APPLICATION OF THE EVIDENCE PROCEDURE TO THE ESTIMATION OF THE NUMBER OF PATHS IN WIRELESS CHANNELS

Dmitriy Shutin

Institute of Signal Processing
and Speech Communication
Graz University of Technology
Inffeldgasse 12, A-8010 Graz, Austria
dshutin@tugraz.at

Bernard H. Fleury

Information and Signals Division
Department of Communication Technology
Aalborg University, Frederik Bajers Vej 7A,
DK-9220 Aalborg, Denmark
fleury@kom.aau.dk

ABSTRACT

This paper addresses application of the Bayesian evidence procedure to the analysis of wireless channels. We use Relevance Vector Machines– a kernel-based technique to locally maximize evidence that turns out to be promising in the context of the wireless channel estimation. This approach not only allows to estimate channel parameters, but also provides a tool to asses the number of multipath components. We show that in the case of channel sounding using pulse-compression technique it is possible to design an optimal kernel, as well as to estimate parameters of the additive noise and base on it a thresholding level to implement model order estimation. The applicability of the proposed scheme is demonstrated with synthetic as well as real channel measurements.

Keywords: Channel parameter estimation, Bayesian analysis, evidence maximization, sparse representations, model order estimation.

1. INTRODUCTION

Deep understanding of the wireless channel is an essential prerequisite to satisfy ever-growing demands for the fast information access in wireless systems. A wireless channel contains explicitly or implicitly all the information about the propagation environment. To ensure reliable communication the transceiver should be constantly aware of the channel state. In order to make this task feasible, accurate channel models that reproduce in a realistic manner the channel behavior are required. However efficient estimation of the channel parameters, e.g., number of the multipath components, their relative delays, Doppler frequencies, directions of the impinging wavefronts, and polarizations is often difficult. Most often a joint estimation is desired, but this results in intractable optimization procedures, and thus separate estimation schemes are used [1, 2]. Joint estimation of the model order (i.e. the number of the multipath components) and other channel parameters is a particularly difficult task. Both underspecifying and overspecifying the model order leads to significant performance degradation: residual intersymbol interference impairs the performance of the decoder in the former case, while additive noise is injected in the channel equalizer in the latter. The classical solution to this problem is found in the spirit of the Occham's principle, i.e., several models are trained and then those that offer the 'simplest' explanation of the data in terms of number of parameters are selected. Examples are Akaike Information Criterion (AIC) and Minimum Description Length [3], that are special cases of the maximum likelihood model selection, or cross-validations, where the estimated models are compared by their performance over the validation set (see, for example [4]). Thus, if we were to use the latter in real-time environments there were a need to train several models in parallel. On the other hand, incorporating a model selection scheme in the estimation algorithm would eliminate this.

In this contribution we propose to use the evidence maximization approach that can be applied both to the model selection (as a Bayesian extension of the Maximum likelihood model selection) as well as to the estimation of the channel parameters. Evidence for a particular model hypothesis $\mathcal{H}_i$ given the data $\mathcal{D}$ is expressed as the following integral:

$$p(\mathcal{D}|\mathcal{H}_i) = \int p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{H}_i)p(\boldsymbol{\theta}|\mathcal{H}_i)d\boldsymbol{\theta}, \tag{1}$$

where $\boldsymbol{\theta}$ describes the parameters of the candidate model. Maximizing this integral with respect to the unknown parameters and models is known as evidence maximization procedure [5]. Although generally closed form solutions to (1) can be difficult, effective maximization schemes can be constructed under some particular assumptions. In this paper we consider the Relevance Vector Machines (RVM) technique proposed by M. Tipping [6] that effectively locally maximizes the evidence integral (1) for linear kernel-based models. Developed originally for general linear problems, this technique can be quite effectively modified for the estimation of the wireless channels, thus resulting in an effective channel parameter estimation and model selection scheme within the Bayesian framework. RVM initially comes up with overcomplete representation of the data with more kernels than it is needed, and then using the evidence procedure finds which of the kernels are 'irrelevant' to prune them.

The presented material is organized as follows: Section 2 introduces the signal model and notation used in the paper, Section 3 explains the application of the RVM technique to the problem of estimating wireless channels, and Section 4 presents some results illustrating the performance of the RVM-based estimator as well as its application to measured channel impulse responses.

