

Comparison between Viterbi Detectors based on Regressive and Autoregressive Noise Models

GTTI 2009 - Sessione Trasmissione

Lorenzo Maggi, Pietro Savazzi

Università degli Studi di Pavia

Dipartimento di Elettronica

Pavia, Italy

Email: lorenzo.maggi@aresys.it, pietro.savazzi@unipv.it

Stefano Valle

STMicroelectronics

Data Storage Division

Cornaredo (MI), Italy

Email: stefano.valle@st.com

Abstract—Recently, a regressive noise model has been proposed for the data-dependent correlated noise at the output of a magnetic recording channel detector. This paper generalizes this channel model, considering digital equalization and a more efficient correlation matrix, in order to make a comparison with the usual detector in a more realistic environment. Simulation results show that the regressive detector performs better when the number of trellis states is lower than needed, while both approaches are comparable when the number of states matches the channel memory.

Index Terms—data-dependent channel noise, perpendicular magnetic recording, regressive and autoregressive models, non-stationary process, Taylor expansion, linear model, Cholesky factorization, Viterbi algorithm.

I. INTRODUCTION

The state-of-the-art of data detection in perpendicular magnetic recording (PMR) channels is based on data-dependent Gauss-Markov detectors. The reader can find in [3] a complete presentation of this approach. According to this approach, data-dependent correlated noise results to be auto-regressive and then it can be whitened by means of data dependent finite impulse response (FIR) filters. This is the approach widely adopted in present-day read channels implementations. Recently, a very interesting paper [2] highlighted that such model implies an exponential increase in the number of states in the detector trellis against a linear increase of the channel density. As a consequence, the continuous request for higher density would result in an unsustainable growth of the detector complexity. More interestingly, the aforementioned paper argues that such growth is mostly due to the inherent non-Markov nature of the media noise. The paper assumes a simple linear model of the read channel derived from the first order Taylor approximation. According to this approximation, data-dependent correlated noise results to be regressive and then whitening can be achieved by means of IIR (*Infinite Impulse Response*) filters. According to [2] the regressive noise model is closer to the actual noise, and IIR whitening would require less predictors and, as a consequence, the corresponding detector would be smaller compared to the AR (*Auto Regressive*) version.

The present work generalizes the noise regressive model

formerly presented by S. Gratrix in [2] by proposing a first-order model for the noise on magnetic recording channel which considers more than three-banded correlation matrix. This extension is more realistic, given the present day channel densities, and it easily includes the effect of the digital equalization (sec. III-D) (sec. III-B). In this paper, the channel conditions under which this model is acceptable are reviewed (sec. III-F). Then the noise regression coefficients, used to solve the ML (*Maximum Likelihood*) problem, are proved to coincide with the model coefficients themselves (sec. III-G). The section III-K suggests a simple method to avoid numerical instabilities during the ML detection. Finally, the performances of the regressive detector are compared to the autoregressive in a realistic context.

II. SIGNAL MODEL

A. Read-Back Waveform

Assuming that transition jitter and AWG (*Additive White Gaussian*) are the two predominant noise components on magnetic channel, the read-back waveform can be described by the following equation:

$$r(t) = \sum_{k \in \mathbb{Z}} a_k h(t - kT - \tau_k) + e(t) \quad (1)$$

where b_k are the information bits, $a_k = b_k - b_{k-1}$, $h(t)$ is the *transition response*, T is the bit interval, $\tau_k \sim \mathcal{N}(0, \sigma_j^2)$ represents the amount of position jitter in $t = kT$ and $e(t)$ is AWG noise.

CBD represents the channel bit density:

$$CBD = \frac{PW_{50}}{T} \quad (2)$$

and the parameter PW_{50} is the width of $dh(t)/dt$ at half height.

Moreover, both the stochastic processes $\{\tau_k\}$ and $\{e_k\}$ are assumed to be strictly stationary, white and mutually independent.

B. SNR definition

The Signal-to-Noise Ratio is defined as:

$$SNR_{dB} = 10 \log_{10} \frac{1}{N_o + M_o} \quad (3)$$

where N_o is the single-sided AWGN power spectral density and M_o is twice the media noise variance; the parameter mix expresses the percentage of AWGN power with respect to the overall noise power:

$$mix = \frac{N_o}{N_o + M_o} \quad (4)$$

so that:

$$N_o = mix \cdot 10^{-SNR_{dB}/10} \quad (5)$$

$$M_o = (1 - mix) \cdot 10^{-SNR_{dB}/10} \quad (6)$$

The variance σ_j^2 of jitter random variable τ can be derived from M_o as follows:

$$\frac{M_o}{2} \cong \frac{1}{2} \sigma_j^2 \sum_{k \in \mathbb{Z}} \left(\frac{d}{dt} h(t - kT) \right)^2 \quad (7)$$

III. REGRESSIVE MODEL OF NOISE

A. Taylor Expansion of Read-Back Signal

Assuming $h(t)$ infinitely differentiable $\forall t$, and $\sigma_j \ll T$, the Taylor expansion of $h(t - \tau)$:

$$h(t + \tau) = h(t) + \sum_{i=1}^{\infty} \frac{\tau^i}{i!} \frac{d^i}{dt^i} h(t) \quad (8)$$

is convergent with very high probability. The first order approximation of channel noise results:

$$r(t) \cong r_{id}(t) + \sum_{k \in \mathbb{Z}} a_k \tau_k \frac{d}{dt} h(t - kT) + e(t) \quad (9)$$

$$\cong r_{id}(t) + \nu(t) \quad (10)$$

where $\nu(t)$ is the noise component.

