BAYESIAN BLIND EQUALIZATION OF TIME-VARYING FREQUENCY-SELECTIVE CHANNELS SUBJECT TO UNKNOWN VARIANCE NOISE

Claudio J. Bordin Jr. and Marcelo G. S. Bruno

Instituto Tecnológico de Aeronáutica
São José dos Campos, 12228-900, Brazil. E-mail: {bordin,bruno}@ita.br

ABSTRACT

We present in this article a novel particle-filter-based blind equalization algorithm suitable for FIR time-varying frequency-selective communication channels corrupted by unknown variance additive Gaussian noise. The proposed method is fully Bayesian, integrating out the unknown parameters via an original recursive method, unlike previous approaches that rely on suboptimal plug-in estimates. We verify via numerical simulations that the proposed method's performance approaches that of the trained MAP equalizer, exceeding that of the linear least squares Kalman equalizer for medium to low noise levels.

Index Terms— Adaptive equalizers, Sequential estimation, Monte Carlo methods, Bayes procedures.

1. INTRODUCTION

The use of particle filters in the solution of blind equalization [1] and related digital communication problems [2] has drawn significant research interest lately. Most of the methods found in the literature rely however on the exact knowledge of the noise variance parameter, or resort to suboptimal estimates of this quantity [3], which may lead to performance issues. In [4], we introduced an algorithm that filled this gap, treating the unknown variance as a nuisance parameter that was analytically integrated out. The algorithm in [4] assumes however that the channel parameters are time-invariant, and cannot be trivially modified to deal with different models.

In this article, we lift this restriction, introducing a novel fully-Bayesian blind equalization algorithm suitable for frequency-selective *time-varying* FIR channels. The unknown filter coefficients are assumed to evolve in time according to an AR model. The proposed scheme relies on an original method for propagating the posterior statistics of the noise variance inspired in [5], and on the adoption of conjugate priors [6] for the unknown parameters. As we verify via numerical simulations, the performance of the proposed algorithm is not critically dependent on the validity of the prior assumptions, achieving near-optimal performance in simulations employing mismatched models.

The first author's work was funded by FAPESP, Brazil.

The remainder of this article is organized as follows: after describing the underlying signal model in Sec. 2, we present the proposed particle filter equalizer in Sec. 3. The performance of the proposed algorithm is assessed via numerical simulations in Sec. 4, and finally, in Sec. 5, we summarize the main contributions of our work.

2. SIGNAL MODEL

Denote by $b_n, \{n \geq 0\}$ the transmitted binary bits, assumed to form an independent, identically distributed (i.i.d) sequence, and let $s_n \in \{\pm 1\}$ be the resulting differentially encoded symbols. The observations $y_{0:n} \triangleq \{y_0, ..., y_n\}$ are assumed to be the output of the additive noise frequency selective FIR channel

$$y_n = \mathbf{h}_n^H \mathbf{S}_n + v_n, \tag{1}$$

where $\mathbf{h}_n \in \mathbb{C}^{L \times 1}$ is a vector that collects the channel impulse response terms, L denotes the (known) channel order, $\mathbf{S}_n \triangleq [s_n \ldots s_{n-L+1}]^T \in \mathbb{R}^{L \times 1}$, and v_n represents the contribution of the additive noise. The unknown channel impulse response vector $\mathbf{h}_n$ is assumed to evolve in time according to a first-order autoregressive model

$$\mathbf{h}_{n+1} = \mathbf{A}\mathbf{h}_n + \mathbf{w}_n, \tag{2}$$

where $\mathbf{A} \in \mathbb{C}^{L \times L}$ is assumed known and $\{\mathbf{w}_n\}, n \geq 0$, is a complex multivariate process. Given the above model, our objective is to compute recursively the MAP smoothed estimate ($d \geq 0$)

$$\hat{b}_{n-d} = \arg\max_{b_{n-d}} p(b_{n-d}|y_{0:n}). \tag{3}$$

where $p(b_{n-d}|y_{0:n})$ denotes the probability mass function (p.m.f) of the symbol b_{n-d} conditioned on the observations $y_{0:n}$.

