MINIMUM MEAN SQUARE DEVIATION IN ZA-NLMS ALGORITHM

Abdullah Al-Shabili, Shihab Jimaa

Department of Electrical and Computer Engineering
Khalifa University, United Arab Emirates
{abdullah.alshabili, saj}@kustar.ac.ae

Luis Weruaga

proactivaudio
Vienna, Austria
weruaga@ieee.org

ABSTRACT

The ZA-NLMS (for zero-attractor) represents arguably the seminal sparsity-aware gradient adaptive algorithm. As it is constraint by the ℓ_1-norm of the filter weights, the underlying problem turns convex, hence with unique solution (in expected sense). Despite these friendly properties, the algorithm convergence and, more important, the best-performing sparsity tradeoff are yet to be effectively studied. This paper presents a comprehensive analytical study on ZA-NLMS' convergence, which results in the optimal (constant) sparsity tradeoff. The value of this decisive hyperparameter from a practitioner point of view turns out related to the $3/2$-power of the adaptive filter length. This outcome, difficult to argue intuitively, as well as the convergence model, have been exhaustively validated with numerical experiments.

***Index Terms*—** Sparsity, NLMS, ℓ_1 norm, modal analysis, mean square deviation, optimal tradeoff.

1. INTRODUCTION

Electrical, acoustic echo plants [1–3] and multi-path wireless communication channels [4–6] are few examples of sparse systems, e.g., those with a small fraction of their coefficients relevant or non-zero. The adaptive identification of sparse systems arouses currently large interest [7–13] and it is supported by solid theoretical foundations [14, 15]. The zero-attracting (ZA) least-mean-square (LMS) [8] and NLMS (for normalized) algorithms represents the seminal work on sparsity-aware adaptive algorithms, founded on the minimization of the ℓ_1 norm on the adaptive system coefficients. For sparse plants, the ZA-NLMS exhibits larger robustness against additive noise than the unconstrained NLMS [10, 16].

Let us consider the output of a noisy linear plant

$$d_n = \mathbf{h}^T \mathbf{x}_n + v_n \tag{1}$$

where n is time, $\mathbf{x}_n = [x_n, \cdots, x_{n-N+1}]^T$ contains the last N samples of the input signal x_n, $\mathbf{h}$ corresponds to the N-length impulse response of the plant, v_n is additive white Gaussian noise of power σ_v^2, and $(\cdot)^T$ denotes transpose. Given a linear filter $\mathbf{w}_n = [w_{0,n}, \cdots, w_{N-1,n}]^T$, the ZA-NLMS algorithm updates the identification filter coefficients according to the well-known rule

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu \frac{e_n \mathbf{x}_n}{\|\mathbf{x}_n\|^2 + \epsilon} - \gamma \operatorname{sgn}(\mathbf{w}_n) \tag{2}$$

where e_n is the error between the noisy plant and the filter

$$e_n = d_n - \mathbf{w}_n^T \mathbf{x}_n. \tag{3}$$

μ is the step size, ϵ is a small positive value to prevent division by zero,[1] $\operatorname{sgn}(w) = w/|w|$, and γ is the tradeoff between estimation error and sparseness of the solution.

Recently, in [16], we proposed an adaptive tradeoff for the ZA-NLMS, a result from equating the convergence modes of significant and negligible taps. Even though the methodology was somewhat unorthodox, the resulting "instantaneous" tradeoff γ_n was proven numerically to perform nearly optimal. In this paper we follow a different approach in the analysis of the (constant) optimal tradeoff γ, with a methodology similar to the one we used recently for the ℓ_0-NLMS algorithm [17],[2] which is based on the minimization of the mean square deviation (MSD) in steady state. Similar convergence analyses have been previously attempted in [18, 19]. However, those related works have failed to deliver not only insightful conclusions therefrom, but also an explicit expression for the optimal sparsity tradeoff. With our present work we aim to bridge that gap.

