Closed-Form Expressions for Channel Shortening Receivers Using *A Priori* Information

Albert Abelló, Jean-Marie Freixe and Damien Roque, *Member, IEEE*

***Abstract*—Channel shortening has been studied in the context of ISI and MIMO channels as a means to compute *a posteriori* probabilities with a BCJR algorithm at a reduced computational complexity. This is done by considering an approximate channel response of reduced length. In a turbo receiver, soft *a priori* information can be linearly combined with the received sequence to form a new input to the BCJR trellis-based processing. In this letter, we provide closed-form expressions for the channel shortening filters using a generalized mutual information objective function. The proposed receiver allows a complexity reduction with respect to numerical optimization approaches which may also present stability, precision and convergence issues.**

***Index Terms*—inter-symbol interference, turbo equalization, channel shortening, iterative receivers.**

I. Introduction

Channel shortening (CS) was first explored in [1] as a means to reduce the complexity of trellis-based detection in the presence of inter-symbol interference (ISI). This approach considers a linear preprocessing followed by a trellis-based nonlinear processing. A complexity reduction with respect to the optimal approach is obtained by considering an approximate channel response (the target response) of reduced length. The receiving filter and the target response are properly optimized. Recent work on channel shortening considers the mutual information as the optimization criterion [2] and closed-form expressions for the receiving and target response filters have been obtained by assuming that input symbols follow a Gaussian distribution. Channel shortening has been extended to the optimization of the transmit filters [3] and a channel shortening receiver with an additional return filter using *a priori* information has been introduced in [4], [5] although no closed-form expressions have been provided for the return filter and the target response so far. In [6], the authors observe that further constraints must be put on the return channel shortening filter in order for the receiver to perform interference cancellation in a turbo equalization scheme. These additional constraints do not allow a closed-form derivation of the channel shortening solution with *a priori* information. In this letter, we consider a more general case in which we do not restrict the choice of the symbol estimator and we consider a Gaussian model for *a priori* information. We first provide both the preprocessing and return filters in closed-form. This result allows us to provide the target response in closed-form by applying the method in [2] to our resulting generalized mutual information. In a turbo receiver, this implies that all filters are obtained in closed-form at each iteration step thus avoiding the use of more complex numerical optimization methods. We validate the proposed closed-form solution by comparing it to a standard numerical optimization approach and we demonstrate practical performance results by considering a typical Gaussian model for *a priori* information.

II. Capacity approaching channel shortening

The received signal at the output of the additive, white, Gaussian noise (AWGN) channel can be described by the following discrete-time model

$$\boldsymbol{x} = \boldsymbol{H}\boldsymbol{s} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{s} = [s_0, \dots, s_{N-1}]^T$ is the sequence of transmitted symbols with $(\cdot)^T$ the transpose operator, $\boldsymbol{x} = [x_0, \dots, x_{N-1}]^T$ is the sequence of received symbols and $\boldsymbol{n} = [n_0, \dots, n_{N-1}]^T$ is the sequence of circularly-symmetric independent complex Gaussian noise samples with variance σ_n^2. The ISI channel is described by a Toeplitz matrix $\boldsymbol{H}$ whose first column has $L+1$ nonzero elements $\boldsymbol{h} = [h_0, h_1, \dots, h_L]^T$ forming the channel impulse response. The conditional probability density function of $\boldsymbol{x}$ is given by [7]

$$p(\boldsymbol{x}|\boldsymbol{s}) \propto \exp\left(\frac{1}{\sigma_n^2}\left(2\Re\left\{\boldsymbol{s}^H\boldsymbol{H}^H\boldsymbol{x}\right\} - \boldsymbol{s}^H\boldsymbol{G}\boldsymbol{s}\right)\right) \tag{2}$$

where $\boldsymbol{G} = \boldsymbol{H}^H\boldsymbol{H}$ is the channel response after the matched filter with $(\cdot)^H$ the conjugate transpose operator.

