Bayesian Foundations of Channel Estimation for Smart Radios

Romain Couillet
ST-Ericsson, Supélec
635 Route des Lucioles
06560 Sophia Antipolis, France
Email: romain.couillet@supelec.fr

Andrea Ancora
ST-Ericsson, Eurécom
635 Route des Lucioles
06560 Sophia Antipolis, France
Email: andrea.ancora@eurecom.fr

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

***Abstract*—In this paper, we revisit the philosophical foundations of the field of channel estimation. Our main intention is to come up with a partial answer to the question: "given some available sensed signals, how should cognitive radios ideally perform channel estimation?". We specifically introduce a general framework to provide optimal channel estimates under any prior knowledge at the sensing device. Our discussion is articulated as a top-down approach, introducing successively (i) a discussion on the philosophical foundations of channel estimation as a simplification means for the general problem of wireless detection, (ii) an information theoretically optimal approach to channel detection assuming the sensing device has infinite memory, and (iii) a derived optimal approach when limited memory size is accounted for. The key mathematical tools used in this discussion emerge from Bayesian probability theory and are known as the maximum entropy principle and the minimum update principle. Derivations are carried out for the particular case of channel estimation in orthogonal frequency division multiplexing (OFDM) systems. While some theoretical results will be proven to match already known techniques, such as Kalman filters, another set of novel results will be shown by simulations to perform better than known channel estimation schemes.**

I. Introduction

Channel estimation, along with most synchronization procedures, is itself a major and historical field of research in the realm of wireless communications . As such, thousands of novel channel estimators are proposed and compared to previous estimators every year. One of the obvious reasons for such an activity around channel estimation is that there does not exist a universal measure of performance to rate any given scheme with respect to any other; instead, several different selection criteria are considered, such as computational complexity, mean square error of the estimate, robustness against outage channel conditions, required processing memory etc. In this work, we wish to propose a framework to encompass the aforementioned selection conditions into a unique channel estimation framework. In some cases, we shall generate the optimal channel estimators for this framework. The sought for a general framework for channel estimation is motivated by the recent trend towards cognitive radios, and in particular by the trend towards developing terminal-centered intelligence in future wireless flexible networks. In the framework of cognitive radios, terminals are required to learn from their environment and take optimal decisions based on a constantly updated knowledge: performing optimal channel estimation from limited side information is part of the requirements demanded of cognitive radios.

The prior requirement for channel estimators is to help signal decoders as best as possible. The decision to consider a particular measure of performance for a given channel estimators should only reflect the subsequent effects on the eventual signal decoding process. For instance, minimizing the mean square error of channel estimators has no theoretical ground; it is merely a convenient mathematical way to compare estimators. In Section II, we shall briefly remind the foundations of signal decoding and channel estimation, which we shall review on a cognitive radio viewpoint. This study naturally follows the theoretical ideas introduced in [5], in which the authors introduce a new look on cognitive radios, and in [6], where the specific problem of blind source detection is addressed. Section II will conclude that today's theoretical and technological advances does not yet allow smart devices to perform optimal signal decoding without considering channel estimation as an independent entity.

In Section IV, we shall therefore treat channel estimation as a self-contained process, independent of the problem of signal detection, as is conventionally the case. We shall introduce a complete framework to derive optimal channel estimators under any prior state of knowledge at the signal receiver. While the conventional approach is to treat specific channel models and develop estimators for those models, we shall here instead consider prior knowledge about the environment at the receiver, and develop consistent estimators for this knowledge. We wish indeed to insist on the fact that smart devices should be able to come up with an ideal channel estimator for any given prior information on the channel. The approach addressed here is based on conventional Bayesian probability theory and on the maximum entropy principle [8]. In this section, we shall essentially remind the results originally derived in [7]. However, the practical finite memory size of the processing devices will not be taken into account in this section.

To answer the problem of optimal channel estimation under finite memory-size constraint, we shall subsequently introduce a novel aspect of the maximum entropy principle, known as the minimal update principle [9]. The major difference between both is the fact that the maximum entropy principle assumes an initial starting point and performs optimal decisions for data collected from this starting point on, while the minimal update principle assumes that one might be oblivious of old data and performs optimal decisions for a finite window of "remembrance". In our specific channel estimation considerations, this means that estimation based on the maximum entropy principle assume infinite memory at the receiving device, while the minimal update principle does only assume finite-time data recollection. From this novel approach, we derive optimal decisions, which will be shown in simulations to be often as good as those provided by the infinite memory size process.

Indeed, while Shannon [10] allows us to derive the capacity of a channel for which all synchronization parameters, plus the noise variance, are perfectly known, no such theory exists when the knowledge of some of these parameters is missing. More precisely, for a scalar communication $y = hx+n$, $x \in \mathcal{X}$, for some codebook $\mathcal{X}$, if h is unknown, then the maximum *a posteriori* estimate for x is

$$\hat{x} = \arg\max_{x\in\mathcal{X}} \int_h \int_n p(x|h,n)p(h)p(n)\mathrm{d}h\mathrm{d}n \tag{1}$$

which requires to have an *a priori* $p(h)$ for h. But this *a priori* is too impractical to obtain and would require to know all possible channel realizations and their respective probability. As a consequence to this strong difficulty, most contributions in the synchronization field have provided various empirical models based on field observations in order either to give an expression to $p(h)$ or, more practically, to propose good channel estimators $\hat{h}$ to h. Among those solutions, we mention [1] [2] [3] [4].

The difficulty of handling estimation problems when little side information is available is treated by Jaynes, through the Bayesian probability field, thanks to the maximum entropy principle [8]. However MaxEnt does not allow to perform updates of probability when new information, such as new pilots in the channel estimation problem, is available. In this case, the complete set of past symbols along with a prior distribution for the channel h_0 at time $t = 0$. This question was treated in [12] for the OFDM framework, when the channel delay spread, the channel time correlation and the signal-to-noise ratio (SNR) are alternatively known or unknown. When these parameters are not perfectly known, MaxEnt provides channel estimates minimizing the estimate mean square error (MMSE estimates) that outperform classical estimates which use empirical (often erroneous) models. Recent contributions in the Bayesian probability field enable one to perform probability updates, in particular based minimum cross entropy considerations [9]. In this work, we will then provide a channel estimation method, using the ME principle, which allows to assign probability distributions for the channel, when the estimator only knows the last past inferred channel distribution and the new received pilot symbols.

The remainder of this article unfolds as follows: in Section II, we discuss the foundations of channel estimation under a Bayesian point of view; in Section III, we introduce the OFDM model that shall be used as a toy example to illustrate in practice the theoretical ideas elaborated in the following sections. In Section IV, we discuss optimal infinite memory channel estimation, while in Section V, we introduce the minimal update principle and extend the previous optimality framework to finite-time memory channel estimation; also in this section, technical comparison is made against classical techniques. In Section VI, simulation and results are proposed, which compare the new method to the aforementioned classical algorithms. Finally, in Section VII, we draw our conclusions.

Notations: In the following, boldface lower case symbols represent vectors, capital boldface characters denote matrices ($\mathbf{I}_N$ is the $N\times N$ identity matrix). The transposition operation is denoted $(\cdot)^\mathsf{T}$. The Hermitian transpose is denoted $(\cdot)^\mathsf{H}$. The operator $\mathrm{diag}(\mathbf{x})$ turns the vector $\mathbf{x}$ into a diagonal matrix. The symbol $\det(\mathbf{X})$ is the determinant of matrix $\mathbf{X}$. The symbol $\mathrm{E}[\cdot]$ denotes expectation. The Kronecker delta function is denoted δ_x that equals 1 if $x = 0$ and equals 0 otherwise.