The organization of the paper follows: The mean bias and mean square deviation of the ZA-NLMS is presented in Sec. 2; the sparsity tradeoff that makes minimum the MSD in steady state is derived in Sec. 3; the numerical validation is brought in Sec. 4; finally, the conclusions close the paper.

2. ZA-NLMS MODAL ANALYSIS

By replacing (3) into (2), and defining the misalignment

$$\mathbf{g}_n = \mathbf{h} - \mathbf{w}_n \tag{4}$$

[1] For the sake of simplicity in the notation, we drop ϵ in what follows.

[2] ℓ_0-NLMS is known to beat ℓ_1- or ZA-NLMS; however, the convergence analysis presented hereby has singular aspects that make such a study very appealing for studying other sparsity-aware NLMS adaptive algorithms.

978-1-5090-4117-6/17/\$31.00 ©2017 IEEE

the deviation update equation of the ZA-NLMS (2) can be written as

$$\mathbf{g}_{n+1} = \left(\mathbf{I} - \mu\frac{\mathbf{x}_n\mathbf{x}_n^T}{\|\mathbf{x}_n\|^2}\right)\mathbf{g}_n - \mu\frac{v_n\mathbf{x}_n}{\|\mathbf{x}_n\|^2} + \gamma\,\mathrm{sgn}(\mathbf{h}-\mathbf{g}_n). \quad (5)$$

In order to study the ZA-NLMS and its performance, rather than starting from (5), we acknowledge that the N plant taps can be classified into negligible (zero) taps and significant taps. Hence, the plant output (1) can be written as

$$d_n = \mathbf{h}'^T\mathbf{x}'_n + \mathbf{h}''^T\mathbf{x}''_n + v_n \quad (6)$$

where vectors $\mathbf{h}'$ and $\mathbf{h}''$ contain the significant and zero taps respectively (hence, $\mathbf{h}'' = \mathbf{0}$), and vectors $\mathbf{x}'_n$ and $\mathbf{x}''_n$ contain the input samples accordingly. We define ρ to be the sparsity degree of the plant; hence the number of elements in $\mathbf{x}'_n$ is ρN, and that length in $\mathbf{x}''_n$ is thus $(1-\rho)N$. The estimation error e_n (3) can be written accordingly

$$e_n = \mathbf{g}'^T_n\mathbf{x}'_n + \mathbf{g}''^T_n\mathbf{x}''_n + v_n \quad (7)$$

where $\mathbf{g}'_n = \mathbf{h}' - \mathbf{w}'_n$ and $\mathbf{g}''_n = \mathbf{h}'' - \mathbf{w}''_n = -\mathbf{w}''_n$.

By using (7), it is simple to deduce that the global deviation update (5) can be split in the following two equations

$$\mathbf{g}'_{n+1} = (\mathbf{I}-\mathbf{A}'_n)\mathbf{g}'_n - \mathbf{B}_n\mathbf{g}''_n - \frac{\mu v_n\mathbf{x}'_n}{\|\mathbf{x}_n\|^2} + \gamma\mathrm{sgn}(\mathbf{h}'-\mathbf{g}'_n) \quad (8)$$

$$\mathbf{g}''_{n+1} = (\mathbf{I}-\mathbf{A}''_n)\mathbf{g}''_n - \mathbf{B}_n^T\mathbf{g}'_n - \frac{\mu v_n\mathbf{x}''_n}{\|\mathbf{x}_n\|^2} - \gamma\mathrm{sgn}(\mathbf{g}''_n) \quad (9)$$

where $\mathbf{A}'_n = \mu\,\mathbf{x}'_n\mathbf{x}'^T_n/\|\mathbf{x}_n\|^2$, $\mathbf{A}''_n = \mu\,\mathbf{x}''_n\mathbf{x}''^T_n/\|\mathbf{x}_n\|^2$, and $\mathbf{B}_n = \mu\,\mathbf{x}'_n\mathbf{x}''^T_n/\|\mathbf{x}_n\|^2$.