This work has been partly done in cooperation with Elektrobit Oy, Oulu, Finland

2. SIGNAL MODEL

Let us consider the discrete-time channel sounding model shown in Fig.1. The transmitted signal $u[n]$ is used to sound the channel $h[n]$ and is designed to have white-noise-like properties. It is common to model the channel as a tapped-delay line with L taps, each with a delay τ_l and a weight ω_l, $l = 1 \ldots L$, representing the corresponding impinging waves. [1] The received signal $y[n]$ is therefore given by:

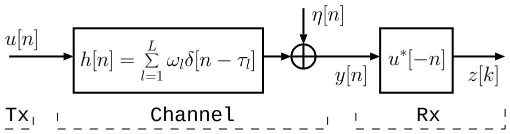

Fig. 1. Model of the radio channel with receiver matched filter front-end.

$$y[n] = \sum_{l=1}^{L} \omega_l u[n - \tau_l] + \eta[n].$$

Here, $\eta[n]$ is an additive white Gaussian noise process with zero mean and variance σ_η^2. The receiver front-end consists of a Matched Filter (MF) matched to the transmitted sequence $u[n]$. The signal $z[k]$ at the output of the MF is then given as:

$$z[k] = \sum_{l=1}^{L} \omega_l R_{uu}[k - \tau_l] + \xi[k], \tag{2}$$

where $R_{uu}[k] = \sum_n u[n]u^*[n+k]$ is the autocorrelation sequence of the transmitted sequence $u[n]$ and $\xi[k] = \sum_n \eta[n] \cdot u^*[n+k]$ is a zero-mean wide-sense stationary noise with autocorrelation function $R_{\xi\xi}[k] = \sigma_\eta^2 R_{uu}[k]$. Equation (2) states that the channel impulse response is a linear combination of L delayed kernel functions $R_{uu}[k - \tau_l]$, observed in the presence of the colored noise $\xi[k]$. By assuming that N samples of $z[k]$ are available, we can rewrite (2) in the vector form:

$$\boldsymbol{z} = \boldsymbol{K}\boldsymbol{\omega} + \boldsymbol{\xi}, \tag{3}$$

where we have defined $\boldsymbol{z} = [z[0], z[1], \ldots, z[N-1]]^T$, $\boldsymbol{\omega} = [\omega_1, \omega_1, \ldots, \omega_L]^T$, $\boldsymbol{\xi} = [\xi[0], \xi[1], \ldots, \xi[N-1]]^T$. The matrix $\boldsymbol{K}$, also called a design matrix, accumulates delayed versions of the kernel function $R_{uu}[k]$ and is constructed as follows: $\boldsymbol{K} = [\boldsymbol{r}_1, \ldots, \boldsymbol{r}_L]$, with $\boldsymbol{r}_l = [R_{uu}[-\tau_l], R_{uu}[1-\tau_l], \ldots, R_{uu}[N-1-\tau_l]]^T$. The goal is to estimate the channel parameters ω_l , τ_l, and noise properties, as well as to determine the order of the model L. It should be stressed that each kernel $\boldsymbol{r}_l$ is associated with the corresponding delay τ_l. It follows that the number of resulting kernels determines the model oder, while their indices determine the corresponding multipath delays. We now show how the RVM technique can be applied to the problem at hands.

[1] In this paper we consider Single-Input-Single-Output channels, thus omitting aspects arising with multiple antennas.

3. APPLICATION OF EVIDENCE PROCEDURE TO THE ANALYSIS OF WIRELESS CHANNELS

First of all we will make some assumptions that are crucial for the further discussion. We assume that the model order is initially overestimated, i.e. our initial hypothesis comprising L basis functions is more complex than the true underlying reality. We will also show that the color of the additive noise at the output of the MF can be effectively accounted for in the algorithm. In fact, it can be shown that the covariance matrix of $\boldsymbol{\xi}$ is given as $\boldsymbol{\Sigma} = \sigma_\eta^2 \boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda}$ is fixed matrix with elements given as $\Lambda_{ij} = R_{uu}[i-j]$. Thus, only the factor σ_η^2 has to be estimated within the proposed framework, since $R_{uu}[k]$ is known.

