Channel Shortening for Nonlinear Satellite Channels

Giulio Colavolpe, *Senior Member, IEEE*, Andrea Modenini, *Student Member, IEEE*, and Fredrik Rusek, *Member, IEEE*

***Abstract*—We design of an efficient channel shortener for nonlinear satellite channels. When the memory of the channel is too large to be taken into account by a full complexity detector, excellent performance can be achieved by properly filtering the received signal followed by a reduced-state detector. This letter derives closed-form expressions for the front-end filter and the target response of the reduced-state detector.**

***Index Terms*—Nonlinear satellite channels, channel shortening, information rate, intersymbol interference, receiver optimization.**

I. Introduction

SATELLITE channels are affected by nonlinear distortions and by intersymbol interference (ISI) of possibly long duration. The former originate from the presence of a high power amplifier (HPA), whereas the latter is introduced by the input and output multiplexing (IMUX and OMUX) filters placed before and after the HPA. During the last decades in the literature, the nonlinear effects and the channel memory have been coped with nonlinear compensation and data predistortion at the transmitter side (see [1] and references therein) or with advanced detection techniques (see [2] and references therein).

When the channel memory is too large to be taken into account at the detector, these advanced detection techniques quickly become unmanageble and low-complexity solutions are required. The conceptually simplest solution is to let the detector work with a truncated version of the channel response. However, as expected, such a strategy often yields poor performance unless the truncated part of the channel response has negligible power.

Channel shortening (CS) is a technique originally proposed by Falconer and Magee in 1973 [3] and recently improved in [4] for general linear channels, such as multiple-input multiple-output (MIMO) and ISI channels. By using an information-theoretic framework, the optimal front-end filter (the *channel shortener*) can be stated in closed form for a given channel memory considered at the detector.

In this paper, we generalize the analysis in [4] to maximum-a-posteriori (MAP) detection for nonlinear satellite channels. In Section II, we briefly review the system model for the satellite channel and the underlying detection algorithm assumed in this letter. In Section III, we extend the channel shortening technique and in Section IV we assess its performance by numerical simulations.

Manuscript received August 27, 2012. The associate editor coordinating the review of this letter and approving it for publication was D. B. da Costa.

G. Colavolpe and A. Modenini are with the Università di Parma, Dipartimento di Ingegneria dell'Informazione, Viale G. P. Usberti, 181A, 43124 Parma, Italy (e-mail: giulio@unipr.it, modenini@tlc.unipr.it).

F. Rusek is with Lund University, Department of Electrical and Information Technology, Box 118, SE-221 00 Lund, Sweden (e-mail: fredrik.rusek@eit.lth.se). The work of F. Rusek was supported by SSF through the distributed antenna project.

Digital Object Identifier 10.1109/LCOMM.2012.12.121929

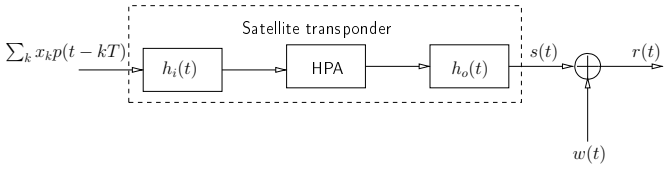

Fig. 1. Block diagram of the satellite channel.

Notation: Matrices are denoted by uppercase boldface letters (e.g., $\boldsymbol{H}$) and vectors by lowercase boldface letters (e.g., $\boldsymbol{x}$). Block matrices and block vectors will be denoted with roman boldface letters (e.g., $\mathbf{H}$ and $\mathbf{x}$). The identity matrix is denoted as $\mathbf{I}$ or $\boldsymbol{I}$ depending on the context. The entry (i,j) of block matrix $\mathbf{H}$ will be denoted as $\boldsymbol{H}_{ij}$ (matrix entry). The transpose of a matrix $\boldsymbol{H}$ will be denoted by $\boldsymbol{H}^T$, its conjugate by $\boldsymbol{H}^*$, and its conjugate and transpose by $\boldsymbol{H}^\dagger$. For a sequence of matrices $\{\boldsymbol{H}_k\}$ we use $\boldsymbol{H}(\omega)$ to denote its discrete-time Fourier transform (DTFT) $\boldsymbol{H}(\omega)=\sum_k \boldsymbol{H}_k e^{-j\omega k}$. The inverse $\boldsymbol{H}(\omega)^{-1}$ is such that $(\boldsymbol{H}(\omega))^{-1}\boldsymbol{H}(\omega)=\boldsymbol{H}(\omega)(\boldsymbol{H}(\omega))^{-1}=\mathbf{I}$ for each ω.