Throughout this study the following assumptions are made: vectors $\mathbf{g}_n$, $\mathbf{x}_n$ and v_n are mutually independent (which extends directly to the mutual independence within the same taps category, i. e. all $\mathbf{g}'_n$, $\mathbf{x}'_n$, $\mathbf{g}''_n$, $\mathbf{x}''_n$, v_n are mutually independent.), the input x_n is white[3] Gaussian of power σ_x^2, for the sake of analytical simplicity we assume the values in the plant taps $\mathbf{h}'$ to be of a similar magnitude σ_h: this assumption guarantees that the random components in $\mathbf{g}'_n$ can be considered independent and identically distributed.

2.1. Mean Analysis

For moderate SNR and for significant taps it can be assumed that $|\mathbf{h}'| > |\mathbf{g}'_n|$, and given that for large N, $\mathrm{E}\{\mathbf{A}'_n\} \simeq \mu\mathrm{E}\{\mathbf{x}'_n\mathbf{x}'^T_n\}/\mathrm{E}\{\|\mathbf{x}_n\|^2\} = (\mu/N)\,\mathbf{I}$, the expectation on (8) and (9) yields

$$\boldsymbol{\delta}'_{n+1} = \left(1-\frac{\mu}{N}\right)\boldsymbol{\delta}'_n + \gamma\,\mathrm{sgn}(\mathbf{h}') \quad (10)$$

where $\boldsymbol{\delta}'_n \triangleq \mathrm{E}\{\mathbf{g}'_n\}$. The previous result has been deduced for moderate SNR and adequate tradeoff γ,[4] and comparable results have been reported in [8, 18, 19].

As zero-mean weight initialization implies

$$\boldsymbol{\delta}'_0 \triangleq \mathrm{E}\{\boldsymbol{g}'_0\} = \mathrm{E}\{\mathbf{h}' - \mathbf{w}'_0\} = \mathbf{h}' - \mathrm{E}\{\mathbf{w}'_0\} = \mathbf{h}' \quad (11)$$

the sign of the misalignment bias (10) turns out equal to the sign of the plant taps, that is

$$\mathrm{sgn}(\boldsymbol{\delta}'_n) = \mathrm{sgn}(\mathbf{h}'). \quad (12)$$

On the other hand, it is simple to show that the misalignment bias in negligible taps is null, $\mathrm{E}\{\mathbf{g}''_n\} = \mathbf{0}$, a statement suggested also in [8, 18, 19].

2.2. MSD Analysis

Let $\Phi'^2_n = \mathrm{E}\{\|\mathbf{g}'_n\|^2\}$ and $\Phi''^2_n = \mathrm{E}\{\|\mathbf{g}''_n\|^2\}$ be the MSD of significant and negligible taps respectively.

2.2.1. Significant Taps

The mean square deviation of (8) yields

$$\begin{aligned}\Phi'^2_{n+1} &= \mathrm{E}\{\mathbf{g}'^T_n(\mathbf{I}-\mathbf{A}'_n)^T(\mathbf{I}-\mathbf{A}'_n)\mathbf{g}'_n\} \\ &\quad + \mathrm{E}\{\mathbf{g}''^T_n\mathbf{B}'^T_n\mathbf{B}'_n\mathbf{g}''_n\} + \frac{\sigma_v^2}{\sigma_x^2}\frac{\mu^2\rho}{N+2} + \gamma^2\rho N \\ &\quad + 2\gamma\mathrm{E}\{\mathbf{g}'^T_n(\mathbf{I}-\mathbf{A}'_n)^T\,\mathrm{sgn}(\mathbf{h}'-\mathbf{g}'_n)\}\,. \quad (13)\end{aligned}$$

The first term in (13) is simplified as

$$\mathrm{E}\{\|(\mathbf{I}-\mathbf{A}'_n)\,\mathbf{g}'_n\|^2\} = \left(1 - \frac{2\mu}{N} + \frac{\mu^2(2+\rho N)}{N(N+2)}\right)\Phi'^2_n. \quad (14)$$

