Bayesian Inference for Multiple Antenna Cognitive Receivers

Romain Couillet, Merouane Debbah

Bayesian Inference
for Multiple Antenna Cognitive Receivers

Romain Couillet
NXP Semiconductors, Supelec
505 Route des Lucioles
06560 Sophia Antipolis, France
Email: romain.couillet@nxp.com

Mérouane Debbah
Alcatel-Lucent Chair, Supelec
Plateau de Moulon, 3 rue Joliot-Curie
91192 Gif sur Yvette, France
Email: merouane.debbah@supelec.fr

***Abstract*—A Bayesian inference learning process for cognitive receivers is provided in this paper. We focus on the particular case of signal detection as an explanatory example to the learning framework. Under any prior state of knowledge on the communication channel, an information theoretic criterion is presented to decide on the presence of informative data in a noisy wireless MIMO communication. We detail the particular cases of knowledge, or absence of knowledge at the receiver, of the number of transmit antennas and noise power. The provided method is instrumental to provide intelligence to the receiver and gives birth to a novel Bayesian signal detector. The detector is compared to the classical power detector and provides detection performance upper bounds. Simulations corroborate the theoretical results and quantify the gain achieved using the proposed Bayesian framework.**

I. Introduction

Since a few years, the idea of smart receiving devices has made its way through the general framework of cognitive radio [2]. The general idea of an ideal cognitive receiver is a device that is capable of inferring on any information it is given to discover by itself the surrounding environment. Such a device should be first able to turn prior information on the transmission channel into a mathematically tractable form. This allows then the terminal to take optimal instantaneous decisions in terms of information to feed back, bandwidth to occupy, transmission power to use etc. It should also be capable of updating its knowledge to continuously adapt to the dynamics of the environment.

In particular, one of the key features of cognitive receivers is their ability to *sense* free spectrum. Indeed, when the cognitive device is switched on, its prior knowledge is very limited but still it is requited to *decide* whether it receives informative data or pure noise due to interfering background electromagnetic fields, on different frequency bands: this is called the *signal detection* procedure.

In the SISO (single input antenna, single output antenna) scenario, the study of the optimal signal detector from the Bayesian viewpoint dates back to the work of Urkowitz [1] on AWGN (additive white Gaussian noise) channels. It was later extended to more realistic channel models [3]-[4]. By *optimal signal detector*, Urkowitz means the process that enjoys the maximum *correct detection rate* (i.e. the odds for an informative signal to be detected as such) for a given low *false alarm rate* (i.e. the odds for a pure noise input to be wrongly declared an informative signal). To the authors' knowledge, the MIMO extension has not been studied, because of the almost prohibitive mathematical complexity of the problem. In tacit accordance among the scientific community, the usual power detection technique from Urkowitz was then simply adapted to the MIMO scenario, e.g. [5].

This raises the interest for new techniques such as cooperative spectrum sensing using multiple antennas [5]. Those techniques propose to improve the signal detection method of Urkowitz by using extra system dimensions (space dimension through cooperation among terminals for example). Unfortunately, the approaches used are highly dependent on the initial assumptions made and have led to many different contributions. For instance, some insightful work emerged which uses eigenspectrum analysis of the received sampled signals [10]-[11]. Those might provide interesting results in their simplicity and their limited need for prior system knowledge; however, the space over which those techniques are valuable is usually difficult to determine (this space can be seen as a multidimensional field spanning from 0 to infinite SNR, from pure void to heavily loaded environment etc.).

In this work, we introduce a general Bayesian framework providing a sound basis for signal detection using information theoretic tools. The methodology is based a *consistent*[1] approach to deal with prior information. This approach follows the work of E. T. Jaynes [9] on *probability theory seen as an extension of logic*. In this theory, the set of information on the environment is encoded into probability assignments using jointly the *maximum entropy principle*[13] and the Bayes' rule.

This paper is structured as follows: In section II we formulate the signal detection model. Then in section III, the optimal Bayesian signal detectors are computed for different levels of knowledge on the system model. Simulations are then presented in section IV. Finally, after a short discussion in Section V on the general framework and its limitations, we provide our conclusions.

Notations: In the following, boldface lowercase and uppercase characters are used for vectors and matrices, respectively.

[1] by *consistent* we mean that two problems defined with the same amount of prior information should lead to the same final solution.

We note $(\cdot)^{\mathsf{H}}$ the Hermitian transpose, $\mathrm{tr}(\cdot)$ denotes the matrix trace. $\mathcal{M}(\mathcal{A}, N, M)$ is the set of matrices of size $N \times M$ over the algebra $\mathcal{A}$. $\mathcal{U}(N)$ is the set of unitary square matrices of size N. The notation $P_X(Y)$ denotes the probability density function of the variable X evaluated in the vicinity of Y. The notation $(x)_+$ equals x if $x > 0$ and 0 otherwise.

