BAYESIAN BLIND AND SEMIBLIND CHANNEL ESTIMATION

Dirk T.M. Slock[*]

Eurecom Institute
2229 route des Crêtes, B.P. 193, 06904 Sophia Antipolis Cedex, FRANCE
Email: dirk.slock@eurecom.fr

ABSTRACT

Blind and semiblind channel estimation techniques are developed and usually evaluated for a given channel realization, i.e. with a deterministic channel model. Such blind channel estimates, especially those based on subspaces in the data, are often only partial and ill-conditioned. On the other hand, in wireless communications the channel is typically modeled as Rayleigh fading, i.e. with a Gaussian (prior) distribution expressing variances of and correlations between channel coefficients. In recent years, such prior information on the channel has started to get exploited in pilot-based channel estimation, since often (as e.g. in 3G WCDMA systems) the pure pilot-based (deterministic) channel estimate is of limited quality. The fading in wireless communications leads indeed often to a poor data to parameter ratio. In this paper we introduce a Bayesian approach to (semi-)blind channel estimation, exploiting a priori information on fading channels. Two case can be considered, either given prior information or joint estimation of channel and prior. In the second case there are still identifiability issues whereas in the first case there are typically none. However, the identifiability issues can be resolved with a reduced amount of training in a semiblind approach. Various models/parameterizations for the channel correlation structure and the fading process can be considered.

1. INTRODUCTION

Blind and semiblind channel estimation techniques have been developed and are usually evaluated for a given channel realization, i.e. with a deterministic channel model, see [1] for an overview of such techniques. Such blind channel estimates, especially those based on subspaces in the data, are often only partial and ill-conditioned. Indeed, only part of the channel is blindly identifiable, especially in the case of MIMO channels. The type of blind channel estimation techniques we are mostly referring to here involve an FIR multichannel and are typically based on the second-order statistics of the received signal. Two types of techniques can be considered, treating the unknown input symbols as either deterministic unknowns or Gaussian white noise. In the first case, the techniques are often based on the subspace structure induced in the data by the multichannel aspect. The part of the channel that can be identified blindly is larger in the Gaussian input model case than in the deterministic input model case, but is in any case incomplete. Many of the deterministic input approaches are also quite sensitive to a number of hypotheses such as correct channel length (filter order) and no channel zeros. In general this means that these blind channel estimates can often become ill-conditioned, when the channel impulse response is tapered (e.g. due to a pulse shape filter) or when the channel is close to having zeros. In fact this means that the blind information on the channel can be substantial, but is limited to only part of the channel.

1.1. Time-Varying Channels

Blind channel estimation techniques have also been developed for time-varying channels, by using so-called Basis Expansion Models (BEMs), in which the time-varying channel coefficients are expanded into known time-varying basis functions, and the unknown deterministic parameters are now no longer the channel coefficients but the combination coefficients in the BEM. The BEM model was introduced by Y. Grenier around 1980 for time-varying filtering, by E. Karlsson in the early 1990's for time-varying channel modeling and by M. Tsatsanis and G. Giannakis in 1996 for blind time-varying channel estimation. A. Sayeed introduced his related canonical coordinates approach around the same time frame.

In wireless communications the channel is typically modeled as Rayleigh fading, i.e. with a Gaussian (prior) distribution expressing variances of and correlations between channel coefficients. Below we shall elaborate on possible models for such a channel distribution. In recent years, such

[*] Eurécom's research is partially supported by its industrial partners: Hasler Foundation, Swisscom, Thales Communications, ST Microelectronics, CEGETEL, France Télécom, Bouygues Telecom, Hitachi Europe Ltd. and Texas Instruments.

prior information on the channel has started to get exploited in pilot-based channel estimation, since often (as e.g. in 3G WCDMA systems) the pure pilot-based (deterministic) channel estimate is of limited quality, see e.g. [2],[3].

Here we propose to combine Gaussian prior information on the channel with (semi-)blind information, in a Bayesian approach. The time-variation can be modeled at two levels.

1.1.1. Frame-based Processing

In a first model, *block-wise variation*, the channel realization is assumed to be piece-wise constant over blocks of data. Within each block, a classical (semi-)blind channel estimate gets performed. The channel estimation errors are then typically independent between blocks. We may have correlation though between the a priori channel in consecutive blocks. This first model is inspired by what can easily be justified in training-based channel estimation, in which case the channel is estimated from concentrated training bursts that occur at a regular pace and that are separated by data bursts. In the case of (semi-)blind channel estimation, all data gets involved in the channel estimation and hence, depending on the time scale, the temporal variation of the channel over the data may be difficult to ignore.

