Sequential Bayesian Algorithms for Identification and Blind Equalization of Unit-Norm Channels

Claudio J. Bordin Jr.[1] **and Marcelo G. S. Bruno**[2]

[1]UFABC, claudio.bordin@ufabc.edu.br [2]Instituto Tecnológico de Aeronáutica, bruno@ita.br

1. Introduction

- In many problems the unknowns reside on spherical manifolds.
- We employ Fisher-Bingham (F-B) prior distributions to the unknowns.
- The F-B priors form a conjugate model that yields closed-form, recursive estimates, naturally constrained to the spherical manifold.
- We apply this model to a communication setup with multiple gain-controlled FIR channels, deriving a MAP channel estimator and a blind equalizer based on Rao-Blackwellized particle filters.

2. Problem Formulation

- Single transmitter, R receivers.
- Let $\{b_n\}$ be the bits and $\{x_n\}$ the corresponding differentially encoded symbols.
- Signal observed at the r−th receiver:

$$y_{r,n} = \mathbf{h}_r^T \mathbf{x}_n + v_{r,n}, \quad (1)$$

where $\mathbf{x}_t \triangleq [x_t \dots x_{t-L+1}]^T$, $v_{r,t} \sim \mathcal{N}(0; \sigma_r^2)$, and $\mathbf{h}_r \in \mathbb{R}^{L\times 1}$ is the (time-invariant) impulse response of the channel to the r−th receiver.

- The random quantities $\mathbf{h}_r$, $\{\mathbf{x}_n\}$, and $\{v_{r,n}\}$, $r \in \{1,...,R\}$, are presumed to be mutually independent a priori.
- Due to automatic gain control (AGC), $\|\mathbf{h}_r\| = 1$; to take this constraint into account, we assume that $\mathbf{h}_r$ has an $L-$variate F-B prior

$$p(\mathbf{h}_r) = \mathcal{FB}(\mathbf{h}_r|\mathbf{a}_{r,0}, \mathbf{B}_{r,0}) \triangleq C(\mathbf{a}_{r,0}, \mathbf{B}_{r,0})^{-1} \times \exp(\mathbf{h}_r^T \mathbf{B}_{r,0} \mathbf{h}_r + \mathbf{h}_r^T \mathbf{a}_{r,0}) \mathcal{I}_{\|\mathbf{h}_r\|_2 = 1}, \quad (2)$$

where $\mathbf{a}_{r,0} \in \mathbb{R}^L$ and $\mathbf{B}_{r,0} \in \mathbb{R}^{L\times L}$ are the hyper-parameters, and $C(\mathbf{a}_{r,0}, \mathbf{B}_{r,0})$ is the normalization constant.

- Accurate numerical estimates to $C(\mathbf{a}_{r,0}, \mathbf{B}_{r,0})$ can be computed via the so-called saddlepoint approximations [1] (see Appendix).

3. Main Result

- As a consequence of (1) and (2), the posterior distribution of $\mathbf{h}_r$ given $\mathbf{X}_n \triangleq \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ and $Y_{r,n} \triangleq \{y_{r,1}, \cdots, y_{r,n}\}$ is F-B, i.e.,

$$\boxed{p(\mathbf{h}_r|\mathbf{X}_n, Y_{r,n}) = \mathcal{FB}(\mathbf{h}_r|\mathbf{a}_{r,n}, \mathbf{B}_{r,n}),} \quad (3)$$

where the parameters $\mathbf{a}_{r,n}$ and $\mathbf{B}_{r,n}$ can be recursively determined via

$$B_{r,n} = B_{r,n-1} - \mathbf{x}_n \mathbf{x}_n^T / 2\sigma_r^2, \quad (4)$$

$$\mathbf{a}_{r,n} = \mathbf{a}_{r,n-1} + \mathbf{x}_n y_{r,n}/\sigma_r^2. \quad (5)$$