II. Signal Model

A. Prior information

We consider a MIMO communication system for which the receiver may have different levels of knowledge. We first define hereafter the minimum channel state of knowledge available to the receiving device:

S-i) the receiver has N antennas.
S-ii) the receiver samples as many as L times the input from the RF interface.
S-iii) naming T_s the sampling period, LT_s is supposed less than the channel coherence time.
S-iv) the signal sent by the transmitter has a constant unit mean power.
S-v) the MIMO channel has a constant mean power.

We similarly define additional information the receiver may have:

V-i) the transmitter possesses (and uses) M antennas.
V-ii) the noise power σ^2 is known.

This list could of course be extended (e.g. knowledge of the transmit signal constellation, number of interferers, channel length...) but our present work shall only treat the enumerated cases.

B. Model

Given a certain amount of sampled signals, the objective of the signal detection methods is to be able to optimally infer on the following hypothesis:

- $\mathcal{H}_0$. Only background noise is received.
- $\mathcal{H}_1$. Informative data added to background noise is received.

Given hypothesis S-iv), the only information on the transmitted signal (under $\mathcal{H}_1$) is their unit variance. The maximum entropy principle claims that, under this limited state of knowledge, the transmitted data must be modeled as i.i.d. Gaussian [9]. The data vector, at time $l \in \{1, \ldots, L\}$, is denoted $\mathbf{s}^{(l)} = (s_1^{(l)}, \ldots, s_M^{(l)})^{\mathsf{T}} \in \mathbb{C}^M$. The data vectors are stacked into the receive matrix $\mathbf{S} = [\mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(L)}]$.

If the noise level σ^2 is known, then either under $\mathcal{H}_0$ or $\mathcal{H}_1$, the background noise must be represented, thanks to the same maximum entropy argument as before, by a *complex standard Gaussian matrix* $\mathbf{\Theta} \in \mathcal{M}(\mathbb{C}, N, L)$ (i.e. a matrix with i.i.d. standard complex Gaussian entries θ_{ij}) [6]. Under $\mathcal{H}_1$, the channel matrix (constant over the LT_s duration) is denoted $\mathbf{H} \in \mathcal{M}(\mathbb{C}, N, M)$ with entries h_{ij} being the link between the j^{th} transmitting antenna and the i^{th} receiving antenna. The model for $\mathbf{H}$ is only described in the following sections since modeling $\mathbf{H}$ is one of the key points in our derivation. The received data at sampling time l are given by the $N \times 1$ vector $\mathbf{y}^{(l)}$ that we stack, over the L sampling periods, into the matrix $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(L)}] \in \mathcal{M}(\mathbb{C}, N, L)$.

This leads for $\mathcal{H}_0$ to the model,

$$\mathbf{Y} = \sigma \mathbf{\Theta} \tag{1}$$

with $\mathbf{Y}$ and $\mathbf{\Theta}$ of size $N \times L$.

And for $\mathcal{H}_1$ to

$$\mathbf{Y} = [\mathbf{H},\ \sigma \mathbf{I}_N] \cdot \begin{bmatrix} \mathbf{S} \\ \mathbf{\Theta} \end{bmatrix} \tag{2}$$

with $\mathbf{Y}$ of size $N \times L$. We also denote by $\mathbf{\Sigma}$ the autocovariance matrix:

$$\begin{aligned} \mathbf{\Sigma} &= \mathrm{E}[\mathbf{Y}\mathbf{Y}^{\mathsf{H}}] && (3)\\ &= L\left(\mathbf{H}\mathbf{H}^{\mathsf{H}} + \sigma^2 \mathbf{I}_N\right) && (4)\\ &= \mathbf{U}\left(L\mathbf{\Lambda}\right)\mathbf{U}^{\mathsf{H}} && (5) \end{aligned}$$

where $\mathbf{\Lambda} = \mathrm{diag}\left(\nu_1 + \sigma^2, \ldots, \nu_N + \sigma^2\right)$, with $\{\nu_i, i \in \{1, \ldots, N\}\}$ the eigenvalues of $\mathbf{H}\mathbf{H}^{\mathsf{H}}$ and $\mathbf{U}$ is a certain unitary matrix.

Our intention is to make a decision on whether, given received data $\mathbf{Y}$, the probability for $\mathcal{H}_1$ is greater than the probability for $\mathcal{H}_0$. This problem is usually referred to as *hypothesis testing* [9]. The decision criterion is based on the ratio

$$C(\mathbf{Y}) = \frac{P_{\mathcal{H}_1|\mathbf{Y}}(\mathbf{Y})}{P_{\mathcal{H}_0|\mathbf{Y}}(\mathbf{Y})} \tag{6}$$

Thanks to Bayes' rule [8], this derives into

$$C(\mathbf{Y}) = \frac{P_{\mathcal{H}_1} \cdot P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y})}{P_{\mathcal{H}_0} \cdot P_{\mathbf{Y}|\mathcal{H}_0}(\mathbf{Y})} \tag{7}$$

Checking our list of prior information, nothing tells us whether $\mathcal{H}_1$ is more or less probable than $\mathcal{H}_0$. Using the maximum entropy principle on this rather obvious example, we must set $P_{\mathcal{H}_1} = P_{\mathcal{H}_0} = \frac{1}{2}$, and then

$$C(\mathbf{Y}) = \frac{P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y})}{P_{\mathbf{Y}|\mathcal{H}_0}(\mathbf{Y})} \tag{8}$$

reduces to a maximum likelihood criterion.