2.1. Prior Model

In this work, we assume a set of conjugate prior distributions for v_n and w_n similar to that employed in [7] in a different context. Namely, we assume that both variables are zero-mean complex circular Gaussian, with moments

$$E \begin{bmatrix} v_n \\ \mathbf{w}_n \end{bmatrix} \begin{bmatrix} v_m^* \ \mathbf{w}_m^H \end{bmatrix} = \begin{bmatrix} \sigma^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma^2\epsilon^2 \end{bmatrix} \delta[m-n], \tag{4}$$

where $\delta[\cdot]$ denotes the Kronecker delta function. The unknown variance parameter σ^2 is assumed to be distributed a priori according to

$$\begin{aligned} p(\sigma^2) = &\quad \mathcal{IG}(\sigma^2|\alpha,\beta) \\ \triangleq &\quad \frac{\beta^\alpha}{\Gamma(\alpha)}(\sigma^2)^{-(\alpha+1)}\exp\left(-\frac{\beta}{\sigma^2}\right)\mathcal{U}(\sigma^2). \end{aligned} \quad (5)$$

where $\mathcal{IG}$ stands for an inverse-gamma probability density function (p.d.f), $\mathcal{U}(\cdot)$ denotes the unit step function, and $\{\alpha, \beta, \epsilon\} \in \mathbb{R}^+$ are the model hyperparameters.

3. PROPOSED ALGORITHM

3.1. Particle Filters

The proposed method is based on particle filters (PFs) [8], now a well-established numerical technique for solving stochastic filtering problems. Basically, PFs operate by approximating the posterior distribution of the inferred variables by a weighted sum of Dirac measures located on the so-called particles, which are in turn random draws from an arbitrarily chosen distribution known as the importance function. The application of this principle to solve the smoothing problem in (3) leads to the approximation [4]

$$p(b_{n-d}|y_{0:n}) \approx \frac{\sum_{p=1}^{P} w_n^{(p)}\delta[b_{n-d}^{(p)} - b_{n-d}]}{\sum_{q=1}^{P} w_n^{(q)}}, \quad (6)$$

where $b_{n-d}^{(p)}$ are elements of the sequences $b_{-L:n}^{(p)}$ (particles), P denotes the number of particles and $w_n^{(p)}$ their respective weights. Adopting the so-called optimal importance function [8], i.e, sampling the particles as

$$b_n^{(p)} \sim p(b_n|b_{-L:n-1}^{(p)}, y_{0:n}), \quad (7)$$

results in the weights being updated according to

$$w_n^{(p)} \propto w_{n-1}^{(p)} p(y_n|b_{-L:n-1}^{(p)}, y_{0:n-1}). \quad (8)$$

Both the p.m.f in (7) and the p.d.f in (8) can be obtained from $p(b_n, y_n|b_{-L:n-1}^{(p)}, y_{0:n-1})$, whose expression we derive in the next section. To verify this fact, observe that

$$p(b_n|b_{-L:n-1}^{(p)}, y_{0:n}) = \frac{p(b_n, y_n|b_{-L:n-1}^{(p)}, y_{0:n-1})}{\sum_{\forall b_n} p(b_n, y_n|b_{-L:n-1}^{(p)}, y_{0:n-1})}, \quad (9)$$

and that the denominator of (9) equals the weight update factor in (8). For convenience, we drop in the sequel the superscript index (p).

3.2. Determination of $p(b_n, y_n|b_{-L:n-1}, y_{0:n-1})$

Initially, observe that as a consequence of Bayes' law, we can write

$$\begin{aligned} p(b_n, y_n|b_{-L:n-1}, y_{0:n-1}) = p(y_n|b_{-L:n}, y_{0:n-1}) \\ p(b_n|b_{-L:n-1}, y_{0:n-1}). \end{aligned} \quad (10)$$

Since b_n is i.i.d, the second factor on the right-hand side (r.h.s) of (10) equals $p(b_n)$. To determine the first factor, note that the deterministic relation between $b_{-L:n}$ and $\mathbf{S}_{0:n}$ implies that [4]

$$p(y_n|b_{-L:n}, y_{0:n-1}) = p(y_n|\mathbf{S}_{0:n}, y_{0:n-1}), \quad (11)$$

where $\mathbf{S}_{0:n}$ denotes the (unique) state sequence corresponding to the bits $b_{-L:n}$. To evaluate the term on the r.h.s of (11), first observe that

$$\begin{aligned} &p(y_n|\mathbf{S}_{0:n}, y_{0:n-1}) = \\ &\int_{\mathbb{R}^+}\int_{\mathbb{C}^L} p(y_n, \mathbf{h}_n, \sigma^2|\mathbf{S}_{0:n}, y_{0:n-1})d\mathbf{h}_n d\sigma^2. \end{aligned} \quad (12)$$