Now, we will redo the major steps of the RVM algorithm with modifications to accommodate communication channels, where needed. For a more detailed treatment of the RVM the interested reader is referred to the original contribution [6]. Estimating parameters of interest consists in considering likelihood function $p(\boldsymbol{z}|\boldsymbol{\omega}, \boldsymbol{\Sigma}) = \mathcal{N}(\boldsymbol{z}|\boldsymbol{K}\boldsymbol{\omega}, \boldsymbol{\Sigma})$ which is a multivariate complex normal distribution with the mean $\boldsymbol{K}\boldsymbol{\omega}$ and covariance matrix $\boldsymbol{\Sigma}$. Model weights $\boldsymbol{\omega}$ are specified by means of the prior $p(\boldsymbol{\omega}|\boldsymbol{\alpha}) = \mathcal{N}(\boldsymbol{\omega}|0, \boldsymbol{A}^{-1})$, where $\boldsymbol{A} = \mathrm{diag}\{\boldsymbol{\alpha}\}$ [6, 5]. We note that there is an individual hyperparameter α_l for each weight ω_l. To complete the specification of the priors, a hyperprior for $\boldsymbol{\alpha}$ is introduced in the form of a Gamma distribution:

$$p(\boldsymbol{\alpha}|a, b) = \prod_{l=1}^{L} \mathcal{G}(\alpha_l|a, b) = \prod_{l=1}^{L} \frac{b^a}{\Gamma(a)} \alpha_l^{a-1} \exp(-b\alpha_l).$$

A similar hierarchy is build to describe the additive noise process $\xi[n]$. In this case the only random parameter is σ_η^2. We define $\beta = \sigma_\eta^{-2}$, with hyperprior $p(\beta|c, d) = \mathcal{G}(\beta|c, d)$. The hyperpriors of $\boldsymbol{\alpha}$ and β are usually made non-informative, i.e., uniform by setting a, b, c, and d to very small values. The presented prior formulation for the given problem is quite common in Bayesian estimation. RVM, as a Bayesian approach, advantageously embodies two modeling steps, that are crucial for us: 1) model fitting (estimation of the $\boldsymbol{\omega}$), and 2) model comparison and selection (estimation of β and $\boldsymbol{\alpha}$). Those steps naturally result in consideration of the posterior over all the unknown: $p(\boldsymbol{\omega}, \boldsymbol{\alpha}, \beta|\boldsymbol{z}) = p(\boldsymbol{\omega}|\boldsymbol{z}, \boldsymbol{\alpha}, \beta)p(\boldsymbol{\alpha}, \beta|\boldsymbol{z})$.

Model fitting consists in finding MAP estimates of the model weights $\boldsymbol{\omega}$ from the posterior $p(\boldsymbol{\omega}|\boldsymbol{z}, \boldsymbol{\alpha}, \beta) \propto p(\boldsymbol{z}|\boldsymbol{\omega}, \beta)p(\boldsymbol{\omega}|\boldsymbol{\alpha})$ with $p(\boldsymbol{z}|\boldsymbol{\omega}, \beta) \equiv p(\boldsymbol{z}|\boldsymbol{\omega}, \boldsymbol{\Sigma})$. The former can be easily evaluated analytically and results in a complex Gaussian distribution:

$$p(\boldsymbol{\omega}|\boldsymbol{z}, \boldsymbol{\alpha}, \beta) = \mathcal{N}(\boldsymbol{\omega}|\boldsymbol{\mu}, \boldsymbol{\Phi}), \tag{4}$$

where $\boldsymbol{\mu}$ is MAP estimate of the model weights $\boldsymbol{\omega}$ with covariance matrix $\boldsymbol{\Phi}$:

$$\boldsymbol{\Phi} = (\boldsymbol{A} + \beta \boldsymbol{K}^H \boldsymbol{\Lambda}^{-1} \boldsymbol{K})^{-1}, \tag{5}$$

$$\boldsymbol{\mu} = \beta \boldsymbol{\Phi} \boldsymbol{K}^H \boldsymbol{\Lambda}^{-1} \boldsymbol{z}. \tag{6}$$

Model selection step is unfortunately a bit more elaborate. It involves analysis of the model posterior $p(\boldsymbol{\alpha}, \beta|\boldsymbol{z})$. Alternative models (hypotheses) are constructed as different subsets of α_l's associated with the column-vectors $\boldsymbol{r}_l$ in the design matrix $\boldsymbol{K}$. Some of the MAP estimates of α_l will concentrate at very large values, so effectively switching off the corresponding column in $\boldsymbol{K}$ since $p(\omega_l|\alpha_l) = \mathcal{N}(\omega_l|0, \alpha_l^{-1})$. Straight-forward optimization of $p(\boldsymbol{\alpha}, \beta|\boldsymbol{z})$ is possible, but computationally prohibitive. In