II. Foundations of Channel Estimation

In 1948, Shannon [10] provided the expression for the capacity of a communication channel between a transmitter and a receiver, modelled as

$$y = x + n \tag{2}$$

where $x \in \mathcal{X}$ is the input signal sent by the transmitter, which the receiver aims at recovering, n some additive noise process and y the effective signal captured by the receiver. In the conventional case when n and x are random variables with zero mean Gaussian distributions, the rate C to which the sequence of x can be decoded with infinitely low decoding error takes the simple form

$$C = \log\left(1 + \frac{\mathrm{E}[|x|^2]}{\mathrm{E}[|n|^2]}\right) \tag{3}$$

but Shannon does not provide any way to achieve such a decoding rate. Although, for appropriate coding schemes, it is possible to get an estimate $\hat{x}$ of x with as low decoding error rate as desired. The estimate $\hat{x}$ is based on the posterior probability $p(x|y)$ of any candidate $x \in \mathcal{X}$ given the output y,

$$p(x|y,I) = \int_n p(x|y,n,I)p(n|I)dn \tag{4}$$

where we denote by I all prior information known by the receiver at the moment it receives y.

In particular, we often take $\hat{x}$ to be the maximum likelihood estimator for x,

$$\hat{x} = \arg\max_{x\in\mathcal{X}} \int_n p(x|y,n,I)p(n|I)dn \tag{5}$$

which is easily derived for Gaussian n and I bringing no information to n, as

$$\hat{x} = \arg\min_{x\in\mathcal{X}} \|y - x\|^2 \tag{6}$$

When the signal x is filtered by a channel h, i.e. $y = hx+n$, the previous derivation is still valid, and we get the posterior probability $p(x|I)$ as,

$$p(x|y, I) = \int_h \int_n p(x|y, n, h, I)p(n|I)p(h|I)dndh \qquad (7)$$

If h is known, this boils down to a scaled version of the previous scenario, and the maximum likelihood estimator for Gaussian n and uninformative I becomes

$$\hat{x} = \arg\min_{x\in\mathcal{X}} \|y - hx\|^2 \qquad (8)$$

However, one rarely has access to the exact value for h and then the true maximum likelihood estimate for x is simply

$$\hat{x} = \arg\max_{x\in\mathcal{X}} \int_h \int_n p(x|y, n, h, I)p(n|I)p(h|I)dndh \qquad (9)$$

When one performs channel estimation, one gets some estimate $\hat{h}$ of the true channel h from previously received pilots, gathered in the information I. Classically, this estimate is then directly used in (9) by replacing the term $p(h|I)$ by $\delta(h - \hat{h})$. This substitution however constitutes a major mathematical flaw, unless some (possibly malevolent) genie ensured the receiver that the true channel is $\hat{h}$ with probability one. As such, the whole field of channel estimation has no information theoretical grounds. However, solving (9) in general is an extremely involved problem, which requires integration over all possible h channels. Note by the way that h might be a multi-dimensional vector channel, so that the integration over h might in truth be a multi-variate integral. It seems therefore natural to approximate the integration (9) by substituting $\hat{h}$ to $p(h|I)$. Or, more exactly, it seems natural to approximate (9) by replacing the function $h \mapsto p(h|I)$ by

$$h \mapsto \frac{\sum_{i=1}^{n} p(h_i|I)1_{h=h_i}(h)}{\sum_{i=1}^{n} p(h_i|I)} \qquad (10)$$

for a finite set of n candidates $h_1, \ldots, h_n$ with high probability.

The question of the choice of n and $h_1, \ldots, h_n$ is a rather involved problem, which, to the authors' knowledge has not yet been addressed. This is however not the purpose of the current work. Instead, we shall focus on the case $n = 1$, where h_1 is an estimate of the true channel h, based on cogent information provided from pilots and previous data, all captured in I. The question that now arises is: what measure should the channel estimator minimize? What are the grounds for performing minimum mean square error (MMSE), maximum likelihood (ML) estimation? The correct estimator for h should be that estimator $\hat{h}$, which is such that the posterior probability $p(x|y, h, I)$ when h is known is "close" to $p(x|y, \hat{h}, I)$ when h is unknown. Taking the classical Kullback-Liebler distance to compare distances between probability distributions, we may then consider

$$\hat{h} = \arg\min_{h_1} p(x|y, h_1, I) \log \frac{p(x|y, h, I)}{p(x|y, h_1, I)} \qquad (11)$$

The computational difficulty of the above expression however leads one to consider more tractable channel estimate minimization functionals, such as MMSE or ML channel estimators. The other difficulty arises from a consistent evaluation of $p(h|I)$ when I encompasses pilots, prior data and overall prior information about the channel at the receiver. It is indeed rather intricate to provide a mathematically sound description of I. The purpose of the subsequent sections will be to cast some light on a general Bayesian framework to evaluate $p(h|I)$, applied to the concrete case of channel estimators for OFDM systems based on pilots. The next section is dedicated to introducing the model for OFDM transmission channels.

III. Case study: OFDM channel model

Consider a single cell OFDM system with N subcarriers. The cyclic prefix (CP) length is N_{CP} samples. In the time-frequency OFDM symbol grid, pilots are found in the symbol positions indexed by the function $\phi_t(n) \in \{0, 1\}$ which equals 1 if a pilot symbol is present at subcarrier n, at symbol time index t, and 0 otherwise. We further denote $\mathbf{P}_t \in \mathbb{R}^{N\times N}$ a diagonal matrix with (i, i) entry $P_{t,ii} = \phi_t(i)$. The time-frequency grid is depicted in Figure 1. Both data and pilots at time t are modeled by the frequency-domain vector $\mathbf{s}_t \in \mathbb{C}^N$ with pilot entries of zero mean and amplitude $|s_{t,k}|^2 = 1$. The transmission channel is denoted $\mathbf{h}_t \in \mathbb{C}^N$ in the frequency-domain and is known only to have overall power 1. The additive noise is denoted $\mathbf{n}_t \in \mathbb{C}^N$ with entries known to have total variance σ^2. The time-domain representation of $\mathbf{h}_t$ is denoted $\boldsymbol{\nu}_t \in \mathbb{C}^L$ with L the channel length, i.e. the channel delay spread expressed in OFDM-sample unit. The frequency-domain received signal $\mathbf{y}_t \in \mathbb{C}^N$ is then

$$\mathbf{y}_t = \text{diag}(\mathbf{h}_\text{t})\mathbf{s}_\text{t} + \mathbf{n}_\text{t} \qquad (12)$$

We will also denote, $\forall k \in \{1, \ldots, N\}$, $h'_k = y_k/s_k = h_k + n_k/s_k$ and $\mathbf{h}' = (h'_1, \ldots, h'_N)^\mathsf{T}$ (here, the time index t is implicit).

The channel $\mathbf{h}_t$ evolves in time with coherence time function $\lambda(\tau)$ such that, independently of the channel delay spread index

$$\text{E}[\nu_{i,t}\nu^*_{i,t+\tau}] = \frac{\lambda(\tau)}{L} \qquad (13)$$

Along this study, we might consider the different system parameters, such as $\lambda(\tau)$ to be either exactly known at the receiver (and then fully part of the prior information I) or only partially known. In the following section, we establish, under partial or total knowledge of the different system parameters, an optimal framework for channel estimation. For the OFDM example, this section mainly recalls the results of [7].