The second term in (13) simplifies in

$$\mathrm{E}\left\{\mathbf{g}''^T_n\mathbf{B}'^T_n\mathbf{B}'_n\mathbf{g}''_n\right\} = \frac{\mu^2\rho}{N+2}\Phi''^2_n\,. \quad (15)$$

For moderate SNR and given the sign relation between plant taps and respective misalignment (12), we are in position to simplify the last term in (13) as follows

$$\mathrm{E}\left\{\mathbf{g}'^T_n(\mathbf{I}-\mathbf{A}'_n)\,\mathrm{sgn}(\mathbf{h}'-\mathbf{g}'_n)\right\} = \left(1-\frac{\mu}{N}\right)\|\boldsymbol{\delta}'_n\|_1\,. \quad (16)$$

Note that it is the absolute norm of $\boldsymbol{\delta}'_n$ that matters in the MSD analysis. This invites us to collect the evolution (10) in form of the global absolute mean $\|\boldsymbol{\delta}'_n\|_1$ as follows

$$\Delta'_{n+1} \triangleq \|\boldsymbol{\delta}'_{n+1}\|_1 = \left(1-\frac{\mu}{N}\right)\Delta'_n + \gamma\rho N\,. \quad (17)$$

Substituting (14)–(17) into (13) results finally in

$$\begin{aligned}\Phi'^2_{n+1} &= \left(1 - \frac{2\mu}{N} + \frac{\mu^2(2+\rho N)}{N(N+2)}\right)\Phi'^2_n + \frac{\mu^2\rho}{N+2}\Phi''^2_n \\ &\quad + \gamma^2\rho N + \mu^2\frac{\sigma_v^2}{\sigma_x^2}\frac{\rho}{N+2} + 2\gamma\left(1-\frac{\mu}{N}\right)\Delta'_n\,. \quad (18)\end{aligned}$$

[3] The results of this paper can be extended to coloured inputs, by using a "whitening" pre-processing such as the subband NLMS [20, 21].

[4] The case of very low SNR levels and large tradeoff γ are not covered by the analysis presented in this paper.

$$\begin{bmatrix}\Phi'^2_{n+1}\\ \Phi''^2_{n+1}\\ \Delta'_{n+1}\end{bmatrix} = \begin{bmatrix}1-a' & b'' & 2\gamma(1-\frac{\mu}{N})\\ b' & 1-a'' & 0\\ 0 & 0 & 1-\frac{\mu}{N}\end{bmatrix}\begin{bmatrix}\Phi'^2_{n}\\ \Phi''^2_{n}\\ \Delta'_{n}\end{bmatrix} + \gamma\begin{bmatrix}0\\ -c''\,\Phi''_n\\ \rho N\end{bmatrix} + \gamma^2 N\begin{bmatrix}\rho\\ 1-\rho\\ 0\end{bmatrix} + \frac{\mu^2}{N+2}\frac{\sigma_v^2}{\sigma_x^2}\begin{bmatrix}\rho\\ 1-\rho\\ 0\end{bmatrix}. \tag{21}$$

2.2.2. *Negligible Taps*

With the previous strategy, it is simple to deduce the MSD update equation for negligible taps. Beside the simplifications similar to (13)–(15), the following expectation becomes

$$\mathrm{E}\big\{\mathbf{g}''^T_n(\mathbf{I}-\mathbf{A}'_n)\,\mathrm{sgn}(\mathbf{g}''_n)\big\} = \Big(1-\frac{\mu}{N}\Big)\mathrm{E}\big\{\|\mathbf{g}''_n\|_1\big\} \tag{19}$$

which can be approximated for moderate SNR as

$$\mathrm{E}\big\{\|\mathbf{g}''_n\|_1\big\} = \kappa\sqrt{(1-\rho)N}\,\Phi''_n \tag{20}$$

where κ is a parameter that depends on distribution of the negligible tap misalignment, namely $1/\sqrt{2}$ and $\sqrt{2/\pi}$ for Laplace and Gaussian respectively.[5] Therefore, the MSD update equation for the zero taps results in the following compact expression

