{\text{CONV},n-1}} \\ + \mu_n \left(d_n - \mathbf{x}_n^T \mathbf{w}_n \right)^2\end{aligned} \tag{26}$$

If we set the derivative of $\xi_{\text{CONV},n}$ with respect to $\mathbf{w}_n$ equal to zero and solve for $\mathbf{w}_n$, we obtain

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \frac{\mu_n e_n \mathbf{t}_n}{1 - \mu_n + \mu_n \tau_n} \tag{27}$$

$$\mathbf{R}_{\text{CONV},n}^{-1} = \frac{1}{1 - \mu_n} \left(\mathbf{R}_{\text{CONV},n-1}^{-1} - \frac{\mu_n \mathbf{t}_n \mathbf{t}_n^T}{1 - \mu_n + \mu_n \tau_n} \right) \tag{28}$$

where (28) was obtained directly from the development of the algorithm and using the inverse matrix lemma in the correlation matrix, $\mathbf{R}_{\text{CONV},n} = (1 - \mu_n) \mathbf{R}_{\text{CONV},n-1} + \mu_n \mathbf{x}_n \mathbf{x}_n^T$.

This algorithm was presented before using different approaches. In [12] this algorithm was developed as a modified RLS algorithm using another convex optimization approach, called optimal bounding ellipsoids (OBE). Interesting enough, this algorithm also becomes equivalent to the RLS algorithm when $1 - \mu_n = \lambda < 1$, which has already been named LMS-Newton in [13]. Indeed, the same algorithm can be obtained if we minimize

$$\xi_{\text{CONV},n} = \sum_{i=1}^{n} \prod_{j=i+1}^{n} (1 - \mu_j) \mu_i \left(d_i - \mathbf{x}_i^T \mathbf{w}_n \right)^2 \tag{29}$$

with respect to $\mathbf{w}_n$.

7. SIMULATION RESULTS

To emphasize the good convergence capabilities of the LLS algorithm and the effect of the constraints applied to it, this section provides a simulation comparison between the LLS, RLS, and QN algorithms. A sinusoidal signal, which is a nonpersistently exciting signal, is used as input signal to an adaptive system in a system identification configuration. The unknown system is an FIR filter with 5 coefficients normalized to unit norm: $H(z) = (1 + z^{-1} + z^{-2} + z^{-3} + z^{-4})/\sqrt{5}$.

The algorithm parameters were adjusted to yield the best performance of each algorithm in terms of misadjustment: $\mu_{\text{RLS}} = 0.99$ for the RLS algorithm; $p_1 = 10$, $p_2 = 1$, $\alpha = 0.0001$ and $\beta = 0.1$ for the LLS algorithm. The simulation results presented were averaged over an ensemble of 100 simulations. The signal-to-noise ratio (SNR) used was constant and equal to 40 dB.

Figure 1 shows the learning curve of mean squared errors (MSE) for the first 250 iterations and after 3000 iterations. As expected, the RLS algorithm diverges after some iterations in the presence of non-persistently exciting signals, whereas the QN and LLS algorithms did not diverge.

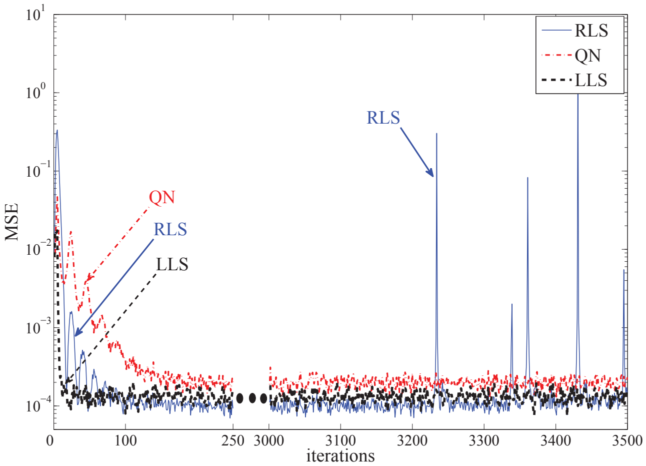

Fig. 1. Learning curves of RLS, QN and LLS algorithms, for the first 250 iterations and after 3000 iterations. The RLS algorithm is diverging around 3200 iterations.