Again, as a result of Bayes' law, we obtain that

$$\begin{aligned} &p(y_n, \mathbf{h}_n, \sigma^2|\mathbf{S}_{0:n}, y_{0:n-1}) = p(y_n|\mathbf{h}_n, \sigma^2, \mathbf{S}_{0:n}, y_{0:n-1}) \\ &p(\mathbf{h}_n|\sigma^2, \mathbf{S}_{0:n}, y_{0:n-1})p(\sigma^2|\mathbf{S}_{0:n}, y_{0:n-1}). \end{aligned} \quad (13)$$

Exploiting conditional independences induced by (1)-(2), one can verify that, see Appendix A,

$$p(\mathbf{h}_n|\sigma^2, \mathbf{S}_{0:n}, y_{0:n-1}) = p(\mathbf{h}_n|\sigma^2, \mathbf{S}_{0:n-1}, y_{0:n-1}), \quad (14)$$

$$p(\sigma^2|\mathbf{S}_{0:n}, y_{0:n-1}) = p(\sigma^2|\mathbf{S}_{0:n-1}, y_{0:n-1}). \quad (15)$$

From the model assumptions, it also follows that

$$p(y_n|\mathbf{h}_n, \sigma^2, \mathbf{S}_{0:n}, y_{0:n-1}) = \mathcal{N}(y_n|\mathbf{h}_n^H\mathbf{S}_n, \sigma^2), \quad (16)$$

and that $\mathbf{h}_n$ is conditional Gaussian, i.e.,

$$p(\mathbf{h}_n|\sigma^2, \mathbf{S}_{0:n-1}, y_{0:n-1}) = \mathcal{N}(\mathbf{h}_n|\hat{\mathbf{h}}_{n|n-1}, \sigma^2\Sigma_{n|n-1}), \quad (17)$$

where $\mathcal{N}$ denotes a complex Gaussian density, and $\hat{\mathbf{h}}_{n|n-1}$ and $\Sigma_{n|n-1}$ can be recursively via the following set of equations

$$\begin{aligned} &\hat{\mathbf{h}}_{n|n-1} = A\hat{\mathbf{h}}_{n-1}, \\ &\Sigma_{n|n-1} = A\Sigma_{n-1}A^H + \mathbf{I}\epsilon^2, \\ &\gamma_n = 1 + \mathbf{S}_n^H\Sigma_{n|n-1}\mathbf{S}_n, \\ &\hat{\mathbf{h}}_n = \hat{\mathbf{h}}_{n|n-1} + \gamma_n^{-1}\Sigma_{n|n-1}\mathbf{S}_n(y_n - \hat{\mathbf{h}}_{n|n-1}^H\mathbf{S}_n)^*, \\ &\Sigma_n = \Sigma_{n|n-1} - \gamma_n^{-1}\Sigma_{n|n-1}\mathbf{S}_n\mathbf{S}_n^H\Sigma_{n|n-1}. \end{aligned} \quad (18)$$

The expressions in (18) can be obtained by designing a Kalman filter to estimate $\mathbf{h}_n$ according to the model (1)-(2), supposing that $\mathbf{S}_{0:n}$ is known. Notice, however, that the actual conditional predictive and filtering covariances in the Kalman filter are $\sigma^2\Sigma_{n|n-1}$ and $\sigma^2\Sigma_n$, respectively. As it will become clear in the sequel, the ability to factor out σ^2 from (18) is fundamental to allow this quantity to be analytically integrated out. To continue the derivation, we employ the following result proven in Appendix B.