II. System model and Considered Detector

We consider a linear modulation with shaping pulse $p(t)$, symbol time T, and uniformly and identically distributed input symbols $\{x_n\}$ belonging to an M-ary constellation, properly normalized such that $\mathrm{E}\{|x_n|^2\}=1$. The nonlinear satellite channel, considering a single-channel-per-transponder scenario, is depicted in Fig. 1. It includes an IMUX filter $h_i(t)$ which removes the adjacent channels, a HPA, and an OMUX filter $h_o(t)$ aimed at reducing the spectral broadening caused by the nonlinear amplifier. Although the HPA is a nonlinear memoryless device, the overall system has memory due to the presence of IMUX and OMUX filters. The received signal is further corrupted by additive white Gaussian noise whose low-pass equivalent $w(t)$ has power spectral density $2N_0$. The complex baseband representation of the received signal has thus the following expression

$$r(t)=s(t)+w(t)\,, \tag{1}$$

where $s(t)$ is the signal at the output of the OMUX filter.

In [2], it is shown that a suitable approximate model for the signal $s(t)$ is based on the following nth-order (with n being any odd integer) *simplified* Volterra-series expansion

$$s(t)\simeq\sum_k \sum_{i=0}^{N_V-1} x_k \left[|x_k|^{2i} h^{(2i+1)}(t-kT)\right], \tag{2}$$

where $N_V=(n+1)/2$, and $h^{(2i+1)}(t)$ are complex waveforms given by linear combinations of the the original N_V Volterra kernels. This simplified Volterra-series expansion is obtained

from the classical one by neglecting some suitable terms. For further details, the reader can refer to [2]. We point out that the approximation (2) is used only for the receiver design and not for generating the received signal $r(t)$.

It is easy to show that MAP symbol detection based on this simplified model can be performed through a bank of filters followed by a conventional Bahl-Cocke-Jelinek-Raviv (BCJR) detector [5] with proper branch metrics and working on a trellis whose number of states exponentially depends on the channel memory. When the actual channel memory is large, we have to resort to complexity reduction techniques. A possible approach is the use of reduced-state techniques (e.g., see [6]) or the use of the graph-based technique described in [2] whose complexity linearly depends on the channel memory. However, to obtain a further complexity reduction, all these techniques can be combined with the CS technique described in [4] properly extended to the channel at hand.

We will separately consider the cases of phase-shift keying (PSK) modulations and amplitude/phase shift keying (APSK) modulations typically employed in satellite transmissions.

PSK modulations: It can be seen that the condition $|x_i|^2 = 1$ implies that the signal (2) simplifies to a linear modulation with shaping pulse $\bar{h}(t) = \sum_{i=0}^{N_V-1} h^{(2i+1)}(t)$ [2]. In this case, detection can be perfomed using the samples $\{y_i\}$ at the output of a filter matched to $\bar{h}(t)$ as described in [7], and the application of CS can be carried out as described in [4] for linear channels.