On the receiver side, optimal turbo equalization algorithms need for the computation of symbol *a posteriori* probabilities $p(s_k|\boldsymbol{x})$, $k = 0, \dots, N-1$ [8] . A suitable factorization of (2) can be used by the BCJR algorithm to compute exact *a posteriori* symbol probabilities [9] although this approach yields a complexity that is exponential in L. To overcome this complexity limitation, channel shortening [2], [4] considers a modified channel law

$$\tilde{p}(\boldsymbol{x}|\boldsymbol{s}) \propto \exp\left(\frac{1}{\sigma_n^2}\left(2\Re\left\{\boldsymbol{s}^H\left(\boldsymbol{F}\boldsymbol{x} - \boldsymbol{B}\hat{\boldsymbol{s}}\right)\right\} - \boldsymbol{s}^H\boldsymbol{G}_{\mathrm{r}}\boldsymbol{s}\right)\right) \tag{3}$$

where $\boldsymbol{F}$ acts as a linear preprocessing of the observation $\boldsymbol{x}$ and the band matrix $\boldsymbol{G}$ is replaced by a target response $\boldsymbol{G}_{\mathrm{r}}$ of reduced length $\nu \le L$

This work was supported by the French armament procurement agency under Grant 10/2015/DGA. The associate editor coordinating the review of this paper and approving it for publication was C. Zhong.

A. Abelló and J-M. Freixe are with Eutelsat S.A., 75015 Paris, FRANCE. email: aabellob@eutelsat.com.

D. Roque is with the Institut Supérieur de l'Aéronautique et de l'Espace (ISAE-SUPAERO), Université de Toulouse, 31055 Toulouse, FRANCE.

Digital Object Identifier 10.1109/LCOMM.2018.2830329

$$(\boldsymbol{G}_{\mathrm{r}})_{i,j} = 0 \quad | i - j | > \nu \tag{4}$$

with $(\boldsymbol{G}_{\mathrm{r}})_{i,j}$ the (i,j)th term of the target response. If soft *a priori* information $\hat{\boldsymbol{s}}$ is available to the channel shortening receiver [4], it can be further preprocessed by $\boldsymbol{B}$ forming the input sequence

$$\boldsymbol{y} = \boldsymbol{F}\boldsymbol{x} - \boldsymbol{B}\hat{\boldsymbol{s}}. \tag{5}$$

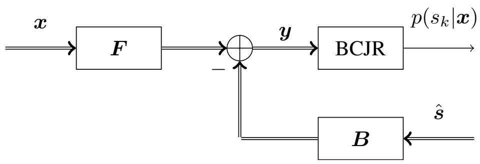

Fig. 1. Channel shortening receiver with soft *a priori* information for the computation of symbol *a posteriori* probabilities (APP).

The modified channel law (3) along with (5) are used by the BCJR algorithm to compute approximate *a posteriori* symbol probabilities. The channel shortening receiver is depicted in Figure 1. In [4], the channel shortening filters are obtained with standard numerical optimization methods by considering a generalized mutual information objective function

$$I_{\mathrm{G}} = \mathrm{E}\left\{\log\left(\tilde{p}\left(\boldsymbol{x}|\boldsymbol{s}\right)\right)\right\} - \mathrm{E}\left\{\log\left(\tilde{p}\left(\boldsymbol{x}\right)\right)\right\} \tag{6}$$

with

$$\tilde{p}\left(\boldsymbol{x}\right) = \int_{-\infty}^{\infty} \tilde{p}\left(\boldsymbol{x}|\boldsymbol{s}\right) p\left(\boldsymbol{s}\right) \mathrm{d}\boldsymbol{s}$$

and $\mathrm{E}\left\{\cdot\right\}$ the expectation operator. $\boldsymbol{F}$ and $\boldsymbol{B}$ are subject to an optimization on (6) of the form

$$(\boldsymbol{F}_{\mathrm{cs}}, \boldsymbol{B}_{\mathrm{cs}}) = \underset{\boldsymbol{F},\boldsymbol{B}}{\operatorname{argmax}}\ I_{\mathrm{G}}. \tag{7}$$

To obtain a closed-form solution, we assume that input symbols follow a circularly-symmetric Gaussian distribution [2] with variance σ_s^2. We assume that input symbols and *a priori* symbols have a correlation matrix $\mathrm{E}\left\{\boldsymbol{s}\hat{\boldsymbol{s}}^H\right\} = \boldsymbol{C}_1$ and that *a priori* symbols have a correlation matrix $\mathrm{E}\left\{\hat{\boldsymbol{s}}\hat{\boldsymbol{s}}^H\right\} = \boldsymbol{C}_0$. We further assume that *a priori* symbols are uncorrelated to the channel noise samples $\mathrm{E}\left\{\boldsymbol{n}\hat{\boldsymbol{s}}^H\right\} = \boldsymbol{0}$. The optimization process is as follows. First, $\boldsymbol{F}$ and $\boldsymbol{B}$ are obtained by maximizing (6) for a given target response.

