ADAPTIVE SPARSITY TRADEOFF FOR ℓ_1 -CONSTRAINT NLMS ALGORITHM

Abdullah Al-Shabili, Luis Weruaga, and Shihab Jimaa

Department of Electrical and Computer Engineering Khalifa University of Science, Technology & Research, UAE. E-mail: abdullah.alshabili@kustar.ac.ae, weruaga@ieee.org, saj@kustar.ac.ae

ABSTRACT

Embedding the ℓ_1 norm in gradient-based adaptive filtering is a popular solution for sparse plant estimation. Even though the foundations are well understood, the selection of the sparsity hyper-parameter still remains today matter of study. Supported on the modal analysis of the adaptive algorithm near steady state, this paper shows that the optimal sparsity tradeoff depends on filter length, plant sparsity and signal-to-noise ratio. In a practical implementation, these terms are obtained with an unsupervised mechanism tracking the filter weights. Simulation results prove the robustness and superiority of the novel adaptive-tradeoff sparsity-aware method.

Index Terms— Sparsity, NLMS, ℓ_1 norm, modal analysis, Gaussian mixture models, expectation-maximization.

1. INTRODUCTION

A system is considered sparse when only a small fraction of its coefficients are relevant, while the rest are negligible. Few examples of sparse systems are multi-path wireless communication channels [1–3] and electrical and acoustic echo plants [4, 5]. In recent years, there has been a resurgent interest in their identification [6–10], largely motivated by the least absolute shrinkage and selection operator (LASSO) [11] and compressive sensing [12]. As a result, the ℓ_1 -norm-constraint least-mean-square (LMS) and NLMS (for normalized) algorithms [6] have been proven to estimate a sparse system with higher accuracy and in less time than the original NLMS, while exhibiting robustness against additive noise [13–16].

Consider the estimation of a linear sparse plant h, such that the output delivered by this system is corrupted by noise, that is, the observed system output is in reality

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

where *n* is discrete time, v_n is the additive white Gaussian noise of power σ_v^2 , and $\mathbf{x}_n = [x_n, \dots, x_{n-N+1}]^T$ denotes the *N*-length vector with the white Gaussian input of power σ_x^2 . A linear filter of coefficients $\mathbf{w}_n = [w_{0,n}, \dots, w_{N-1,n}]^T$ is used to model the plant, such that the difference between the observed system output and the filter output

$$e_n = d_n - \mathbf{w}_n^T \mathbf{x}_n \tag{2}$$

is a measure of the estimation error. The minimization of the square error (2) corresponds to the traditional maximumlikelihood solution for estimating the linear plant.

However, in order to take advantage of the system inherent sparsity, the classical cost function is extended by the ℓ_1 norm penalty of the filter weights

$$J_n(\gamma) = \frac{1}{2}(e_n)^2 + \gamma \|\mathbf{w}_n\|_1$$
(3)

where $\|\mathbf{w}\|_p = (\sum_{i=0}^{N-1} |w_i|^p)^{1/p}$, and $\gamma > 0$ is the regularization hyper-parameter to tradeoff between estimation error and sparseness of the solution. The gradient-descent method over the ℓ_1 -constraint cost function (3) results in the popular zero-attracting (ZA) NLMS algorithm [6]

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu \frac{e_n \mathbf{x}_n}{\mathbf{x}_n^T \mathbf{x}_n + \epsilon} - \mu \gamma \operatorname{sgn}(\mathbf{w}_n)$$
(4)

where sgn(x) = x/|x|, $\epsilon \gtrsim 0$, and μ is the step size. The second term in (4) attracts the weights to zero, which yields faster convergence and lower misalignment when the majority of channel taps are negligible. Because the cost (3) is convex, the ZA-NLMS algorithm convergence is guaranteed under the suitable selection of the step size μ and tradeoff γ [7,9].