APSK modulations: The samples at the output of a bank of filters matched to the pulses $h^{(2i+1)}(t)$, $i = 0, ..., N_V - 1$ form a set of sufficient statistics for detection. Namely, considering an nth-order expansion, we have N_V matched filters whose output, sampled at discrete time k can be collected in a $N_V \times 1$ vector that can be expressed as

$$\boldsymbol{y}_k = \sum_i \boldsymbol{G}_i \boldsymbol{x}_{k-i} + \boldsymbol{\eta}_k \,, \tag{3}$$

where $\boldsymbol{x}_k = \left[x_k,\ x_k|x_k|^2, ..., x_k|x_k|^{n-1}\right]^T$,

$$\boldsymbol{G}_i = \begin{pmatrix} g_i^{(1,1)} & g_i^{(1,3)} & \cdots & g_i^{(1,n)} \\ g_{-i}^{(1,3)*} & g_i^{(3,3)} & \cdots & g_i^{(3,n)} \\ \vdots & & \ddots & \vdots \\ g_{-i}^{(1,n)*} & g_{-i}^{(3,n)*} & \cdots & g_i^{(n,n)} \end{pmatrix}, \tag{4}$$

having defined $g_i^{(m,\ell)} = \int_{-\infty}^{\infty} h^{(\ell)}(t) h^{(m)*}(t - iT)\mathrm{d}t$, and $\boldsymbol{\eta}_k$ is a Gaussian vector with

$$\mathrm{E}\{\boldsymbol{\eta}_{k+i}\boldsymbol{\eta}_k^\dagger\} = \boldsymbol{G}_i \,. \tag{5}$$

Vectors $\{\boldsymbol{y}_k\}$ can be collected into a single vector

$$\mathbf{y} = \mathbf{G}\mathbf{x} + \boldsymbol{\eta} \,, \tag{6}$$

where $\mathbf{G}$ is a block Toeplitz matrix constructed from the matrices $\{\boldsymbol{G}_i\}$, whereas $\mathbf{x}$ and $\boldsymbol{\eta}$ are block vectors from $\{\boldsymbol{x}_k\}$ and $\{\boldsymbol{\eta}_k\}$. The channel is fully described through its conditional probability density function of the output given the input symbols, which reads

$$p(\mathbf{y}|\mathbf{x}) \propto \exp\left(\frac{2\mathcal{R}(\mathbf{x}^\dagger\mathbf{y}) - \mathbf{x}^\dagger\mathbf{G}\mathbf{x}}{2N_0}\right) \tag{7}$$

where $\mathcal{R}(\cdot)$ is the real part.

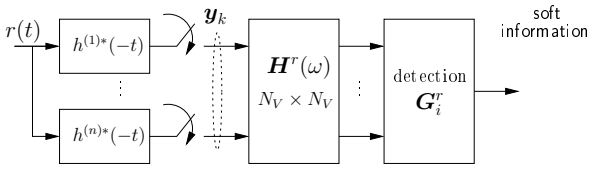

Fig. 2. Block diagram of the suboptimal receiver for the nonlinear satellite channel.

According to the CS approach, a low-complexity detector works on a mismatched channel law [4]

$$\tilde{p}(\mathbf{y}|\mathbf{x}) \propto \exp\left(2\mathcal{R}(\mathbf{x}^\dagger\mathbf{H}^r\mathbf{y}) - \mathbf{x}^\dagger\mathbf{G}^r\mathbf{x}\right) \,, \tag{8}$$

where, by using the notation in [4] as far as the superscript r is concerned, $\mathbf{H}^r$,$\mathbf{G}^r$ are block Toeplitz matrices constructed from the sequences $\{\boldsymbol{H}_i^r\}$ and $\{\boldsymbol{G}_i^r\}$, respectively, being $\{\boldsymbol{H}_i^r\}$ the channel shortener operating on $\mathbf{y}$, and $\{\boldsymbol{G}_i^r\}$ the target response, to be properly designed. Without loss of generality we absorb the noise variance $2N_0$ into the two matrices in (8). In order to reduce the detection complexity, we constrain $\{\boldsymbol{G}_i^r\}$ to

$$\boldsymbol{G}_i^r = \mathbf{0} \quad |i| > L \tag{9}$$

which implies that the memory after CS is L instead of the true memory of the channel. The resulting receiver is suboptimal since it assumes (8) rather than the actual law (7), and is depicted in Fig. 2.