[6] it was proposed to represent $p(\boldsymbol{\alpha}, \beta|\boldsymbol{z})$ by the delta-function at its mode $\{\boldsymbol{\alpha}_{MP}, \beta_{MP}\}$. Thus, RVM learning is the search for the hyperparameter posterior mode, i.e. maximization of the $p(\boldsymbol{\alpha}, \beta|\boldsymbol{z}) \propto p(\boldsymbol{z}|\boldsymbol{\alpha}, \beta)p(\boldsymbol{\alpha})p(\beta)$. This can be further simplified for the case of uniform hyperpriors: only the term $p(\boldsymbol{z}|\boldsymbol{\alpha}, \beta)$ has to be maximized. This term is referred to as the evidence for the hyperparameters and its maximization as the evidence procedure[7]. Unfortunately, the maximizing values $\boldsymbol{\alpha}_{MP}$ and β_{MP} can not be found in a closed form and iterative approaches are needed to solve this optimization task. Below, we provide the complete RVM algorithm (Table 1) along with corresponding re-estimation equations.

Table 1 RVM application algorithm

Initialize $\boldsymbol{\alpha}^{[0]}$, $(\sigma_\eta^2)^{[0]}$; Construct $\boldsymbol{K}$ for $\boldsymbol{\tau} = \{\tau_1, \tau_1, \ldots, \tau_L\}$
Initial hypothesis: $\mathcal{H}_0 = \{\alpha_1, \ldots, \alpha_L\}$

while $\boldsymbol{\alpha}$ and σ_η^2 have not converged

1) Model fitting: Compute $\boldsymbol{\mu}^{[i]}$ and $\boldsymbol{\Phi}^{[i]}$ from (6) and (5), respectively.

2) Model selection:

for all $\alpha_l \in \mathcal{H}_i$

$$(\alpha_l^{-1})^{[i+1]} = \Phi_{ll}^{[i]} + |\mu_l^{[i]}|^2 \tag{7}$$

endfor

$$(\sigma_\eta^2)^{[i+1]} = \frac{\mathrm{tr}[\boldsymbol{\Phi}^{[i]}\boldsymbol{K}^H\boldsymbol{\Lambda}^{-1}\boldsymbol{K}]}{N} + \\ + \frac{1}{N}(\boldsymbol{z} - \boldsymbol{K}\boldsymbol{\mu}^{[i]})^H\boldsymbol{\Lambda}^{-1}(\boldsymbol{z} - \boldsymbol{K}\boldsymbol{\mu}^{[i]}) \tag{8}$$

% — New hypothesis —

for all $\alpha_l \in \mathcal{H}_i$

if $\alpha_l \geq \alpha_{\mathrm{threshold}}$ then $\mathcal{H}_i = \mathcal{H}_i \setminus \{\alpha_l\}$

endfor

endwhile

Here, Φ_{ll} is the lth element on the main diagonal of the $\boldsymbol{\Phi}$ matrix, and μ_l is the lth element in the parameter vector $\boldsymbol{\mu}$. We also used $\sigma_\eta^2 \equiv \beta^{-1}$ in the final expressions.

3.1. Selection of the thresholding level $\alpha_{\mathrm{threshold}}$

One of the main advantages of the RVM algorithm is the ability to estimate parameters and select the best representation of the data in terms of the model order. This requires two ingredients: knowledge of the noise parameters and thresholding level $\alpha_{\mathrm{threshold}}$, which is in general a function of the former. RVM has a superiority of being able to estimate the noise parameters from the data. Thus, assuming that the noise can be estimated we devise the thresholding by analyzing the steady-state behavior of (7). Based on this analysis we can infer that α_l^{-1}'s that correspond to the noise rather than multipath component, $\alpha_{\mathrm{noise}}^{-1}$, are coming from a Gamma distribution $\mathcal{G}(\alpha_{\mathrm{noise}}^{-1}|M, C)$. Values of M and C are inferred by the analysis of the resulting steady state and by means of Monte-Carlo simulations. It was found that the desired distribution can be approximated with $C = 4/\mathrm{median}_l(\alpha_l^{-1})$ and $M = 1/2$. Then, $\alpha_{\mathrm{threshold}}$ is defined such that probability $P[\alpha_{\mathrm{noise}}^{-1} < \alpha_{\mathrm{threshold}}^{-1}] = \rho$ where ρ is a predefined probability of α_l corresponding to the noise rather than signal.