Proposition 1: the optimization (7) yields optimal filters

$$\boldsymbol{F}_{\mathrm{cs}} = (\boldsymbol{G}_{\mathrm{r}} + \boldsymbol{I})\left(\boldsymbol{H}^H\boldsymbol{H} + \boldsymbol{\Gamma}\right)^{-1}\boldsymbol{H}^H, \tag{8}$$

$$\boldsymbol{B}_{\mathrm{cs}} = (\boldsymbol{G}_{\mathrm{r}} + \boldsymbol{I})\left(\left(\boldsymbol{H}^H\boldsymbol{H} + \boldsymbol{\Gamma}\right)^{-1}\boldsymbol{H}^H\boldsymbol{H} - \boldsymbol{I}\right)\boldsymbol{C}_1\boldsymbol{C}_0^{-1} \tag{9}$$

with

$$\boldsymbol{\Gamma} = \sigma_n^2\left(\boldsymbol{I} - \boldsymbol{C}_1\boldsymbol{C}_0^{-1}\boldsymbol{C}_1^H\right)^{-1}.$$

Proof is given in Appendix A. Due to the Hermitian property of $\boldsymbol{G}_{\mathrm{r}}$ and the optimization (7), $\boldsymbol{F}_{\mathrm{cs}}$ and $\boldsymbol{B}_{\mathrm{cs}}$ are band Toeplitz matrices. A filter processing can thus be implemented on the receiver side with a complexity quasi linear in the length of the receiving filters. With optimum $\boldsymbol{F}_{\mathrm{cs}}$ and $\boldsymbol{B}_{\mathrm{cs}}$, the objective function becomes

$$I_{\mathrm{G}} = \log\left(\det\left\{\boldsymbol{G}_{\mathrm{r}} + \boldsymbol{I}\right\}\right) + N + \mathrm{Tr}\left\{\left(\boldsymbol{G}_{\mathrm{r}} + \boldsymbol{I}\right)\boldsymbol{A}\right\}, \tag{10}$$

$$\boldsymbol{A} = \sigma_n^2\boldsymbol{H}^H\boldsymbol{\Gamma}^{-1}\boldsymbol{H}\left(\boldsymbol{H}^H\boldsymbol{H} + \boldsymbol{\Gamma}\right)^{-1}. \tag{11}$$

The optimal $\boldsymbol{G}_{\mathrm{r}}$ can be provided in closed-form by applying the method proposed in [2] to the resulting generalized mutual information (10) which takes into account *a priori* information. This is done by taking the derivative of (10) and by considering a Cholesky factorization of the term $\boldsymbol{G}_{\mathrm{r}} + \boldsymbol{I}$. The optimization process is summarized in Appendix B.

It can be shown that the linear minimum mean square error (MMSE) approach using *a priori* information is a particular case of rate maximizing channel shortening reception. The MMSE filters $\boldsymbol{F}_{\mathrm{mse}}$ and $\boldsymbol{B}_{\mathrm{mse}}$ are well known [10] and can be found by setting

$$(\boldsymbol{F}_{\mathrm{mse}}, \boldsymbol{B}_{\mathrm{mse}}) = \underset{\boldsymbol{F},\boldsymbol{B}}{\operatorname{argmin}}\ \mathrm{E}\left\{\|\boldsymbol{y} - \boldsymbol{s}\|^2\right\}.$$

The optimum filters can be put on the form

$$\boldsymbol{F}_{\mathrm{mse}} = \left(\boldsymbol{H}^H\boldsymbol{H} + \boldsymbol{\Gamma}\right)^{-1}\boldsymbol{H}^H,$$

$$\boldsymbol{B}_{\mathrm{mse}} = \left(\left(\boldsymbol{H}^H\boldsymbol{H} + \boldsymbol{\Gamma}\right)^{-1}\boldsymbol{H}^H\boldsymbol{H} - \boldsymbol{I}\right)\boldsymbol{C}_1\boldsymbol{C}_0^{-1}.$$

Therefore, the channel shortening filters equal the MMSE filters weighted by $\boldsymbol{G}_{\mathrm{r}} + \boldsymbol{I}$. The two approaches are equivalent when $\nu = 0$ (*i.e.*, $\boldsymbol{G}_{\mathrm{r}} = \boldsymbol{I}$, no trellis-based processing is performed). The full complexity receiver is found by setting $\nu = L$ (*i.e.*, $\boldsymbol{G}_{\mathrm{r}} = \boldsymbol{H}^H\boldsymbol{H}$) which yields the matched filter $\boldsymbol{F}_{\mathrm{cs}} = \boldsymbol{H}^H$ and $\boldsymbol{B}_{\mathrm{cs}} = -\boldsymbol{\Gamma}\boldsymbol{C}_1\boldsymbol{C}_0^{-1}$.