III. Channel shortening

The achievable information rate (AIR) I_R of a mismatched detector that works with (8) is given by [8, Section VI]

$$I_R = \lim_{N\to\infty} \frac{1}{N} \mathrm{E}_{\mathbf{y},\mathbf{x}} \left\{ \log_2 \frac{\tilde{p}(\mathbf{y}|\mathbf{x})}{\tilde{p}(\mathbf{y})} \right\} \text{ [bit/ch.use]} \tag{10}$$

where N is the number of transmitted symbols and the average is carried out w.r.t. $\mathbf{y}$, $\mathbf{x}$ according to the actual channel.

The CS technique finds the optimal $\mathbf{H}^r$,$\mathbf{G}^r$ solving the following optimization problem [4]

$$\arg \max_{\mathbf{H}^r, \mathbf{G}^r} I_R \tag{11}$$

under the constraints specified in (9). Problem (11) for a discrete alphabet is a complicated task. However it can be solved in closed form under the assumption that $\mathbf{x}$ is composed of Gaussian random variables. Although this assumption is not even approximately true, since the actual symbols are functions of each other, we will show in the simulation results that a very good performance is still achieved.

Defining the correlation matrix $\boldsymbol{V} = \mathrm{E}\{\boldsymbol{x}_k\boldsymbol{x}_k^\dagger\}$ the optimal matrix-valued front-end filter $\{\boldsymbol{H}_i^r\}$ and target response $\{\boldsymbol{G}_i^r\}$ are obtained in closed form through the following steps:

- Compute the DTFT matrix $\boldsymbol{G}(\omega)$ of $\boldsymbol{G}_i$ and use the spectral decomposition to find $\boldsymbol{L}(\omega)$, i.e., decompose $\boldsymbol{G}(\omega) = \boldsymbol{L}^\dagger(\omega)\boldsymbol{L}(\omega)$. Compute

$$\boldsymbol{B}(\omega) = 2N_0\boldsymbol{V}\boldsymbol{L}^\dagger(\omega) \cdot \left[\boldsymbol{L}(\omega)\boldsymbol{V}\boldsymbol{L}^\dagger(\omega) + 2N_0\boldsymbol{I}\right]^{-1} (\boldsymbol{L}^\dagger(\omega))^{-1} \,. \tag{12}$$

The anti trasform yields the matrix sequence $\{\boldsymbol{B}_k\}$.

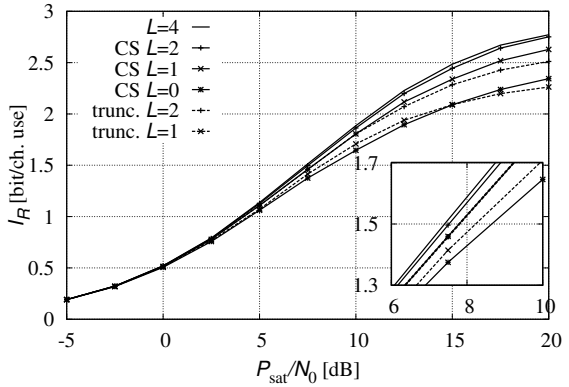

Fig. 3. AIR for 8PSK modulation on the nonlinear satellite channel with IBO=0 dB.

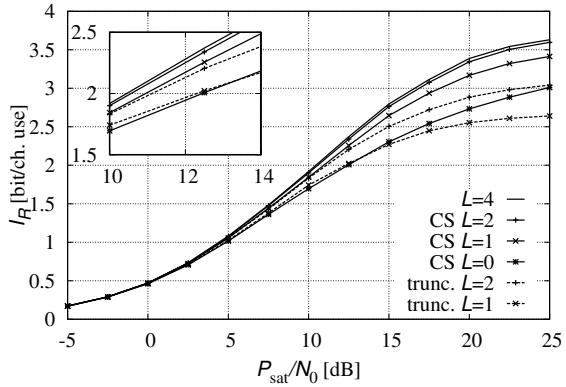

Fig. 4. AIR for 16APSK modulation on the nonlinear satellite channel with IBO=3 dB.