All the mentioned ZA-NLMS benefits can be unleashed through the proper choice of the tradeoff γ . However, the optimal (or even an adequate) value that matches a given scenario still remains matter for study, as its selection is currently carried out without a formal methodology [17] or by means of cross-validation [18]. In [19] we found heuristically that such an optimal value must be somewhat related to the filter length N, plant sparsity, and noise power. As these terms are in principle not available, choosing blindly an (inadequately large) tradeoff value is likely to be counterproductive, and the ZA-NLMS may not be a better option than the plain NLMS.

This paper reveals the optimal tradeoff through the modal analysis of the ZA-NLMS near steady state, a study conducted in Sec. 2; as proposed in Sec. 3, the formula of the optimal adaptive tradeoff is obtained by equalizing the main two convergence modes; Sec. 3 contains also the practical implementation based on unsupervised tracking the filter weights; the numerical validation is brought in Sec. 4; finally, the conclusions and suggestions to extend the proposed methodology to other sparsity-aware NLMS algorithms close the paper.

2. ZA-NLMS MODAL ANALYSIS

Let \mathbf{g}_n be the misalignment vector

$$\mathbf{g}_n = \mathbf{h} - \mathbf{w}_n \tag{5}$$

composed of independent zero-mean Gaussian variables of variance $\sigma_{i,n}^2$, that is, its *i*th component $g_{i,n} \sim \mathcal{N}(0, \sigma_{i,n}^2)$. Note that $g_{i,n}$ (its variance) depends on time *n*. For convenience, we write the update (4) in a sequential fashion as

$$\tilde{\mathbf{w}}_n = \mathbf{w}_n - \gamma_n \operatorname{sgn}(\mathbf{w}_n)$$
 (6a)

$$\mathbf{w}_{n+1} = \tilde{\mathbf{w}}_n + \frac{e_n \mathbf{x}_n}{\mathbf{x}_n^T \mathbf{x}_n + \epsilon}$$
(6b)

where the step size has been set to a fast speed, $\mu = 1$, and the tradeoff γ_n is allowed to change over time.

2.1. Zero-Attracting Modes

The nature of the attractive force exerted by the ℓ_1 norm over the filter taps in the update (6a) is proportional to γ_n and opposite to the sign of each weight tap, that is, the resulting misalignment for the *i*th tap upon the sparsity update is

$$\varepsilon_{i,n} = h_i - \left(w_{i,n} - \gamma_n \operatorname{sgn}(w_{i,n}) \right)$$

= $g_{i,n} + \gamma_n \operatorname{sgn}(h_i - g_{i,n})$. (7)

We obviously wish the power of $\varepsilon_{i,n}$ to be smaller than that of the initial misalignment $g_{i,n}$, which indicates convergent behaviour. However, as we will soon see, the ZA update increases the misalignment (of significant taps). For the sake of simplicity in the notation, we omit subindex n, but we will get it back whenever strictly needed. In the subsequent analysis we will need the second moment of a (truncated) normal distribution $\mathcal{N}(\mu, \sigma^2)$ within the interval [a, b].¹

1. Significant plant taps, $|h_j| \gg 0$. Assuming $|h_j| > |g_j|$, hence $\operatorname{sgn}(w_j) = \operatorname{sgn}(h_j)$, we can write

$$\varepsilon_i = g_i + \gamma \operatorname{sgn}(h_i) \tag{9}$$

that is, the (9) is a normal variable centred on either γ or $-\gamma$. Fig. 1 illustrates its probability density function (PDF). As both cases yield the same second moment, we can claim that the misalignment upon update (9) corresponds to a $\mathcal{N}(\gamma, \sigma_j^2)$. Based on (8), such that $a = -\infty, b = \infty$, and $\mu = \gamma$

$$\mathbb{E}\{\varepsilon_j^2\} = \sigma_j^2 + \gamma^2 \tag{10}$$

which reveals a divergent behaviour, $E\{\varepsilon_j^2\} > E\{g_j^2\}$, that is, the misalignment in the significant taps grows.