- Sketch of the Proof:
  - Exploiting Markovian properties, it follows that:

$$p(\mathbf{h}_r|\mathbf{X}_n, Y_{r,n}) = \frac{p(y_{r,n}|\mathbf{x}_n, \mathbf{h}_r) p(\mathbf{h}_r|\mathbf{X}_{n-1}, Y_{r,n-1})}{\int_{\mathbf{h}_r \in \mathcal{S}^{L-1}} p(y_{r,n}|\mathbf{x}_n, \mathbf{h}_r) p(\mathbf{h}_r|\mathbf{X}_{n-1}, Y_{r,n-1})\, d\mathcal{S}^{L-1}(\mathbf{h}_r)}$$

  - Induction hypothesis:

$$p(\mathbf{h}_r|\mathbf{X}_{n-1}, Y_{r,n-1}) = \mathcal{FB}(\mathbf{h}_r|\mathbf{a}_{r,n-1}, \mathbf{B}_{r,n-1}).$$

  - Observing that $p(y_{r,n}|\mathbf{x}_n, \mathbf{h}_r) = \mathcal{N}(y_{r,n}|\mathbf{h}_r^T \mathbf{x}_n; \sigma_r^2)$, we get after some algebra that

$$p(y_{r,n}|\mathbf{x}_n, \mathbf{h}_r) p(\mathbf{h}_r|\mathbf{X}_{n-1}, Y_{r,n-1}) = \frac{1}{\sqrt{2\pi\sigma_r^2}} \times \frac{C(\mathbf{a}_{r,n}, \mathbf{B}_{r,n})}{C(\mathbf{a}_{r,n-1}, \mathbf{B}_{r,n-1})} \exp\left(-\frac{y_{r,n}^2}{2\sigma_r^2}\right) \mathcal{FB}(\mathbf{h}_r|\mathbf{a}_{r,n}, \mathbf{B}_{r,n}). \quad (6)$$

4. Channel Identification

- The local trained MAP estimate of the channel parameters at the $r-th$ receiver is given by

$$\hat{\mathbf{h}}_{r,n} \triangleq \arg\max_{\mathbf{h}_r} p(\mathbf{h}_r|\mathbf{X}_n, Y_{r,n}). \quad (7)$$

- The constrained optimization problem (7) can be recast as

$$\hat{\mathbf{h}}_{r,n} = \arg\max_{\mathbf{h}_r} \exp(\mathbf{h}_r^T \mathbf{B}_{r,n} \mathbf{h}_r + \mathbf{h}_r^T \mathbf{a}_{r,n}), \quad \text{subject to } \|\mathbf{h}_r\|_2 = 1$$

as the normalization constant $C(\mathbf{a}_{r,n}, \mathbf{B}_{r,n})$ does not depend on $\mathbf{h}_r$.

- The corresponding Lagrange function is given by

$$\Lambda(\mathbf{h}_r, \lambda) \triangleq \exp(\mathbf{h}_r^T \mathbf{B}_{r,n} \mathbf{h}_r + \mathbf{h}_r^T \mathbf{a}_{r,n}) + \lambda(\mathbf{h}_r^T \mathbf{h}_r - 1). \quad (8)$$

- Taking the gradient of (8) with respect to $\mathbf{h}_r$, dividing the result by the exponential term and equating it to zero, it follows that

$$\boxed{\hat{\mathbf{h}}_{r,n} = -\frac{1}{2}\left(\mathbf{B}_{r,n} + \tilde{\lambda}\mathbf{I}\right)^{-1} \mathbf{a}_{r,n}.} \quad (9)$$

where $\tilde{\lambda} \triangleq \lambda \exp(-\hat{\mathbf{h}}_{r,n}^T \mathbf{B}_{r,n} \hat{\mathbf{h}}_{r,n} - \hat{\mathbf{h}}_{r,n}^T \mathbf{a}_{r,n})$.