- Find

$$\boldsymbol{C} = \boldsymbol{B}_0 - \underline{\mathbf{B}}\mathbf{B}^{-1}\underline{\mathbf{B}}^{\dagger} \tag{13}$$

  where we defined the block matrix $\underline{\mathbf{B}} = [\boldsymbol{B}_1, ..., \boldsymbol{B}_L]$ with size $1 \times L$ and the block Toeplitz $\mathbf{B}$ with size $L \times L$ constructed on $\{\boldsymbol{B}_k\}$.
- Define the sequence $\{\boldsymbol{U}_k\}$ where $\boldsymbol{U}_0$ is the Cholesky decomposition of $\boldsymbol{C}$, namely $\boldsymbol{C} = \boldsymbol{U}_0^{\dagger}\boldsymbol{U}_0$, and $\boldsymbol{U}_k$ for $1 \le k \le L$ is the $(1,k)$ entry of $\underline{\mathbf{U}} = -\boldsymbol{U}_0\underline{\mathbf{B}}\mathbf{B}^{-1}$.
- Set

$$\boldsymbol{G}_k^r = \sum_{i=\max(0,k)}^{\min(L,L+k)} \boldsymbol{U}_{i-k}^{\dagger}\boldsymbol{U}_i - \boldsymbol{V}\delta_k \tag{14}$$

  where δ_k is the Kronecker delta.
- The optimal front-end filter is given by

$$\boldsymbol{H}^r(\omega) = \left(\boldsymbol{G}^r(\omega) + \boldsymbol{V}^{-1}\right) \cdot \boldsymbol{V}\boldsymbol{L}^{\dagger}(\omega)\left[\boldsymbol{L}(\omega)\boldsymbol{V}\boldsymbol{L}^{\dagger}(\omega) + 2N_0\boldsymbol{I}\right]^{-1}\left(\boldsymbol{L}^{\dagger}(\omega)\right)^{-1}. \tag{15}$$

For the proof see the appendix. We point out that by analogy to [4] for linear channels, when $L = 0$ the optimal channel shortener equals the MMSE filter of [9] applied to (2).

IV. Numerical results

We consider 8-PSK and 16-APSK modulations. The shaping pulse $p(t)$ has a root-raised-cosine (RRC) spectrum with roll-off 0.05. The IMUX and OMUX filters have frequency characteristics specified in [10] with a 3dB bandwidth of $0.94/T$ and $0.85/T$ respectively. The nonlinear transfer characteristic is the Saleh model [11] with parameters $\alpha_a = 2.1322$, $\alpha_\phi = 1.7054$, $\beta_a = 1.0746$, and $\beta_\phi = 1.5072$. A 5th-order Volterra expansion is considered at the receiver. We report all results as functions of ratio between the normalized power at the saturation $P_{\rm sat}$ and the noise power spectral density N_0.

The AIR in eq. (10) can be computed using the Monte Carlo method described in [8]. Fig. 3 shows the AIR values when CS is employed in combination with a 8PSK modulation, and an input back-off (IBO) equal to zero. Results are shown for different values of the detector memory L and an optimization of the noise variance at receiver has been carried out to further improve the approximate model. For comparison, we show also the AIR values when a simple truncation of the ISI at the detector is adopted. The detector with $L = 4$ can be considered as effective as a full complexity one, since most of the ISI is taken into account[1]. It can be seen that CS has higher AIR than a simple truncation of the ISI response, even though it is designed for a vector $\mathbf{x}$ with Gaussian components. CS with memory $L = 2$ gives only a minimal performance degradation for all $P_{\rm sat}/N_0$ values. Similar conclusions hold for the 16APSK, depicted in Fig. 4. We found similar CS gains also with other modulations (QPSK and 32APSK) and other transponder characteristics (e.g., the HPA in [10]).

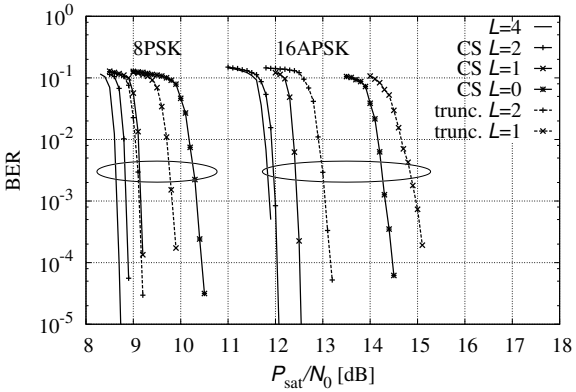

Fig. 5. BER for 8PSK and 16APSK modulation, with the DVB-S2 LDPC code having rate 1/2.

