

ADAPTIVE VARIATIONAL SPARSE BAYESIAN ESTIMATION

Konstantinos E. Themelis, Athanasios A. Rontogiannis, Konstantinos D. Koutroumbas

IAASARS, National Observatory of Athens, GR-15236, Penteli, Greece
 {themelis, tronto, koutroum}@noa.gr

ABSTRACT

This paper presents an online version of the widely used sparse Bayesian learning (SBL) algorithm. Exploiting the variational Bayes framework, an efficient online SBL algorithm is constructed, that acts as a fully automatic learning method for the adaptive estimation of sparse time-varying signals. The new method is based on second order statistics and comprises a simple, automated sparsity-imposing mechanism, different from that of other known schemes. The effectiveness of the proposed online Bayesian algorithm is illustrated using experimental results conducted on synthetic data. These results show that the proposed scheme achieves faster initial convergence and superior estimation performance compared to other related state-of-the-art schemes.

Index Terms— adaptive estimation, sparse Bayesian learning, variational Bayes

1. INTRODUCTION

Adaptive estimation of time-varying signals is a research field that has received considerable attention in the signal processing literature and is met in an immense number of applications, [1]. A plethora of adaptive estimation techniques have been developed and analyzed during the past decades, a large portion of which belongs to the least squares (LS) family of algorithms. Nowadays, the advances of compressive sensing have sparked new interest in the field of adaptive estimation; the exploitation of signal *sparsity* in a time varying environment with the aim to improve the initial convergence rate and estimation performance of adaptive techniques.

It is not surprising that the majority of sparse adaptive estimation methods stems from the deterministic framework. In this context, a soft-thresholding, coordinate descent type algorithm for solving the adaptive formulation of the lasso operator is proposed in [2]. A regularized recursive LS (RLS) type algorithm that utilizes the expectation maximization (EM) algorithm as a low-complexity solver is derived in [3]. In [4] and [5], projection-based adaptive algorithms are developed that induce sparsity either via projections on weighted ℓ_1 balls or generalized thresholding. Adaptive greedy variable selection schemes have been also recently reported, e.g. [6]. However, these algorithms require -at least- a rough knowledge of the signal sparsity level and work effectively for sufficiently high signal sparsity.

In this paper, unlike previous studies, we view the adaptive estimation problem from a Bayesian perspective. In this realm, the paper focuses on the design of an adaptive variational Bayes algorithm based on a slightly modified version of the hierarchical Bayesian

model of the sparse Bayesian learning algorithm (SBL) proposed in [7]. First, an appropriate factorization of the approximating joint posterior distribution of the model parameters is considered, and a variational Bayes algorithm for the batch estimation problem is derived. Next, we move to the adaptive framework, where the updating of the parameters takes place after the presentation of each new observation. Time-dependency on the parameters of the variational scheme is now introduced and low-complexity time-recursions are derived to update the parameters of the approximating posterior distributions. This gives rise to a sparse adaptive variational Bayesian learning algorithm of $\mathcal{O}(N^2)$ complexity, similar to that of all adaptive schemes which are based on second order statistics. In contrast to deterministic optimization algorithms, the proposed scheme directly infers all model parameters from the data, and hence, the need for parameter fine-tuning is eliminated. The performance of the new algorithm is assessed via computer simulations which show that it outperforms all related state-of-the-art adaptive deterministic algorithms under a signal sparsity constraint.

To the best of our knowledge, this is the first time a sparse adaptive variational Bayes algorithm is presented. A Bayesian approach to adaptive filtering has been previously described in [8]. However, in [8] a type-II maximum likelihood method is adopted that leads to a regularized RLS-type scheme. The algorithm developed in this paper can be considered as a generalization of a recently proposed algorithm, (cf. [9]), in the context of variational mean field theory. Note further that compared to [9] a different Bayesian model is adopted here and a different derivation procedure is followed, which facilitates the extension of the proposed adaptive scheme for other sparsity-inducing Bayesian models, such as models encompassing Laplace priors, e.g. [10].