IV. Maximum entropy channel estimation

The essential derivations of this section will consist in establishing, at time t, the posterior probability

$$p(\mathbf{h}_t|\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1, I) \qquad (14)$$

For readability we only treat the case $t = 2$, the general case being a trivial extension. Assume only pilots are transmitted

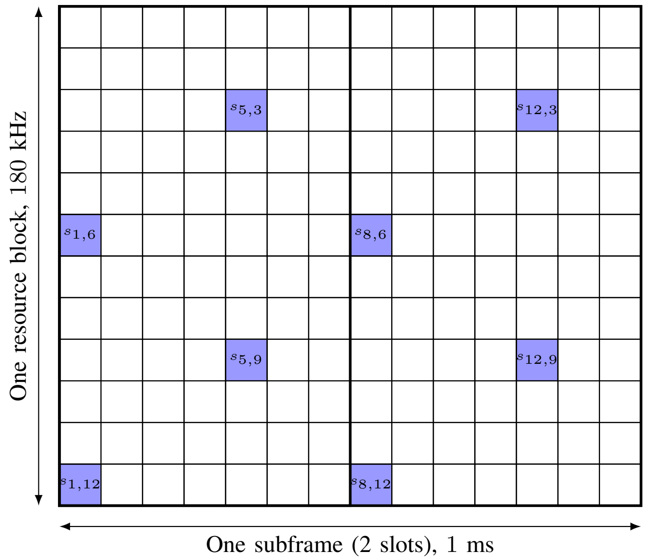

Fig. 1. Time-frequency OFDM grid with pilot positions enhanced

or, at least, that the information about I carried by the non-pilot signals are rather uninformative. Discarding the terms I for readability, we have in that case

$$
\begin{aligned}
&p(\mathbf{h}_2|\mathbf{y}_2,\mathbf{y}_1,I) \\
&= p(\mathbf{h}_2|\mathbf{h}'_2,\mathbf{h}'_1,I) && (15)\\
&= \frac{p(\mathbf{h}_2)p(\mathbf{h}'_2|\mathbf{h}_2)p(\mathbf{h}'_1|\mathbf{h}_2)}{p(\mathbf{h}'_1\mathbf{h}'_2)} && (16)\\
&= \frac{p(\mathbf{h}_2)p(\mathbf{h}'_2|\mathbf{h}_2)\int_{\mathbf{h}_1} p(\mathbf{h}'_1|\mathbf{h}_1)p(\mathbf{h}_1|\mathbf{h}_2)d\mathbf{h}_1}{p(\mathbf{h}'_1\mathbf{h}'_2)} && (17)
\end{aligned}
$$

When the exact time evolution model for $\mathbf{h}_t$ is known, $p(\mathbf{h}_1|\mathbf{h}_2)$ has an explicit expression which allows then to perform the above calculus. In practice however, it is rarely the case that such a time-evolution model be perfectly known. One then needs here an automatic method to provide a consistent expression for $p(\mathbf{h}_1|\mathbf{h}_2,I)$ when little is known about the interaction between $\mathbf{h}_1$ and $\mathbf{h}_2$. The method we shall use here is referred to as the *maximum entropy principle* [8].

Consider a given parameter x, whose knowledge is limited to the information contained in I. The maximum entropy principle allows one to assign a unique probability distribution $p(x|I)$ as follows,

1) among the set of all probability distributions, consider those distributions that satisfy the constraints about x given in I. This is, we exclude all distributions that do not satisfy the prior information I. The remaining set of probability distributions is denoted $\mathcal{Q}$.
2) in this remaining set of acceptable distributions $\mathcal{Q}$, $p(x|I)$ is assign the distribution which has maximum entropy, i.e.

$$p(x|I) = \arg\max_{q\in\mathcal{Q}} -\int q(y)\log q(y)dy \tag{18}$$

Choosing the distribution that has maximum entropy allows one not to make undesired assumptions on the unknown system variables, and as such allows one to remain neutral with regard to unaccessible information; see [11] for further details on the maximum entropy principle. In the model described in Section III, since only the total power is known about both the channel delay profile and the additive noise, both are assigned Gaussian distributions with zero mean and variance consistent with prior knowledge, as required by the maximum entropy principle. We then have $\mathbf{n}_t \sim \mathcal{CN}(0,\sigma^2\mathbf{I}_N)$ and $\nu_t \sim \mathcal{CN}(0,\frac{1}{L}\mathbf{I}_L)$, which in the frequency domain, applying Fourier transform, translates into $\mathbf{h}_t \sim \mathcal{CN}(0,\mathbf{Q})$, with $\mathbf{Q}$ defined as

$$Q_{nm} = \mathrm{E}\left[\sum_{k=0}^{L-1}\sum_{l=0}^{L-1}\nu_k\nu_l^* e^{-2\pi i\frac{kn-lm}{N}}\right] = \frac{1}{L}\sum_{k=0}^{L-1} e^{-2\pi ik\frac{n-m}{N}} \tag{19}$$

Note that $\mathbf{Q}$ is singular, since $L<N$ as OFDM requires.

If the coherence time function $\lambda(\tau)$ is perfectly known, the maximum entropy principle then assign to $p(\mathbf{h}_t|\mathbf{h}_{t+\tau})$ a Gaussian distribution of mean $\lambda(\tau)\mathbf{h}_{t+\tau}$ and variance

$$\mathrm{E}[\mathbf{h}_t\mathbf{h}_{t+\tau}^{\mathsf{H}}] = (1-\lambda(\tau)^2)\mathbf{Q} \tag{20}$$

In the problem with $t=2$, we denote $\lambda=\lambda(1)$. We then find that $p(\mathbf{h}_1|\mathbf{h}_2,I)$ is Gaussian and satisfies

$$p(\mathbf{h}_1|\mathbf{h}_2) = \lim_{\tilde{\mathbf{\Phi}}\to\mathbf{\Phi}} \frac{1}{\pi^N\det(\tilde{\mathbf{\Phi}})} e^{-(\mathbf{h}_1-\lambda\mathbf{h}_2)^{\mathsf{H}}\tilde{\mathbf{\Phi}}^{-1}(\mathbf{h}_1-\lambda\mathbf{h}_2)} \tag{21}$$

for $\{\tilde{\mathbf{\Phi}}\}$ any sequence converging to $\mathbf{\Phi}$. This allows then to compute the full expression of $p(\mathbf{h}_2|\mathbf{y}_2,\mathbf{y}_1,I)$ and a closed-form expression of some conventional estimators. In particular, the MMSE estimator $\hat{\mathbf{h}}_2^{(\mathrm{MMSE})}$ for $\mathbf{h}_2$, defined as

$$\hat{\mathbf{h}}_2^{(\mathrm{MMSE})} = \mathrm{E}\left[\mathbf{h}_2|\mathbf{y}\right] \tag{22}$$

in this case expresses as [7]

$$\hat{\mathbf{h}}_2^{(\mathrm{MMSE})} = \mathbf{M}_2^{-1}\left(\frac{\mathbf{P}_2\mathbf{h}'_2}{\sigma^2} + (\mathbf{I}_N + \frac{1-\lambda^2}{\sigma^2}\mathbf{P}_1\mathbf{Q})^{-1}\frac{\lambda}{\sigma^2}\mathbf{P}_1\mathbf{h}'_1\right) \tag{23}$$

with $\mathbf{M}_2$ satisfying

$$\mathbf{Q}\mathbf{M}_2 = \frac{\mathbf{I}_N}{1-\lambda^2} - \frac{\lambda^2}{1-\lambda^2}(\mathbf{I}_N + \frac{1-\lambda^2}{\sigma^2}\mathbf{Q}\mathbf{P}_1)^{-1} + \frac{\mathbf{Q}\mathbf{P}_2}{\sigma^2} \tag{24}$$

This expression generalizes to $t\geq 1$, for which we have the MMSE estimator $\hat{\mathbf{h}}_t^{(\mathrm{MMSE})}$ given in Equation (25).