The AIRs can be approached in practice with proper modulation and coding (MODCODs) formats. In Fig. 5 we report the bit error rate (BER) of some MODCODs based on the DVB-S2 low-density parity-check code (LDPC) rate 1/2 [10]. We performed iterative detection and decoding with a maximum of 50 global iterations. We note that the MODCODs performance reflects the AIRs well.

V. Conclusion

We generalized the CS technique to the case of nonlinear satellite channels. We showed that when the memory L is lower than the channel memory, an optimization of the mismatched channel law at the detector yields significantly better performance than a truncation of the channel impulse response.

Appendix

In this appendix we derive the CS solution for APSK modulations. The channel in (6) can be whitened using the

[1]The pulse with RRC spectrum gives an infinite memory of the channel. However, based on investigations beyond those presented in this letter, we may assume that $L = 4$ is almost optimal.

$N_V \times N_V$ whitening filter (WF) $(\boldsymbol{L}^\dagger(\omega))^{-1}$. Using matrix notation, the whitened observable $\mathbf{r}$ for N transmitted symbols provides a set of sufficient statistics and reads [12]

$$\mathbf{r} = \mathbf{L}\mathbf{x} + \mathbf{w}\,, \tag{16}$$

where $\mathbf{L}$ is an $N \times N$ block Toeplitz matrix built from $\{\boldsymbol{L}_k\}$. If $\mathbf{x}$ is a complex vector with mean zero and covariance matrix $\mathbf{V} = \mathrm{E}\{\mathbf{x}\mathbf{x}^\dagger\}$ we have

$$\begin{aligned}\tilde{p}(\mathbf{r}) &= \frac{1}{\pi^{N_V N}\det(\mathbf{V})}\int \tilde{p}\left(\mathbf{r}|\mathbf{x}\right)\exp\left\{-\mathbf{x}^\dagger\mathbf{V}^{-1}\mathbf{x}\right\}\mathrm{d}\mathbf{x} \\ &= \frac{1}{\det(\mathbf{G}^r\mathbf{V}+\mathbf{I})}\exp\{\mathbf{d}^\dagger\left(\mathbf{G}^r+\mathbf{V}^{-1}\right)^{-1}\mathbf{d}\} \quad (17)\end{aligned}$$

where $\mathbf{d} = \mathbf{L}^r\mathbf{r}$ and $\mathbf{L}^r = \mathbf{L}(\mathbf{H}^r)^\dagger$. Therefore,

$$\begin{aligned}&\mathrm{E}_{\mathbf{r}}\{-\log\tilde{p}(\mathbf{r})\} = \log\det(\mathbf{G}^r\mathbf{V}+\mathbf{I}) \\ &-\mathrm{Tr}\left((\mathbf{L}^r)^\dagger\left[\mathbf{L}\mathbf{V}\mathbf{L}^\dagger + 2N_0\mathbf{I}\right]\mathbf{L}^r(\mathbf{G}^r+\mathbf{V}^{-1})^{-1}\right) \quad (18)\end{aligned}$$

and

$$\mathrm{E}_{\mathbf{r},\mathbf{x}}\{-\log\tilde{p}(\mathbf{r}|\mathbf{x})\} = \mathrm{Tr}(\mathbf{G}^r\mathbf{V}) - 2\mathcal{R}\left(\mathrm{Tr}\left((\mathbf{L}^r)^\dagger\mathbf{L}\mathbf{V}\right)\right). \tag{19}$$

