Bayesian Inference for Multiple Antenna Cognitive Receivers
Romain Couillet, Merouane Debbah

To cite this version:
Romain Couillet, Merouane Debbah. Bayesian Inference for Multiple Antenna Cognitive Receivers. IEEE Wireless Communications and Networking Conference, Apr 2009, Budapest, Hungary. 6 p. hal-00392392

HAL Id: hal-00392392
https://hal-supelec.archives-ouvertes.fr/hal-00392392
Submitted on 7 Jun 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
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 required 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 on consistent estimation 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.

1by consistent we mean that two problems defined with the same amount of prior information should lead to the same final solution.
We note $(\cdot)^H$ the Hermitian transpose, $\text{tr}(\cdot)$ denotes the matrix trace. $\mathcal{M}(\mathbb{C}, N, M)$ is the set of matrices of size $N \times M$ over the algebra $\mathbb{C}$. $\mathcal{U}(N)$ is the set of unitary square matrices of size $N$. The notation $\mathcal{P}_X (Y)$ 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 $\sigma^2$ 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 $s^{(l)} = (s_1^{(l)}, \ldots, s_M^{(l)})^T \in \mathbb{C}^M$. The data vectors are stacked into the receive matrix $S = [s^{(1)}, \ldots, s^{(L)}]$. If the noise level $\sigma^2$ 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 $\Theta \in \mathcal{M}(\mathbb{C}, N, L)$ (i.e. a matrix with i.i.d. standard complex Gaussian entries $\theta_{ij}$) [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 $y^{(l)}$ that we stack, over the $L$ sampling periods, into the matrix $\mathbf{Y} = [y^{(1)}, \ldots, y^{(L)}] \in \mathcal{M}(\mathbb{C}, N, L)$.

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

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

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

And for $\mathcal{H}_1$ to

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

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

$$\Sigma = \mathbb{E}[\mathbf{YY}^H] = L \left( \mathbf{HH}^H + \sigma^2 \mathbf{I}_N \right)$$

$$= \mathbf{U} \left( \Lambda \right) \mathbf{U}^H$$

where $\Lambda = \text{diag} (\nu_1 + \sigma^2, \ldots, \nu_N + \sigma^2)$, with $\{\nu_i, i \in \{1, \ldots, N\}\}$ the eigenvalues of $\mathbf{HH}^H$ 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

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

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

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

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

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

reduces to a maximum likelihood criterion.

III. OPTIMAL SIGNAL DETECTION

A. Complete set of knowledge

1) Derivation of $P_{\mathcal{H}_1}|\mathbf{Y}$ in SIMO case: Let us analyze the situation when the noise level $\sigma^2$ (hypothesis V-ii)) and the number $M$ of transmit antennas are known to the receiver (hypothesis V-ii)) 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 \( P_{Y|\Theta_0} \): In this first scenario, \( \Theta \) is a Gaussian matrix with independent entries. The distribution for \( Y \), that can be seen as a random vector with \( NL \) entries, is then \( NL \) multivariate uncorrelated complex Gaussian with covariance \( \sigma^2 I_{NL} \),
\[
P_{Y|\Theta_0}(Y) = \frac{1}{(\pi \sigma^2)^{NL}} e^{-\frac{1}{2} \sigma^2 Y Y^H}
\] (9)
by denoting \( x = (x_1, \ldots, x_N)^T \) the eigenvalue distribution of \( YY^H \), (11) only depends on \( \sum_{i=1}^N x_i \),
\[
P_{Y|\Theta_0}(x) = \frac{1}{(\pi \sigma^2)^N} e^{-\frac{1}{2} \sum_{i=1}^N x_i}
\] (10)

b) Informative data likelihood \( P_{Y|\Theta_1} \): In scenario \( \Theta_1 \), the problem is more involved. The maximum entropy principle shows that our best guess is for \( H \) 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 \( H \) to be of unit mean power (i.e. \( \forall i, j \in [1,M] \), \(|h_{ij}|^2 = 1/M\)). Therefore, since \( M = 1, H \in \mathbb{C}^{N \times 1} \) and \( \Sigma = HH^H + \sigma^2 I_N \) has \( N-1 \) eigenvalues equal to \( \sigma^2 \) and another distinct eigenvalue \( \lambda_1 = \nu_1 + \sigma^2 = \left(\sum_{i=1}^N |h_{i1}|^2\right) + \sigma^2 \). The density of \( \lambda_1 - \sigma^2 \) is a complex \( \chi^2_N \) distribution (which is, up to a scaling factor 2, equivalent to a real \( \chi^2_{2N} \) distribution). Hence the uncondensed eigenvalue distribution of \( \Sigma \) [18]
\[
P_{\Lambda}(\Lambda)d\Lambda = \frac{1}{N(\lambda_1 - \sigma^2)^{N-1}} e^{-(\lambda_1 - \sigma^2)} \prod_{i=2}^N \delta(\lambda_i - \sigma^2) d\lambda_i \ldots d\lambda_N
\] (14)