However it often occurs that the assumption that L and $\lambda(1),\lambda(2),\ldots$ are known *a priori* at the receiver is not realistic. In truth, rather limited information is known *a priori* on these parameters. We may then reconsider (17) to include the uncertainty on L and/or $\lambda(1),\lambda(2),\ldots$. In the previous $t=2$ setup, we have in particular

$$
\begin{aligned}
\hat{\mathbf{h}}_2^{(\mathrm{MMSE})} &= \mathrm{E}[\mathbf{h}_2|\mathbf{y}_2,\mathbf{y}_1,I] && (26)\\
&= \int_\lambda p(\lambda|I)p(L|I)\mathrm{E}[\mathbf{h}_2|\mathbf{y}_2,\mathbf{y}_1,\lambda,L,I]d\lambda dL && (27)
\end{aligned}
$$

which leads to yet other expressions derived thoroughly in [7].

Through the OFDM example, we therefore developed a rather automatic method to derive consistent estimators under

$$\hat{\mathbf{h}}_t^{(\mathrm{MMSE})} = \left(\left(\left(1 + \sum_{k=1}^{t} \frac{\lambda(k)^2}{1-\lambda(k)^2}\right) \mathbf{I}_N - \sum_{k=1}^{t} \frac{\lambda(k)^2}{1-\lambda(k)^2} \left(\mathbf{I}_N + \frac{1-\lambda(k)^2}{\sigma^2} \mathbf{Q}\mathbf{P}_k \right)^{-1} \right)^{-1} \mathbf{Q} \left(\sum_{k=1}^{t} \lambda(k) \left(\mathbf{I}_N + \frac{1-\lambda(k)^2}{\sigma^2} \mathbf{P}_k \mathbf{Q} \right)^{-1} \frac{1}{\sigma^2} \mathbf{P}_k \mathbf{h}'_k \right) \tag{25}$$

any prior knowledge at the receiver which we claim optimal on an information theoretic viewpoint, i.e. those estimators are taking into consideration all prior information I and are made such that no *ad-hoc* assumption is taken regarding imperfectly known parameters, while being compliant with the Bayesian principles.

However, an underlying assumption of the previous approach is that infinite storage is available at the receiver. Indeed, for large t, we still need to consider all events from time instants 1 to t; if we decide to discard the oldest data, we then depart from the optimality of the proposed scheme. We then need to reconsider the whole framework to include an additional feature: the receiver is oblivious to part of the past events. The natural way to handle this modification of the current maximum entropy framework is to consider *updated* distribution assignments for posterior probabilities instead of *absolute* distribution assignments. This is the topic of the next and our main section.

V. Minimal channel estimation update

A. Introduction to Bayesian minimal update

When it comes to update probability assignments, Caticha proposes an extension of the maximum entropy principle, namely the minimum cross entropy principle (ME) [9]. When $p(\mathbf{h}_t|I_1)$ has been assigned for some side information I_1, and new cogent information I_2 is later available, then the ME principle consists in assigning to $p(\mathbf{h}_t|I_2)$ the distribution

$$p(\mathbf{h}_t|I_2) = \arg\min_q S[q, p(\mathbf{h}_t|I_1)] \tag{28}$$

where

$$S[q,p] = \int q(x) \log \left(\frac{p(x)}{q(x)} \right) dx \tag{29}$$

The functional $S[q,p]$ is referred to as the cross-entropy between the probability distributions q and p.

This method is based on a *minimal update* requirement, which in essence assigns to $p(\mathbf{h}_t|I_2)$ the unique distribution which minimizes the changes brought to $p(\mathbf{h}_t|I_1)$ while satisfying the new constraints given by I_2. In the following, additional side information on $\mathbf{h}_t$ (which possibly varies over time) will come from new available pilots at later time positions.

B. Perfect system parameters knowledge

We assume here that channel estimation is performed at different time instants $t = 1, 2, \ldots$. Denote I_k the knowledge at time k. Since memory restrictions impose to discard past received data, we decide here only to consider at time k the last received pilot data symbols, the last assigned probability $p(\mathbf{h}_k|I_{k-1})$, and the supposedly known time correlation $\lambda = \lambda(1)$ between the current channel $\mathbf{h}_k$ and the past channel $\mathbf{h}_{k-1}$.

Assume prior assigned distribution $p(\mathbf{h}_{k-1}|I_{k-1})$ at time index $k-1$. We have in general

$$p(\mathbf{h}_k|\mathbf{y}_k, I_k) \tag{30}$$
$$= \frac{p(\mathbf{y}_k|\mathbf{h}_k, I_k) \cdot p(\mathbf{h}_k|I_k)}{p(\mathbf{y}_k|I_k)} \tag{31}$$
$$= \frac{p(\mathbf{y}_k|\mathbf{h}_k, I_k) \cdot \int p(\mathbf{h}_k|\mathbf{h}_{k-1}, I_k) \cdot p(\mathbf{h}_{k-1}|I_k) \mathrm{d}\mathbf{h}_{k-1}}{p(\mathbf{y}_k|I_k)} \tag{32}$$

We use Caticha's ME principle [9] and set $p(\mathbf{h}_{k-1}|I_k)$ to the previous $p(\mathbf{h}_{k-1}|\mathbf{y}_{k-1}, I_{k-1})$. The reason lies in the *minimal update principle*: if no additional information is given in I_k, compared to I_{k-1}, then $p(\mathbf{h}_{k-1}|\mathbf{y}_{k-1}, I_{k-1})$ is the distribution q that minimizes the cross-entropy $S[q, p(\mathbf{h}_{k-1}|\mathbf{y}_{k-1}, I_{k-1})]$[1].

Let us now perform a recursive reasoning over the channel estimates at time indexes $k \in \mathbb{N}$. Assume that $p(\mathbf{h}_{k-1}|\mathbf{y}_{k-1}, I_{k-1})$ is Gaussian $\mathcal{CN}(\mathbf{k}_{k-1}, \mathbf{M}_{k-1})$. We will show that this implies $p(\mathbf{h}_k|\mathbf{y}_k, I_k)$ is still Gaussian. This will therefore be denoted $\mathcal{CN}(\mathbf{k}_k, \mathbf{M}_k^{-1})$. We have

$$p(\mathbf{h}_k|\mathbf{y}_k, I_k) \tag{33}$$
$$= \alpha_1 p(\mathbf{y}_k|\mathbf{h}_k, I_k) \cdot \int p(\mathbf{h}_k|\mathbf{h}_{k-1}, I_k) \cdot p(\mathbf{h}_{k-1}, I_k) \mathrm{d}\mathbf{h}_{k-1} \tag{34}$$
$$= \lim_{\tilde{\mathbf{Q}} \to \mathbf{Q}} e^{(\mathbf{h}_k - \mathbf{h}'_k)^{\mathsf{H}} \frac{\mathbf{P}_k}{\sigma^2} (\mathbf{h}_k - \mathbf{h}'_k)} \int e^{(\mathbf{h}_k - \lambda \mathbf{h}_{k-1})^{\mathsf{H}} \frac{\tilde{\mathbf{Q}}^{-1}}{1-\lambda^2} (\mathbf{h}_k - \lambda \mathbf{h}_{k-1})}$$
$$\times \alpha_2 e^{(\mathbf{h}_{k-1} - \mathbf{k}_{k-1})^{\mathsf{H}} \mathbf{M}_{k-1}^{-1} (\mathbf{h}_{k-1} - \mathbf{k}_{k-1})} \mathrm{d}\mathbf{h}_{k-1} \tag{35}$$

where the $\tilde{\mathbf{Q}}$'s are taken from a set of invertible matrices in the neighborhood of $\mathbf{Q}$, and the α_i's are constants.