The I_R for complex Gaussian symbols is

$$I_R = \mathrm{E}_{\mathbf{r}}\{-\log\tilde{p}(\mathbf{r})\} - \mathrm{E}_{\mathbf{r},\mathbf{x}}\{-\log\tilde{p}(\mathbf{r}|\mathbf{x})\} \tag{20}$$

whose derivative w.r.t. $(\mathbf{L}^r)^\dagger$ is

$$\frac{\partial I_R}{\partial(\mathbf{L}^r)^\dagger} = (\mathbf{L}\mathbf{V})^T - \left(\left[\mathbf{L}\mathbf{V}\mathbf{L}^\dagger + 2N_0\mathbf{I}\right]\mathbf{L}^r(\mathbf{G}^r+\mathbf{V}^{-1})^{-1}\right)^T \tag{21}$$

which gives that the optimal filter $\mathbf{H}^r$ is

$$\left(\mathbf{G}^r+\mathbf{V}^{-1}\right)\mathbf{V}\mathbf{L}^\dagger\left[\mathbf{L}\mathbf{V}\mathbf{L}^\dagger + 2N_0\mathbf{I}\right]^{-1}(\mathbf{L}^\dagger)^{-1}. \tag{22}$$

Using (22), the I_R is

$$I_R = \frac{1}{N}\left(\log(\det(\mathbf{U}^\dagger\mathbf{U}\mathbf{V})) - \mathrm{Tr}\left(\mathbf{U}\mathbf{B}\mathbf{U}^\dagger\right) + N_V N\right) \tag{23}$$

where $\mathbf{U}$ is obtained from the Cholesky decomposition $\mathbf{G}^r + \mathbf{V}^{-1} = \mathbf{U}^\dagger\mathbf{U}$ and

$$\mathbf{B} = \mathbf{V} - \mathbf{V}\mathbf{L}^\dagger\left[\mathbf{L}\mathbf{V}\mathbf{L}^\dagger + 2N_0\mathbf{I}\right]^{-1}\mathbf{L}\mathbf{V}\,.$$

Since $\det(\mathbf{U}^\dagger\mathbf{U}\mathbf{V})$ depends only on the diagonal elements of $\mathbf{U}$, we can optimize I_R over the $\mathbf{U}$ diagonal and the off-diagonal elements separately. We define $\underline{\mathbf{U}}_n = [\boldsymbol{U}_{n\,n+1},..,\boldsymbol{U}_{n\,\min(n+L,N)}]$, $\underline{\mathbf{B}}_n = [\boldsymbol{B}_{n\,n+1},..,\boldsymbol{B}_{n\,\min(n+L,N)}]$,

$$\mathbf{B}_n = \begin{bmatrix} \boldsymbol{B}_{(n+1)\,(n+1)} & \cdots & \boldsymbol{B}_{\min(n+1,L)\,(n+L)} \\ \vdots & \ddots & \vdots \\ \boldsymbol{B}_{\min(n+L,N)\,(n+1)} & \cdots & \boldsymbol{B}_{\min(n+1,L)\,\min(n+1,L)} \end{bmatrix}$$

and finally

$$\boldsymbol{C}_n = \boldsymbol{B}_{nn} - \underline{\mathbf{B}}_n\mathbf{B}_n^{-1}(\underline{\mathbf{B}}_n)^\dagger\,.$$

Now the trace $\mathrm{Tr}\left(\mathbf{U}\mathbf{B}\mathbf{U}^\dagger\right)$ can be rewritten as

$$\sum_n \mathrm{Tr}\left([\boldsymbol{U}_{nn}\,\underline{\mathbf{U}}_n]\begin{bmatrix}\boldsymbol{B}_{nn} & \underline{\mathbf{B}}_n \\ \underline{\mathbf{B}}_n^\dagger & \mathbf{B}_n\end{bmatrix}\begin{bmatrix}\boldsymbol{U}_{nn}^\dagger \\ \underline{\mathbf{U}}_n^\dagger\end{bmatrix}\right).$$