III. Optimal Signal Detection

A. Complete set of knowledge

1) Derivation of $P_{\mathbf{Y}|\mathcal{H}_i}$ in SIMO case: Let us analyze the situation when the noise level σ^2 (hypothesis V-ii)) and the number M of transmit antennas are known to the receiver (hypothesis V-i)) and let us assume in this first scenario that $M = 1$. Consider also the case when $L > N$ (this is a commonly an obvious assumption that the quantity of sampled periods is large compared to any problem dimension).

a) Pure noise likelihood $P_{\mathbf{Y}|\mathcal{H}_0}$: In this first scenario, $\mathbf{\Theta}$ is a Gaussian matrix with independent entries. The distribution for $\mathbf{Y}$, that can be seen as a random vector with NL entries, is then NL multivariate uncorrelated complex Gaussian with covariance $\sigma^2\mathbf{I}_{NL}$,

$$P_{\mathbf{Y}|\mathcal{H}_0}(\mathbf{Y}) = \frac{1}{(\pi\sigma^2)^{NL}} e^{-\frac{1}{\sigma^2}\operatorname{tr}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}} \tag{9}$$

$$\tag{10}$$

by denoting $\mathbf{x} = (x_1, \ldots, x_N)^{\mathsf{T}}$ the eigenvalue distribution of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$, (11) only depends on $\sum_{i=1}^N x_i$,

$$P_{\mathbf{Y}|\mathcal{H}_0}(\mathbf{x}) = \frac{1}{(\pi\sigma^2)^{NL}} e^{-\frac{1}{\sigma^2}\sum_{i=1}^N x_i} \tag{11}$$

b) Informative data likelihood $P_{\mathbf{Y}|\mathcal{H}_1}$: In scenario $\mathcal{H}_1$, the problem is more involved. The maximum entropy principle shows that our best guess is for $\mathbf{H}$ to be jointly uncorrelated Gaussian distributed [7]. Up to a scaling factor at the signal reception, the noise level knowledge allows us to constrain the rows of $\mathbf{H}$ to be of unit mean power 1 (i.e. $\forall i, j\ E[|h_{ij}|^2] = 1/M$). Therefore, since $M = 1$, $\mathbf{H} \in \mathbb{C}^{\mathbf{N}\times\mathbf{1}}$ and $\mathbf{\Sigma} = \mathbf{H}\mathbf{H}^{\mathsf{H}} + \sigma^2\mathbf{I}_N$ has $N-1$ eigenvalues equal to σ^2 and another distinct eigenvalue $\lambda_1 = \nu_1 + \sigma^2 = (\sum_{i=1}^N |h_{i1}|^2) + \sigma^2$. The density of $\lambda_1 - \sigma^2$ is a complex χ^2_N distribution (which is, up to a scaling factor 2, equivalent to a real χ^2_{2N} distribution). Hence the *unordered* eigenvalue distribution of $\mathbf{\Sigma}$ [18]

$$P_{\mathbf{\Lambda}}(\mathbf{\Lambda})d\mathbf{\Lambda} =$$
$$\frac{1}{N}(\lambda_1 - \sigma^2)_+^{N-1} \frac{e^{-(\lambda_1-\sigma^2)}}{(N-1)!} \prod_{i=2}^N \delta(\lambda_i - \sigma^2) d\lambda_1 \ldots d\lambda_N \tag{14}$$

Given model (2), for a fixed $\mathbf{H}$ channel, $\mathbf{Y}$ is distributed as a correlated Gaussian matrix,

$$P_{\mathbf{Y}|\mathbf{\Sigma}I_1}(\mathbf{Y}, \mathbf{U}, L\mathbf{\Lambda}) = \frac{1}{\pi^{LN}\det(\Lambda)^L} e^{-\operatorname{tr}\left(\mathbf{Y}\mathbf{Y}^{\mathsf{H}}\mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^{\mathsf{H}}\right)} \tag{15}$$

where I_k denotes the prior information "$\mathcal{H}_1$ and $M = k$".