Given model (2), for a fixed \( H \) channel, \( Y \) is distributed as a correlated Gaussian matrix,
\[
P_{Y|\Sigma H_1}(Y, U, LA) = \frac{1}{\pi^{NL} \det(L)^{1/2}} e^{-\frac{1}{2} (YY^H U U'^{-1} Y)^H}
\] (15)
where \( J_k \) denotes the prior information “\( \Theta_1 \) and \( M = k \)”. Since the channel \( H \) 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 \( \Sigma \) over the space of such covariance matrices
\[
P_{Y|\Theta_1}(Y) = \int_{\Sigma} P_{Y|\Sigma H_1}(Y, \Sigma) P_{\Sigma}(\Sigma)d\Sigma
\] (16)

Using the eigenvalue factorization (5), one can move from the space of covariance matrices \( \Sigma \) to the space of diagonal matrices \( \Lambda \), which is isomorphic to the real positive half-linig that carries \( \Lambda_1 \). Also, for any unitary matrix \( U \) and any standard i.i.d. Gaussian vector \( h \), the product \( U[h, \sigma^2 I_N] \) results in another matrix \( [h', \sigma^2 I_N] \) with \( h' \) standard Gaussian thanks to the unitary linear product. More generally, for any zero mean i.i.d. Gaussian vector \( h \), the set \( \{Uh, \ U \in \mathbb{U}(N)\} \) is uniformly distributed on the ensemble of zero mean i.i.d. Gaussian vectors and of variance \( hh^H \). This property leads to the independence of the respective distributions of \( U \) and \( \Lambda \) in (3), which implies
\[
P_{Y|\Theta_1}(Y) = \int_{\Sigma} P_{Y|\Sigma H_1}(Y, \Sigma) P_{\Sigma}(\Sigma)d\Sigma
\] (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 \( \mathbb{U}(N) \) 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
\[
J_k(x, y) = \int_{x}^{+\infty} t^k e^{-t - \frac{y}{2}} dt
\] (18)
the final result expresses as equation (12).

2) Derivation of \( P_{Y|\Theta_1} \) in MIMO case: In the MIMO configuration, \( P_{Y|\Theta_0} \) 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, \( H \in \mathbb{C}(N, M) \) 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 \( E[|h_{ij}|^2] = 1/M \). Therefore \( MHH^H \) is distributed as a standard Wishart matrix [6]. Hence, observing that \( \Sigma - \sigma^2 I_N \) is the diagonal matrix of eigenvalues of \( HH^H \),
\[
\Sigma = U \cdot \text{diag}(\nu_1 + \sigma^2, \ldots, \nu_M + \sigma^2, \sigma^2, \ldots, \sigma^2) \cdot U^H
\] (19)
the uncondensed eigenvalue distribution of \( \Lambda \) can be derived [6]
\[
P_{\Lambda}(\lambda_1, \ldots, \lambda_M)d\Lambda = d\lambda_1 \ldots d\lambda_M \frac{(N-M)!M^{MN}}{N!} \prod_{i=1}^M e^{-M \sum_{i=1}^M (\lambda_i - \sigma^2)} \frac{(\lambda_i - \sigma^2)^{N-M}}{(M-i)!(N-i)!} \prod_{i<j}^M (\lambda_i - \lambda_j)^2
\] (20)
which is defined on the set \( \{\lambda_i | \lambda_i > \sigma^2, i \in \{1, \ldots, M\}\} \). 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 \( \{1, \ldots, k\} \), \( \text{sgn}(b) \) the sign of the permutation \( b \) and
\[
\alpha = \frac{(N-M)!M^{2L-M+1}M/2}{N!\pi^NL \sigma^2 (N-M)(L-M)} \prod_{j=1}^{M-1} j!
\] (21)
The complete derivation of this solution is proposed in [19].