The paper is structured as follows. Section 2 defines the mathematical formulation of the adaptive estimation problem. In Section 3 the adopted hierarchical Bayesian model is described. Section 4 presents a variational Bayes for the batch estimation problem. Section 5 presents the adaptive formulation of the batch variational Bayes algorithm, while experimental results are given in Section 6. Concluding remarks are reported in Section 7.

Notation: Vectors are represented as boldface lowercase letters, e.g. \mathbf{x} , and matrices as boldface uppercase letters, e.g. \mathbf{X} , $(\cdot)^T$ denotes transposition, $\|\cdot\|$ stands for the standard ℓ_2 -norm, $\mathcal{N}(\cdot)$ is the Gaussian distribution, $|\cdot|$ denotes the determinant of a matrix or absolute value in case of a scalar, $\mathcal{G}(\cdot)$ is the Gamma distribution, $\Gamma(\cdot)$ is the Gamma function, $\mathbb{E}[\cdot]$ is the expectation operator, and $\text{diag}(\mathbf{x})$ denotes a diagonal matrix whose diagonal entries are the elements of \mathbf{x} .

2. PROBLEM FORMULATION

Let $\mathbf{w} = [w_1, w_2, \dots, w_N]^T$ denote an unknown signal vector that may be time varying and is assumed to be sparse, i.e. ξ of its N el-

This research is cofinanced by the European Union (European Social Fund ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) Research Funding Program THALIS.

ements (whose positions are not known) are non-zero, with $\xi \ll N$. The unknown vector \mathbf{w} can be sequentially estimated in time utilizing a set of observations and data $\{y(n), \mathbf{x}(n)\}$ that are assumed to be associated by the linear regression model

$$y(n) = \mathbf{x}^T(n)\mathbf{w} + \epsilon(n), \quad (1)$$

where n is the time index, $\mathbf{x}(n) = [x_1(n), x_2(n), \dots, x_N(n)]^T$ is a known signal vector, and $\epsilon(n)$ is assumed to be zero mean Gaussian noise with precision β , i.e. $\epsilon(n) \sim \mathcal{N}(\epsilon(n)|0, \beta^{-1})$. To estimate \mathbf{w} in the least squares (LS) sense, the following minimization criterion can be used, typically expressed as the exponentially weighted sum of the instantaneous estimation errors, i.e.,

$$\min_{\mathbf{w}(n)} \sum_{k=1}^n \lambda^{n-k} |y(k) - \hat{\mathbf{w}}^T(n)\mathbf{x}(k)|^2, \quad (2)$$

where $0 \ll \lambda < 1$ is the forgetting factor. Let $\mathbf{X}(n)$ be the $n \times N$ data matrix whose k th row is $\mathbf{x}^T(k)$, i.e., $\mathbf{X}(n) = [\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(n)]^T$, $\mathbf{y}(n) = [y(1), y(2), \dots, y(n)]^T$ be the observations vector, and $\mathbf{\Lambda}(n) = \text{diag}([\lambda^{n-1}, \lambda^{n-2}, \dots, 1]^T)$. Then, (2) can be written in vector notation as

$$\min_{\mathbf{w}(n)} \|\mathbf{\Lambda}^{1/2}(n)\mathbf{y}(n) - \mathbf{\Lambda}^{1/2}(n)\mathbf{X}(n)\hat{\mathbf{w}}(n)\|^2. \quad (3)$$

Our goal in this paper is, using the pairs of observations and data vectors $y(n), \mathbf{x}(n)$, to design an adaptive estimator for \mathbf{w} , that is constrained to be sparse. To this end, we adopt a slightly modified version of the hierarchical Bayesian model of [7] and propose a fast online variational Bayes algorithm.