Since the channel $\mathbf{H}$ is unknown, we need to integrate out all possible channels of the model (2) over the probability space of $N \times M$ matrices with Gaussian i.i.d. distribution. This is equivalent to integrating out all possible covariance matrices $\mathbf{\Sigma}$ over the space of such covariance matrices

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \int_{\mathbf{\Sigma}} P_{\mathbf{Y}|\mathbf{\Sigma}\mathcal{H}_1}(\mathbf{Y}, \mathbf{\Sigma}) P_{\mathbf{\Sigma}}(\mathbf{\Sigma}) d\mathbf{\Sigma} \tag{16}$$

Using the eigenvalue factorization (5), one can move from the space of covariance matrices $\mathbf{\Sigma}$ to the space of diagonal matrices $\mathbf{\Lambda}$, which is isomorphic to the real positive half-lign that carries λ_1. Also, for any unitary matrix $\mathbf{U}$ and any standard i.i.d. Gaussian vector $\mathbf{h}$, the product $\mathbf{U}\left[\mathbf{h},\ \sigma^2\mathbf{I}_N\right]$ results in another matrix $\left[\mathbf{h}',\ \sigma^2\mathbf{I}_N\right]$ with $\mathbf{h}'$ standard Gaussian thanks to the unitary linear product. More generally, for any zero mean i.i.d. Gaussian vector $\mathbf{h}$, the set $\{\mathbf{U}\mathbf{h},\ \mathbf{U} \in \mathcal{U}(N)\}$ is uniformly distributed on the ensemble of zero mean i.i.d. Gaussian vectors and of variance $\mathbf{h}^{\mathsf{H}}\mathbf{h}$. This property leads to the independence of the respective distributions of $\mathbf{U}$ and $\mathbf{\Lambda}$ in (3), which implies

$$\begin{aligned}
&P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) \\
&= \int_{\mathbf{\Sigma}} P_{\mathbf{Y}|\mathbf{\Sigma}\mathcal{H}_1}(\mathbf{Y}, \mathbf{\Sigma}) P_{\mathbf{\Sigma}}(\mathbf{\Sigma}) d\mathbf{\Sigma} \\
&= \int_{\mathcal{U}(N)\times\mathbb{R}^{+N}} P_{\mathbf{Y}|\mathbf{\Sigma}\mathcal{H}_1}(\mathbf{Y}, \mathbf{U}, L\mathbf{\Lambda}) P_{L\mathbf{\Lambda}}(L\mathbf{\Lambda}) d\mathbf{U} d(L\mathbf{\Lambda}) \\
&= \int_{\mathcal{U}(N)\times\mathbb{R}^{+N}} P_{\mathbf{Y}|\mathbf{\Sigma}\mathcal{H}_1}(\mathbf{Y}, \mathbf{U}, L\mathbf{\Lambda}) P_{\mathbf{\Lambda}}(\mathbf{\Lambda}) d\mathbf{U} d\mathbf{\Lambda}
\end{aligned} \tag{17}$$

For a precise demonstration of how this is obtained, refer to [7].

The integrands of (17) are given by both equations (14) and (15). The complete derivation requires recent tools from random matrix theory [6], among which the Harish-Chandra identity [14] which allows to integrate (17) over the space $\mathcal{U}(N)$ of unitary $N \times N$ matrices. The complete derivation is provided by the authors in an extended version of the current article [19]. Denoting J_k the integral

$$J_k(x, y) = \int_x^{+\infty} t^k e^{-t-\frac{y}{t}} dt \tag{18}$$

the final result expresses as equation (12).

2) Derivation of $P_{\mathbf{Y}|\mathcal{H}_i}$ in MIMO case: In the MIMO configuration, $P_{\mathbf{Y}|\mathcal{H}_0}$ remains unchanged and equation (11) is still correct. For the subsequent derivations, we only treat the situation when $M \leq N$ but the case $M > N$ is a trivial extension.

In this scenario, $\mathbf{H} \in \mathcal{M}(\mathbb{C}, N, M)$ is, as already mentioned, distributed as a Gaussian i.i.d. matrix according to the maximum entropy principle. The mean variance of every row is $\mathrm{E}[\sum_{j=1}^M |h_{ij}|^2] = 1$. Therefore $M\mathbf{H}\mathbf{H}^{\mathsf{H}}$ is distributed as a standard Wishart matrix [6]. Hence, observing that $\mathbf{\Sigma} - \sigma^2\mathbf{I}_N$ is the diagonal matrix of eigenvalues of $\mathbf{H}\mathbf{H}^{\mathsf{H}}$,

$$\mathbf{\Sigma} = \mathbf{U} \cdot \operatorname{diag}(\nu_1 + \sigma^2, \ldots, \nu_M + \sigma^2, \sigma^2, \ldots, \sigma^2) \cdot \mathbf{U}^{\mathsf{H}} \tag{19}$$

the *unordered* eigenvalue distribution of $\mathbf{\Lambda}$ can be derived [6]

$$P_{\mathbf{\Lambda}}(\lambda_1, \ldots, \lambda_M) d\mathbf{\Lambda} = d\lambda_1 \ldots d\lambda_M \frac{(N-M)! M^{MN}}{N!} \times$$
$$\prod_{i=1}^M e^{-M\sum_{i=1}^M(\lambda_i - \sigma^2)} \frac{(\lambda_i - \sigma^2)^{N-M}}{(M-i)!(N-i)!} \prod_{i<j}^M (\lambda_i - \lambda_j)^2 \tag{20}$$

which is defined on the set $\left\{\lambda_i | \lambda_i > \sigma^2,\ i \in \{1, \ldots, M\}\right\}$.