First we need to write the exponents of the Gaussian products in the integrand in a single Gaussian exponent form of the vector $\mathbf{h}_{k-1}$ times a constant independent of $\mathbf{h}_{k-1}$. By expansion and simplification, this is

$$(\mathbf{h}_k - \lambda \mathbf{h}_{k-1})^{\mathsf{H}} \frac{\tilde{\mathbf{Q}}^{-1}}{1-\lambda^2} (\mathbf{h}_k - \lambda \mathbf{h}_{k-1})$$
$$+ (\mathbf{h}_{k-1} - \mathbf{k}_{k-1})^{\mathsf{H}} \mathbf{M}_{k-1}^{-1} (\mathbf{h}_{k-1} - \mathbf{k}_{k-1})$$
$$= (\mathbf{h}_{k-1} - \mathbf{l})^{\mathsf{H}} \mathbf{N} (\mathbf{h}_{k-1} - \mathbf{l}) + C(\mathbf{h}_k) \tag{36}$$

[1] if new statistical information comes in, the minimum cross-entropy distribution with $p(\mathbf{h}_{k-1}|I_k)$ *satisfying those new constraints* should be computed and used instead.

with

$$\begin{cases} \mathbf{N} & = \frac{\lambda^2\tilde{\mathbf{Q}}^{-1}}{1-\lambda^2} + \mathbf{M}_{k-1}^{-1} \\ \mathbf{l} & = \mathbf{N}^{-1}(\frac{\lambda}{1-\lambda^2}\tilde{\mathbf{Q}}^{-1}\mathbf{h}_k + \mathbf{M}_{k-1}^{-1}\mathbf{k}_{k-1}) \\ C & = \mathbf{h}_k^{\mathsf{H}}\frac{\tilde{\mathbf{Q}}^{-1}}{1-\lambda^2}\mathbf{h}_k + \mathbf{k}_{k-1}^{\mathsf{H}}\mathbf{M}_{k-1}^{-1}\mathbf{k}_{k-1} - \\ & (\mathbf{h}_k + \frac{1-\lambda^2}{\lambda}\tilde{\mathbf{Q}}\mathbf{M}_{k-1}^{-1}\mathbf{k}_{k-1})^{\mathsf{H}}\frac{\lambda^2}{(1-\lambda^2)^2}\tilde{\mathbf{Q}}^{-1} \\ & \left(\frac{\lambda^2\tilde{\mathbf{Q}}^{-1}}{1-\lambda^2} + \mathbf{M}_{k-1}^{-1}\right)^{-1}\tilde{\mathbf{Q}}^{-1}(\mathbf{h}_k + \frac{1-\lambda^2}{\lambda}\tilde{\mathbf{Q}}\mathbf{M}_{k-1}^{-1}\mathbf{k}_{k-1}) \end{cases}$$

The integral (35) is then a constant times e^C, which depends on $\mathbf{h}_k$. The term C must then be written again into a quadratic expression of $\mathbf{h}_k$. This is

$$C = (\mathbf{h}_k - \mathbf{j})^{\mathsf{H}}\mathbf{R}(\mathbf{h}_k - \mathbf{j}) + B \tag{37}$$

with

$$\begin{cases} \mathbf{R} & = \left(\lambda^2\mathbf{M}_{k-1} + (1-\lambda^2)\tilde{\mathbf{Q}}\right)^{-1} \\ \mathbf{j} & = \lambda\mathbf{k}_{k-1} \\ B & = 0 \end{cases} \tag{38}$$

Together with the term outside the integral (35), this is

$$p(\mathbf{h}_k|\mathbf{y}_k, I_k) = \alpha \cdot e^{(\mathbf{h}_k-\mathbf{k}_k)^{\mathsf{H}}\mathbf{M}_k^{-1}(\mathbf{h}_k-\mathbf{k}_k)} \tag{39}$$

with $\alpha = (\int p(\mathbf{h}_k|\mathbf{y}_k, I_k)\mathrm{d}\mathbf{h}_k)^{-1}$. Finally, after some arithmetic derivation, in the limit $\tilde{\mathbf{Q}} \to \mathbf{Q}$,

$$\begin{cases} \mathbf{M}_k & = \lambda^2\left(\mathbf{M}_{k-1} + \frac{1-\lambda^2}{\lambda^2}\mathbf{Q}\right) \\ & \quad \times \left(\frac{\lambda^2}{\sigma^2}\mathbf{P}_k(\mathbf{M}_{k-1} + \frac{1-\lambda^2}{\lambda^2}\mathbf{Q}) + \mathbf{I}_N\right)^{-1} \\ \mathbf{k}_k & = \lambda\mathbf{k}_{k-1} + \frac{1}{\sigma^2}\mathbf{M}_k\mathbf{P}_k(\mathbf{h}'_k - \lambda\mathbf{k}_{k-1}) \end{cases} \tag{40}$$

And the MMSE estimator $\hat{\mathbf{h}}_k$ for the channel at time index k is the first order moment of a Gaussian distribution centered in $\mathbf{k}_k$, which is $\hat{\mathbf{h}}_k = \mathbf{k}_k$. At initial time instant t_0, if nothing but the channel delay spread L is known, $\mathbf{M}_0 = \mathbf{Q}$ from the maximum entropy principle, as shown in [12], and $\mathbf{k}_0 = 0$. Therefore, we prove by the above recursion that, under this state of initial knowledge, for all $k \in \mathbb{N}$, $p(\mathbf{h}_k|I_k)$ is Gaussian with mean $\mathbf{k}_k$ and variance $\mathbf{M}_k$, and $\hat{\mathbf{h}}_k = \mathbf{k}_k$. Note that, while regularized inverses of $\mathbf{Q}$ were used along the derivations, the final formulas are properly conditioned with respect to $\mathbf{Q}$.

Note that the solution in (40) coincides with the classical structure of adaptive filters [16] such as the Kalman filter [15], for the problem of dynamic estimation of $\mathbf{h}_k$ from the observed $\mathbf{y}_k$, when $\mathbf{h}_k$ is assumed to evolve in time as an order-1 auto-regressive model. As such, Kalman filters are compliant with our current framework and are developped in section V-D for the sake of comparison. However, when some system parameters are not perfectly known, Kalman filters depart from our general minimal update framework as it will be further detailed.

C. Imperfect system parameter knowledge

In practical applications, contrary to what was stated in Section V-B, the different parameters λ, σ^2 and L especially, are not perfectly known. For simplicity we assume those parameters are constant over the duration of the channel estimation process. Following the maximum entropy principle, these parameters must be assigned an *a priori* distribution. Let us focus on the time correlation λ, which is typically the most difficult parameter to track. In this respect, one has

$$p(\mathbf{h}_k|\mathbf{y}_k, I_k) = \int p(\mathbf{h}_k|\mathbf{y}_k, \lambda, I_k)p(\lambda|\mathbf{y}_k, I_k)\mathrm{d}\lambda \tag{41}$$

Since $\mathbf{y}_k$ cannot bring alone any cogent information on λ, $p(\lambda|\mathbf{y}_k, I_k) = p(\lambda|I_k)$. The probability $p(\mathbf{h}_k|\mathbf{y}_k, \lambda, I_k)$ was computed in Section V-B and is given by the right-hand side of Equation (39), in which α depends on λ and must therefore be made explicit.

Further computation leads to

$$p(\mathbf{h}_k|\mathbf{y}_k, I_k) = \int p(\lambda|I_k)\alpha(\lambda)e^{(\mathbf{h}_k-\mathbf{k}_k^{(\lambda)})^{\mathsf{H}}(\mathbf{M}_k^{(\lambda)})^{-1}(\mathbf{h}_k-\mathbf{k}_k^{(\lambda)})}\mathrm{d}\lambda \tag{42}$$

with $\mathbf{M}_k^{(\lambda)}$ and $\mathbf{k}_k^{(\lambda)}$ given by Equation (40) for the λ in question and

$$\alpha(\lambda) = \beta e^{-x(\lambda)}\det[\mathbf{X}(\lambda)] \tag{43}$$

with

$$\begin{cases} \mathbf{X}(\lambda) & = \left(\mathbf{I} + \frac{\mathbf{P}_k}{\sigma^2}(\lambda^2\mathbf{M}_{k-1}^{(\lambda)} + (1-\lambda^2)\mathbf{Q})\right)^{-1} \\ x(\lambda) & = (\lambda\mathbf{k}_{k-1}^{(\lambda)} - \mathbf{h}'_k)^{\mathsf{H}}\mathbf{X}(\lambda)\frac{\mathbf{P}_k}{\sigma^2}(\lambda\mathbf{k}_{k-1}^{(\lambda)} - \mathbf{h}'_k) \end{cases} \tag{44}$$

and $\beta = (\int p(\mathbf{h}_k|\mathbf{y}_k, I_k)\mathrm{d}\mathbf{h}_k)^{-1}$, independent of λ.