3. BAYESIAN MODELING

We consider first the estimation problem defined in (3), in its batch mode, where there is no dependence on n . Thus, we temporarily drop the time index n , which will be retrieved when introducing time-recursions in Section 5. Due to the exponentially weighted data window used in (3), we consider the alternative (with respect to (1)) observation model (cf. [3])

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \boldsymbol{\varepsilon}, \quad (4)$$

where $\boldsymbol{\varepsilon}$ is assumed to be zero mean Gaussian noise with precision $\beta\mathbf{\Lambda}$, $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{\varepsilon}|0, \beta^{-1}\mathbf{\Lambda}^{-1})$. From a probabilistic point of view, the exact form of the additive noise gives rise to the likelihood function

$$p(\mathbf{y}|\mathbf{w}, \beta) = (2\pi)^{-\frac{n}{2}} \beta^{\frac{n}{2}} |\mathbf{\Lambda}|^{\frac{1}{2}} \exp \left[-\frac{\beta}{2} \|\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{y} - \mathbf{\Lambda}^{\frac{1}{2}}\mathbf{X}\mathbf{w}\|^2 \right]. \quad (5)$$

Notice that the maximum likelihood (ML) estimate from (5) coincides with the LS solution of (3). However, as it was previously mentioned, the estimator is constrained to be sparse. To this end, in the Bayesian framework, the likelihood in (5) should be complemented by suitable *conjugate* priors over the parameters \mathbf{w} and β . Specifically, a nonnegative Gamma prior with parameters ρ and δ is adopted for β ,

$$p(\beta; \rho, \delta) = \mathcal{G}(\beta; \rho, \delta) = \frac{\delta^\rho}{\Gamma(\rho)} \beta^{\rho-1} \exp[-\delta\beta]. \quad (6)$$

Next, the prior knowledge of the sparsity of \mathbf{w} is embedded in our Bayesian model via a heavy-tailed Student-t distribution, (see also

[7]). The Student-t distribution is introduced as a two-level conjugate hierarchical prior. In the first level, a Gaussian prior is selected for \mathbf{w} ,

$$p(\mathbf{w}|\boldsymbol{\alpha}, \beta) = \mathcal{N}(\mathbf{w}|0, \beta^{-1}\mathbf{A}^{-1}) = \prod_{i=1}^N p(w_i|\alpha_i, \beta) \\ = \prod_{i=1}^N (2\pi)^{-\frac{1}{2}} \beta^{\frac{1}{2}} \alpha_i^{\frac{1}{2}} \exp \left[-\frac{\beta}{2} w_i^2 \alpha_i \right], \quad (7)$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_N]^T$ and $\mathbf{A} = \text{diag}(\boldsymbol{\alpha})$. In the second level of hierarchy, a Gamma distribution is assigned to the precision parameters α_i , i.e.,

$$p(\alpha_i; a, b) = \mathcal{G}(\alpha_i; a, b) = \frac{b^a}{\Gamma(a)} \alpha_i^{a-1} \exp[-b\alpha_i]. \quad (8)$$

The difference between the proposed model and the hierarchical Bayesian model of [7] is the normalization of the variances of w_i 's by β in (7). This modification ensures the unimodality of the posterior joint distribution, [11], and leads to simpler and more compact update equations, as will be shown later.

4. VARIATIONAL BAYESIAN INFERENCE

The variational approach to the proposed hierarchical Bayesian model involves the maximization of the log-evidence $\log p(\mathbf{y})$ of the model via a variational lower bound, [12]. This operation is equivalent to minimizing the Kullback-Leibler distance between the true posterior pdf $p(\mathbf{w}, \beta, \boldsymbol{\alpha}|\mathbf{y})$ (which cannot be expressed in closed form) and an approximating distribution $q(\mathbf{w}, \beta, \boldsymbol{\alpha})$, which here is assumed to accept the following mean field factorization

$$q(\mathbf{w}, \beta, \boldsymbol{\alpha}) = q(\mathbf{w})q(\beta)q(\boldsymbol{\alpha}) = \prod_{i=1}^N q(w_i)q(\beta) \prod_{i=1}^N q(\alpha_i) \quad (9)$$

Equation (9) assumes posterior independence over the disparate coordinates of \mathbf{w} , which (a) differentiates the proposed approach from the variational SBL setting adopted in [13], and (b) is vital for the development of the adaptive scheme that will be introduced in the next Section.