$$\begin{aligned}\Phi''^2_{n+1} =& \left(1-\frac{2\mu}{N}+\frac{\mu^2(2+(1-\rho)N)}{N(N+2)}\right)\Phi''^2_n\\ &+\frac{\mu^2(1-\rho)}{N+2}\Phi'^2_n+\gamma^2(1-\rho)N+\mu^2\frac{\sigma_v^2}{\sigma_x^2}\frac{(1-\rho)}{N+2}\\ &-2\gamma\kappa\left(1-\frac{\mu}{N}\right)\sqrt{(1-\rho)N}\,\Phi''_n\,.\end{aligned} \tag{21}$$

It is important to note the negative sign in the last term in (21), which comes to say that the ZA update reduces the misalignment in zero taps (as proven in [16]). In contrast the positive sign of the γ-dependent terms in the update for significant taps (18), which obviously has the opposite effect.

3. OPTIMAL TRADEOFF PARAMETER

We can rewrite the previous mean and MSD dynamic equations (17), (18) and (21) in matrix form as in (21), where the constants a', a'', b', b'' and c'' can be easily deduced therefrom. It is worth remarking that this dynamic rule (21) possesses a quadratic character with respect to Φ''_n.

In the steady state ($n\to\infty$), $\Phi'_{n+1}=\Phi'_n$ and $\Phi''_{n+1}=\Phi''_n$, (21) yields the solution to the steady-state misalignment

$$\Phi^2_\infty = \Phi'^2_\infty + \Phi''^2_\infty \tag{22}$$

where Φ'_∞ and Φ''_∞ result from solving the quadratic system

$$\begin{bmatrix}a' & -b'' & -\gamma c'\\ -b' & a'' & 0\\ 0 & 0 & \mu/N\end{bmatrix}\begin{bmatrix}\Phi'^2_\infty\\ \Phi''^2_\infty\\ \Delta'_\infty\end{bmatrix} + \gamma c''\begin{bmatrix}0\\ \Phi''_\infty\\ 0\end{bmatrix} = \begin{bmatrix}d'\\ d''\\ \gamma\rho N\end{bmatrix} \tag{23}$$

and where c', d', and d'' can be deduced from (21). It becomes obvious that the bias magnitude in the steady-state is

$$\Delta'_\infty = \frac{\gamma\rho N^2}{\mu} \tag{24}$$

which simplifies (23) in

$$\begin{bmatrix}a' & -b''\\ -b' & a''\end{bmatrix}\begin{bmatrix}\Phi'^2_\infty\\ \Phi''^2_\infty\end{bmatrix} + \gamma c''\begin{bmatrix}0\\ \Phi''_\infty\end{bmatrix} = \begin{bmatrix}d'+\gamma c'\Delta'_\infty\\ d''\end{bmatrix}. \tag{25}$$

Equation (25) has an explicit real solution, which due to space constraints we must omit hereby. By minimizing the mean-square deviation (MSD) (22) resulting from (25) with respect to γ such that

$$\frac{\partial\,\Phi^2_\infty}{\partial\gamma} = 0 \tag{26}$$

and by assuming large N,[6] the optimal tradeoff can be closely approximated in the following readily expression

$$\gamma_{\mathrm{ZA}} \triangleq \frac{\alpha_{\mu,\rho}}{2}\left(\frac{1-\rho}{\sqrt{\rho}+0.1\sqrt{\mu}}\right)\left(\frac{\mu}{N}\right)^{3/2}\left(\frac{\sigma_v}{\sigma_x}\right) \tag{27}$$

where

$$\alpha_{\mu,\rho} = \mu^{0.3}(1-\rho)^{0.1} \tag{28}$$

acts as correction term for low μ values.

The sparsity degree ρ has a big impact on the tradeoff (27): the smaller ρ is, the larger the tradeoff; for a regular plant, $\rho\to 1$, the tradeoff tends to zero, hence the optimal algorithm becomes the plain NLMS. An interesting result, hard to argue with intuition, is the dependance on the $3/2$ power of μ/N;[7] this power rule appears to result from the global steady-state bias magnitude (24), which depends on the square of N.