Similarly to the previous SIMO derivation, the integration (17) is carried out and ends up to the generalized MIMO result of equation (13) in which $\mathcal{P}(k)$ is the set of permutations of $\{1, \ldots, k\}$, $\operatorname{sgn}(\mathbf{b})$ the sign of the permutation $\mathbf{b}$ and

$$\alpha = \frac{(N-M)! M^{(2L-M+1)M/2}}{N! \pi^{NL} \sigma^{2(N-M)(L-M)} \prod_{j=1}^{M-1} j!} \tag{21}$$

The complete derivation of this solution is proposed in [19].

Decisions regarding the signal detection are then carried out by computing the ratio $C(\mathbf{Y})$ between equation (13) and equation (11).

$$P_{\mathbf{Y}|I_1}(\mathbf{Y}) = \frac{e^{\sigma^2 - \frac{1}{\sigma^2}\sum_{i=1}^N x_i}}{N\pi^{LN}\sigma^{2(N-1)(L-1)}} \sum_{l=1}^{N} \frac{e^{\frac{x_l}{\sigma^2}}}{\prod_{\substack{i=1\\ i\neq l}}^N (x_l - x_i)} J_{N-L-1}(\sigma^2, x_l) \tag{12}$$

$$P_{\mathbf{Y}|I_M}(\mathbf{Y}) = \alpha \cdot e^{M^2\sigma^2 - \frac{\sum_{i=1}^N x_i}{\sigma^2}} \sum_{\mathbf{a}\subset[1,N]} \frac{e^{\frac{\sum_{i=1}^M x_{a_i}}{\sigma^2}}}{\prod\limits_{a_i} \prod\limits_{\substack{j\neq a_1\\ \dots\\ j\neq a_i}} (x_{a_i} - x_j)} \sum_{\mathbf{b}\in\mathcal{P}(M)} (-1)^{\mathrm{sgn}(\mathbf{b})+1} \prod_{l=1}^{M} J_{N-L-2+b_l}(M\sigma^2, Mx_{a_i}) \tag{13}$$

B. Incomplete set of knowledge

1) Unknown SNR: Efficient signal detection when the noise level is unknown is highly desirable. Indeed, if the noise level were exactly known, some prior noise detection mechanism would be required. The difficulty here, is handily avoided thanks to *ad-hoc* methods that are independent of the noise level [10]-[11]. Instead, we shall consider some prior information about the noise level. Establishing prior information of variables defined in a continuum is still a controverted debate of the maximum entropy theory. However, a few solutions are classically considered that are based on *desirable* properties. Those are successively detailed in the following.

Two classical cases are usually encountered,

- the noise level is known to belong to a continuum $[\sigma_-^2,\ \sigma_+^2]$. If no more information is known, then it is desirable to take a uniform prior for σ^2 and then

$$P_{\sigma^2|I_M}(\sigma^2)d\sigma^2 = \frac{1}{\sigma_+^2 - \sigma_-^2} d\sigma^2 \tag{22}$$

  However, a questionable issue of invariance to variable change arises. Indeed, if $P_{\sigma^2|I_M}$ is uniform, $\sigma = \sqrt{\sigma^2}$ is not uniform. This old problem is partially answered by Jeffreys [12] who suggests that an *uninformative prior* should be any distribution that does not add information to the posterior distribution $P_{\sigma^2|\mathbf{Y},I_M}$ (for recent developments, see also [16]). However, in our problem, the uninformative prior is rather involved so we only consider uniform prior distribution (22) for σ^2 (we denote I'_M = "$\mathcal{H}_1, \sigma^2 \in [\sigma_-^2,\ \sigma_+^2]$") and therefore

$$P_{\mathbf{Y}|I'_M} = \frac{1}{\sigma_+^2 - \sigma_-^2} \int_{\sigma_-^2}^{\sigma_+^2} P_{\mathbf{Y}|\sigma^2, I'_M}(\mathbf{Y}, \sigma^2) d\sigma^2 \tag{23}$$

- one has no information concerning the noise power. The only information about σ^2 is $\sigma^2 > 0$. Again, we might want to subjugate σ^2 to Jeffreys' *uninformative prior*. However, computing this prior is again rather involved. The other alternative is to take the limit of (23) when σ_- tends to zero and σ_+ tends to infinity. This limiting process produces an improper integral form. This would be, with I''_M the updated background information,

$$P_{\mathbf{Y}|I''_1} = \lim_{x\to\infty} \frac{1}{x - \frac{1}{x}} \int_{\frac{1}{x}}^{x} P_{\mathbf{Y}|\sigma^2}(\mathbf{Y}, \sigma^2) d\sigma^2 \tag{24}$$

The computational difficulty raised by the integrals $J_k(x, y)$ does not allow for any satisfying closed-form formulas for