¹The second moment m_2 is given by the closed formula

$$m_2 = \mu^2 + \sigma^2 - \sigma \frac{(\mu+b)\phi(\frac{b-\mu}{\sigma}) - (\mu+a)\phi(\frac{a-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \tag{8}$$

where $\phi(x)$ is the standard probability density function (PDF) and $\Phi(x)$ its cumulative density function (CDF).



Fig. 1. PDF of misalignment upon sparsity update (7) for negative- (solid) and postive-valued (dashed) filter taps in case of actual significant (top) and zero plant taps (bottom).

2. *Negligible plant taps,* $h_k = 0$, such that we can thus simplify (7) as follows

$$\varepsilon_k = g_k - \gamma \operatorname{sgn}(g_k). \tag{11}$$

Fig. 1 illustrates also its PDF. For symmetry, we can state that the effective misalignment (11) is equivalent to a (truncated) normal variable $\mathcal{N}(\gamma, \sigma_k^2)$ defined in $[-\infty, \gamma]$. Based on (8), and given that here $a = -\infty$, $b = \gamma$, and $\mu = \gamma$, it is simple to deduce that

$$\mathcal{E}\{\varepsilon_k^2\} = \sigma_k^2 + \gamma^2 - 4\sigma_k\gamma/\sqrt{2\pi}$$
(12)

is a convergent mode $\mathbb{E}\{\varepsilon_i^2\} < \mathbb{E}\{g_k^2\}^2$.

Supported visually on Fig. 1, the ZA update reduces the misalignment power in zero-valued taps, but increases it on the significant ones. Note that we will be using subindexes j and k to denote the significant and zero taps respectively.

2.2. NLMS Modes

The previous results (10) and (12) represent the first half of the ZA-NLMS. The plain NLMS update (6b) has a different effect than the sparsity update. It is worth recalling that the plain NLMS acts on the weight misalignment according to the well-known rule [20,21]

$$\sigma_{n+1}^2 = (1 - 1/N)\sigma_n^2 + \sigma_{\max}^2/N$$
(13)

where *n* is time and σ_{\max}^2 refers to the excess steady-state misalignment of the NLMS, $\sigma_{\max}^2 = \sigma_v^2/\sigma_x^2$. During convergence, $\sigma_{\max}^2 \ll \sigma_n^2$, the speed of convergence is exponential; in steady state, equation (13) reflects the convergent and disturbing forces of the NLMS update. Unlike the sparsity update, which acts on each weight individually, the NLMS acts globally on all weights from a common error source (2).

²For $0 < \gamma < 4\sigma_k / \sqrt{2\pi}$.

3. OPTIMAL ADAPTIVE TRADEOFF

3.1. Equalizing Global Modes Near Steady State

The global modes of the ZA-NLMS algorithm *near* steady state are deduced in what follows. Needless to say that we use the common assumption that the sparsity and the NLMS updates are statistically uncorrelated.

1. Significant taps, $|h_j| \gg 0$. As the ZA update is divergent (10) on significant taps, the NLMS update is still somewhat found in convergence, hence the misalignment rate can be considered to be

$$\sigma_{j,n+1}^2 \simeq \sigma_{j,n}^2 (1 - 1/N) + \gamma^2.$$
 (14)

2. Negligible taps, $h_k = 0$. Unlike the previous case, the lower misalignment (12) on negligible taps makes them to be in steady state, hence vulnerable to the additive noise from the NLMS mechanism (13), resulting in

$$\sigma_{k,n+1}^2 = \sigma_{k,n}^2 (1 - 1/N) + \gamma^2 - \frac{4}{\sqrt{2\pi}} \sigma_{k,n} \gamma + \frac{\sigma_{\max}^2}{N}.$$
(15)