[5] It has been widely accepted that the misalignment $\mathbf{g}''_n$ follows a Gaussian distribution. However, invoking the *central limit theorem* here is not a valid argument because the ZA update is based on a single random source (the sign of the weight itself). In fact, for useful γ values, e.g. those that give an edge versus the plain NLMS, the misalignment turns long-tailed with positive excess kurtosis (according to our empirical observation). We acknowledge this issue by considering the Laplace distribution in our analysis as well.

[6] As the statement (26) results in a very long equation, and given that sparsity-aware algorithms are meant to be used with a large number of taps N, we present hereby the asymptotic formula of the sparsity tradeoff.

[7] It is worth comparing this result (27) with the optimal tradeoff for ℓ_0-NLMS, which we deduced in [17] and we bring hereby

$$\gamma_{\ell_0} \triangleq \sqrt{\frac{2}{2-\mu\rho}}\left(\frac{\mu}{N}\right)\left(\frac{\sigma_v}{\sigma_x}\right). \tag{29}$$

4. SIMULATION RESULTS

In this section, the empirical evaluation of the steady-state performance (22) of the ZA-NLMS algorithm is conducted, with emphasis on the validation of the optimal sparsity-tradeoff (27). In all experiments the location of the significant taps are selected randomly, their magnitude generated from a Gaussian of variance σ_h^2 (this setting supersedes the initial assumption of significant taps being nearly equal in amplitude), $N = 100$, and both input signal with $\sigma_x^2 = 1$ and noise are white Gaussian. The step size was set $\mu = 1$ for faster convergence, and $\kappa = 1/\sqrt{2}$ to model the zero taps misalignment as Laplacian distribution. The performance evaluation is carried out with the steady-state mean square deviation SS-MSD $= \|\mathbf{h} - \mathbf{w}_\infty\|_2^2$ over 200 Monte Carlo simulations.