I	L_{ch}	$\tau_l, \times 10^{-7} sec$	ω_l
127	256	$\tau_l \in \mathcal{U}(7.65, 15.3)$	$\omega_l = e^{j\phi}$, $\phi \in \mathcal{U}(0, 2\pi)$

Table 2: Parameter settings used in the Monte-Carlo simulations

4. NUMERICAL EXAMPLES

The performance of the proposed algorithm is demonstrated with simulated as well as measured channel impulse responses. The channels are simulated as shown in Fig.1. An I-chip-long pseudo-noise PN sequence is used to sound the channel with L multipath components. It is assumed that the channel realization is corrupted with complex zero-mean white Gaussian noise with spectral density σ_η^2. The MF output is sampled with frequency $F_s = 100$MHz that results in L_{ch} samples with one sample-per-chip resolution. The design matrix $\boldsymbol{K}$ is composed of the shifted autocorrelations of the PN sequence as explained in Section 2, with initial hypothesis $\mathcal{H}_0$ comprising all of the sampling instances.

The performance of the algorithm is evaluated versus Signal-to-Noise Ratio (SNR) at the output of the MF, defined as:

$$\mathrm{SNR}_{\mathrm{out}} = 10\log_{10}\left(\frac{IA^2}{\sigma_\eta^2}\right), \quad A^2 = \frac{1}{L}\sum_l \omega_l^2.$$

4.1. Simulated channels

In this subsection we present some multipath detection results for synthetic channels performed with Monte-Carlo simulations. The parameters involved in the simulations are specified in Table 2. We denote $\mathcal{U}(a, b)$ to be a uniform distribution in the interval $[a, b]$. Initialization of the RVM algorithm consists in specifying initial variance $(\sigma_\eta^2)^{[0]}$ and hyperparameters $\boldsymbol{\alpha}^{[0]}$. The components of the $\boldsymbol{\alpha}^{[0]}$ all have identical values set to the inverse of the second non-central moment of $z[n]$. The initial value of the input noise $(\sigma_\eta^2)^{[0]}$ has to be somehow known or measured. We also used $\rho = 0.999$ to compute $\alpha_{\mathrm{threshold}}$. In order to evaluate the performance of the proposed scheme it is necessary to come up with an appropriate quality criteria. There are in fact two types of errors the algorithm can make: (a) *insertion error*– it erroneously detects a non-existing component; (b) *deletion error*– it misses an existing component. The situation when a component has been found exactly on its position is a *hit*. We use the number of hits, deletions and insertions as the measure of the algorithm performance. Figure 2 shows the obtained results versus different $\mathrm{SNR}_{\mathrm{out}}$ values. The presented results are averaged over 500 Monte-Carlo runs. It can be seen that RVM algorithm is able to estimate the model order, provided the $\mathrm{SNR}_{\mathrm{out}}$ is high enough. We will stress that the case shown represents the situation when the the design matrix contains the kernel $\boldsymbol{r}_l$ positioned at the simulated delay τ_l, which corresponds to the discrete-time model (2) matching the measurement. Misalignment will result in the increased number of components, and as the result, overall performance degradation. In fact the problem lies in the nature of kernel-based models: the design matrix $\boldsymbol{K}$ acts as a fixed alphabet used to reconstruct the MF output signal. If the latter has a contribution not present in the alphabet, let us say a multipath between sampling instances, it will be represented as a linear combination of the kernels in $\boldsymbol{K}$, thus leading to artefacts. To amend this problem, it is necessary to re-design the matrix $\boldsymbol{K}$ to higher resolution after one complete RVM cycle, thus allowing any desired accuracy in the channel representation. Those investigations are not shown here due to the space limitations.

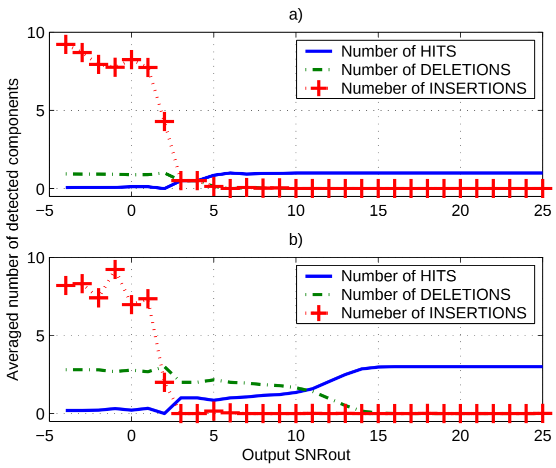

Fig. 2. Number of hits, deletiprobabilityons and insertions for a channel with respectively one (a) and three (b) components.