1.1.2. Continuous Processing

This leads to a second model, *continuous variation*, in which the time-varying channel gets expanded into a BEM and prior information gets formulated on the basis expansion coefficients. A simple BEM can correspond to a subsampling and temporal interpolation operation, which is motivated by the fact that the channel will show a maximum Doppler spread.

1.2. Bayesian Channel Estimation

The Bayesian (semi-)blind channel estimation problem can be formulated at two levels.

1.2.1. Known Prior

In the *given prior* case, the issue becomes simply one of properly combining deterministic and prior information, e.g. in the form of

$$\|\mathbf{h}+\mathbf{b}(\mathbf{h})-\widehat{\mathbf{h}}\|^2_{C^{-1}_{\widehat{\mathbf{h}}\widehat{\mathbf{h}}}} + \|\mathbf{h}-\mathbf{h}^o\|^2_{C^{-1}_{\mathbf{hh}}} \qquad (1)$$

where $\mathbf{h}$ contains the whole channel impulse reponse in a vector, and we assumed for the estimate $\widehat{\mathbf{h}} \sim \mathcal{N}(\mathbf{h}+\mathbf{b}, C_{\widehat{\mathbf{h}}\widehat{\mathbf{h}}})$, where $\mathbf{b}(\mathbf{h})$ is a potential bias, and for the prior $\mathbf{h} \sim \mathcal{N}(\mathbf{h}^o, C_{\mathbf{hh}})$. The given prior case becomes particularly simple when combined with the block-wise variation model. For a proper choice of prior, channel identifiability is no longer an issue since the prior information by itself already makes the channel covariance matrix bounded. Hence this Bayesian approach can in principle be appplied to channels with limited blind identifiability, such as even SISO channels (with Gaussian symbol model).

1.2.2. Prior with Unknown Parameters

The second Bayesian level would be *joint channel and prior estimation*. In this (realistic) case we assume that we observe the channel over a number of fades so that we have ergodicity for the presumed stationary prior channel distribution. The parameters to be estimated are now both parameters related to instantaneous realizations of the channel and parameters in the Gaussian prior distribution. Some of the techniques in [2],[3] can be invoked here. Parameter identifiability also becomes again an issue.

2. (MIMO WIRELESS) CHANNEL MODELS

In order to improve channel estimation and reduce MFB loss, it is advantageous to exploit correlations in the channel, if present. For time-varying channel, two channel models can be considered according to two transmission modes:

1. continuous transmission: in this case the vectorized channel impulse response can be modeled as a (locally) stationary vector signal; limited bandwidth usually allows downsampling w.r.t. symbol rate; stationarity can only be local due to slow fading

2. bursty transmission: in this case, the time axis is cut up in bursts, the channel (down)samples within each burst can be rerepresented in terms of Basis Expansion Models (BEMs); limited bandwidth leads to limited BEM terms.

Both models are equivalent as long as the temporal correlation structure in the continuous mode gets properly transformed to intra and inter burst correlation between BEM coefficients.

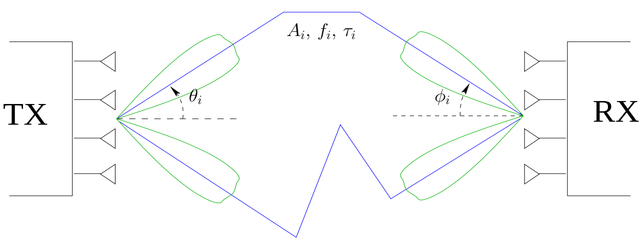

Fig. 1. MIMO transmission with N_T transmit and N_R receive antennas.

2.1. Specular Wireless MIMO Channel Model

Now consider a MIMO transmission configuration as depicted in Fig. 1. We get for the impulse response of the time-varying channel $\mathbf{h}(t,\tau)$ [4]

$$\mathbf{h}(t,kT)=\sum_{i=1}^{N_P} A_i(t)\, e^{j2\pi\, f_i\, t}\, \mathbf{a}_R(\phi_i)\, \mathbf{a}_T^T(\theta_i)\, p(kT-\tau_i) \quad . \tag{2}$$

The channel impulse response $\mathbf{h}$ has per path a rank 1 contribution in 3 dimensions; there are N_P pathwise contributions where

- A_i: complex attenuation
- f_i: Doppler shift
- θ_i: angle of departure
- ϕ_i: angle of arrival
- τ_i: path delay
- $\mathbf{a}(.)$: antenna array response
- $p(.)$: pulse shape (TX filter)

The fast variation of the phase in $e^{j2\pi\, f_i\, t}$ and possibly the variation of the A_i correspond to the fast fading. All the other parameters (including the Doppler frequency) vary on a slower time scale and correspond to slow fading.