Claim 1 *For the adopted signal model, one can show that*

$$p(\sigma^2|\mathbf{S}_{0:n-1}, y_{0:n-1}) = \mathcal{IG}(\sigma^2|\alpha_{n-1}, \beta_{n-1}), \quad (19)$$

where

$$\begin{aligned} &\alpha_n = \alpha_{n-1} + 1, \\ &\beta_n = \beta_{n-1} + \gamma_n^{-1}\|y_n - \hat{\mathbf{h}}_{n|n-1}^H\mathbf{S}_n\|^2, \end{aligned} \quad (20)$$

with $\alpha_{-1} = \alpha$ and $\beta_{-1} = \beta$.

Substituting (16)-(19) into (13) leads to

$$p(y_n, \mathbf{h}_n, \sigma^2|\mathbf{S}_{0:n}, y_{0:n-1}) = \mathcal{N}(y_n|\mathbf{h}_n^H\mathbf{S}_n, \sigma^2)$$
$$\mathcal{N}(\mathbf{h}_n|\hat{\mathbf{h}}_{n|n-1}, \sigma^2\Sigma_{n|n-1})\mathcal{IG}(\sigma^2|\alpha_{n-1}, \beta_{n-1}). \quad (21)$$

After a long algebraic manipulation, one can verify that

$$\mathcal{N}(y_n|\mathbf{h}_n^H\mathbf{S}_n, \sigma^2)\mathcal{N}(\mathbf{h}_n|\hat{\mathbf{h}}_{n|n-1}, \sigma^2\Sigma_{n|n-1}) =$$
$$\mathcal{N}(y_n|\hat{\mathbf{h}}_{n|n-1}^H\mathbf{S}_n, \sigma^2\gamma_n)\mathcal{N}(\mathbf{h}_n|\hat{\mathbf{h}}_n, \sigma^2\Sigma_n). \quad (22)$$

The r.h.s of (22) is a Gaussian density in $\mathbf{h}_n$. Substituting it in (21) and noticing that the remaining parameters are not functions of $\mathbf{h}_n$, this variable can be integrated out, resulting that

$$\begin{aligned} &p(y_n, \sigma^2|\mathbf{S}_{0:n}, y_{0:n-1}) = \\ &= \mathcal{N}(y_n|\hat{\mathbf{h}}_{n|n-1}^H\mathbf{S}_n, \sigma^2\gamma_n)\mathcal{IG}(\sigma^2|\alpha_{n-1}, \beta_{n-1}) \\ &= \left[\beta_{n-1}^{\alpha_{n-1}}(\sigma^2)^{-(\alpha_{n-1}+2)}\right] / \left[\Gamma(\alpha_{n-1})\pi\gamma_n\right] \\ &\quad\times \exp\left\{-\sigma^{-2}\left[\gamma_n^{-1}\|\hat{\mathbf{h}}_{n|n-1}^H\mathbf{S}_n - y_n\|^2 + \beta_{n-1}\right]\right\} \\ &= \left[\beta_{n-1}^{\alpha_{n-1}}\Gamma(\alpha_n)\right] / \left[\beta_n^{\alpha_n}\Gamma(\alpha_{n-1})\pi\gamma_n\right]\mathcal{IG}(\sigma^2|\alpha_n, \beta_n). \end{aligned} \quad (23)$$

As none of the parameters of (23) are dependent on σ^2, integrating out this variable is equivalent to discarding the inverse-gamma distribution, from which we finally get that

$$p(y_n|\mathbf{S}_{0:n}, y_{0:n-1}) = \frac{\Gamma(\alpha_n)}{\Gamma(\alpha_{n-1})} \cdot \frac{1}{\pi\gamma_n} \cdot \frac{\beta_{n-1}^{\alpha_{n-1}}}{\beta_n^{\alpha_n}}, \quad (24)$$

which via (11) allows one to evaluate (7) and (8). The parameters α_n and β_n can be obtained via (20), which in turn is dependent on the outcomes of the Kalman filter (18).

4. RESULTS

In order to assess the performance of the proposed equalization scheme, we carried out numerical simulations evaluating the BER (bit error rate) as a function of the signal-to-noise ratio E_b/N_0. Simulations consisted of 400 independent realizations, in each of which a block of 300 i.i.d binary symbols was transmitted. BER estimation was made after discarding the first 100 symbols to allow for algorithm convergence. For all algorithms, the 1% extreme results were discarded. The filter employs $P = 300$ particles and smoothing lag of $d = 2$ samples. A resampling step was carried out at every iteration using the residual algorithm [9].