Decisions regarding the signal detection are then carried out by computing the ratio \( C(Y) \) between equation (13) and equation (11).
\[
P_{Y|I_1}(Y) = \frac{e^{2 - \frac{x_2}{\sigma^2}} \sum_{i=1}^{N} x_i}{N^L \sigma^2 (N-1)(L-1)} \sum_{l=1}^{N} \prod_{i \neq l}^{N} (x_l - x_i) J_{N-L-1}(\sigma^2, x_l)
\]

\[
P_{Y|I_M}(Y) = \alpha \cdot e^{M^2 \sigma^2 - \frac{\sum_{i=1}^{N} x_i^2}{\sigma^2}} \sum_{a \in [1,N]} \prod_{j \neq a}^{N} (x_{a_j} - x_j) \prod_{b \in Y(M)} (-1)^{\text{sgn}(b)+1} \prod_{l=1}^{M} J_{N-L-2+b}(M \sigma^2, M x_{a_i})
\]  

(12)

(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 was 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 \(\sigma^2\) and then

\[
P_{\sigma^2|I_M}(\sigma^2)\, d\sigma^2 = \frac{1}{\sigma^2_+ - \sigma^2_-} d\sigma^2
\]

(22)

However, a questionable issue of invariance to variable change arises. Indeed, if \(P_{\sigma^2|I_M}(\sigma^2)\) 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 \(P_{\sigma^2|Y,I_M}\) (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 \(\sigma^2\) (we denote \(I_M = \mathcal{H}_1, \sigma^2 \in [\sigma^2_-, \sigma^2_+]\)) and therefore

\[
P_{Y|I_M} = \frac{1}{\sigma^2_+ - \sigma^2_-} \int_{\sigma^2_-}^{\sigma^2_+} P_{Y|\sigma^2, I_M}(Y, \sigma^2)\, d\sigma^2
\]

(23)

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

\[
P_{Y|I_M''} = \lim_{x \to \infty} \lim_{x \to -\infty} \frac{1}{x} \int_{-\frac{1}{x}}^{\frac{1}{x}} P_{Y|\sigma^2, I_M}(Y, \sigma^2)\, d\sigma^2
\]

(24)

The computational difficulty raised by the integrals \(J(x, y)\) does not allow for any satisfying closed-form formulas for (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 \(M_{\text{max}}\) for \(M\) is known, a uniform prior for \(M\) is brought by the maximum entropy principle and the probability distribution of \(Y\) 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

\[
P(Y|I_0) = \sum_{i=1}^{M_{\text{max}}} P(Y|\text{"M = i"}, I_0) \cdot P(\text{"M = i"}|I_0)
\]

(25)

\[
= \frac{1}{M_{\text{max}}} \sum_{i=1}^{M_{\text{max}}} P(Y|\text{"M = i"}, I_0)
\]

(26)

which does not meet any computational difficulty.
that leaves little probability to deep fading channels. Those
detector, compared to the SIMO situation. This is explained by
classical power detector closes in the gap with the Bayesian

$$= \text{no less information than the Bayesian detector.}$$

theoretic framework since the power detector is provided with
those presented in the model of section II.

cerning incoming data, channel aspect and noise figure are
the Bayesian estimator, obtained on

$$= 2 \text{ (Bayesian detector)} \quad \text{SNR} = 10 \text{ dB}$$

Detection amplitude comparison in SIMO - $M = 1$, $N = 8$, $L = 20$, $\text{SNR} = -10 \text{ dB}$

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

IV. SIMULATION AND RESULTS

In the following, we present results obtained for the afore-
mentioned SIMO and MIMO scenarios, using formulas (12) and (13) respectively. In the simulations, the hypothesis con-
cerning 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 $\text{SNR} = -10 \text{ dB}$. 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 $C(Y) = 0 \text{ dB}$. 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
$\text{SNR} = -10 \text{ dB}$, 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
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 $\sigma^2$
is only known to belong to the interval $[\sigma_2^2, \sigma_4^2]$. The two-
dimensional integration of equation (22) is prohibitive for
producing numerical results. Nonetheless, the continuum of
$[\sigma_2^2, \sigma_4^2]$ can be broken down in a finite number of $K$
subsets $[\sigma_2^2 + k \Delta(\sigma^2), \sigma_2^2 + (k + 1) \Delta(\sigma^2)]$, for $k \in \{0, \ldots, K - 1\}$
and $\Delta(\sigma^2) = (\sigma_2^2 - \sigma_4^2)/K$. 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 $5\text{SNR} = 2.5 \text{ dB}$. Comparison is made
between the cases of exact SNR knowledge, short SNR range
$[\sigma_2^2, \sigma_4^2] = [-1.25, 6.75] \text{ dB}$ discretized as a set $\{0, 2.5, 5\} \text{ dB}$
and large SNR range $[\sigma_2^2, \sigma_4^2] = [-6.75, 6.75] \text{ dB}$ discretized
as a set $\{-5, -2.5, 0, 2.5, 5\} \text{ dB}$. 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. 3. Detection amplitude comparison in MIMO - $M = 2$, $N = 8$, $L = 10$, $\text{SNR} = -10 \text{ dB}$
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