2.2. MIMO Channel Prediction

Consider vectorizing the impulse response coefficients

$$(N\times 1) \quad \underline{\mathbf{h}}(t)=\text{vec}\{\mathbf{h}(t,.)\}=\sum_{i=1}^{N_P} \underline{\mathbf{h}}_i\, A_i(t)\, e^{j2\pi\, f_i\, t} \tag{3}$$

where $\underline{\mathbf{h}}_i=\text{vec}\{\mathbf{a}_R(\phi_i)\,\mathbf{a}_T^T(\theta_i)\,p(.-\tau_i)\}$ and the total number of coefficients becomes $N=N_TN_RN_\tau$ = number of TX antennas times number of RX antennas times delay spread. Due to the Doppler shift, the phase of the path complex amplitude is varying rapidly. The actual path amplitude is not varying rapidly unless what we consider to be a specular path is already the superposition of multiple paths that are not resolvable in delay, Doppler and angles. With $f_i\in(-f_d,f_d)$, the Doppler shift for path i, the (fast fading) variation is bandlimited and hence the channel should be perfectly predictible! (not so due to the slow fading: the slow parameters such as delays and angles will vary eventually). When only the fast fading is taken into account as temporal variation, the matrix spectrum $S_{\underline{\mathbf{hh}}}(f)$ of the vectorized channel can be doubly singular:

1. if $A_i(t)\equiv A_i$ and N_P finite: spectral support singularity: sum of cisoids!
2. if $N_p<N$: matrix singularity, limited source of randomness (limited diversity)

When the channel spectral support becomes singular, the channel becomes perfectly predictible. Hence channel prediction should play an important role in channel estimation.

2.3. Subspace AR Channel Model

After sampling the temporal variation at $t=kT$, the vectorized impulse response can be represented as

$$\underbrace{\underline{\mathbf{h}}[k]}_{N\times 1}=\underbrace{\mathbf{H}}_{N\times N_P}\;\underbrace{\underline{\mathbf{A}}[k]}_{N_P\times 1} \tag{4}$$

where $\underline{\mathbf{A}}[k]=[A_1(kT)\,e^{j2\pi\,f_1\,kT}\cdots A_{N_P}(kT)\,e^{j2\pi\,f_{N_P}\,kT}]^T$ contains the fast fading part and $\mathbf{H}=[\underline{\mathbf{h}}_1\cdots\underline{\mathbf{h}}_{N_P}]$.

The important issue here is that the spectral modeling of the channel coefficient temporal variation should be done in a transform domain and not on the channel impulse response coefficients themselves. Since each such coefficient can be the result of the contributions of many paths, the dynamics of the temporal variation of the coefficients are necessarily of higher order, compared to the variation of $A_i(kT)\,e^{j2\pi\,f_i\,kT}$ which can be of an order as low as one (when $A_i(kT)$ is constant; the cisoid $e^{j2\pi\,f_i\,kT}$ is perfectly predictible with first-order linear prediction). Also, if the impulse response coefficients are modeled directly, then their (spatial and delay-wise) correlation has to be taken into account: $S_{\underline{\mathbf{hh}}}(f)$ cannot be modeled accurately as diagonal, whereas $S_{\underline{\mathbf{AA}}}(f)$ can.

So the diagonal elements of $\underline{\mathbf{A}}[k]$ are modeled as decorrelated stationary scalar processes. The channel distribution is typically taken to be complex Gaussian. If the fast parameters $\underline{\mathbf{A}}[k]$ are not too predictible, then the estimation errors of the slow parameters $\mathbf{H}$ should be negligible (change only with slow fading, hence their estimation error should be small). From (4) we obtain the spectrum

$$S_{\underline{\mathbf{hh}}}(f)=\mathbf{H}\;S_{\underline{\mathbf{AA}}}(f)\;\mathbf{H}^H\;,\;S_{\underline{\mathbf{AA}}}(f)\text{ diagonal.} \tag{5}$$

The components of $\underline{\mathbf{A}}[k]$ can conveniently be modeled as AR processes, each spanning only a fraction of the Doppler range $(-f_d,f_d)$. In fact, a subsampled version of the fast parameters $\underline{\mathbf{A}}[k]$ could be introduced, with the subsampled rate corresponding to the (maximum) Doppler spread. A stationary (AR) model can be taken for the subsamples and the other samples can be obtained by linear interpolation from the subsamples. This is the case of a BEM with a single basis function: the interpolation filter response.

