Adaptive Combination of Normalized Filters for Robust System Identification

J. Arenas-García and A. R. Figueiras-Vidal

The performance of square norm adaptive filters degrades in the presence of impulsive noise. The sign algorithm overcomes this problem by minimizing the absolute value of the error, at the cost of a slower convergence. Here, we present an adaptive combination of two independent filters, using the absolute and quadratic error criteria, respectively, so that the resulting scheme gets a fast convergence and remains robust to impulsive noise.

Introduction: Stochastic gradient filters play an essential role in many adaptive signal processing applications, such as channel equalization and system identification [1]. The Least Mean Square (LMS) algorithm has probably become the most widely used adaptive filtering scheme, due to its simplicity and stability. However, when the filtering scenario is contaminated with impulsive or heavy tailored noise, LMS performance significantly degrades. The Sign Algorithm (SA) is not affected by these kinds of noise, but it converges slower than LMS [1].

Recently, mixed-norm adaptive filters have been proposed in an attempt to put together the best properties of LMS and SA [2, 3]. Robust Mixed-Norm (RMN) filters minimize the cost function

$$L(n) = \lambda(n)E\{e^2(n)\} + [1 - \lambda(n)]E\{|e(n)|\} \tag{1}$$

so that, in general, a minimum quadratic error is not guaranteed.

Instead of combining cost functions, we propose to use a convex combination of two filters minimizing the absolute and quadratic errors. Each filter is adapted according to its own criterion, while their combination is adapted to minimize the overall quadratic error. Using Normalized LMS (NLMS) and SA (NSA) algorithms [4] will provide additional insensitivity with respect to the input signal level. We will present below our combination scheme and illustrate its effectiveness at joining the fast convergence of NLMS with the robustness against impulsive noise of NSA.

Normalized Adaptive Filters Performance: To know how NSA and NLMS perform is useful for combining them. We will consider a plant identification setup, where the objective is to identify an unknown system from some measurements at its input and output. The output is given by $d(n) = \mathbf{w}_0^T\mathbf{u}(n) + e_0(n)$, where $\mathbf{w}_0$ is an M length vector characterizing the plant, $\mathbf{u}(n)$ is the input vector which is taken from the input signal $u(n)$ as $\mathbf{u}(n) = [u(n), u(n-1), \cdots, u(n-M+1)]^T$, and $e_0(n)$ is an i.i.d. measurement noise, independent of $\mathbf{u}(n)$ and with variance σ_0^2.

The outputs of the adaptive filters are $y_i(n) = \mathbf{w}_i^T(n)\mathbf{u}(\mathbf{n})$, where $\mathbf{w}_1$ and $\mathbf{w}_2$ are the NSA and NLMS weights, respectively, and are adapted via

$$\mathbf{w}_i(n+1) = \mathbf{w}_i(n) + \mu_i g_i(n)\mathbf{u}(n) \tag{2}$$

μ_1 and μ_2 being the step-sizes, and

$$g_1(n) = \frac{\text{sign}[e_1(n)]}{\varepsilon_1 + \|\mathbf{u}(n)\|^2} \quad \text{and} \quad g_2(n) = \frac{e_2(n)}{\varepsilon_2 + \|\mathbf{u}(n)\|^2} \tag{3}$$

where $e_i(n) = d(n) - y_i(n)$ are the errors incurred by the filters and ε_i are small constants to avoid division by 0.

To measure the stationary performance of adaptive filters, it is customary to use the limiting value of the excess mean-square error (EMSE) [1, Ch. 6]:

$$J_{ex,i}(\infty) = \lim_{n\to\infty} E\{e_{a,i}(n)\} \tag{4}$$

where we have introduced the a-priori errors of the filters, $e_{a,i}(n) = e_i(n) - e_0(n)$.

Applying energy conservation arguments, it can be shown that the stationary behaviour of (2) is characterized by [1, Eq. (6.4.9)]

$$\mu_i E\{\|\mathbf{u}(n)\|^2 g_i^2(n)\} = 2E\{e_{a,i}(n)g_i(n)\}; \quad n \to \infty \tag{5}$$

from which NLMS misadjustment is obtained as [1, Eq. (6.6.7)]:

$$\boxed{J_{ex,2}(\infty) = \mu_2\sigma_0^2/(2-\mu_2)} \tag{6}$$

For NSA, we get a similar result by first introducing $g_1(n)$ into (5), arriving to

$$\mu_1 E\left\{\frac{\|\mathbf{u}(n)\|^2}{[\varepsilon_1 + \|\mathbf{u}(n)\|^2]^2}\right\} = 2E\left\{\frac{e_{a,1}(n)\text{sign}[e_{a,1}(n)+e_0(n)]}{\varepsilon_1 + \|\mathbf{u}(n)\|^2}\right\}; \quad n \to \infty \tag{7}$$