If no *a priori* information is available to the receiver, then $\boldsymbol{C}_1 = \boldsymbol{C}_0 = \boldsymbol{0}$, $\boldsymbol{\Gamma} = \sigma_n^2\boldsymbol{I}$. In this case, $\boldsymbol{F}_{\mathrm{cs}}$ equals the classical MMSE receiver. On the other hand, with perfect *a priori* information $\hat{\boldsymbol{s}} = \boldsymbol{s}$ and if we consider $\sigma_s^2 = 1$ then $\boldsymbol{C}_1 = \boldsymbol{C}_0 = \boldsymbol{I}$ and $\boldsymbol{\Gamma} = \infty\boldsymbol{I}$. In this situation, $\boldsymbol{F}_{\mathrm{cs}} = \boldsymbol{0}$ and $\boldsymbol{B}_{\mathrm{cs}} = -\left(\boldsymbol{G}_{\mathrm{r}} + \boldsymbol{I}\right)$. The noiseless case $\sigma_n^2 = 0$ yields the zero forcing solution weighted by the target response $\boldsymbol{F}_{\mathrm{cs}} = \boldsymbol{G}_{\mathrm{r}}\boldsymbol{H}^{-1}$ and $\boldsymbol{B}_{\mathrm{cs}} = \boldsymbol{0}$.

III. Numerical results

The transmitted uncoded symbols are obtained using a binary phase-shift keying (BPSK) mapping. For numerical simulation, we assume $\boldsymbol{C}_1 = c_1\boldsymbol{I}$ and $\boldsymbol{C}_0 = c_0\boldsymbol{I}$ so that $\boldsymbol{\Gamma} = \gamma\boldsymbol{I}$ with $\gamma = \sigma_n^2\left(1 - |c_1|^2/c_0\right)^{-1}$. Performance of the proposed channel shortening receiver has been evaluated by assuming an additive white Gaussian noise model for soft *a priori* information $\hat{\boldsymbol{s}} = \boldsymbol{s} + \boldsymbol{e}$ with $\boldsymbol{e}$ a sequence of circularly-symmetric independent complex Gaussian noise samples with

TABLE I
SE BETWEEN CLOSED-FORM AND NUMERICAL FILTER OPTIMIZATION

γ	$\boldsymbol{g}_{\mathrm{r}}$	$\boldsymbol{f}_{\mathrm{cs}}$	$\boldsymbol{b}_{\mathrm{cs}}$
	$\nu = 2$		
0.2	1.5832e-06	1.3836e-06	1.8215e-06
0.4	3.0729e-06	2.3371e-06	3.3412e-06
0.6	7.1155e-06	2.3204e-06	2.7197e-06
0.8	2.2225e-05	3.6139e-06	2.6077e-06
1.0	7.1042e-05	8.3007e-06	4.7152e-06
	$\nu = 3$		
0.2	2.7235e-06	7.5011e-06	8.7205e-06
0.4	4.5422e-06	1.7081e-05	2.0377e-05
0.6	4.2465e-06	1.5438e-05	1.8783e-05
0.8	7.5227e-06	1.3169e-05	1.5589e-05
1.0	2.4116e-05	1.3224e-05	1.3622e-05

variance σ_e^2. This yields correlation coefficients $c_1 = 1$, $c_0 = 1+\sigma_e^2$. In a turbo receiver [8], *a priori* information varies at each iteration step and the receiving filters $\boldsymbol{F}_{\mathrm{cs}}$, $\boldsymbol{B}_{\mathrm{cs}}$ and $\boldsymbol{G}_{\mathrm{r}}$ can be recalculated accordingly in closed-form as described in Section II.