Setting its derivative w.r.t. $\underline{\mathbf{U}}_n^\dagger$ to zero gives

$$\frac{\partial}{\partial\underline{\mathbf{U}}_n^\dagger}\mathrm{Tr}\left(\mathbf{U}\mathbf{B}\mathbf{U}^\dagger\right) = (\boldsymbol{U}_{nn}\underline{\mathbf{B}}_n)^T + (\underline{\mathbf{U}}_n\mathbf{B}_n)^T = 0$$

which gives

$$\underline{\mathbf{U}}_n = -\boldsymbol{U}_{nn}\underline{\mathbf{B}}_n\mathbf{B}_n^{-1}\,. \tag{24}$$

Replacing (24) in (23) we find

$$\begin{aligned}I_R &= \frac{1}{N}\log\det(\mathbf{V}) + N_V \\ &+ \frac{1}{N}\sum_n \log(\det(\boldsymbol{U}_{nn}^\dagger\boldsymbol{U}_{nn})) - \mathrm{Tr}\left(\boldsymbol{U}_{nn}\boldsymbol{C}_n\boldsymbol{U}_{nn}^\dagger\right) \quad (25)\end{aligned}$$

that can be maximized by setting its derivative w.r.t. $\boldsymbol{U}_{nn}^\dagger$ equal to zero. This gives that

$$\frac{\partial I_R}{\partial\boldsymbol{U}_{nn}^\dagger} = (\boldsymbol{U}_{nn}^*)^{-1} - (\boldsymbol{U}_{nn}\boldsymbol{C}_n)^T = 0$$

and the optimal $\boldsymbol{U}_{nn}$ is given by the Cholesky decomposition

$$\boldsymbol{C}_n^{-1} = \boldsymbol{U}_{nn}^\dagger\boldsymbol{U}_{nn}\,. \tag{26}$$

Inserting (26) into (25), the AIR for Gaussian symbols is

$$I_R = \frac{1}{N}\log\det(\mathbf{V}) + \frac{1}{N}\sum_n \log(\det(\boldsymbol{C}_n^{-1}))\,. \tag{27}$$

When $N \to \infty$, $\mathbf{B}_n$ and $\underline{\mathbf{B}}_n$ are the same for all n, with elements given by the anti trasform of (12). Thus, using Szegö's Theorem [13], the I_R tends to be the integral of the stationary solution in (12)-(15).

REFERENCES

[1] A. N. D'Andrea, V. Lottici, and R. Reggiannini, "RF power amplifier linearization through amplitude and phase predistortion," *IEEE Trans. Commun.*, vol. 44, pp. 1477–1484, Nov. 1996.

[2] G. Colavolpe and A. Piemontese, "Novel SISO detection algorithms for nonlinear satellite channels," *IEEE Wireless Commun. Lett.*, vol. 1, pp. 22–25, Feb. 2012.

[3] D. D. Falconer and F. Magee, "Adaptive channel memory truncation for maximum likelihood sequence estimation," *Bell System Tech. J.*, vol. 52, pp. 1541–1562, Nov. 1973.

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

[5] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. 20, pp. 284–287, Mar. 1974.

[6] D. Fertonani, A. Barbieri, and G. Colavolpe, "Reduced-complexity BCJR algorithm for turbo equalization," *IEEE Trans. Commun.*, vol. 55, pp. 2279–2287, Dec. 2007.

[7] G. Colavolpe and A. Barbieri, "On MAP symbol detection for ISI channels using the Ungerboeck observation model," *IEEE Commun. Lett.*, vol. 9, pp. 720–722, Aug. 2005.

[8] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3498–3508, Aug. 2006.

[9] S. Benedetto and E. Biglieri, "Nonlinear equalization of digital satellite channels," *IEEE J. Sel. Areas Commun.*, vol. 1, pp. 57–62, Jan. 1983.

[10] ETSI EN 301 307 Digital Video Broadcasting (DVB); V1.1.2 (2006-06), Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other Broadband satellite applications, 2006. Available: http://www.etsi.org.

[11] A. A. M. Saleh, "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers," *IEEE Trans. Commun.*, vol. 29, pp. 1715–1720, Nov. 1981.

[12] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 1991.

[13] R. M. Gray, *Toeplitz and Circulant Matrices: A Review*. Now Publishers Inc., 2006.