In this paper we follow a novel approach to optimize the convergence of the ZA-NLMS. Rather than working with the excess misalignment error, which results in equations very difficult to manage [7], we use the global modes near steady state (14) and (15). It is well known that the performance of any adaptive system is conditioned by its slowest mode. Our proposed methodology to obtain an optimal tradeoff γ is thus based on equalizing the existing two modal types, that is, to look for the value of γ that makes (14) and (15) equal, assuming homogeneous misalignment $\sigma_j = \sigma_k$.³ This approach yields the solution equation

$$\sigma_k^2 - \frac{\sigma_k^2}{N} + \gamma^2 = \sigma_k^2 - \frac{\sigma_k^2}{N} + \gamma^2 - \frac{4}{\sqrt{2\pi}}\sigma_k\gamma + \frac{\sigma_{\max}^2}{N}$$
(16)

which is solved for

$$\gamma = \frac{\sqrt{2\pi}\sigma_{\max}^2}{4N\sigma_k} = \frac{\sqrt{2\pi}\kappa\sigma_k}{4N}.$$
 (17)

In the previous step, we posit the existence of $\kappa > 1$ such that $\sigma_k^2 = \sigma_{\max}^2/\kappa$ in steady state. In order to determine the value of factor κ , we obtained numerically the minimum excess steady-state misalignment from Monte-Carlo simulations for a range of the system sparsity ρ . The result throws an extremely accurate fitting⁴ with the rule $\kappa = 1/\sqrt{\rho}$, such that the optimal tradeoff is readily given by

$$\gamma \triangleq \frac{\sqrt{2\pi}\,\sigma_k}{4\sqrt{\rho}\,N}.\tag{18}$$

The numerical methodology to obtain factor κ was chosen against the direct analysis of [7], as this last turns to be extremely challenging (work [7] does not tackle the optimal tradeoff γ). Finally, note that the tradeoff (18) is built with the misalignment of the zero taps σ_k . This unknown term and the system sparsity ρ , also unknown, must be estimated blindly during the execution of the ZA-NLMS algorithm.

3.2. Implementation

The optimal tradeoff parameter (18) depends on both the zero-tap misalignment in standard deviation σ_k and the system sparsity ρ . In order to unleash the features of the ℓ_1 -constraint NLMS, these two parameters must be obtained at any instant n. The signal plus noise model (5) implies $h \sim \rho \mathcal{G}(h, 0, \eta^2) + (1 - \rho)\delta(h)$ and $g \sim \mathcal{G}(g, 0, \sigma_k^2)$, where η the standard deviation of the significant taps, $\delta(x)$ is the Dirac delta, and $\mathcal{G}(x, \mu, \theta)$ is the μ -centered θ -variance Gaussian function. Hence, the filter coefficients in \mathbf{w}_n follow the probabilistic model

$$w_{i,n} \sim \rho \mathcal{G}(w, 0, \eta^2 + \sigma_k^2) + (1 - \rho) \mathcal{G}(w, 0, \sigma_k^2).$$
 (19)

The estimation of the parameters in (19) can be approached with a Gaussian mixture model (GMM) trained by the expectation maximization (EM) algorithm [22]. In the scenario of the paper only two Gaussian units are required: the Gaussian unit with the lowest variance corresponds to the zero-tap misalignment σ_k^2 , while its probability equals $1 - \rho$.⁵

We bring the proposed adaptive-tradeoff ZA-NLMS algorithm in what follows

$$\{\sigma_n, \rho_n, \eta_n\} = \text{EM-GMM}(\mathbf{w}_n, \sigma_{n-1}, \rho_{n-1}, \eta_{n-1}) \quad (20a)$$

$$\gamma_n = \frac{\sqrt{2\pi}\,\sigma_n}{4\sqrt{\rho_n + \delta}\,N}\tag{20b}$$

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu \frac{e_n \mathbf{x}_n}{\mathbf{x}_n^T \mathbf{x}_n + \varepsilon} - \mu \gamma_n \operatorname{sgn}(\mathbf{w}_n)$$
(20c)

where $\delta \gtrsim 0$ is a small bias to prevent division by zero.