We can simplify this expression assuming that, in steady-state $\|\mathbf{u}(n)\|^2$ is independent of $e_{a,1}(n)$ [1]. Accepting also that ε_1 is small enough, we have

$$\mu_1 = 2E\{e_{a,1}(n)\text{sign}[e_{a,1}(n) + e_0(n)]\}; \quad n \to \infty \tag{8}$$

Finally, for small μ_1 the expectation can be further simplified by means of Price's Theorem. After some elementary manipulations we arrive to

$$\boxed{J_{ex,1}(\infty) = \frac{\mu_1'}{2}\left(\mu_1' + \sqrt{\mu_1'^2 + 4\sigma_0^2}\right)} \tag{9}$$

with $\mu_1' = \mu_1\sqrt{\pi/8}$.

We can see that both NSA and NLMS stationary behaviours are independent of the input signal power. Regarding the influence of the measurement noise, $J_{ex,2}(\infty)$ linearly increases with σ_0^2 while the NSA misadjustment is approximately proportional to σ_0. We have checked that there exists a very close agreement between these theoretical results and the real performance of the filters.

Adaptive Combination of Filters: To put together the best properties of the NSA and NLMS filters we propose to use an adaptive convex combination of the outputs of both schemes:

$$\boxed{y(n) = \eta(n)y_1(n) + [1 - \eta(n)]y_2(n)} \tag{10}$$

where $\eta(n) \in (0, 1)$. If $\eta(n)$ takes an appropriate value at each iteration, the combination will retain the fast convergence of NLMS and the robustness against impulsive noise of NSA. $\mathbf{w}_1(n)$ and $\mathbf{w}_2(n)$ are independently adapted following their own rules, while $\eta(n)$ is adapted to minimize the overall quadratic error. However, instead of directly adapting $\eta(n)$, we will adapt a new variable $a(n)$ which defines $\eta(n)$ via $\eta(n) = \text{sigm}[a(n)] = \{1 + \exp[-a(n)]\}^{-1}$. Consequently, the adaptation rule for $a(n)$ becomes

$$\begin{aligned} a(n+1) &= a(n) - \frac{1}{2}\frac{\partial[d(n)-y(n)]^2}{\partial a(n)} \qquad (11) \\ &= a(n) + \mu_a[d(n)-y(n)][y_1(n)-y_2(n)]\eta(n)[1-\eta(n)] \end{aligned}$$

Note that the application of (11) guarantees that the overall filter minimizes the quadratic error, independently of the norm used by the components.

Using the sigmoid activation assures that $\eta(n)$ remains in the desired range (0,1). Furthermore, factor $\eta(n)[1-\eta(n)]$ in (11) reduces the adaptation speed near $\eta(n) = 0$ and $\eta(n) = 1$, what guarantees that the steady-state error of the combination is at least as low at that of the best filter [5]. Nevertheless, note that the update of $a(n)$ could stop whenever $\eta(n)$ is too close 0 or 1. To circumvent this problem, we impose $a(n) \in [-4, 4]$ at each iteration.

To make the resulting filter robust against impulsive noise, we should select μ_1 and μ_2 so that $J_{ex,1}(\infty) < J_{ex,2}(\infty)$. From our analysis in [5], such a setting will make the combination perform like the NSA filter along stationary intervals, thus inheriting its robustness properties in these situations. Regarding convergence rate, the combination converges as fast as the NLMS filter.

If the component filters remained completely decoupled, NSA would take a long time before getting a complete convergence, thus degrading the performance of the overall scheme. To alleviate this problem, we will step-by-step transfer a part of $\mathbf{w}_2$ to $\mathbf{w}_1$ when the NSA error is significantly larger than that of NLMS. More specifically, we apply to $\mathbf{w}_1$ the modified adaption rule:

$$\mathbf{w}_1(n+1) = \alpha\left[\mathbf{w}_1(n) + \mu_1 g_1(n)\mathbf{u}(n)\right] + (1-\alpha)\mathbf{w}_2(n+1) \qquad (12)$$

provided that $e_{1,f}(n)/e_{2,f}(n) > \gamma$, where $e_{i,f}(n)$ are rough estimates of the quadratic error of each filter (for instance, $e_{i,f}(n) = 0.9e_{i,f}(n) + 0.1e_i^2(n)$).

Applying (12) with a high α has a negligible effect on NSA performance at each iteration. However, step-by-step weight transfer will speed up the convergence of this filter. The selection of the “speeding up” parameters is very easy and we have checked that $\alpha = 0.95$ and $\gamma = 1.5$ seem to work well in most situations.