It is known, [12], that the general expression for the approximating distributions $q(\theta_i)$, where θ_i is any of the model parameters in $\boldsymbol{\theta} = [w_1, \dots, w_N, \beta, \alpha_1, \dots, \alpha_N]^T$, is given by

$$q(\theta_i) = \frac{\exp(\mathbb{E}_{j \neq i} [\log p(\mathbf{y}, \boldsymbol{\theta})])}{\int \exp(\mathbb{E}_{j \neq i} [\log p(\mathbf{y}, \boldsymbol{\theta})]) d\theta_i} \quad (10)$$

where $\mathbb{E}_{j \neq i} [\cdot]$ denotes expectation w.r.t. all $q(\theta_j)$'s except for $q(\theta_i)$. Applying (10) to the proposed model, the approximating pdf for each coordinate w_i is found to be Gaussian given by

$$q(w_i) = \mathcal{N}(w_i; \mu_i, \sigma_i) = (2\pi)^{-\frac{1}{2}} \sigma_i^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \frac{(w_i - \mu_i)^2}{\sigma_i} \right], \quad (11)$$

with

$$\sigma_i = \langle \beta \rangle^{-1} (\mathbf{x}_i^T \mathbf{\Lambda} \mathbf{x}_i + \langle \alpha_i \rangle)^{-1}, \quad (12)$$

$$\mu_i = \langle \beta \rangle \sigma_i \mathbf{x}_i^T \mathbf{\Lambda} (\mathbf{y} - \mathbf{X}_{-i} \mathbf{w}_{-i}). \quad (13)$$

In (13), \mathbf{x}_i is the i -th column of \mathbf{X} , \mathbf{X}_{-i} is the data matrix \mathbf{X} after removing its i -th column and $\langle \cdot \rangle$ denotes expectation w.r.t. the

distributions of the parameters appearing within the brackets. Next, for the noise precision β we get that $q(\beta)$ is a Gamma distribution expressed as

$$q(\beta) = \mathcal{G}(\beta; \tilde{\rho}, \tilde{\delta}) \quad (14)$$

with $\tilde{\rho} = \frac{n+N}{2} + \rho$ and $\tilde{\delta} = \delta + \frac{1}{2} \left\langle \|\Lambda^{\frac{1}{2}} \mathbf{y} - \Lambda^{\frac{1}{2}} \mathbf{X} \mathbf{w}\|^2 \right\rangle + \frac{1}{2} \langle \mathbf{w}^T \mathbf{A} \mathbf{w} \rangle$. The precision parameters α_i 's are also Gamma distributed, i.e.,

$$q(\alpha_i) = \frac{\left(\frac{\langle \beta \rangle \langle w_i^2 \rangle}{2} + b \right)^{a + \frac{1}{2}}}{\Gamma(a + \frac{1}{2})} \alpha_i^{a - \frac{1}{2}} \exp \left[- \left(\frac{\langle \beta \rangle \langle w_i^2 \rangle}{2} + b \right) \alpha_i \right]. \quad (15)$$

Notice that each parameter of the pdfs in (11), (14), and (15) is expressed in terms of expectations of quantities that involve the remaining ones and, thus, all parameters are interrelated. More specifically, as far as the expected values of the parameters of the above pdf's are concerned, these are given by (13) for w_i and

$$\langle \beta \rangle = \frac{\frac{n+N}{2} + \rho}{\delta + \frac{1}{2} \left\langle \|\Lambda^{\frac{1}{2}} \mathbf{y} - \Lambda^{\frac{1}{2}} \mathbf{X} \mathbf{w}\|^2 \right\rangle + \frac{1}{2} \sum_{i=1}^N (\mu_i^2 + \sigma_i) \langle \alpha_i \rangle} \quad (16)$$

$$\langle \alpha_i \rangle = \frac{a + \frac{1}{2}}{b + \frac{1}{2} \langle \beta \rangle (\mu_i^2 + \sigma_i)}, \quad (17)$$