It can also be seen that the LLS algorithm has the best performance, followed by that of the QN algorithm, which has a 3 dB misadjustment due to normalization. The LLS algorithm, on the other hand, converges faster than both QN and RLS algorithms, and with a misadjustment error lower than that of the QN algorithm and similar to that of the RLS algorithm.

8. CONCLUSIONS AND CRITIQUE

This work further investigated the derivation of adaptation algorithms via a coefficient-vector minimum-disturbance approach combined with constraints involving the *a posteriori* output error and estimate of the Hessian matrix. Although a similar approach has been used before to present alternative derivations of the NLMS algorithm and its variations, we have showed that a proper choice of the quadratic norm used to measure the *disturbance* of the coefficients can yield different LS-like or QN-like algorithms. Besides giving us tools to derive new algorithms, the approach described herein puts all algorithms under the same framework and provides us insight on their expected behavior.

The framework just described allows us to introduce different constraints to the adaptation algorithms together with the cost function, as was the case studied for the positive definiteness of the correlation matrix. This work showed that with proper choice of the constraints, robust algorithms that also have fast convergence speed can be derived. Another interesting aspect of this framework is the presentation of an alternative deterministic objective function for LS-like algorithms that does not include n instances of the coefficient vector. Although a detail that might easily pass unnoticed, the alternative deterministic objective function allows us to optimize with respect to the convergence factor, μ_n, without the difficulty imposed by the implicit dependence of the n instances of $\mathbf{w}_n$ on μ_n.

This paper also introduces a new robust leakage algorithm that maintains its convergence independently of the type of input signals present in the system. Simulation results showed that the LLS algorithm can be an option where fast and robust adaptation algorithms are needed.

The derivation of adaptation algorithms and variable forgetting factors are a subject of further study, and this work offers a different perspective that may attract and amuse the expert and help the novice.

9. REFERENCES

[1] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, Springer, New York, 3rd edition, 2008.

[2] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, New Jersey, 4th edition, 2001.

[3] A. H. Sayed, *Fundamentals of Adaptive Filtering*, John Wiley & Sons, 2003.

[4] M. L. R. de Campos and A. Antoniou, "A new quasi-Newton adaptive filtering algorithm," in *IEEE Transactions on Circuits and Systems*, November 1997, vol. 44, pp. 924–934.

[5] G. C. Goodwin and Kwai Sang Sin, *Adaptive Filtering Prediction and Control*, Prentice-Hall, Englewood Cliffs, New Jersey, 1984.

[6] K. H. Afkhamie, Z.-Q. Luo, and K. M. Wong, "Adaptive linear filtering using interior point optimization techniques," in *IEEE Transactions on Signal Processing*, June 2000, vol. 48, pp. 1637–1647.

[7] R. Fletcher, *Practical Methods of Optimization*, Wiley, New York, 2nd edition, 1987.

[8] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, New York, 2004.

[9] F. T. Castoldi and M. L. R. de Campos, "A leakage quasi-Newton adaptation algorithm," in *XXV Simpósio Brasileiro de Telecomunicações*, September 2007.

[10] S. Nagaraj, S. Gollamudi, S. Kapoor, and Y.-F. Huang, "Beacon: An adaptive set-membership filtering technique with sparse updates," in *IEEE Transactions on Signal Processing*, November 1999, vol. 47, pp. 2928–2941.

[11] F. T. Castoldi and M. L. R. de Campos, "On the step-size of leakage QN and leakage LS algorithms," in *XXVI Simpósio Brasileiro de Telecomunicações*, September 2008.

[12] S. Dasgupta and Y.-F. Huang, "Asymptotically convergent modified recursive least-squares with data-dependent updating and forgetting factor for systems with bounded noise," in *IEEE Transactions on Information Theory*, May 1987, vol. 33, pp. 383–392.

[13] M. L. R. de Campos, P. S. R. Diniz, and A. Antoniou, "Analysis of LMS-Newton adaptive filtering algorithms with variable convergence factor," *IEEE Transactions of Signal Processing*, vol. 43, pp. 617–627, 1995.