Simulation Results: We will illustrate the performance of our method in a system identification task. The plant is given by $\mathbf{w}_0 = [1,2,3,4,5,4,3,2,1]^T$ ($M = 9$), normalized to get $\|\mathbf{w}_0\| = 1$. At $n = 4000$, the weights are circularly shifted 5 positions to the right and multiplied by $\sqrt{10}$. Both $u(n)$ and $e_0(n)$ are i.i.d. Gaussian sequences with variances $\sigma_u^2 = 1$ and $\sigma_0^2 = 0.1$. In addition to $e_0(n)$, every 1000 iterations the output is corrupted with an impulsive noise with variance $10^4/12$ generated using the model in [2].

We have used $\mu_1 = 0.005$, $\mu_2 = 0.1$ (so $J_{ex,1}(\infty) < J_{ex,2}(\infty)$ for $\sigma_0^2 > 0.004$) and $\mu_a = 10$, initializing with zeros the weights of both filters and $a(0)$. All results have been averaged over 10000 runs.

In Figure 1 we depict the behaviour of the conventional NSA and NLMS filters, as well as that of our combination approach. As expected, NLMS shows a fast convergence, but it significantly degrades as a consequence of the impulsive samples. NSA is able to achieve a very reduced misadjustment, but its convergence is very slow. Furthermore, NSA is not affected by the impulsive noise.

The combined filter keeps the fast convergence of NLMS and, taking advantage of the speeding up mechanism, it very quickly achieves an EMSE of about -19 dB. The degradation due to impulsive noise is very reduced, both in amplitude and duration, if compared with NLMS. As we could expect, the influence of impulsive samples is more significant at the beginning of the convergence, when $\eta(n)$ is still close to 0. If these peaks should be completely removed, an immediate solution would be to use the (speeded up) NSA component alone (see Figure 2) with a slightly slower convergence.

We have compared our proposal with the RMN filter [2], exploring different values for its step-size. In Figure 2 we show the result for $\mu_{RMN} = 0.01$. Although RMN remains robust to impulsive moise, it converges slower than the combined scheme, getting also a higher misadjustment. A larger value of μ_{RMN} would accelerate the convergence of the filter at the cost of a higher EMSE, and viceversa. This balance, however, does not occur with the combined scheme: the convergence rate and misadjustment are always dictated by the NLMS and NSA components, respectively.

Conclusions: Adaptive combination of adaptive filters is an interesting approach to improve adaptive filter performance. Here we have presented a combination of NLMS and NSA filters, where each component is adapted following its own criterion, while the combination minimizes the overall quadratic error. We have shown in a plant identification context that the combination keeps the fast convergence of NLMS, remaining robust to impulsive noise. Furthermore, using a small value for the step-size of the NSA component results in a reduced misadjustment along stationary intervals.

Acknowledgment

This work was partly supported by CICYT grant TIC2002-03713.

References

[1] A. H. Sayed, *Fundamentals of Adaptive Filtering*, Wiley, New York, NY, 2003.

[2] J. Chambers and A. Avlonitis, “A robust mixed-norm adaptive filter algorithm,” *IEEE Signal Proc. Lett.*, vol. 4, pp. 46–48, 1997.

[3] E. V. Papoulis and T. Stathaki, “A normalized robust mixed-norm adaptive algorithm for system identification,” *IEEE Signal Proc. Lett.*, vol. 11, pp. 56–59, 2004.

[4] S.'i. Koike, “Variable step size normalized sign algorithm for fast convergent adaptive filters with robustness against impulsive noise,” *NEC Research & Development*, vol. 41, 2000.

[5] J. Arenas-Garcia, A. R. Figueiras-Vidal, and A. H. Sayed, “Steady-state performance of convex combinations of adaptive filters,” in *Proc. IEEE ICASSP*, vol. IV, pp. 33-36, Philadelphia, PA, March 2005.

Authors’ affiliation:

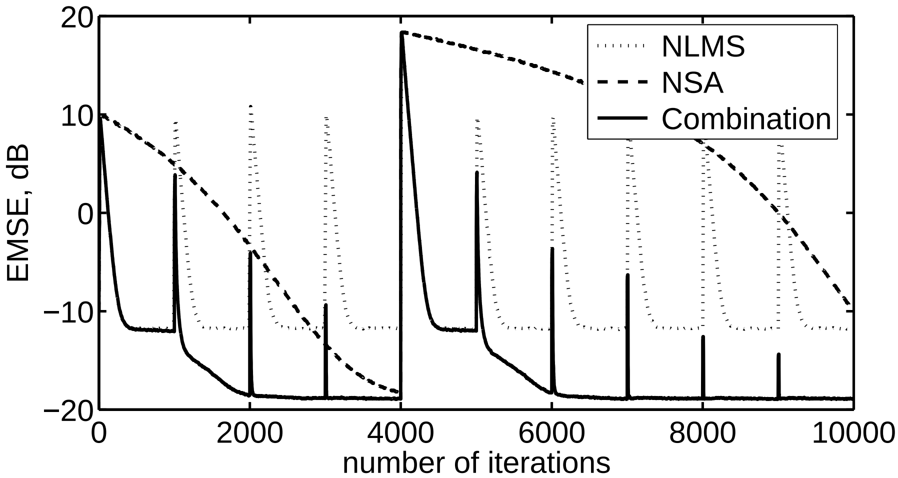
J. Arenas-García and A. R. Figueiras-Vidal (Department of Signal Theory and Communications, Universidad Carlos III de Madrid, 28911 Leganés, Spain).

Figure captions:

Fig. 1: Identification of a system corrupted by impulsive noise. EMSEs achieved by NSA ($\mu_1 = 0.005$), NLMS ($\mu_2 = 0.1$), and by the adaptive combination of these filters ($\mu_a = 10$, $\alpha = 0.95$ and $\gamma = 1.5$).

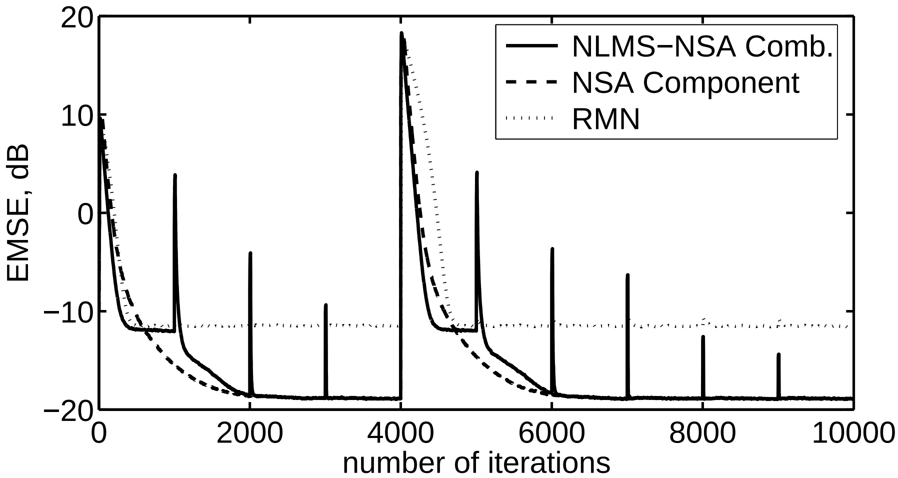
Fig. 2: Identification error achieved by the combination of the NSA and NLMS filters, by the “speeded up” NSA component, and by the RMN filter of [2] ($\mu_{RMN} = 0.01$).

Figure 1



20
10
0
−10
−20
EMSE, dB
0
2000
4000
6000
8000
10000
number of iterations
NLMS
NSA
Combination

Figure 2



NLMS−NSA Comb.
NSA Component
RMN
EMSE, dB
20
10
0
−10
−20
0
2000
4000
6000
8000
10000
number of iterations

Figure 1 (unlettered)

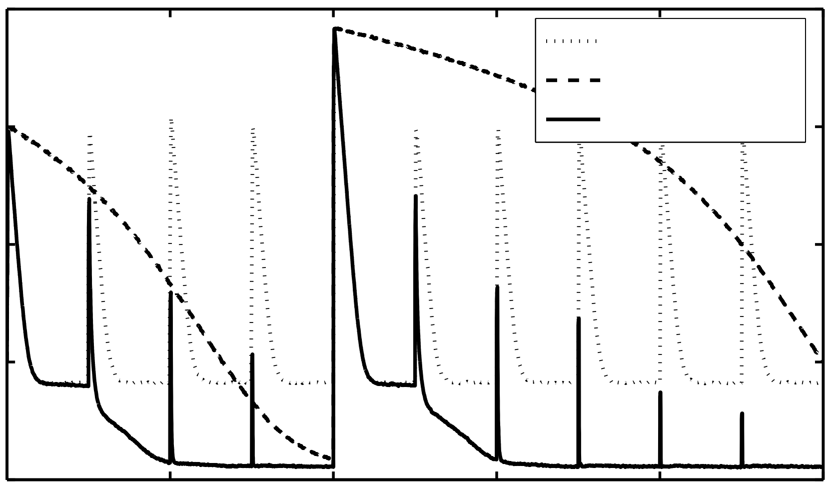


Figure 2 (unlettered)