The EM-GMM step involves a number of iterations of the expectation and maximization algorithm. The EM is known to converge rapidly to the solution when sufficiently close to it [23]. Hence, given that NLMS-type algorithms are slow, one EM iteration per sample, or even less than that, turns out to be sufficient. The suggested partial update is a beneficial aspect for the computational complexity. Finally, an obliged question regards the convergence properties of the proposed algorithm (20). A thorough study thereto is beyond the length of this paper, but worth undertaking. In the next section, we present simulation results that support the excellent convergence properties of the algorithm.

 $^{^{3}}$ It is important to remark that we are not claiming the misalignment near steady state on significant and negligible taps to be equal, but in the event that they are the same, both modes are forced to converge at the same speed.

⁴Details on this numerical analysis are given in Sec. 4.

⁵Non-Gaussian plant tap distributions can be also approached as a GMM driven by the EM algorithm, for instance by including additional Gaussian units that may be non-zero centred. In any configuration, the fundamental condition of the EM-GMM step is that the zero-mean Gaussian unit of lowest variance always corresponds to the misalignment of the zero taps.

4. SIMULATION RESULTS

In this section, the performance evaluation of the proposed adaptive sparsity-tradeoff NLMS is conducted. The scenario chosen corresponds to a noisy sparse plant whose sparsity and SNR change abruptly. For a given sparsity, the location of the significant taps are selected randomly and their magnitude is generated from a Gaussian distribution. The input signal and the noise are white. Three methods are selected in the comparison, namely, the regular NLMS, the original ZA-NLMS (20). In all methods $\mu = 1$ and $\epsilon = 0.01$. In the second method, the constant tradeoff is selected with cross-validation to perform best during the first phase. The performance assessment is carried out with the mean square deviation MSD = $\|\mathbf{h} - \mathbf{w}_n\|_2^2$ over 100 Monte Carlo simulations.

Fig. 2 shows the results of this experiment. Note that the simulation scenario consider different combinations of sparsity, SNR, and two different plant lengths N. It is worth recalling that the MSD performance of the NLMS is not affected by the plant sparsity, and it is bounded by the current noise floor $\sigma_{\max}^2 = \sigma_v^2/\sigma_x^2$ in steady state [20]. As the constant- γ ZA-NLMS is tailored to the first phase (of high sparsity and noise), it shows the typical improvement of ℓ_1 -constraint NLMS algorithms. However, with different scenar-



Fig. 2. Performance of NLMS (dotted), constant- γ (dashed) and adaptive- γ ZA-NLMS (solid) for a filter length N = 100 (top) and N = 200 (bottom). Every 1000 samples n abrupt change to either the sparsity and/or the noise power.



Fig. 3. ZA-NLMS empirical misalignment $\log \sigma_k$ in steady state with respect to sparsity ρ and tradeoff γ .

ios (lower sparsity and/or noise) the performance deteriorates, often below the NLMS. On the contrary, the adaptive- γ algorithm delivers outstanding performance at any situation. The tracking capabilities of the EM-GMM step and its impact on the tradeoff γ to sudden scenario changes reveal the proposed method as a good candidate in real scenarios.

The numerical analysis used to deduce (18) from (17) is addressed at last. The ZA-NLMS with constant γ was executed with 1000 Monte Carlo runs to evaluate empirically the steady-state misalignment in the zero taps for a plant with sparsity ρ . The resulting isolevel map is shown in Fig. 3: the dashed line corresponds to $\kappa = 1/\sqrt{\rho}$, which marks the locus of minimum misalignment according to rule (17). The scenario under analysis corresponds to a filter length N = 100and SNR equal to 10 dB. In case of other values, as (18) accounts for N and σ_k^2 , the results follow the same trend. It is worth recalling that the weight deviation σ_k^2 is an outcome of the EM-GMM mechanism and proportional to the noise power level.