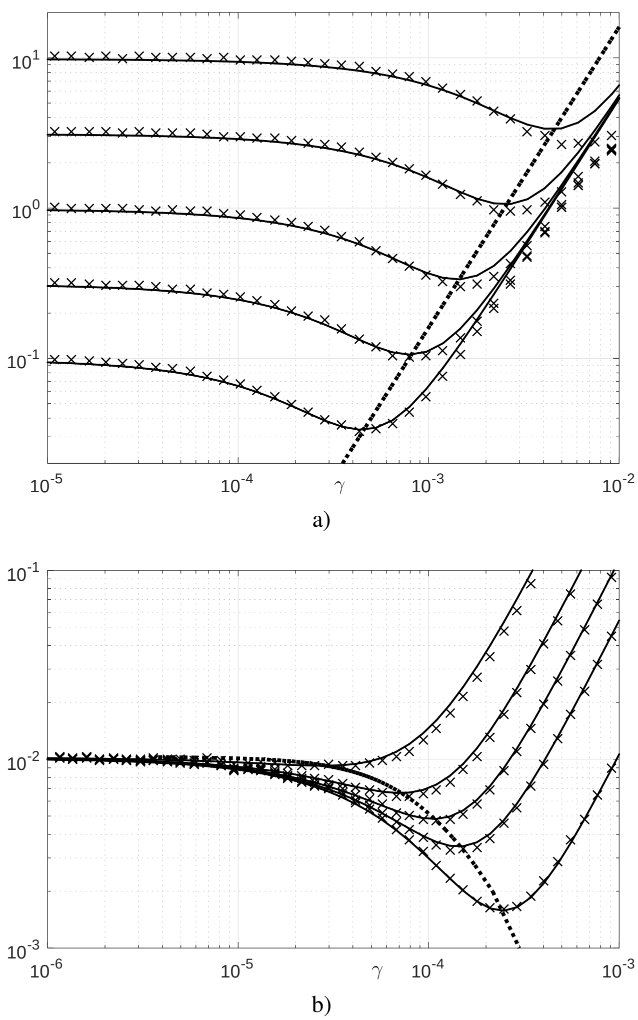

Fig. 1. Analytical (22) (solid) and empirical ($\times$) SS-MSD with respect to tradeoff γ for: a) different noise levels (bottom to top) $\sigma_v^2 = \{0.1, \sqrt{0.1}, 1, \sqrt{10}, 10\}$ with $\rho = 5\%$, and b) different sparsity (bottom to top) $\rho = \{1, 5, 10, 20, 50\}\%$ with $\sigma_v^2 = 0.01$; locus (dotted) of optimal sparsity tradeoff (27).

In the first experiment, the SS-MSD is exhaustively evaluated against the tradeoff γ for different noise levels σ_v^2 and a small sparsity degree. The results, brought in Fig. 1.a), throw a very good agreement with the analytical predictions (22). Under low SNR and large sparsity tradeoff the empirical results and the theoretical model suffer an expected mismatch, a situation that is not covered by our study because it has little interest. It is very relevant that the optimal tradeoff (27) follows closely the empirical trend of minimum SS-MSD. The previous methodology is repeated this time for different sparsity degrees ρ and one noise level. Fig. 1.b) shows the match between the empirical and the analytical model (22). The optimal tradeoff outlines the minimum SS-MSD locus, such that for regular plants ($\rho > 50\%$) the optimal ZA-NLMS becomes essentially the regular NLMS.

The last experiment deals with the validation of the $3/2$-power rule of the optimal tradeoff (27) with respect to N. Different cases of filter length N were considered, while sparsity degree and noise level were kept constant. For each case, the best performing tradeoff corresponds to the locus of the minimum empirical SS-MSD in a dense γ-grid. Fig. 2 reveals the results of this experiment, throwing an extremely accurate match between theory and empirical validation. This result is probably the main surprising outcome of the present work.

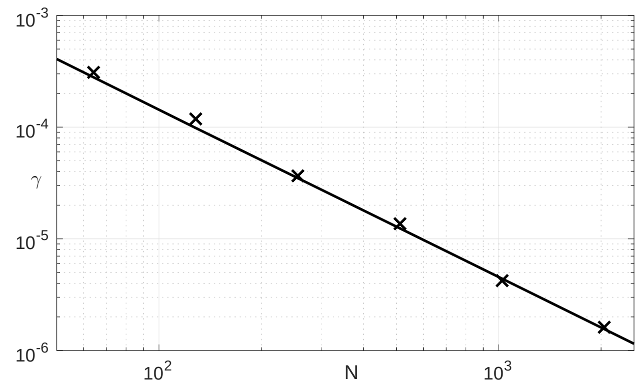

Fig. 2. Empirical ($\times$) and theoretical (solid) optimal tradeoff. Sparsity degree $\rho = 5\%$ and noise $\sigma_v^2 = 0.01$.

5. CONCLUSIONS

The concept of significant and negligible filter taps have allowed us to derive the theoretical convergence model of the ZA-NLMS algorithm. This approach facilitated the derivation of the optimal sparsity tradeoff that acomplishes to the minimum mean deviation in steady state. The resulting optimal tradeoff depends on the environment parameters such as the filter length, step size, plant sparsity and SNR. The extension of this work into very low SNR scenarios, as well as developing the numerical framework for estimating the sparsity degree in the practice are in our current research agenda.

6. REFERENCES

[1] K. Helwani, H. Buchner, and S. Spors, "Multi-channel adaptive filtering with sparseness constraints," *Intl. Wksp. Acoust. Signal Enhanc. 2012*, pp. 1–4.

[2] 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.

[3] Y. Kopsinis, S. Chouvardas, and S. Theodoridis, "Sparse models in echo cancellation: When the old meets the new," *Trends in Digital Signal Processing*, CRC Press, pp. 175–200, 2015.

[4] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374–377, Aug. 2002.