- As $\mathbf{B}_{r,n}$ is symmetrical, we can plug its eigenvalue decomposition $\mathbf{B}_{r,n} \triangleq \mathbf{U}_{r,n} \mathbf{D}_{r,n} \mathbf{U}_{r,n}^T$ into (9) and rewrite the constraint $\|\hat{\mathbf{h}}_{r,n}\|_2^2 = 1$ as

$$\mathbf{a}_{r,n}^T \mathbf{U}_{r,n} \left(\mathbf{D}_{r,n} + \tilde{\lambda}\mathbf{I}\right)^{-2} \mathbf{U}_{r,n}^T \mathbf{a}_{r,n} = 4. \quad (10)$$

- Equation (10) is equivalent to

$$\sum_{k=1}^{L} \left(\frac{\left[\mathbf{a}_{r,n}^T \mathbf{U}_{r,n}\right]^{[k]}}{\left[\mathbf{D}_{r,n}\right]^{[k]} + \tilde{\lambda}}\right)^2 = 4 \quad (11)$$

which can be rewritten as a $2L$ degree polynomial equation and numerically solved for $\tilde{\lambda}$.

- We verified experimentally that the most negative real $\tilde{\lambda}$ that solves (11) *almost* always leads to the MAP solution when replaced in (9).

5. Multichannel Blind Equalization

- We wish now to obtain the joint MAP estimate

$$\hat{b}_n = \arg\max_{b_n} p(b_n|Y_n),$$

that employs the observations available at all receivers, $Y_n \triangleq \{Y_{1,n}, \cdots, Y_{r,n}\}$, via a particle filtering method

$$p(\mathbf{X}_n|Y_n) \approx \sum_{p=1}^{P} w_n^{(p)} \mathcal{I}_{\{\mathbf{X}_n - \mathbf{X}_n^{(p)}\}},$$

where P denotes the number of particles $\mathbf{X}_n^{(p)}$ and $w_n^{(p)}$ the normalized weights.

- Adopting the prior importance function, the particles are extended as $\mathbf{x}_n^{(p)} \sim p(\mathbf{x}_n|\mathbf{x}_{n-1}^{(p)})$ and the weights updated as

$$w_n^{(p)} \propto w_{n-1}^{(p)}\, p(y_n|\mathbf{X}_n^{(p)}, Y_{n-1}).$$

- The a priori independence of the channel parameters and noise samples at each receiver imply that

$$p(y_n|\mathbf{X}_n^{(p)}, Y_{n-1}) = \prod_{r=1}^{R} p(y_{r,n}|\mathbf{X}_n^{(p)}, Y_{r,n-1}).$$

- Likewise, conditional independence relations induced by the model result that

$$p(y_{r,n}|\mathbf{X}_n^{(p)}, Y_{r,n-1}) = \int_{\mathbf{h}_r \in \mathcal{S}^{L-1}} p(y_{r,n}|\mathbf{x}_n^{(p)}, \mathbf{h}_r) \times p(\mathbf{h}_r|\mathbf{X}_{n-1}^{(p)}, Y_{r,n-1})\, d\mathcal{S}^{L-1}(\mathbf{h}_r). \quad (12)$$

- The integrand in (12) is identical to (6), which depends on $\mathbf{h}_r$ only through a F-B density. Therefore

$$p(y_{r,n}|\mathbf{X}_n^{(p)}, Y_{r,n-1}) = \frac{1}{\sqrt{2\pi\sigma_r^2}} \cdot \frac{C(\mathbf{a}_{r,n}^{(p)}, \mathbf{B}_{r,n}^{(p)})}{C(\mathbf{a}_{r,n-1}^{(p)}, \mathbf{B}_{r,n-1}^{(p)})} \exp\left(\frac{-y_{r,n}^2}{2\sigma_r^2}\right), \quad (13)$$

where $\mathbf{a}_{r,n}^{(p)}$ and $\mathbf{B}_{r,n}^{(p)}$ are defined as in (4)-(5) replacing $\mathbf{x}_n$ with $\mathbf{x}_n^{(p)}$.