Table I shows the squared error between the receiving filters obtained in closed-form and obtained iteratively using the gradient projection method [11] on the generalized mutual information derived in (14) as a function of γ and ν for $\sigma_n^2 = 2$. The receiving filters $\boldsymbol{f}_{\mathrm{cs}}$, $\boldsymbol{b}_{\mathrm{cs}}$ and $\boldsymbol{g}_{\mathrm{r}}$ correspond to a row of $\boldsymbol{F}_{\mathrm{cs}}$, $\boldsymbol{B}_{\mathrm{cs}}$ and $\boldsymbol{G}_{\mathrm{r}}$ respectively. We define the error as $\mathrm{SE} = k^{-1}\|\boldsymbol{f}_{\mathrm{cs}} - \boldsymbol{f}_{\mathrm{cs}}^{\mathrm{i}}\|^2$ where the superscript $(\cdot)^{\mathrm{i}}$ means that the iterative method was considered and k is the filter length. Results confirm that a numerical optimization of $\boldsymbol{F}$, $\boldsymbol{B}$ and $\boldsymbol{G}_{\mathrm{r}}$ converges toward the optimal closed-form solution.

To better illustrate the complexity reduction obtained with the closed-form solution, please note that we consider in general $N \times N$ band Toeplitz matrices and we focus on the case $\boldsymbol{C}_1 = c_1\boldsymbol{I}$ and $\boldsymbol{C}_0 = c_0\boldsymbol{I}$. The direct method for computing the optimal shortening filters yields a complexity in $O(N^2)$. On the other hand, the gradient projection method, that we used to validate the closed-form solution, has a complexity in $O(\epsilon \cdot N^2)$ with ϵ the number of iterations performed by the algorithm. The number of iterations needed depends on γ, σ_n^2 and ν. In the worst case simulation scenario, we needed 2000 iterations to achieve convergence. For a variety of numerical optimization methods the complexity remains $O(\epsilon \cdot N^2)$. Other than the complexity advantage, the proposed closed-form solution avoids stability, convergence and precision issues that must be dealed with when numerical optimization approaches are used.

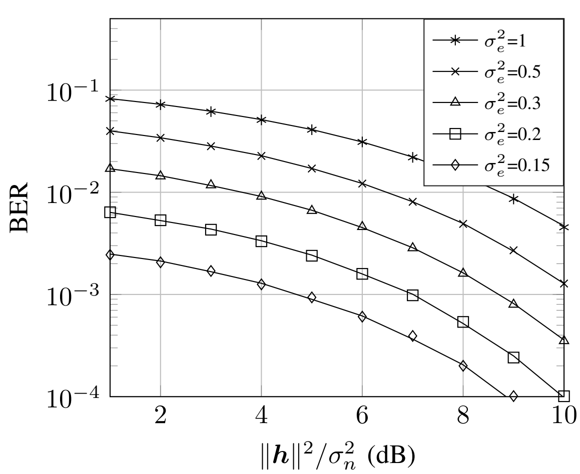

Fig. 2. Performance results of the channel shortening receiver for $\nu = 2$ over the partial response channel $\boldsymbol{h} = [0.033, \text{-}0.131, \text{-}0.037, 0.446, 0.750, 0.446, \text{-}0.037, \text{-}0.131, 0.033]^T$.

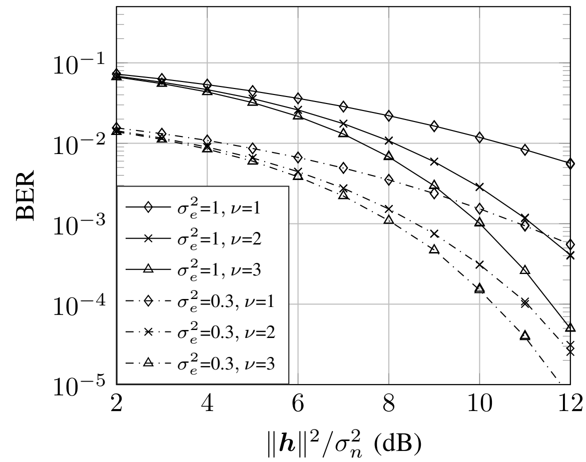

Fig. 3. Performance results of the channel shortening receiver over the EPR4 channel $\boldsymbol{h} = [0.5, 0.5, -0.5, -0.5]^T$.

Receiver performance is depicted in Figures 2 and 3. The lines correspond to the optimal closed-form solution. The markers correspond to the iterative solution after convergence. The squared errors obtained between the closed-form and the iterative approaches prove to be sufficiently low in terms of bit error rate degradation.