[5] W. F. Schreiber, "Advanced television systems for terrestrial broadcasting: Some problems and some proposed solutions," *Proc. of the IEEE*, vol. 83, no. 6, pp. 958–981, Jun. 1995.

[6] F. Al-Ogaili, H. Elayan, L. Alhalabi, A. Al-Shabili, B. Taha, L. Weruaga and S. Jimaa, "Leveraging the ℓ_1-LS criterion for OFDM sparse wireless channel estimation," *IEEE Wireless Mobile Comput. Netw. Commun. (WiMob)*, 2015, pp. 845–849.

[7] K. T. Wagner, and M. I. Doroslovacki, "Towards analytical convergence analysis of proportionate-type NLMS algorithms," *IEEE Intl. Conf. Acoust. Speech Signal Proces. (ICASSP)*, 2008, pp. 3825–3828.

[8] Y. Chen, Y. Hu, and A. O. Hero, "Sparse LMS for system identification," *IEEE Intl. Conf. Acoust. Speech Signal Proces. (ICASSP)*, 2009, pp. 3125–3128.

[9] O. Taheri and S. A. Vorobyov, "Sparse channel estimation with ℓ_p-norm and ℓ_1-norm reweighted penalized least mean squares," *IEEE Intl. Conf. Acoust. Speech Signal Proces. (ICASSP)*, 2011, pp. 2864–2867.

[10] B. C. Gwun, J. W. Han, K. M. Kim, and J. Jung, "MIMO underwater communication with sparse channel estimation," *Intl. Conf. Ubiq. Future Netw. (ICUFN)*, 2013, pp. 32–36.

[11] G. Gui, W. Peng, and F. Adachi, "Improved adaptive sparse channel estimation based on the least mean square algorithm," *IEEE Wireless Comm. Netw. Conf.*, 2013, pp. 3105–3109.

[12] L. Weruaga and S. Jimaa, "Exact NLMS algorithm with ℓ_p-norm constraint," *IEEE Signal Proces. Lett.*, vol. 22, no. 3, pp. 366–370, Mar. 2015.

[13] A. Al-Shabili, B. Taha, H. Elayan, F. Al-Ogaili, L. Alhalabi, L. Weruaga, and S. Jimaa, "Sparse NLMS adaptive algorithms for multipath wireless channel estimation," *IEEE Wireless Mobile Comput. Netw. Commun. (WiMob)*, 2015, pp. 839–844.

[14] R. Tibshirani, "Regression shinkage and selection via the LASSO," *J. Roy. Statist. Soc.*, vol. 58, no. 1, pp. 267–288, 1996.

[15] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[16] A. Al-Shabili, L. Weruaga, and S. Jimaa, "Adaptive sparsity tradeoff for ℓ_1-constraint NLMS algorithm," *IEEE Intl. Conf. Acoust. Speech Signal Proces. (ICASSP)*, 2016.

[17] A. Al-Shabili, L. Weruaga, and S. Jimaa, "Optimal sparsity tradeoff in ℓ_0-NLMS algorithm," *IEEE Signal Proces. Lett.*, vol. 23, no. 8, pp. 1121–1125, Aug. 2016.

[18] K. Shi and P. Shi , "Convergence analysis of sparse LMS algorithms with ℓ_1-norm penalty based on white input signal," *Signal Processing*, vol. 90, no. 12, pp. 3289–3293, Dec. 2010.

[19] S. Zhang and J. Zhang, "Transient analysis of zero attracting NLMS algorithm without Gaussian inputs assumption," *Signal Processing*, vol. 97, pp. 100–109, Apr. 2014.

[20] Y. S. Choi, "Subband adaptive filtering with l_1-norm constraint for sparse system identification," *Mathematical Problems Engineering*, vol. 2013, 7 pages, 2013.

[21] S. S. Pradhan and V. U. Reddy, "A new approach to subband adaptive filtering," *IEEE Trans. Signal Proces.*, vol. 47, no. 3, pp. 655–664, Mar. 1999.