where in (16)

$$\left\langle \|\Lambda^{\frac{1}{2}} \mathbf{y} - \Lambda^{\frac{1}{2}} \mathbf{X} \mathbf{w}\|^2 \right\rangle = \|\Lambda^{\frac{1}{2}} \mathbf{y} - \Lambda^{\frac{1}{2}} \mathbf{X} \boldsymbol{\mu}\|^2 + \sum_{i=1}^N \sigma_i \mathbf{x}_i^T \Lambda \mathbf{x}_i$$

and $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_N]^T$. Hence, a cyclic iterative procedure can be established involving (12), (13), (16) and (17)¹. Due to the convexity of the factors $q(w_i)$, $q(\beta)$ and $q(\alpha_i)$, the variational Bayes algorithm procedure converges and solves the batch estimation problem by providing the posterior mean $\boldsymbol{\mu}$ as a sparse estimate for \mathbf{w} .

5. ADAPTIVE VARIATIONAL SBL (AVSBL)

In this Section, the variational scheme presented previously for solving the batch estimation problem is properly adjusted in order to perform online processing in a computationally efficient manner. To this end, we first define the following time-dependent quantities:

$$\mathbf{R}(n) = \mathbf{X}^T(n) \Lambda(n) \mathbf{X}(n) + \mathbf{A}(n), \quad (18)$$

$$\mathbf{z}(n) = \mathbf{X}^T(n) \Lambda(n) \mathbf{y}(n), \quad (19)$$

$$d(n) = \mathbf{y}^T(n) \Lambda(n) \mathbf{y}(n), \quad (20)$$

Note that all three of them can be efficiently time-updated as follows:

$$\mathbf{R}(n) = \lambda \mathbf{R}(n-1) + \mathbf{x}(n) \mathbf{x}^T(n) - \lambda \mathbf{A}(n-1) + \mathbf{A}(n) \quad (21)$$

$$\mathbf{z}(n) = \lambda \mathbf{z}(n-1) + \mathbf{x}(n) y(n), \quad (22)$$

$$d(n) = \lambda d(n-1) + y^2(n). \quad (23)$$

¹In (13) the first $i-1$ elements of the vector \mathbf{w}_{-i} are μ_1, \dots, μ_{i-1} computed in the current cycle, while its remaining elements are μ_{i+1}, \dots, μ_N computed in the previous cycle of the iterative scheme.

Initialize $\beta(0), \hat{\mathbf{w}}(0), \mathbf{A}(0), \mathbf{R}(0), \mathbf{z}(0), d(0)$

Set a, b, ρ, δ to very small values

for $n = 1, 2, \dots$

for $i = 1, 2, \dots, N$

$$\alpha_i(n) = \frac{a + \frac{1}{2}}{b + \frac{1}{2} (\beta(n-1) \hat{w}_i^2(n-1) + r_{ii}^{-1}(n-1))}$$

end for

$$\mathbf{R}(n) = \lambda \mathbf{R}(n-1) + \mathbf{x}(n) \mathbf{x}^T(n) - \lambda \mathbf{A}(n-1) + \mathbf{A}(n)$$

$$\mathbf{z}(n) = \lambda \mathbf{z}(n-1) + \mathbf{x}(n) y(n)$$

$$d(n) = \lambda d(n-1) + y^2(n)$$

$$\beta(n) = ((1-\lambda)^{-1} + \rho) / (d(n) - \mathbf{z}^T(n) \hat{\mathbf{w}}(n) + \delta)$$

for $i = 1, 2, \dots, N$

$$\hat{w}_i(n) = r_{ii}^{-1}(n) (z_i(n) - \mathbf{r}_{-i}^T(n) \hat{\mathbf{w}}_{-i}(n))$$