IV. CONCLUSION

In this letter, we have provided a generalization of the rate-maximizing channel shortening receiver using *a priori* information. The receiving filters have been provided in closed-form at each turbo iteration step thus avoiding the use of more complex numerical optimization methods. Additionally, this result enables the derivation of the linear MMSE turbo approach as a particular case of channel shortening when *a priori* information is available to the receiver. Numerical results show that standard iterative optimization methods converge toward the optimal closed-form solution. Future work may study a suitable soft-symbol estimator for the optimal closed-form channel shortening solution.

APPENDIX

A. Proof of proposition 1

We assume $\boldsymbol{s} \sim \mathcal{CN}\left(\boldsymbol{0}, \sigma_s^2\boldsymbol{I}\right)$. For ease of reading and without loss of generality, we take $\sigma_s^2 = 1$ and we introduce σ_n^2 in $\boldsymbol{F}, \boldsymbol{B}$ and $\boldsymbol{G}_{\mathrm{r}}$ respectively, yielding

$$\tilde{p}(\boldsymbol{x}) = \frac{K(\boldsymbol{x})}{\pi^N}\int_{\boldsymbol{s}} \exp\left(2\Re\left\{\boldsymbol{s}^H\boldsymbol{y}\right\} - \boldsymbol{s}^H\left(\boldsymbol{G}_{\mathrm{r}} + \boldsymbol{I}\right)\boldsymbol{s}\right)\mathrm{d}\boldsymbol{s},$$

$$K(\boldsymbol{x}) = \frac{1}{(\pi\sigma_n^2)^N} \exp\left(-\frac{\boldsymbol{x}^H\boldsymbol{x}}{\sigma_n^2}\right).$$

We restrict $\boldsymbol{G}_\mathrm{r}$ to being Hermitian so that rearranging into squared forms yields

$$\tilde{p}(\boldsymbol{x}) = K(\boldsymbol{x}) \det\left\{(\boldsymbol{G}_\mathrm{r}+\boldsymbol{I})^{-1}\right\} \exp\left(\boldsymbol{y}^H (\boldsymbol{G}_\mathrm{r}+\boldsymbol{I})^{-1} \boldsymbol{y}\right).$$

The mutual information is

$$\begin{aligned} I_\mathrm{G} =& \mathrm{E}\left\{\log\left(\tilde{p}\left(\boldsymbol{x}|\boldsymbol{s}\right)\right)\right\} - \mathrm{E}\left\{\log\left(\tilde{p}\left(\boldsymbol{x}\right)\right)\right\} \\ =& 2\Re\left\{\mathrm{E}\left\{\boldsymbol{s}^H\boldsymbol{y}\right\}\right\} - \mathrm{E}\left\{\boldsymbol{s}^H\boldsymbol{G}_\mathrm{r}\boldsymbol{s}\right\} \\ &+ \log\left(\det\left\{\boldsymbol{G}_\mathrm{r}+\boldsymbol{I}\right\}\right) - \mathrm{E}\left\{\boldsymbol{y}^H\left(\boldsymbol{G}_\mathrm{r}+\boldsymbol{I}\right)^{-1}\boldsymbol{y}\right\}, \end{aligned} \quad (12)$$

$$\begin{aligned} &\mathrm{E}\left\{\boldsymbol{s}^H\boldsymbol{y}\right\} = \mathrm{Tr}\left\{\boldsymbol{F}\boldsymbol{H} - \boldsymbol{B}\boldsymbol{C}_1^H\right\}, \\ &\mathrm{E}\left\{\boldsymbol{s}^H\boldsymbol{G}_\mathrm{r}\boldsymbol{s}\right\} = \mathrm{Tr}\left\{\boldsymbol{G}_\mathrm{r}\right\}, \\ &\mathrm{E}\left\{\boldsymbol{y}^H\left(\boldsymbol{G}_\mathrm{r}+\boldsymbol{I}\right)^{-1}\boldsymbol{y}\right\} = \mathrm{Tr}\left\{\left(\boldsymbol{F}\left(\boldsymbol{H}\boldsymbol{H}^H + \sigma_n^2\boldsymbol{I}\right)\boldsymbol{F}^H \right.\right. \\ &\left.\left. +\boldsymbol{B}\boldsymbol{C}_0\boldsymbol{B}^H - 2\Re\left\{\boldsymbol{F}\boldsymbol{H}\boldsymbol{C}_1\boldsymbol{B}^H\right\}\right)\left(\boldsymbol{G}_\mathrm{r}+\boldsymbol{I}\right)^{-1}\right\}. \end{aligned} \quad (13)$$