2.4. Separable Correlation Channel Model

The subspace channel model is appropriate when the channel is fairly specular, with limited diversity so that the number of paths is not large w.r.t. the total number of channel coefficients. Now consider the other extreme of rich diversity, when $N_P \gg N$, in which case the dynamics of all paths get mixed up and the spatial-delay correlations between the channel impulse response elements become separable [5]. The spectrum of the temporal variation of the in this case diffuse channel can then be written as

$$S_{\underline{\mathbf{hh}}}(f) = R_\tau \otimes R_T \otimes R_R \; S_d(f) \tag{6}$$

where
R_τ: correlation matrix between delays, typically diagonal with power delay profile
R_T: TX side spatial correlation matrix
R_R: RX side spatial correlation matrix
$S_d(f)$: scalar common Doppler spectrum of all impulse response coefficients. Such a completely separable 4D correlation model is not very realistic though. Though separability in spatial correlation between transmit and receive side can be argued to some extent, both spatial correlation and Doppler spectrum are typically delay-dependent.

Other cases of singular channel correlation as in (4) when $N > N_P$ may occur. For instance, correlation in delay is due to the pulse shaping filter, which extends the duration of the channel impulse response without adding fading sources and makes the impulse response taper off in both directions, causing ill-conditioning for some blind channel estimation approaches. To handle the pulse shape induced correlations, one may write the (delay dimension aspect of) the impulse response as in (4) where now $\mathbf{H}$ is a tall banded Toeplitz pulse shape convolution matrix and $\underline{\mathbf{A}}$ would contain mutually uncorrelated components (discrete-time representation of the actual propagation channel).

3. BLIND AND SEMIBLIND BAYESIAN CHANNEL ESTIMATION

Blind and semiblind versions of training based Bayesian channel estimation such as in [2],[3] can in fact be formulated in a fairly straightforward fashion. In [2],[3], training based channel estimates are obtained at regular time intervals, let's say slots. So the temporal sampling period for the channel variation would be one slot. The training based, unbiased channel estimate $\widehat{\mathbf{h}}_k$ provides a measurement equation for the channel $\mathbf{h}_k$ in the slot k in which the measurement noise $\widetilde{\mathbf{h}}_k$ is the channel estimation error

$$\widehat{\mathbf{h}}_k \;=\; \mathbf{h}_k + \widetilde{\mathbf{h}}_k \tag{7}$$

where training is typically designed for the channel estimation error $\widetilde{\mathbf{h}}_k$ to be white in time and between components: $\mathrm{E}\,\widetilde{\mathbf{h}}_k\widetilde{\mathbf{h}}_n^H \;=\; \sigma^2_{\tilde{h}}\, I\;\; \delta_{kn}$ (Kronecker delta). The Bayesian aspects comes from the fact that now $\mathbf{h}_k$ gets modeled typically with a Gaussian distribution with zero mean (Rayleigh fading) in which the complete statistical description resides in the second-order moments. Two variations are possible depending on whether the temporal variations are described over a limited time frame with a BEM, to handle one frame of a number of slots, or the time period over which channel coding gets performed. The other approach is to model $\mathbf{h}_k$ as a stationary process with a spectral density matrix $S_{\mathbf{hh}}(z) \;=\; \sum_{n=-\infty}^{\infty}(\,\mathrm{E}\,\mathbf{h}_{k+n}\mathbf{h}_k^H\,)\,z^{-n}$ which can be structured in a number of ways, as already alluded to earlier. The Bayesian estimation problem then becomes the joint estimation on the basis of the process $\widehat{\mathbf{h}}_k$ the process $\mathbf{h}_k$ and the parameters in $S_{\mathbf{hh}}(z)$ which are typically treated themselves as deterministic parameters. Although the Bayesian approach tries to improve the estimation of the process $\mathbf{h}_k$ by invoking other parameters (in the description of $S_{\mathbf{hh}}(z)$) that need to be estimated, but these other parameters (which are subject to the slow fading) evolve much more slowly (than the fastly fading $\mathbf{h}_k$) so that their estimation is possible and beneficial.