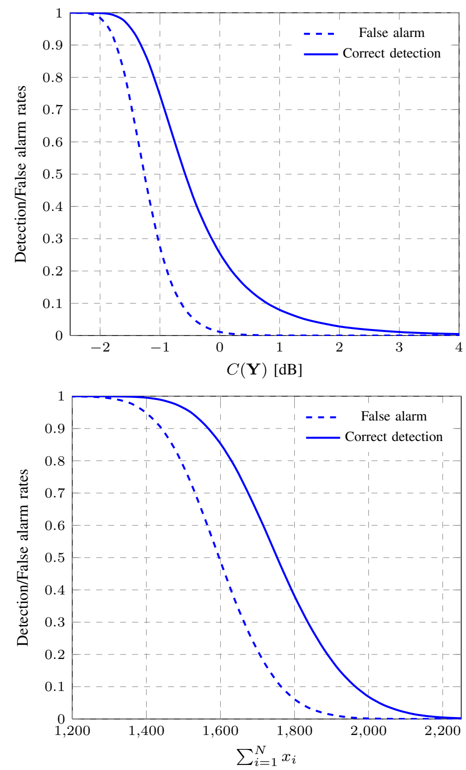


Fig. 1. Power detection performance in SIMO - $M = 1$, $N = 8$, $L = 20$, SNR $= -10$ dB

(23) and (24). In the following, we only consider the bounded continuum scenario.

C. Unknown M

In practical cases, the number of transmitting antennas is known to be finite. If only an upper bound value M_{max} for M is known, a uniform prior for M is brought by the maximum entropy principle and the probability distribution of $\mathbf{Y}$ under hypothesis I_0 which gathers all the system prior information (under $\mathcal{H}_0$ or $\mathcal{H}_1$), excluding the knowledge of M, reads

$$P(\mathbf{Y}|I_0) = \sum_{i=1}^{M_{max}} P(\mathbf{Y}|\text{“}M = i\text{”}, I_0) \cdot P(\text{“}M = i\text{”}|I_0) \tag{25}$$

$$= \frac{1}{M_{max}} \sum_{i=1}^{M_{max}} P(\mathbf{Y}|\text{“}M = i\text{”}, I_0) \tag{26}$$

which does not meet any computational difficulty.

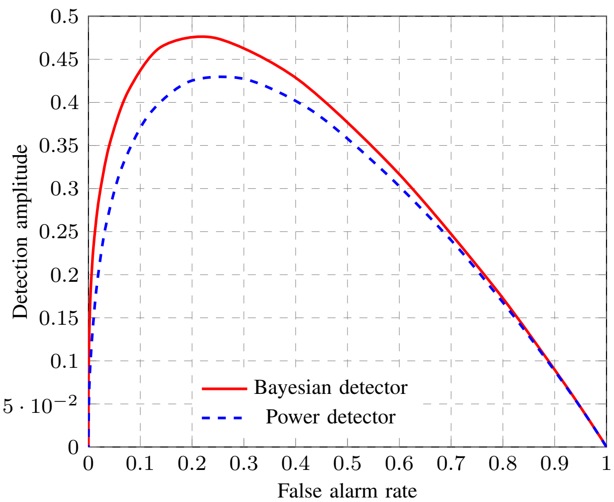


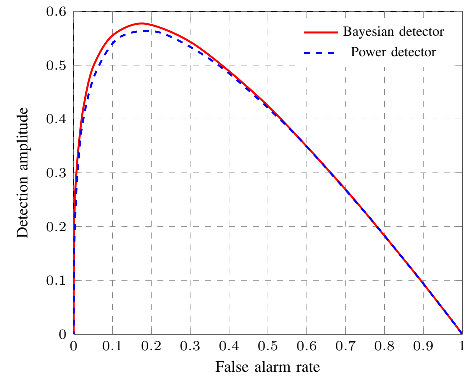
Fig. 2. Detection amplitude comparison in SIMO - $M = 1$, $N = 8$, $L = 20$, SNR $= -10$ dB

IV. Simulation and Results

In the following, we present results obtained for the aforementioned SIMO and MIMO scenarios, using formulas (12) and (13) respectively. In the simulations, the hypothesis concerning incoming data, channel aspect and noise figure are those presented in the model of section II.

As a first example, we consider a SIMO channel with $N = 8$ antennas at the receiver, $L = 20$ sampling periods and a signal to noise ratio SNR $= -10$ dB. For fair comparison with classical signal detection algorithms, we stick to the *false alarm rate* (FAR) against *correct detection rate* (CDR) performance evaluation. Figure 1 presents the respective FAR and CDR for the classical power detector and for the novel Bayesian estimator, obtained on $50,000$ channel realizations. The decision threshold for the power detector is somewhere around the total mean cumulated power over the antenna array while the threshold for the Bayesian approach is somewhere around $C(\mathbf{Y}) = 0$ dB. Since both algorithms scale very differently, fair comparison is obtained by plotting the *CDR minus FAR* gap (which is an objective performance criterion and that we call *detection amplitude*) against the FAR. This is depicted in figure 2. A significant performance gain is observed in this single transmit antenna scenario. This seems to imply that second order statistics of the incoming signal are far from bringing sufficient statistics to represent the complete information status I_1. This also demonstrates that the power detector is not in fact "optimal" in our Bayesian information-theoretic framework since the power detector is provided with no less information than the Bayesian detector.