end for

end for

Table 1. The AVSBL algorithm

It is easy to realize that $\mathbf{R}(n)$ is the sample auto-correlation matrix of $\mathbf{x}(n)$, regularized by the diagonal matrix $\mathbf{A}(n)$, $\mathbf{z}(n)$ is the sample cross-correlation vector between $\mathbf{x}(n)$ and $y(n)$, and $d(n)$ is the energy of the observation vector $\mathbf{y}(n)$. Now, by substituting (12) in (13) and using (18) and (19) it is straightforward to show that the adaptive weights $\hat{w}_i(n) (= \mu_i(n))$ can be efficiently time updated as follows

$$\hat{w}_i(n) = \frac{1}{r_{ii}(n)} \left(z_i(n) - \mathbf{r}_{-i}^T(n) \hat{\mathbf{w}}_{-i}(n) \right), \quad (24)$$

for $i = 1, 2, \dots, N$. In (24) $r_{ii}(n)$ is the i -th diagonal element of $\mathbf{R}(n)$, $\mathbf{r}_{-i}(n)$ is the i -th column of $\mathbf{R}(n)$ after removing its i -th element, and $\hat{\mathbf{w}}_{-i}^T(n) = [\hat{w}_1(n), \dots, \hat{w}_{i-1}(n), \hat{w}_{i+1}(n-1), \dots, \hat{w}_N(n-1)]^T$. Note that each weight estimate $\hat{w}_i(n)$ depends on the most recent estimates of the other weights². Moreover, it can be shown from (16) that the precision parameter $\beta(n)$ can be very well approximated by the following expression (a detailed derivation is omitted here due to space limitations)

$$\beta(n) = \frac{1/(1-\lambda) + \rho}{d(n) - \mathbf{z}^T(n) \hat{\mathbf{w}}(n) + \delta}, \quad (25)$$

where the term $1/(1-\lambda)$ stands for the active time window in an exponentially weighted LS setting. Finally from (12) and (17) we get for $i = 1, 2, \dots, N$,

$$\alpha_i(n) = \frac{a + \frac{1}{2}}{b + \frac{1}{2} (\beta(n-1) \hat{w}_i^2(n-1) + r_{ii}^{-1}(n-1))}. \quad (26)$$

The main steps of the proposed adaptive variational sparse Bayesian learning (AVSBL) algorithm are given in Table 1. The AVSBL algorithm is based on second-order statistics and has an $\mathcal{O}(N^2)$ complexity, similar to the classical RLS. The most computationally costly steps of the algorithm requiring $\mathcal{O}(N^2)$ operations are those related to the update rule of $\mathbf{R}(n)$ and the weight vector $\hat{\mathbf{w}}(n)$. As shown in the simulations of Section 6, the AVSBL algorithm converges very fast to a sparse estimate for \mathbf{w} and offers much lower steady-state estimation error compared to other competing sparse adaptive schemes.

²It can be shown that, for n fixed, the recursion in (24) over i coincides with one Gauss-Seidel iteration for solving the system of equations $\mathbf{R}(n) \hat{\mathbf{w}}(n) = \mathbf{z}(n)$, see also [9].

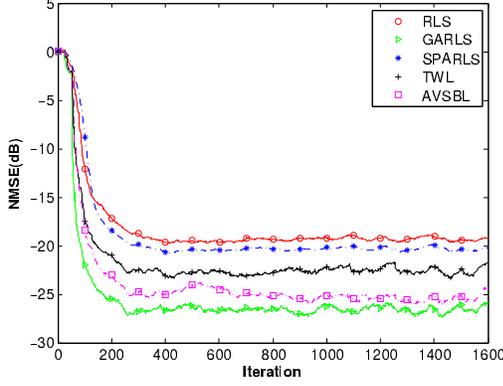


Fig. 1. NMSE of sparse adaptive algorithms.

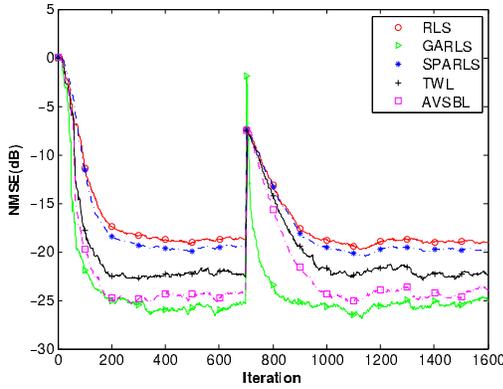


Fig. 2. Tracking performance of sparse adaptive algorithms.