4.2. Application of the RVM method to the measured data

The proposed RVM algorithm has been also applied to the measurement data collected in indoor environments. Channel measurements were done with the MIMO-capable channel sounder PropSound manufactured by Elektrobit Oy. A single measurement realization from one sub-channel has been selected to illustrate the algorithm behavior. The channel sounding setup is equivalent to the block-diagram shown in Fig. 1. The design matrix $\boldsymbol{K}$ was constructed using the autocorrelation sequence obtained from calibration measurement. Components of $\boldsymbol{\alpha}^{[0]}$ are all set equal to the inverse of the averaged power of the MF output signal. We used tail of the IR to estimate the inital noise variance, which resulted in $\sigma_\eta^2 = 5.0 \cdot 10^{-10}$. Thresholding level $\alpha_{\text{threshold}}$ was selected based on $\rho = 0.999$. The estimated channel impulse response along with the original MF output signal are shown in Fig. 3. Usually it requires from 3 to 10 iterations for the algorithm to converge. Visual inspection of the impulse response shows that the algorithm is capable of selecting areas in the MF output signal with significant power, i.e. where multipaths are likely to happen.

5. CONCLUSIONS

Relevance vector machine is a powerful kernel-based Bayesian method that is used to find sparse solutions to general linear problems by applying evidence maximization approach. In this contribution we have shown how the problem of channel estimation can be posed as a general linear problem, which makes the application of the RVMs possible. We have also developed modifications of the original RVM algorithm that are specific for the analysis of MF output signals in a pulse-compression-based channel sounder, which altogether allow to estimate the number of multipaths as well as channel parameters within the Bayesian framework.

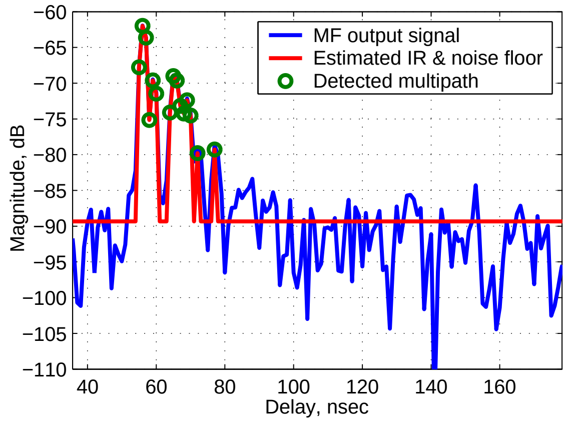

Fig. 3. Estimated channel impulse response and noise floor.

Proposed scheme performs well even when a single channel snapshot is available, i.e., in situations when it is difficult to construct a channel correlation matrix to use AIC/MDL based criteria. We also found a way to approximate the distribution of the hyperparameters associated with noise contributions, thus allowing to select a non-empirical thresholding based on a certain required confidence level. However, this analysis has been done only for a single channel snapshot. Similar analysis of the multiple channel snapshots or multiple-antenna cases is still an open issue.

We have also applied the method to the analysis of synthetic as well as measured impulse responses. The obtained results support our conjecture that the method is quite promising and can lead to the significant improvement of channel estimation.

6. REFERENCES

[1] H. Krim and M. Viberg, "Two decades of the array signal processing research," *IEEE Signal Processing Mag.*, p. 6794, July 1996.

[2] B.H. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, and K. Ingeman Pedersen, "Channel parameter estimation in mobile radio environments using the SAGE algorithm," *Selected Areas in Communications, IEEE Journal on*, vol. 17, no. 3, pp. 434–450, March 1999, Channel parameters estimation.

[3] T. Wax, M. Kailath, "Detection of signals by information theoretic criteria," *Acoustics, Speech, and Signal Processing , IEEE Transactions on*, vol. Volume: 33, no. 2, pp. 387– 392, April 1985.

[4] Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern classification*, John Wiley & Sons, Inc., second edition, 2000.

[5] D. J. C. MacKay, "Bayesian methods for backpropagation networks," in *Models of Neural Networks III*, E. Domany, J. L. van Hemmen, and K. Schulten, Eds., chapter 6, pp. 211–254. Springer-Verlag, New York, 1994.

[6] Michael Tipping, "Sparse Bayesian learning and the Relevance Vector Machine," *Journal on Machine Learning Research*, , no. 1, pp. 211–244, June 2001.

[7] D. J. C. MacKay, "Bayesian interpolation," *Neural Computation*, vol. 4, no. 3, pp. 415–447, 1992.