In the following simulations, we assumed that $\mathbf{A} = 0.99\mathbf{I}$, $\epsilon^2 = 10^{-2}$, $\alpha = 1$, $\beta = 0.1$, and $L = 3$. Since the assumed prior model does not lead to constant E_b/N_0, the channel coefficients were obtained according to the mismatched model

$$\mathbf{h}_{n+1} = \frac{\mathbf{A}\mathbf{h}_n + \mathbf{w}_n}{\|\mathbf{A}\mathbf{h}_n + \mathbf{w}_n\|}, \quad (25)$$

where $\mathbf{w}_n \sim \mathcal{N}(0; \mathbf{I}\sigma^2\epsilon^2)$. The value of σ^2 was determined so as to lead to the desired signal-to-noise ratio according to the relation

$$E_b/N_0 = \|\mathbf{h}_n\|^2/\sigma^2 = 1/\sigma^2. \quad (26)$$

In Figure 1, we display the results obtained for the proposed algorithm (solid line). For comparison, the same figure depicts the performance of the MAP equalizer (BCJR) (◯) and that of the linear equalizer based on the Kalman filter (▽). The Kalman filter was designed to estimate $\mathbf{S}_n$; smoothed estimates with a lag of $d = 2$ samples were obtained via the relation $\hat{s}_{n-d} = [\hat{\mathbf{S}}_n]_{d+1}$, where $[\cdot]_l$ denotes the $l-$th element of the vector in brackets. Both alternative algorithms operate on blocks of 300 samples with exact knowledge of channel and noise variance parameters, being therefore not subject to model mismatches. For the alternative algorithms, differential decoding was performed separately from equalization.

As one might verify, the proposed blind algorithm outperforms the linear least squares (Kalman) approach for E_b/N_0 levels greater than 6dB, exhibiting a performance penalty of $2 - 3$dB in comparison to the optimal MAP estimate.

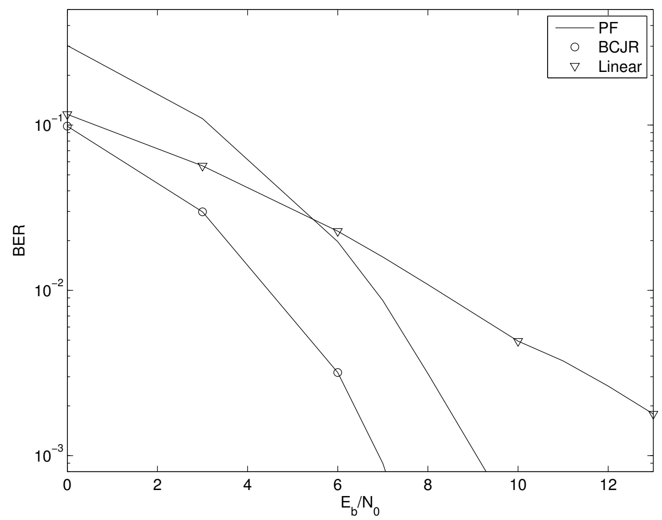

Fig. 1. Performance of the proposed blind equalization algorithm (PF) as a function of E_b/N_0 compared to the optimal (BCJR) and to the Kalman equalizer (Linear).

5. CONCLUSIONS

We presented in this article a new blind equalization algorithm for time-varying frequency-selective channels corrupted by unknown variance additive Gaussian noise. Our main contribution was to introduce a new method for integrating out the unknown noise variance σ^2, which relies on the choice of a dynamic conjugate prior model and on a new scheme for recursively determining the noise posterior distribution. As we verified via numerical simulations, the proposed method exhibits a performance gap of only $2 - 3$dB in comparison to the MAP equalizer based on the BCJR algorithm, and outperforms a Kalman filter based equalizer for $E_b/N_0 > 6$dB. As a final remark, note that the proposed method requires that an independent Kalman filter be run for each particle, resulting in a computational complexity of $\mathcal{O}(PL^2)$, roughly equivalent to that of the pioneering algorithm in [1], which assumed a time-invariant channel with known noise variance.