Introducing (13) in (12) yields

$$\begin{aligned} &I_\mathrm{G} = \log\left(\det\left\{\boldsymbol{G}_\mathrm{r}+\boldsymbol{I}\right\}\right) - \mathrm{Tr}\left\{\boldsymbol{G}_\mathrm{r}\right\} \\ &+ 2\Re\left\{\mathrm{Tr}\left\{\boldsymbol{F}\boldsymbol{H} - \boldsymbol{B}\boldsymbol{C}_1^H\right\}\right\} - \mathrm{Tr}\left\{\left(\boldsymbol{F}\left(\boldsymbol{H}\boldsymbol{H}^H + \sigma_n^2\boldsymbol{I}\right)\boldsymbol{F}^H \right.\right. \\ &\left.\left. +\boldsymbol{B}\boldsymbol{C}_0\boldsymbol{B}^H - 2\Re\left\{\boldsymbol{F}\boldsymbol{H}\boldsymbol{C}_1\boldsymbol{B}^H\right\}\right)\left(\boldsymbol{G}_\mathrm{r}+\boldsymbol{I}\right)^{-1}\right\}. \end{aligned} \quad (14)$$

The optimal filter $\boldsymbol{F}_\mathrm{cs}$ is obtained from (14) by setting $\nabla_{\boldsymbol{F}_\mathrm{cs}} I_\mathrm{G} = \boldsymbol{0}$ which yields after some calculations

$$\boldsymbol{F}_\mathrm{cs} = \left(\boldsymbol{B}\boldsymbol{C}_1^H + (\boldsymbol{G}_\mathrm{r}+\boldsymbol{I})\right)\boldsymbol{H}^H\left(\boldsymbol{H}\boldsymbol{H}^H + \sigma_n^2\boldsymbol{I}\right)^{-1}. \quad (15)$$

We now set $\nabla_{\boldsymbol{B}_\mathrm{cs}} I_\mathrm{G} = \boldsymbol{0}$ which yields after some calculations

$$\boldsymbol{B}_\mathrm{cs} = \left(\boldsymbol{F}\boldsymbol{H} - (\boldsymbol{G}_\mathrm{r}+\boldsymbol{I})\right)\boldsymbol{C}_1\boldsymbol{C}_0^{-1}. \quad (16)$$

Introducing (16) in (15) yields optimum filters (8) and (9).

B. *Optimization of $\boldsymbol{G}_r$*

$\boldsymbol{G}_\mathrm{r}+\boldsymbol{I}$ is assumed to be a Hermitian positive-definite matrix allowing the Cholesky factorization $\boldsymbol{G}_\mathrm{r}+\boldsymbol{I} = \boldsymbol{U}\boldsymbol{U}^H$ with $\boldsymbol{U}$ an upper triangular matrix

$$(\boldsymbol{U})_{m,n} = \begin{cases} u_{m,n} & \text{if } n \geq m \\ 0 & \text{otherwise.} \end{cases}$$

Then, (10) can be rewritten as

$$I_\mathrm{G} = \left(2\sum_{n=1}^{N}\log(u_{n,n}) - \mathrm{Tr}\left\{\boldsymbol{U}\boldsymbol{A}\boldsymbol{U}^H\right\} + N\right). \quad (17)$$

Since the first term in (17) does not contain off-diagonal values, I_G can be optimized over the diagonal and off-diagonal elements separately [2]

$$\begin{aligned} \boldsymbol{U} = \underset{u_{n,n}}{\mathrm{argmax}} & \left(2\sum_{n=1}^{N}\log(u_{n,n}) + N \right. \\ & \left. - \underset{\{u_{m,n}\}_{m+1\leq n\leq \min(m+\nu,N)}}{\mathrm{argmin}} \mathrm{Tr}\left\{\boldsymbol{U}\boldsymbol{A}\boldsymbol{U}^H\right\}\right). \end{aligned} \quad (18)$$