It is interesting to note that in contrast to other recently reported deterministic sparse adaptive estimation techniques, e.g. [2–4], the proposed sparse learning method is fully automated; there is no need for fine-tuning or cross-validating parameters, which is very important from a practical point of view. Most adaptive deterministic schemes are sequential alternatives to the lasso estimator, and hence, variable selection is performed via soft-thresholding, e.g. [2]. Instead, a completely different sparsity-inducing mechanism is used in the AVSBL algorithm, which is fully automatic. More specifically, as the algorithm progresses in time, many of the exponentially distributed precision parameters α_i 's are driven to very large values, forcing also the corresponding diagonal elements $r_{ii}(n)$ of $\mathbf{R}(n)$ to become large (Eq. (18)). As a result, according to (24), many weight parameters are forced to become nearly zero, thus imposing sparsity.

6. EXPERIMENTAL RESULTS

In this Section we conduct computer simulations in an adaptive filtering setup to assess the performance of existing state-of-the-art deterministic adaptive algorithms and the proposed AVSBL algorithm. In all experiments noise is assumed to be white Gaussian, the input sequence is a random ± 1 sequence of length 1600 and the forgetting factor λ is set to 0.99. A sparse Rayleigh fading channel of length $N = 64$ is considered, where each nonzero coefficient follows the so-called Jakes' model with a normalized Doppler frequency

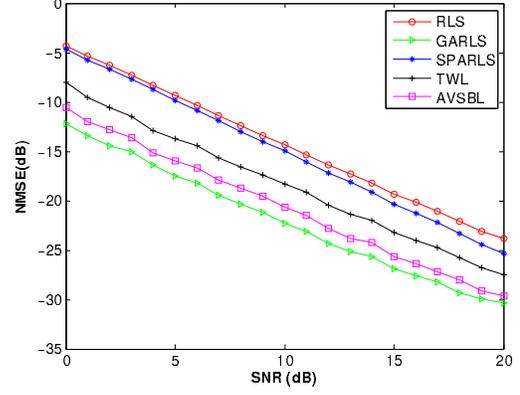


Fig. 3. NMSE versus SNR for sparsity $\xi = 12$.

$f_d T_s = 5 \times 10^{-5}$. Hyperparameters a, b, ρ and δ are set to 10^{-6} . The normalized mean square error (NMSE) used to assess the performance of the algorithms is defined as $\text{NMSE} = \mathbb{E} [\|\mathbf{w} - \hat{\mathbf{w}}\|^2] / \mathbb{E} [\|\mathbf{w}\|^2]$. The proposed AVSBL algorithm is compared to the RLS, the time-weighted Lasso (TWL) coordinate descent type algorithm of [2] and the sparse RLS (SPARLS) algorithm of [3]. In addition, the genie-aided RLS (GARLS), which operates only on the nonzero coefficients of the parameter vector \mathbf{w} , is also included as a benchmark. Note that both TWL and SPARLS incorporate soft-thresholding operations, and multiple runs were first performed to estimate the optimal threshold parameter values in terms of NMSE.

In the first experiment, the signal to noise ratio (SNR) is set to 15dB and a sparsity level of $\xi = 12$ is assumed. Fig. 1 shows the NMSE values of the considered algorithms, averaged over 30 input and noise sequence realizations. The proposed AVSBL algorithm exhibits significantly better estimation performance in comparison to TWL and SPARLS and approaches the optimum channel cognizant GARLS.