6. REFERENCES

[1] J. S. Liu and R. Chen, "Blind Deconvolution via Sequential Imputations," *J. of the Am. Stat. Assoc.*, vol. 90, no. 430, pp. 567–576, June 1995.

[2] P. M. Djurić, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Míguez, "Particle Filtering," *IEEE Signal Process. Mag.*, vol. 20, no. 5, pp. 19–38, September 2003.

[3] M.J. Daly, J.P. Reilly, and M.R. Morelande, "Rao-Blackwellised Particle Filtering for Blind System Identification," in *Proc. of ICASSP*, 2005, vol. IV, pp. 321–324.

[4] C. J. Bordin Jr. and M. G. S. Bruno, "A Rao-Blackwellized Particle Filter for Blind Equalization of Frequency-Selective Channels with Unknown Order and Noise Variance," in *Proc. of ICASSP*, 2007, vol. III, pp. 1273–1276.

[5] J. Míguez and P. M. Djurić, "Blind Equalization of Frequency-Selective Channels by Sequential Importance Sampling," *IEEE Trans. Signal Process.*, vol. 52, no. 10, October 2004.

[6] A. B. Gelman, *Bayesian Data Analysis*, Chapman & Hall / CRC Press, 2004.

[7] J. R. Hopgood and P. J. W. Rayner, "Blind Single Channel Deconvolution using Nonstationary Signal Processing," *IEEE Trans. Signal Process.*, vol. 11, no. 5, pp. 476–488, 2003.

[8] A. Doucet, S. Godsill, and C. Andrieu, "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering," *Statistics and Computing*, vol. 10, pp. 197–208, 2000.

[9] T. Higuchi, "Monte Carlo Filter using the Genetic Algorithm Operators," *J. of Statistical Computation and Simulation*, vol. 59, no. 1, pp. 1–23, 1997.

A. PROOF OF EQUATIONS (14)-(15)

To verify (14)-(15), note that as a consequence of the definition of conditional p.d.f, it follows that

$$\frac{p(\mathbf{h}_n|\sigma^2, \mathbf{S}_{0:n}, y_{0:n-1})}{p(\mathbf{h}_n|\sigma^2, \mathbf{S}_{0:n-1}, y_{0:n-1})} = \frac{p(\mathbf{S}_n|\mathbf{h}_n, \sigma^2, \mathbf{S}_{0:n-1}, y_{0:n-1})}{p(\mathbf{S}_n|\sigma^2, \mathbf{S}_{0:n-1}, y_{0:n-1})}, \tag{27}$$

$$\frac{p(\sigma^2|\mathbf{S}_{0:n}, y_{0:n-1})}{p(\sigma^2|\mathbf{S}_{0:n-1}, y_{0:n-1})} = \frac{p(\mathbf{S}_n|\sigma^2, \mathbf{S}_{0:n-1}, y_{0:n-1})}{p(\mathbf{S}_n|\mathbf{S}_{0:n-1}, y_{0:n-1})}. \tag{28}$$

As $\{\mathbf{S}_n\}$ is Markovian, the densities on the numerators and denominators of the right-hand sides of (27) and (28) equal $p(\mathbf{S}_n|\mathbf{S}_{0:n-1})$. This implies that the ratios on the left-hand sides of (27) and (28) equal 1, leading to the results stated in (14)-(15).

B. PROOF OF EQUATIONS (19)-(20)

Initially, observe that

$$p(\sigma^2|\mathbf{S}_{0:n}, y_{0:n}) = \frac{p(\mathbf{S}_n, y_n, \sigma^2|\mathbf{S}_{0:n-1}, y_{0:n-1})}{\int_{\mathbb{R}+} p(\mathbf{S}_n, y_n, \sigma^2|\mathbf{S}_{0:n-1}, y_{0:n-1})d\sigma^2}. \tag{29}$$

The density on the r.h.s of (29) factors as

$$p(\mathbf{S}_n, y_n, \sigma^2|\mathbf{S}_{0:n-1}, y_{0:n-1}) = p(\sigma^2|\mathbf{S}_{0:n-1}, y_{0:n-1}) \times p(\mathbf{S}_n|\mathbf{S}_{0:n-1}, y_{0:n-1}, \sigma^2)p(y_n|\mathbf{S}_{0:n}, y_{0:n-1}, \sigma^2), \tag{30}$$