By defining the submatrix

$$\boldsymbol{A}_n^\nu = \begin{pmatrix} a_{n+1,n+1} & \cdots & a_{n+1,\min(N,n+\nu)} \\ \vdots & \ddots & \vdots \\ a_{\min(N,n+\nu),n+1} & \cdots & a_{\min(N,n+\nu),\min(N,n+\nu)} \end{pmatrix}$$

and the row vectors $\boldsymbol{a}_n^\nu = \left[a_{n,n+1}, \cdots, a_{n,\min(N,n+\nu)}\right]$ and $\boldsymbol{u}_n^\nu = \left[u_{n,n+1}, \cdots, u_{n,\min(N,n+\nu)}\right]$ then

$$\mathrm{Tr}\left\{\boldsymbol{U}\boldsymbol{A}\boldsymbol{U}^H\right\} = \sum_{n=1}^{N}\begin{pmatrix} u_{n,n} & \boldsymbol{u}_n^\nu \end{pmatrix}\begin{pmatrix} a_{n,n} & \boldsymbol{a}_n^\nu \\ \boldsymbol{a}_n^{\nu\,H} & \boldsymbol{A}_n^\nu \end{pmatrix}\begin{pmatrix} u_{n,n} \\ \boldsymbol{u}_n^{\nu\,H} \end{pmatrix}.$$

Setting $\nabla_{\boldsymbol{u}_n^\nu}\mathrm{Tr}\left\{\boldsymbol{U}\boldsymbol{A}\boldsymbol{U}^H\right\} = \boldsymbol{0}$ yields $\boldsymbol{u}_n^\nu = -u_{n,n}\boldsymbol{a}_n^\nu(\boldsymbol{A}_n^\nu)^{-1}$. Introducing $\boldsymbol{u}_n^\nu$ in (18) yields

$$\boldsymbol{U} = \underset{u_{n,n}}{\mathrm{argmax}}\left(2\sum_{n=1}^{N}\log(u_{n,n}) + N - \sum_{n=1}^{N}u_{n,n}^2 c_n\right) \quad (19)$$

where $c_n = a_{n,n} - \boldsymbol{a}_n^\nu(\boldsymbol{A}_n^\nu)^{-1}\boldsymbol{a}_n^{\nu\,H}$. Setting the derivative of (19) to zero yields $u_{n,n} = \sqrt{1/c_n}$.

References

[1] D. D. Falconer and F. R. Magee, "Adaptive channel memory truncation for maximum likelihood sequence estimation," *Bell Syst. Techn. J.*, vol. 52, no. 9, pp. 1541–1562, Nov. 1973.

[2] F. Rusek and A. Prlja, "Optimal channel shortening for MIMO and ISI channels," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 810–818, Feb. 2012.

[3] A. Modenini, F. Rusek, and G. Colavolpe, "Optimal transmit filters for constrained complexity channel shortening detectors," in *2013 IEEE Int. Conf. Commun. (ICC)*, Jun. 2013, pp. 3095–3100.

[4] F. Rusek, N. Al-Dhahir, and A. Gomaa, "A rate-maximizing channel-shortening detector with soft feedback side information," in *2012 IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2012, pp. 2256–2261.

[5] S. Hu, H. Kröll, Q. Huang, and F. Rusek, "Optimal channel shortener design for reduced- state soft-output viterbi equalizer in single-carrier systems," *IEEE Trans. Commun.*, vol. 65, no. 6, pp. 2568–2582, Jun. 2017.

[6] S. Hu and F. Rusek, "On the design of channel shortening demodulators for iterative receivers in MIMO and ISI channels," *CoRR*, vol. abs/1506.07331, 2015.

[7] G. Ungerboeck, "Adaptive maximum-likelihood receiver for carrier-modulated data-transmission systems," *IEEE Trans. Commun.*, vol. 22, no. 5, pp. 624–636, May 1974.

[8] M. Tüchler and A. Singer, "Turbo equalization: An overview," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 920–952, Feb. 2011.

[9] G. Colavolpe and A. Barbieri, "On MAP symbol detection for ISI channels using the Ungerboeck observation model," *IEEE Commun. Lett.*, vol. 9, no. 8, pp. 720–722, Aug. 2005.

[10] M. Tüchler, A. C. Singer, and R. Koetter, "Minimum mean squared error equalization using a priori information," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 673–683, Mar. 2002.

[11] J. Rosen, "The gradient projection method for nonlinear programming. part I. linear constraints," *J. Soc. Ind. Appl. Mathematics*, vol. 8, no. 1, pp. 181–217, Mar. 1960.