5. CONCLUSIONS

Modal analysis *near* steady state has allowed us to offer an answer to the optimal sparsity tradeoff for ℓ_1 -norm-constraint NLMS algorithms, unlike previous analytical works dealing with the final steady-state excess misalignment. As the plant sparsity and the noise level are part of the equation, in order to be able to unleash performance and robustness, a sparsityaware NLMS algorithm must be consciously aware of both terms at any time. In this paper, we use an unsupervised learning methodology to obtain the relevant terms and build the optimal tradeoff. Simulation results have emphatically proven the validity of our approach. The extension of this methodology to ℓ_0 -like and ℓ_p norms embedded in the NLMS mechanism is in our current research agenda.

6. REFERENCES

- 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.
- [2] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Comm.*, vol. 50, no. 3, pp. 374–377, Mar. 2002.
- [3] 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.
- [4] K. Helwani, H. Buchner, and S. Spors, "Multichannel adaptive filtering with sparseness constraints," *Intl. Wksp. Acoust. Signal Enhanc. (IWAENC)*, 2012, pp. 1–4.
- [5] D. L. Duttweiler, "Proportionate normalized leastmean-squares adaptation in echo cancellers," *IEEE Trans. Speech Audio Proces.*, vol. 8, no. 5, pp. 508–518, Sep. 2000.
- [6] 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.
- [7] 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.
- [8] F. Y. Wu and F. Tong, "Gradient optimization *p*-normlike constraint LMS algorithm for sparse system estimation," *Signal Processing*, vol. 93, no. 4, pp. 967–971, Apr. 2013.
- [9] G. Gui, W. Peng, and F. Adachi, "Improved adaptive sparse channel estimation based on the least mean square algorithm," *IEEE Wireless Comm. Netw. Conf.* (WCNC), 2013, pp. 3105–3109.
- [10] O. Taheri and S. A. Vorobyov, "Sparse channel estimation with *l_p*-norm and *l₁*-norm reweighted penalized least mean squares," *IEEE Intl. Conf. Acoust. Speech Signal Proces. (ICASSP)*, 2011, pp. 2864–2867.
- [11] R. Tibshirani, "Regression shrinkage and selection via the LASSO," J. Roy. Statist. Soc., vol. 58, no. 1, pp. 267– 288, 1996.
- [12] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006,

- [13] J. F. Andrade, M. L. R. Campos, and J. A. Apolinário, "An ℓ₁-constrained normalized LMS algorithm and its application to thinned adaptive antenna arrays," *IEEE Intl. Conf. Acoust. Speech Signal Proces. (ICASSP)*, May 2013.
- [14] J. F. Andrade, M. L. R. Campos, and J. A. Apolinário, "ℓ₁-constrained normalized LMS algorithms for adaptive beamforming," *IEEE Trans. Signal Proces.*, pp. 1– 16, in Press.
- [15] 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.
- [16] 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.
- [17] W. Ma, H. Qu, J. Zhao, B. Chen, and J. C. Principe, "Sparsity aware minimum error entropy algorithms," *IEEE Intl. Conf. Acoust. Speech Signal Proces. (ICASSP)*, 2015, pp. 2179–2183.
- [18] 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.
- [19] 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.
- [20] B. Widrow and S. D. Stearns, Adaptive signal processing. Prentice-Hall, 1985.
- [21] S. Haykin, *Adaptive filter theory (4th Ed)*. Prentice-Hall, 2001.
- [22] J. A. Bilmes, "A gentle tutorial of the EM algorithm", Tech. Rep. 97-021, *Intl. Comp. Sci. Instit.*, Berkeley, 1997.
- [23] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Proces. Mag.*, vol. 13, no. 6, pp. 47–60, Nov. 1996.