In particular, the conventional MMSE estimate $\hat{\mathbf{h}}_k^{(\mathrm{MMSE})}$ is then the weighted sum

$$\hat{\mathbf{h}}_k^{(\mathrm{MMSE})} = \frac{\int p(\lambda|I_k)e^{-x(\lambda)}\det[\mathbf{X}(\lambda)]\mathbf{k}_k^{(\lambda)}\mathrm{d}\lambda}{\int p(\lambda|I_k)e^{-x(\lambda)}\det[\mathbf{X}(\lambda)]\mathrm{d}\lambda} \tag{45}$$

This integral is however very involved. In practice, it must be broken into a finite sum over a set of potential values for λ. Denoting $\mathcal{S}$ this set and $|\mathcal{S}|$ its cardinality, the recursive algorithm that provides the successive estimates $\hat{\mathbf{h}}_k$, $k = 1, \ldots, K$, requires that at every step, the values for $\mathbf{M}_k^{(\lambda)}$ and $\mathbf{k}_k^{(\lambda)}$, $\lambda \in \mathcal{S}$ are kept in memory.

Note that the MMSE estimator (45) is no longer linear in $\mathbf{h}'_k$ and, as such, does no longer enter the conventional linear Kalman filters. In the next section, we propose simulation results for the proposed minimum update channel estimators, and compare them to maximum entropy channel estimators derived in [7]. We study hereafter classical approaches and show how they differ from or are special cases of our derived techniques.

D. Comparison with classical channel estimation techniques

The channel estimation problem is related to the channel model assumed, mainly determined by the electromagnetic propagation characteristics of the wireless transmission such as transmission bandwidth, carrier frequency, relative speed and spatial configuration of the propagation environment which itself rules the multipath.

These conditions characterize the channel correlation function in a two-dimensional space comprising frequency and time domains. In the general case, each multipath channel component can experience different but related spatial scattering conditions leading to a full bi-dimensional correlation function across these domains.

Nevertheless, classical Clarke and Jakes derivations [18] [19] are based on the assumption that the physical scattering environment is chaotic and therefore the angle of arrival of the electromagnetic wave at the receiver is a uniformly distributed random variable in the angular domain. As a consequence, the time-correlation function is strictly real-valued and governed by the well known expression $r_\nu(\Delta t) = J_0(2\pi f_d \Delta t)$ where J_0 is the zero$^{\text{th}}$-order Bessel function. In addition, the Doppler spectrum is symmetric and interestingly there is a delay-temporal separability property in the general bi-dimensional scattering function.

Under this light, the Wide-Sense Stationary Uncorrelated Scattering (WSSUS) channel model has been proposed [17] and commonly employed for the multipath channels experienced in mobile communications.

This framework might be suboptimal in general. For example, when the mobile is moving in a fixed and known direction, as for example in rural or suburban areas, the WSSUS model would be non applicable. Instead, it can be considered to be separable when the direction of motion averages out because each multipath component is the result of omnidirectional scattering from objects surrounding the mobile, as one would expect in urban and indoor propagation scenarios. Separability is a very important assumption for reducing the complexity of channel estimation, allowing the problem to be separated into two one-dimensional operations.

Hence, for the sake of comparison with the methods proposed in previous paragraphs and which do not make the separability assumption, the channel can be estimated using a two step approach. First, pilot sub-carriers are used to estimate the whole channel impulse response (CIR) performing frequency-smoothing on each OFDM symbol where pilots are present. Secondly, the smoothed impulse response functions corresponding to a set of of OFDM symbols is used in order to improve the channel transfer (CTF) function estimate at the symbol of interest.

Even-though TD filtering could be applied remaining in the frequency domain to CTF estimates rather than to the CIR in the time domain because of the linearity relationship between the two, we prefer this option to limit the complexity of the operation.

Thus, for the first step, the Frequency-Domain (FD) optimal MMSE estimator under the assumption of uniform channel power delay profile of known length L is linear and given by

$$\hat{\boldsymbol{\nu}}_k^{(FD)} = (\mathbf{F}_L^{\mathsf{H}} \mathbf{P}_k^{\mathsf{H}} \mathbf{P}_k \mathbf{F}_L + \sigma^2 \mathbf{I}_L)^{-1} \mathbf{F}_L^{\mathsf{H}} \mathbf{P}_k^{\mathsf{H}} \mathbf{h}'_k \qquad (46)$$

For the second step, Time-Domain (TD) filtering to exploit time correlation with the channel at previous OFDM symbols containing pilots can be approximated in the form of a finite impulse response filter.

The channel CIR at the l^{th} tap position and at time instant k is estimated as

$$\hat{\nu}_{l,k} = \mathbf{w}_l^{\mathsf{H}} \hat{\boldsymbol{\nu}}_{l,k}^M \qquad (47)$$

where we exploit the vector $\hat{\boldsymbol{\nu}}_{l,k}^M = [\hat{\nu}_{l,k}^{(FD)}, \ldots, \hat{\nu}_{l,k-M+1}^{(FD)}]^{\mathrm{T}}$ of length M of l-th tap estimates across M time instants.

Finite length filter approximation seems reasonable as the correlation between consecutive symbols decreases as the terminal speed increases. The fact that the TD correlation is inversely proportional to the terminal speed sets a limit on the possibilities for TD filtering in high-mobility conditions.

The statistical TD filter which is optimal in terms of Mean Square Error (MSE) [16] is the $M \times 1$ vector $\mathbf{w}_l$ given by

$$\mathbf{w}_l = (\mathbf{R}_{\nu_k} + \sigma^2 \mathbf{I})^{-1} \mathbf{r}_{\nu_k} \qquad (48)$$

where $\mathbf{R}_{\nu_k} = \mathrm{E}[\boldsymbol{\nu}_{l,k}^M (\boldsymbol{\nu}_{l,k}^M)^{\mathrm{H}}]$ is the l^{th} channel tap $M \times M$ correlation matrix and $\mathbf{r}_{\nu_k} = \mathrm{E}[\boldsymbol{\nu}_l^M \nu_{l,n}^*]$ the $M \times 1$ correlation vector between the l^{th} tap of the current channel tap realization and M previous realizations including the current one.

In practical cases, the FIR filter length M is dimensioned according to a performance-complexity trade-off as a function of the terminal speed.

As an alternative to FIR TD-MMSE channel smoothing coefficient computation, an adaptive estimation approach can be considered which does not require knowledge of second-order statistics of both channel and noise. A feasible solution is the Normalized Least-Mean-Square (NLMS) estimator.