6. Simulation Results

- To evaluate the performance of the proposed algorithms, we ran numerical experiments with $L = 3$. We drew $\mathbf{h}_r$ by sampling independently in each realization from a uniform distribution on the unit sphere. Accordingly, the hyperparameters $\mathbf{a}_{r,0}$ and $\mathbf{B}_{r,0}$ were set to zero.

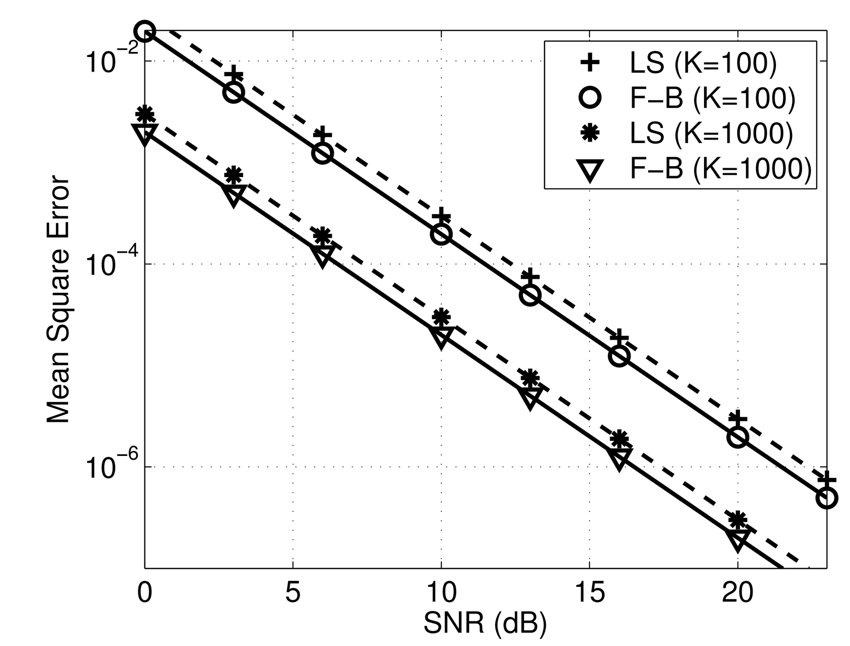

Figure 1: *Average m.s.e. in the identification of the parameter* $\mathbf{h}_r$ *as a function of the number of observed samples* (K) *and the SNR.*

- As one may verify (Figure 1), the algorithm employing F-B prior led to a lower m.s.e. (about 66% of the LS estimator m.s.e.) for all SNR levels; this remains true when $K = 1,000$.
- To assess the mean BER of the proposed blind equalization algorithm, the SNR was set equal on all receivers.
- In each realization, an i.i.d. sequence of 250 differentially encoded binary symbols was transmitted, being the first 150 bits discarded. The simulated system has $R = 4$. $\mathbf{h}_r$ was sampled independently for each r. The particle filter uses $P = 300$ and performs systematic resampling at each time step.
- Figure 2 displays the mean BER of the proposed algorithm (F-B) and that of the equivalent method that employs mismatched Gaussian priors [2, Sec. 3.1].