In figure 3, we took $M = 2$, $N = 8$, $L = 10$ and SNR $= -10$ dB, using then formula (13). In this scenario the classical power detector closes in the gap with the Bayesian detector, compared to the SIMO situation. This is explained by the *channel hardening* effect [17] of multiple antenna systems that leaves little probability to deep fading channels. Those



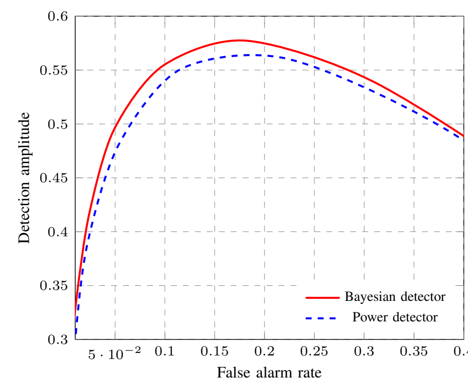


Fig. 3. Detection amplitude comparison in MIMO - $M = 2$, $N = 8$, $L = 10$, $SNR = -10$dB

deep fades, which are seen as absence of informative signal from the classical power detector can be correctly interpreted by the Bayesian detector. Therefore, for a given SNR, the more antennas are added to the system, the closer to optimal the power detector.

Consider now the scenario when the noise variance σ^2 is only known to belong to the interval $[\sigma_-^2, \sigma_+^2]$. The two-dimensional integration of equation (22) is prohibitive for producing numerical results. Nonetheless, the continuum of $[\sigma_-^2, \sigma_+^2]$ can be broken down in a finite number of K subsets $[\sigma_-^2 + k\Delta(\sigma^2), \sigma_-^2 + (k+1)\Delta(\sigma^2)]$, for $k \in \{0, \ldots, K-1\}$ and $\Delta(\sigma^2) = (\sigma_+^2 - \sigma_-^2)/K$. If $\Delta(\sigma^2)$ is chosen small enough, this should produce a rather good approximation of (22). This is experimented in figure **??** which demonstrates the effect of an inaccurate knowledge of the noise power in terms of CDR and FAR. In this simulation, $M = 1$, $N = 8$, $L = 10$ and SNR $= 2.5$ dB. Comparison is made between the cases of exact SNR knowledge, short SNR range $[\sigma_-^2, \sigma_+^2] = [-1.25, 6.75]$ dB discretized as a set $\{0, 2.5, 5\}$ dB and large SNR range $[\sigma_-^2, \sigma_+^2] = [-6.75, 6.75]$ dB discretized as a set $\{-5, -2.5, 0, 2.5, 5\}$ dB. While the short SNR range provides slightly poorer detection abilities than the perfect scenario, the large SNR range shows performance impairment. This suggests that, if the SNR range is totally unknown from the start, the first signal detection process (before information update [15]) does not lead to any valuable inference.

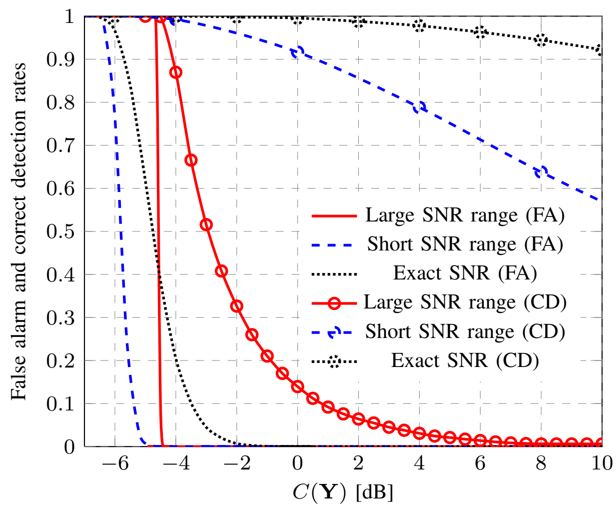


Fig. 4. False alarm (FA) and correct detection (CD) rates for exact, short discrete range ($\{0, 2.5, 5\}$) and large discrete range ($\{-5, -2.5, 0, 2.5, 5\}$) SNR - $M = 1$, $N = 8$, $L = 10$, SNR $= 2.5$ dB

V. Discussion

In the previous framework, we extensively used the maximum entropy principle and Bayes' rule in order to perform adequate signal detection. The provided Bayesian solutions, derived from the channel state of knowledge available at the receiver, are optimal in the proposed probability framework. However, some limitations can be raised.

First, as stated in III-B1, uninformative priors modeling is still an incomplete and controversial theory. When one faces parameters about which no information is known, one might come up with *ad-hoc* priors. However, if the quantity of data provided to the receiver is sufficient enough, then the effect of an approximate prior usually fades out. Also, the maximum entropy principle only deals with statistical knowledge and never deterministic knowledge. One must wait for future tools to be able to model, for instance, the deterministic knowledge of the presence of scatterers in the environment.