Taking into consideration conditional independences induced by (1)-(2), we can rewrite (30) as

$$p(\mathbf{S}_n, y_n, \sigma^2|\mathbf{S}_{0:n-1}, y_{0:n-1}) = p(\sigma^2|\mathbf{S}_{0:n-1}, y_{0:n-1}) \times p(\mathbf{S}_n|\mathbf{S}_{0:n-1})p(y_n|\mathbf{S}_{0:n}, y_{0:n-1}, \sigma^2), \tag{31}$$

The second factor on the r.h.s of (31) is a discrete distribution that assumes an equal value for all valid $\mathbf{S}_n$. To determine the third term, observe that

$$p(y_n|\mathbf{S}_{0:n-1}, y_{0:n-1}, \sigma^2) = \int_{\mathbb{C}^L} p(y_n, \mathbf{h}_n|\mathbf{S}_{0:n}, y_{0:n-1}, \sigma^2)d\mathbf{h}_n. \tag{32}$$

Similarly to Sec. 3.2, one can verify that

$$\begin{aligned} &p(y_n, \mathbf{h}_n|\mathbf{S}_{0:n}, y_{0:n-1}, \sigma^2) = \\ &= p(y_n|\mathbf{h}_n, \mathbf{S}_n, \sigma^2)p(\mathbf{h}_n|\mathbf{S}_{0:n-1}, y_{0:n-1}, \sigma^2) \\ &= \mathcal{N}(y_n|\mathbf{h}_n^H\mathbf{S}_n, \sigma^2)\mathcal{N}(\mathbf{h}_n|\hat{\mathbf{h}}_{n|n-1}, \sigma^2\Sigma_{n|n-1}) \\ &= \mathcal{N}(y_n|\hat{\mathbf{h}}_{n|n-1}^H\mathbf{S}_n, \sigma^2\gamma_n)\mathcal{N}(\mathbf{h}_n|\hat{\mathbf{h}}_n, \sigma^2\Sigma_n) \end{aligned} \tag{33}$$

which implies that

$$p(y_n|\mathbf{S}_{0:n-1}, y_{0:n-1}, \sigma^2) = \mathcal{N}(y_n|\hat{\mathbf{h}}_{n|n-1}^H\mathbf{S}_n, \sigma^2\gamma_n). \tag{34}$$

Assume now, as an induction hypothesis, that (19)-(20) are valid for $n-1$, i.e.,

$$p(\sigma^2|\mathbf{S}_{0:n-1}, y_{0:n-1}) = \mathcal{IG}(\sigma^2|\alpha_{n-1}, \beta_{n-1}). \tag{35}$$

The induction step consists then in verifying whether (35) is valid for n. For this purpose, we substitute (34)-(35) into (31), obtaining that

$$p(\mathbf{S}_n, y_n, \sigma^2|\mathbf{S}_{0:n-1}, y_{0:n-1}) = p(\mathbf{S}_n|\mathbf{S}_{0:n-1}) \times\mathcal{N}(y_n|\hat{\mathbf{h}}_{n|n-1}^H\mathbf{S}_n, \sigma^2\gamma_n)\mathcal{IG}(\sigma^2|\alpha_{n-1}, \beta_{n-1}). \tag{36}$$

Algebraic manipulations similar to those of (23) now lead to

$$p(\mathbf{S}_n, y_n, \sigma^2|\mathbf{S}_{0:n-1}, y_{0:n-1}) = p(\mathbf{S}_n|\mathbf{S}_{0:n-1}) \left[\beta_{n-1}^{\alpha_{n-1}}\Gamma(\alpha_n)\right] / \left[\beta_n^{\alpha_n}\Gamma(\alpha_{n-1})\pi\gamma_n\right]\mathcal{IG}(\sigma^2|\alpha_n, \beta_n). \tag{37}$$

Substituting (37) in (29) finally results that

$$p(\sigma^2|\mathbf{S}_{0:n}, y_{0:n}) = \mathcal{IG}(\sigma^2|\alpha_n, \beta_n). \tag{38}$$

As (35)-(38) are valid for $n \geq 0$, the principle of finite induction confirms (19)-(20).