It can be expressed exactly as in Equation (47) but with the $M \times 1$ vector of filter coefficients $\mathbf{w}$ updated according to

$$\mathbf{w}_{l,k} = \mathbf{w}_{l,k-1} + \mathbf{u}_{l,k-1} e_{l,k} \qquad (49)$$

where M here denotes the NLMS filter order. The $M \times 1$ update gain vector is computed according to the well-known NLMS adaption

$$\mathbf{u}_{l,k} = \frac{\mu}{\|\hat{\boldsymbol{\nu}}_{l,k}^M\|^2} \hat{\boldsymbol{\nu}}_{l,k}^M \qquad (50)$$

where μ is an appropriately-chosen step adaptation and

$$e_{l,k} = \hat{\nu}_{l,k} - \hat{\nu}_{l,k-1} \qquad (51)$$

It can be observed that the TD-NLMS estimator requires much lower complexity compared to TD-MMSE as no matrix inversion is required, as well as not requiring any a priori statistical knowledge.

Finally, using both MMSE or NLMS approaches, the CTF channel estimate at k^{th} symbol can then be retrieved by

$$\hat{\mathbf{h}}_k = \mathbf{F}_L \hat{\boldsymbol{\nu}}_k \qquad (52)$$

When the channel $\mathbf{h}_k$ is modelled in a similar manner as previous paragraph, i.e. the channel evolution across time is expressed by the following state-space model

$$\begin{aligned} \mathbf{h}_k &= \lambda \mathbf{h}_{k-1} + \sqrt{1-\lambda^2} \mathbf{w}_k \\ \mathbf{h}'_k &= \mathbf{h}_k + \mathbf{n}_k \end{aligned} \qquad (53)$$

where $\mathbf{w}_k \sim \mathcal{CN}(0, \mathbf{Q})$ is known as the channel innovation term and $\mathbf{n}_k \sim \mathcal{CN}(0, \sigma^2 I_N)$ is the additive white Gaussian noise on pilot (and data) symbols. It is to be noted that this 1^{st} order auto-regressive model still complies with the statistical assumption on $\mathbf{h}_k \sim \mathcal{CN}(0, \mathbf{Q})$ made so far.

Under these assumptions, one can easily come up with the expression of a channel estimator according to the classical Kalman form [15]. In fact, letting $(\mathbf{F}_L)_{nm} = \mathrm{e}^{-2\pi \mathrm{i} \frac{nm}{N}}$ with $0 \leq n \leq \mathrm{N}-1$ and $0 \leq m \leq \mathrm{L}-1$, this can be written as

$$\begin{cases} \mathbf{M}_k &= \mathbf{F}_L \left(\lambda^2 \mathbf{C}_{k-1} + \frac{1-\lambda^2}{L}\mathbf{I}_L\right) \mathbf{F}_L^{\mathsf{H}} \mathbf{P}_k^{\mathsf{H}} \\ & \quad \times \left(\mathbf{P}_k \mathbf{F}_L \left(\lambda^2 \mathbf{C}_{k-1} + \frac{1-\lambda^2}{L}\mathbf{I}_L\right) \mathbf{F}_L^{\mathsf{H}} \mathbf{P}_k^{\mathsf{H}} + \sigma^2 \mathbf{I}_N\right)^{-1} \\ \mathbf{k}_k &= \lambda \mathbf{k}_{k-1} + \mathbf{M}_k \mathbf{P}_k (\mathbf{h}'_k - \lambda \mathbf{k}_{k-1}) \\ \mathbf{C}_k &= (\mathbf{I}_L - \mathbf{K}_k \mathbf{P}_k \mathbf{F}_L) \left(\lambda^2 \mathbf{C}_{k-1} + \frac{1-\lambda^2}{L}\mathbf{I}_L\right) \end{cases} \tag{54}$$

Interestingly, as previously pointed out in Section V-B, Equation (54) is somewhat consistent with (40) derived using the minimal update approach. Nevertheless, both expressions differ in the adaption mechanism which, in the case of the Kalman estimation algorithm, relies on the Kalman-gain $\mathbf{M}_k$ and the error estimate covariance matrix $\mathbf{C}_k$ update. Notice importantly that the estimation process assumes the knowledge of noise statistics σ^2 and the channel length L. In case the latter is not provided, it would be necessary to assume L as the largest channel length allowed by the OFDM system parameters in use. Therefore in practice, one should dimension it as $L = \lfloor N/M \rfloor$ [2] in case of imperfect knowledge. In spite of these limitations, the Kalman estimator is often chosen because of the well know robustness against non-stationarity of the signal statistics via the adaptation of the estimate covariance matrix. In order to counter the intrinsic need of parameter knowledge, one could think of using Expectation Maximization in conjunction with Kalman or plain MMSE techniques. Indeed, with an additional complexity cost, any of such channel estimator can be coupled with parameter estimation (speed or channel length) in an iterative fashion. Nevertheless, contrary to the original methods presented based on Maximum Entropy principle and then constructed to be robust with respect to parameter knowledge, they would need the necessary amount of data to converge to construct the correct a-priori information. Such methods are then well suited only in those cases where the channel is stationary.

Note that other classical adaptive estimators such as (normalized) least mean squares and recursive least squares, that discard most *a priori* knowledge, perform much less accurately than optimal 2-D MMSE optimal filter [16].

VI. Simulation and Results

In this section, we provide simulation plots to compare, at time t, the minimal channel estimation update method against (i) the one-dimensional MMSE [3], [2], taken at time t, which takes only into account the last past pilot symbols and uses a fixed empirical covariance matrix, (ii) the optimum two-dimensional MMSE provided in [12], (iii) the 1D+1D optimum MMSE, (iii) the 1D+NLMS and (v) the Kalman provided as reference in the section of classical channel estimation techniques, with $K = 4$ pilot time indexes. The OFDM DFT size is $N = 64$, the channel length $L = 6$ is known to the receiver, the vehicular speed is $v = 120$ km/h, pilot sequences are transmitted every 0.29 ms (as in 3GPP-LTE [14]), and the induced Jake's time correlation λ between t and the past pilot sequence arrival time is known to the receiver. In scenario (ii), all K past received pilot sequences and time correlations are perfectly known. The channel time correlation model is a K-order autoregressive model following [13]. A performance comparison is proposed in Figure 2. We notice here that the minimal update algorithm does not show significant performance decay compared to the optimal two-dimensional MMSE estimator, while the one-dimensional MMSE estimator, also relying on the last past pilot sequence, shows large performance impairment. Kalman estimation shows to be comparable in performance only when the channel length parameter is perfectly known but heavily impaired when the maximum channel length assumption is taken instead. The 1D+1D MMSE shows to behave exactly as optimal 2D as well, only when perfect knowledge of parameters is assumed. Interestingly, the NLMS method shows to fail because of the extremely little adaptation lag used in this comparison. Anyway, results not presented here show that NLMS can only be useful if allowed to train over long periods of hundreds of symbols.

In Figure 3, with the same assumptions as previously, we consider the hypothesis where the vehicular speed v is *a priori* known to be (with equal probabilities) either $5,\ 50,\ 120$ km/h. The performance is compared against the optimal 2-D algorithm where v is known but erroneously estimated ($v = 5,\ 50,\ 100$ km/h). It is observed that, again, even when λ, or equivalently v, is *a priori* unknown, the Bayesian minimal update framework manages to ideally recover the channel with no performance decay. On the opposite, when λ is erroneously estimated, the performance decay of the optimal estimator might be dramatic.

In Figure 4, we show the Block Error Rate (BLER) performance comparison of a realistic LTE OFDM setup with Turbo Codes and actual signal detection and channel decoding. The BLER plots are obtained for classical OFDM detection performed using the minimal update and Kalman (with imperfect knowledge of channel length parameter) channel estimation. The case where detection is performed using ideal channel knowledge is also presented for the sake of completeness. The minimal update channel estimation provides performances that lies in between the other two cases. Moreover, it shows to offer the same performance as for the optimal 2D estimation although the plot has been omitted for clarity. Hence, the robust minimal update estimation method reveals to be an excellent choice with respect to a method of similar structure and complexity such as Kalman but avoiding the bargain of estimating side information.