Also, it was noticed, already in this simple situation, that the marginalisation computation can easily turn out mathematically prohibitive. Especially, since random matrix theory [6] is not a mature theory yet, multi-dimensional problems still lead to a high computational complexity.

However, Jaynes' probability theory allows for an easily extensible and optimal framework in an information theoretic viewpoint. For instance, if interferers had to be taken into account in a more general model, then our previous derivations consist in the particular case of interferers of null power ($P_{int} = 0$). Integrating out the new variable P_{int} allows for the introduction of interfering sources in the model. This again increases the modeling complexity but provides at least some upper bound on the achievable performance.

Also, the present probability framework only allows to gather a fixed amount of information from which inference is performed to assign static probabilities. Dynamics in the system model are not easy to capture. Nonetheless, updating Bayesian probabilities given dynamic knowledge at the receiver is a recent and active research topic [15]; this would be appropriate for the cognitive receiver to assign time-varying probabilities. This is envisioned as one of the next fundamental steps in the characterization of cognitive receivers.

VI. Conclusion

In this work, we introduced a general Bayesian framework for learning in cognitive receivers. This framework is based on a consistent treatment of the available system information. Signal detection is treated as an explanatory case of this framework. The performance of the novel Bayesian signal detector in SIMO and MIMO systems are derived and are shown to outperform the classical detection techniques. We observed in particular that in a MIMO system with many antennas the classical energy detector performs close-to-optimally, while in SIMO setups, significant gain is provided by the Bayesian detector. Extensions to other frameworks than the signal detection one are being conducted.

References

[1] H. Urkowitz, "Energy detection of unknown deterministic signals," Proc. of the IEEE, vol. 55, no. 4, pp. 523-531, Apr. 1967

[2] Mitola III, J., and Maguire Jr, GQ, "Cognitive radio: making software radios more personal", Personal Communications, IEEE [see also IEEE Wireless Communications] 6(4), 13-18, 1999.

[3] V. I. Kostylev, "Energy detection of a signal with Random Amplitude", Proc IEEE Int. Conf. on Communications (ICC'02). New York City, pp. 1606-1610, May 2002.

[4] Simon, M.K., F. F. Digham, M.-S. Alouini, "On the Energy Detection of Unknown Signals over Fading Channels," ICC 2003 Conference Record, Anchorage, Alaska, May 11-15, 2003.

[5] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, "Spatial-spectral joint detection for wideband spectrum sensing in cognitive radio networks," Proc. ICASSP, Las Vegas, April 2008.

[6] A. M. Tulino, S. Verdú, "Random Matrix Theory and Wireless Communications", Now Publishers, vol. 1, Issue 1, 2004.

[7] M. Guillaud, M. Debbah, A. L. Moustakas, "Maximum Entropy MIMO Wireless Channel Models", Submitted to IEEE Trans. Information Theory Dec. 2006, http://arxiv.org/abs/cs.IT/0612101

[8] R. T. Cox, "Probability, frequency and reasonable expectation", American journal of physics, vol. 14 (1946), pp. 1-13.

[9] E. T. Jaynes, "Probability Theory: The Logic of Science", Cambridge University Press, June 2003.

[10] L. S. Cardoso, M. Debbah, P. Bianchi, J. Najim, "Cooperative Spectrum Sensing Using Random Matrix Theory", International Symposium on Wireless Pervasive Computing 2008, Santorini, Greece.

[11] Y. Zeng, Y.-Ch. Liang, "Eigenvalue based Spectrum Sensing Algorithms for Cognitive Radio", arXiv:0804.2960

[12] H. Jeffreys, "An Invariant Form for the Prior Probability in Estimation Problems", Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 186 (1007): 453-461, 1946.

[13] E. T. Jaynes, "On the rationale of maximum-entropy methods", Proc. IEEE, vol. 70, no. 9, pp. 939-952, 1982.

[14] A. B. Balantekin, "Character expansions, Itzykson-Zuber integrals, and the QCD partition function", *Phys. Rav. D,* vol. 62, no. 8, Oct. 2000.

[15] A. Caticha, "Lectures on Probability, Entropy and Statistical Physics", arXiv:0808.0012v1 [physics.data-an], 2008.

[16] A. Caticha, "Maximum Entropy, fluctuations and priors", A.C., in Maximum Entropy and Bayesian Methods in Science and Engineering, ed. by A. Mohammad-Djafari, A.I.P. Vol. 568, 94 (2001); arXiv.org/abs/math-ph/0008017.

[17] B. Hochwald, T. Marzetta, V. Tarokh, "Multiple-Antenna Channel Hardening and Its Implications for Rate Feedback and Scheduling", IEEE Trans. on Information Theory, vol. 50, no. 9, pp. 1893-1909, 2004.

[18] M. Mehta, "Random Matrices", Academic Press, 2004.

[19] R. Couillet, M. Debbah, "Multi-Antenna Cognitive Receivers and Signal Detection", *to be submitted to IEEE Trans. on Information Theory.*