The tracking capability of the proposed AVSBL algorithm is shown in Fig. 2. This time a new Rayleigh distributed nonzero tap is added to the channel at time instant 700. It is easy to verify that the AVSBL algorithm adapts more rapidly to sudden changes and converges faster than the other algorithms. In the third experiment, the sparsity level $\xi = 12$ is kept fixed and the SNR ranges from 0dB to 20dB. Figure 3 shows the resulting NMSE curves versus SNR. For each SNR level, the NMSE values after convergence are averaged over a total of 100 realizations of input and noise sequences. We observe that the proposed AVSBL algorithm achieves the best performance for all SNR values and is relatively close to the optimal NMSE bound of GARLS.

7. CONCLUSION

In this paper, a variational framework for the SBL algorithm is developed, which allows the SBL to operate in an adaptive estimation setting, imposing also the sparsity constraint on the parameter vector. Experimental results verify the enhanced variable selection capability of the variational adaptive SBL. A proof of convergence of the proposed adaptive algorithm is currently under investigation. Possible extensions of this work include the incorporation of the sparsity promoting Laplace distribution, as utilized in the Bayesian models of [10] and [14].

8. REFERENCES

- [1] S. O. Haykin, *Adaptive Filter Theory*, Springer, 4th edition, 2002.
- [2] D. Angelosante, J.A. Bazerque, and G.B. Giannakis, "Online adaptive estimation of sparse signals: Where RLS meets the ℓ_1 -norm," *Signal Processing, IEEE Transactions on*, vol. 58, pp. 3436–3447, July 2010.
- [3] B. Babadi, N. Kalouptsidis, and V. Tarokh, "SPARLS: The sparse RLS algorithm," *Signal Processing, IEEE Transactions on*, vol. 58, pp. 4013–4025, Aug. 2010.
- [4] Y. Kopsinis, K. Slavakis, and S. Theodoridis, "Online sparse system identification and signal reconstruction using projections onto weighted ℓ_1 balls," *Signal Processing, IEEE Transactions on*, vol. 59, no. 3, pp. 936–952, Mar 2011.
- [5] K. Slavakis, Y. Kopsinis, S. Theodoridis, and S. McLaughlin, "Generalized thresholding and online sparsity-aware learning in a union of subspaces," *Signal Processing, IEEE Transactions on*, vol. 61, no. 15, pp. 3760–3773, Aug 2013.
- [6] G. Mileounis, B. Babadi, N. Kalouptsidis, and V. Tarokh, "An adaptive greedy algorithm with application to nonlinear communications," *Signal Processing, IEEE Transactions on*, vol. 58, pp. 2998–3007, June 2010.
- [7] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.
- [8] H. Koepl, G. Kubin, and G. Paoli, "Bayesian methods for sparse RLS adaptive filters," in *Signals, Systems and Computers, 2004 Conference Record of the Thirty-Seventh Asilomar Conference on*, Nov. 2003, pp. 1273–1277.
- [9] K.E. Themelis, A.A. Rontogiannis, and K.D. Koutroumbas, "Variational Bayesian sparse adaptive filtering using a Gauss-Seidel recursive approach," in *EUSIPCO*, Sep. 2013.
- [10] S.D. Babacan, R. Molina, and A.K. Katsaggelos, "Bayesian compressive sensing using laplace priors," *Image Processing, IEEE Transactions on*, vol. 19, no. 1, pp. 53–63, 2010.
- [11] T. Park and G. Casella, "The Bayesian Lasso," *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 681–686, Jun 2008.
- [12] D.G. Tzikas, C.L. Likas, and N.P. Galatsanos, "The variational approximation for Bayesian inference," *Signal Processing Magazine, IEEE*, vol. 25, no. 6, pp. 131–146, 2008.
- [13] D. Shutin, T. Buchgraber, S.R. Kulkarni, and H.V. Poor, "Fast variational sparse Bayesian learning with automatic relevance determination for superimposed signals," *Signal Processing, IEEE Transactions on*, vol. 59, no. 12, pp. 6257–6261, 2011.
- [14] K.E. Themelis, A.A. Rontogiannis, and K.D. Koutroumbas, "A novel hierarchical Bayesian approach for sparse semi-supervised hyperspectral unmixing," *Signal Processing, IEEE Transactions on*, vol. 60, no. 2, pp. 585–599, Feb. 2012.