VII. Conclusion

In this paper, we proposed a novel framework to channel estimation, applied to OFDM-based systems. We successively discussed the fundamental nature of channel estimation, under a Bayesian point of view. This approach allowed us to redefine channel estimation as a technique allowing one

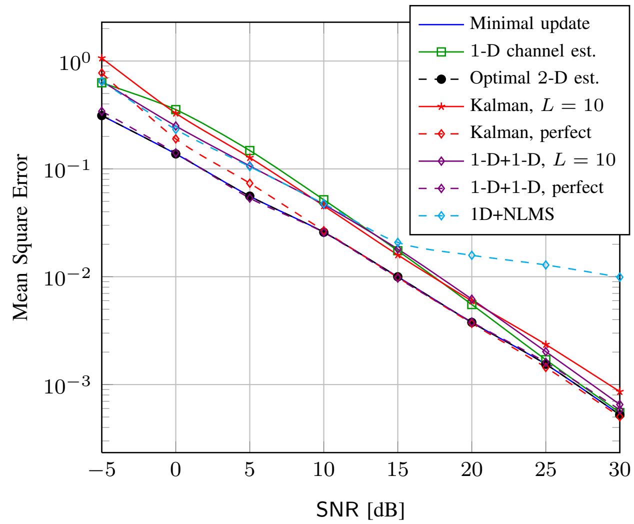

Fig. 2. MSE performance of the minimal update channel estimation against one-dimensional and two-dimensional (optimal) MMSE with $K = 4$ past pilot symbols, vehicular speed 120 km/h, LTE OFDM model.

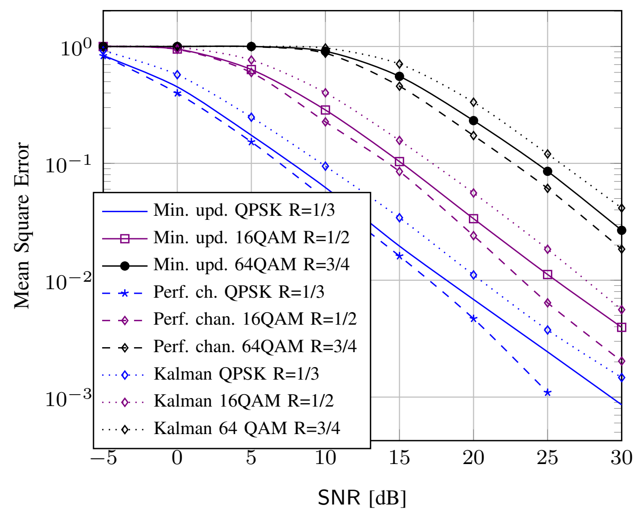

Fig. 4. Turbo Coding BLER performance comparison using the minimal updateand Kalman channel estimation against perfect channel knowledge, vehicular speed 120 km/h, LTE OFDM model.

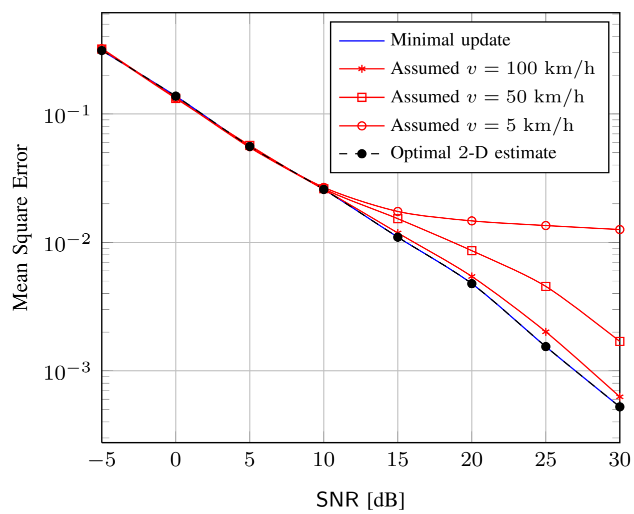

Fig. 3. MSE performance of the minimal update channel estimation with unknown time correlation, $K = 4$ past pilot symbols, vehicular speed $v = 120$ km/h, hypothesis space $v = 5, 50, 120$ km/h, LTE OFDM model.

to infer the posterior probability distribution $p(h|y, I)$ of a channel h given some input data y and prior information I. Assuming the receiver is allowed to store as much past information as desired, we then discussed optimal channel estimators under various levels of prior information at the receiver; the optimality emerges from a systematic usage of the maximum entropy principle. Then we proposed a novel approach to extend the maximum entropy setup when the receiver is oblivious of past received data. In a particular case, the latter was shown to be equivalent to the classical Kalman filter. Simulations suggest that the proposed novel technique is indistinguishable in performance from the optimal infinite time maximum entropy approach.

VIII. Acknowledgement

This work was partially supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless Communications NEWCOM++.

References

[1] O. Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson, and P. O. Börjesson, "OFDM channel estimation by singular value decomposition", IEEE Trans. Commun., vol. 46, pp. 931-939, July 1998.
[2] M. Morelli et al., "A comparison of pilot-aided Channel estimation methods for OFDM systems", IEEE Transactions on Signal Processing, vol. 49, pp. 3065–3073, 2001.
[3] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems",IEEE Trans. Veh. Technol., vol. 49, pp. 1207–1215, July 2000.
[4] Y. Li, L.J. Cimini, N.R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels", IEEE Transactions on Communications, Vol. 46, Issue 7, pp. 902–915, July 1998
[5] R. Couillet, M. Debbah, "Mathematical Foundations of Cognitive Radios," Journal of Telecommunications and Information Technology, no. 4, 2009.
[6] R. Couillet, M. Debbah, "A Bayesian Framework for Collaborative Multi-Source Signal Detection," IEEE Trans. on Signal Processing, submitted.
[7] R. Couillet, M. Debbah, "A maximum entropy approach to OFDM channel estimation," Proceedings of IEEE SPAWC Conference, Perugia, Italy, 2008.
[8] E. T. Jaynes, "Information Theory and Statistical Mechanics", Physical Review, APS, vol. 106, no. 4, pp. 620-630, 1957.
[9] A. Caticha, "Lectures on Probability, Entropy and Statistical Physics", arXiv:0808.0012v1 [physics.data-an], 2008.
[10] C. E. Shannon, "A mathematical theory of communications", Bell System Technical Journal, vol. 27, no. 7, pp. 379-423, 1948.
[11] E. T. Jaynes, "Probability Theory: The Logic of Science", Cambridge University Press, 2003.
[12] R. Couillet, M. Debbah, "A maximum entropy approach to OFDM channel estimation", arxiv Preprint http://arxiv.org/pdf/0811.0778
[13] K. E. Baddour and , N. C. Beaulieu, "Autoregressive modeling for fading channel simulation," IEEE Transaction on Wireless Communications, July 2005.
[14] S. Sesia, I. Toufik and M. Baker, "LTE, The UMTS Long Term Evolution: From Theory to Practice", Wiley & Sons, 2009.
[15] J. Cai, X. Shen and J.W. Mark, "Robust channel estimation for OFDM wireless communication systems-an H∞ approach", IEEE Trans. Wireless Commun., vol. 3,no. 6, pp. 2060–2071, 2004.

[16] D. Schafhuber, G. Matz. and F. Hlawatsch, “Adaptive Wiener filters for time-varying channel estimation in wireless OFDM systems”,IEEE ICASSP, vol. 4, pp. 688–691, 2003
[17] J. G. Proakis, *Digital Communications*. New York: McGraw Hill, 1995.
[18] R. H. Clarke, ‘A Statistical Theory of Mobile-radio Reception’. *Bell Syst. Tech. J.*, pp. 957–1000, July 1968.
[19] W. C. Jakes, *Microwave Mobile Communications*. New York: John Wiley & Sons, Ltd/Inc., 1974.