The difference between (semi-)blind and training based Bayesian channel estimation is not very substantial. Indeed, if the training information in the semiblind approach suffices to provide channel identifiability in a slot (unbiased channel estimate), then the slot-wise measurement equation is again of the form (7) with only some correlation between the components of $\widetilde{\mathbf{h}}_k$ being introduced: $\mathrm{E}\,\widetilde{\mathbf{h}}_k\widetilde{\mathbf{h}}_n^H \;=\; \sigma^2_{\tilde{h}}\, R_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}\;\; \delta_{kn}$ where the correlation matrix $R_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}$ depends on the semi-blind approach.

In the case of blind Bayesian channel estimation, the measurement equation is more profoundly affected. The measurement noise remains temporally white, but only part of the channel impulse response gets observed, due to the blind inidentifiability. The particular form of the measurement equation depends on the blind method used. In general it will not be possible to attain joint identifiablity of all parameters involved, unless the prior information is structured in a particular fashion (e.g. the power delay profile is known to be monotonically decreasing).

The Bayesian aspect tends to reduce blind inidentifiability to a minimum, so that less training information in a semiblind approach is required compared to semiblind deterministic channel estimation. For instance, consider the blind estimation of a FIR SIMO channel with the input signal modeled as deterministic unkown. Then the channel may become more inidentifiable if its vector transfer function exhibits zeros. However, those zeros occur with zero probability if the power delay profile does not vanish over the channel impulse response duration and if the subchannels are not perfectly correlated. So inidentifiability in this case gets reduced to a minimum, i.e. a global complex scale factor. However, if the power delay profile support gets overestimated, then so will be the channel impulse responses, and vice versa. This last instance is another example of the next observation.

Bayesian model parameters inherit the inidentiability aspects of the blind channel estimate. Consider e.g. a SISO channel of length 2: $h(z) = h_0 + h_1 z^{-1}$, where the two coefficients have a priori variances (power delay profile) σ_0^2, σ_1^2 and are uncorrelated. Consider blind channel estimation in the case of a Gaussian model for the unknown input (and suppose we know the additive white noise variance). Then h_0, h_1 can only be identified up to a common phase factor of the form $e^{j\theta}$. This is the minimal inidentifiability in this Gaussian case. However, we can also not identify blindly whether the channel is minimum-phase or maximum-phase (one bit of uncertainty). If we knew that either $\sigma_0^2 > \sigma_1^2$ or $\sigma_0^2 < \sigma_1^2$ then in the Bayesian blind approach, we could jointly blindly estimate the channel (up to the minimum inidentifiability) and the power delay profile. However, the knowledge of the order of σ_0^2 and σ_1^2 is again one bit of information. For a longer SISO channel and Gaussian blind channel estimation, the Bayesian problem should be identifiable if it would be known that e.g. the power delay profile is monotonically decreasing.

4. FURTHER OBSERVATIONS

The Bayesian framework may be/is the proper tool to handle bias/variance trade-offs in channel estimation. In a Bayesian approach, the bias corresponds to the prior information (mean). If the blind information is ill-conditioned, the ill-conditioned parts will automatically be overridden by the prior info. How does the cleaning of unreliable information as in semiblind estimation [1] fit in here? Is the denoising of the blind information still useful?

The Bayesian blind channel estimation formulation may alo be the right framework for analyzing the oversampling versus excess bandwidth issue, an issue in blind multichannel estimation that has so far eluded a convincing analysis.

5. REFERENCES

[1] E. de Carvalho and D.T.M. Slock. Semi–Blind Methods for FIR Multichannel Estimation. In G.B. Giannakis, Y. Hua, P. Stoica, and L. Tong, editors, *Signal Processing Advances in Wireless & Mobile Communications*. Prentice Hall, 2001.

[2] M. Lenardi and D.T.M. Slock. Estimation of Time-Varying Wireless Channels and Application to the UMTS W-CDMA FDD Downlink. In *Proc. European Wireless (EW) Conf.*, Florence, Italy, Feb. 2002.

[3] G. Montalbano and D.T.M. Slock. Joint Common-Dedicated Pilots Based Estimation of Time-Varying Channels for W-CDMA Receivers. In *Proc. Vehic. Tech. Conf. (VTCfall)*, Orlando, FL, USA, Sept. 2003.

[4] B. Fleury, P. Jourdan and A. Stucki “High-Resolution Channel Parameter Estimation for MIMO Applications using the SAGE Algorithm”. In Proc. *International Zürich Seminar on Broadband Communications*, pp. 301-309, Feb. 2002.

[5] S. Visuri and D.T.M. Slock “Colocated Antenna Arrays: Design Desiderata for Wireless Communications”. In Proc. *IEEE Workshop on Sensor Array and Multichannel (SAM) processing*, Aug. 2002.