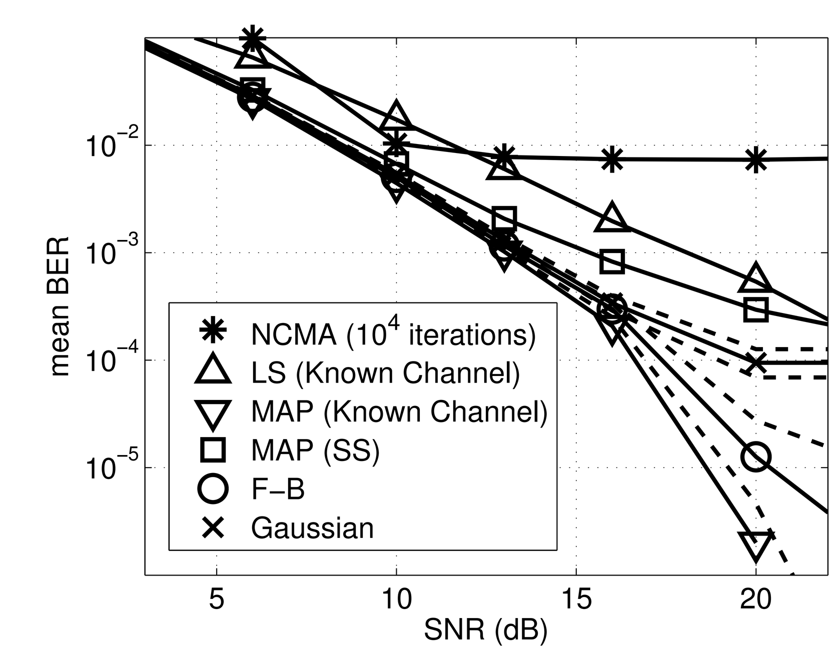

Figure 2: *Mean BER estimated in 5,000 Monte Carlo runs as a function of the SNR. The dashed lines surrounding the solid ones display 95% confidence intervals.*

- As one may note, the mean BER of the new (F-B) algorithm is equivalent to the optimal MAP equalizer and Gaussian particle filter method for low SNR. For high SNR, the new method outperforms the others.

7. Conclusions

- We introduced new algorithms for channel identification and blind equalization using a Fisher-Bingham prior model for the unknown parameters.
- F-B priors lead to a conjugate model that results in closed-form expressions for the parameters of the posterior densities, dropping with the need for approximations performed by previous works that employed sphere-constrained distributions.
- As we assessed via Monte Carlo simulations, the new channel identification and blind equalization algorithms outperformed conventional algorithms that adopt mismatched Gaussian priors, at the cost of increased computational complexity.

References

[1] A. Kume, "Saddlepoint approximations for the Bingham and Fisher-Bingham normalising constants," *Biometrika*, n. 2, vol. 92, pp. 465-476, 2005.

[2] C. J. Bordin Jr. and M. G. S. Bruno, "Consensus-Based Distributed Particle Filtering Algorithms for Cooperative Blind Equalization in Receiver Networks," *Proc. of ICASSP*, pp. 3968–3971, May 2011.

Appendix: Saddlepoint Approximation

- Kume [1] developed a method to compute the normalization constant

$$C(\mathbf{a}, \mathbf{B}) = \int_{\mathbf{h} \in \mathcal{S}^{L-1}} \exp(\mathbf{h}^T \mathbf{B} \mathbf{h} + \mathbf{h}^T \mathbf{a})\, d\mathcal{S}^{L-1}(\mathbf{h})$$

that exploits the fact that the F-B density arises when an $L-$variable Gaussian r.v. $\mathbf{x}$ is conditioned to have unit norm.

- Assuming (without loss of generality) that B is diagonal and introducing the change of variables $\nu \triangleq \mathbf{x}^T\mathbf{x}$ and $\mathbf{h} \triangleq \mathbf{x}/\nu^{1/2}$, it follows that $C(\mathbf{a}, \mathbf{B})$ depends on $p(\nu)$ (i.e., the p.d.f. of ν), evaluated at $\nu = 1$.
- As ν can be shown to be a linear combination of noncentral χ_1^2 r.v.'s, there is a closed-form expression for $K(t)$, the cumulant generating function of the distribution of ν, from which one can derive the saddlepoint approximation

$$\hat{p}(\nu) \triangleq \left(2\pi K''(\hat{t})\right)^{-\frac{1}{2}} \exp\left(K(\hat{t}) - \hat{t}\nu\right),$$

where $\hat{t}$ denotes the solution to $K'(\hat{t}) = \nu$, and $K'(t)$ and $K''(t)$ are the 1^{st} and 2^{nd} derivatives of $K(t)$, respectively.

ICASSP 2016, Shanghai - China.