Let \dot{h}_k be the first derivative of $h(t)$ evaluated in $t = kT$. The signal $r(t)$ sampled at $t = kT$ results to be:

$$r_k \cong r_{id,k} + \sum_{i=-\infty}^{\infty} a_{k-i} \tau_{k-i} \dot{h}_i + e_k \quad (11)$$

The accuracy of this approximation is already demonstrated in [2].

It is reasonable to assume that

$$\dot{h}(t) \cong 0 \quad \text{for } |t| > \lambda T \quad (12)$$

where λ depends on *CBD*. As a result, the first order approximation of the noise becomes:

$$\nu_k = \sum_{i=-\lambda}^{\lambda} a_{k-i} \tau_{k-i} \dot{h}_i + e_k \quad (13)$$

B. Equalization

Typically, the sampled read-back signal is equalized by a digital¹ filter \mathbf{f} ; here, let us suppose that the equalization is applied to the symbol-time sampled signal \mathbf{r} :

$$\mathbf{r}_{eq} = (\mathbf{r}_{id} + \boldsymbol{\nu}) \star \mathbf{f} \quad (14)$$

$$r_k^{eq} = r_{id,k}^{eq} + \sum_{i=-\lambda_f}^{\lambda_p} a_{k-i} \tau_{k-i} \dot{h}_i^{eq} + e_k^{eq} \quad (15)$$

where:

$$\dot{\mathbf{h}}_{eq} = \dot{\mathbf{h}} \star \mathbf{f} \quad (16)$$

$$\mathbf{e}_{eq} = \mathbf{e} \star \mathbf{f} \quad (17)$$

On the contrary, if the equalizer filter \mathbf{f} is defined over an oversampled grid and $\rho \in \mathbb{N}$ is the oversampling factor, the expression of the equivalent time-invariant impulse response $\dot{\mathbf{h}}_{eq}$ associated to transition jitter noise becomes:

$$\dot{h}_k^{eq} = \left\{ \left[\dot{h}(t) \cdot \sum_i \delta(t - iT/\rho) \right] \star f(t) \right\} \Big|_{t=kT} \quad (18)$$

where:

$$f(t) = \sum_i f_i \delta(t - iT/\rho) \quad (19)$$

The parameters λ_f, λ_p are the so-called pre- and post-cursor of the equalized transition response:

$$\dot{h}_k^{eq} \cong 0 \quad \text{for } k > \lambda_p, k < -\lambda_f \quad (20)$$

Finally, the first-order approximation of the noise when digital equalization is taken into account is:

$$n_k = \sum_{i=-\lambda_f}^{\lambda_p} a_{k-i} \tau_{k-i} \dot{h}_i^{eq} + e_k^{eq} \quad (21)$$

The stochastic process \mathbf{n} is non-stationary and, accordingly, noise sample n_k is *data-dependent* and its value depends on the pattern $\boldsymbol{\alpha}_k$:

$$\boldsymbol{\alpha}_k = [b_{k-\lambda_p-1} b_{k-\lambda_p} \dots b_{k+\lambda_f}] \quad (22)$$

C. Noise Mean

When the Taylor series in (8) is not limited to the first term, the noise mean is given by:

$$E[n_k | \boldsymbol{\alpha}_k] = \sum_{i=-\lambda_f}^{\lambda_p} a_{k-i} \sum_{j=1}^{\infty} \frac{E[\tau^{2j}]}{(2j)!} \frac{d^{2j}}{dt^{2j}} h^{eq}(iT) \quad (23)$$

where:

$$\frac{d^{2j}}{dt^{2j}} h^{eq}(iT) = \left\{ \left[\frac{d^{2j}}{dt^{2j}} h(t) \cdot \sum_{k \in \mathbb{Z}} \delta(t - \frac{kT}{\rho}) \right] \star f(t) \right\} \Big|_{t=iT} \quad (24)$$

In fact,

$$E[\tau^{2j+1}] = 0 \quad \text{for } j \geq 0 \quad (25)$$

because the zero-mean normal function is even.

Remarkably, *under the first order approximation*, noise on magnetic channel is *zero-mean*.

¹digital signals are named with bold font

D. Noise Autocorrelation

Noise autocorrelation has a finite length and depends both on the stored binary sequence \mathbf{b} and on the time instant $t = kT$:

$$R(0)^{(k)} = E[n_k^2 | \mathbf{b}] = \sigma_j^2 \sum_{i=-\lambda_f}^{\lambda_p} (a_{k-i} \dot{h}_i^{eq})^2 + \sigma_e^2$$

$$R(\gamma)^{(k)} = E[n_k n_{k-\gamma} | \mathbf{b}] = \sigma_j^2 \sum_{i=-\lambda_f+\gamma}^{\lambda_p} a_{k-i}^2 \dot{h}_i^{eq} \dot{h}_{i-\gamma}^{eq}$$

$$R(\gamma)^{(k)} = E[n_k n_{k-\gamma} | \mathbf{b}] = 0 \quad \text{for } \gamma \geq \varpi \quad (26)$$

where:

$$\varpi = \lambda_p + \lambda_f + 1 \quad (27)$$

It is clear that every correlation coefficient does not depend on the whole binary sequence \mathbf{b} , but only on the pattern $\beta_k^{(\gamma)}$:

$$E[n_k n_{k-\gamma} | \mathbf{b}] = E[n_k n_{k-\gamma} | \beta_k^{(\gamma)}] \quad (28)$$

$$\beta_k^{(\gamma)} = [b_{k-\lambda_p-1} \dots b_{k+\lambda_f-\gamma}] \quad (29)$$

More precisely, correlation coefficients hinges on $\{a_k^2\}$, so it is assumed that:

$$E[n_k n_{k-\gamma} | \beta_k^{(\gamma)}] = E[n_k n_{k-\gamma} | \bar{\beta}_k^{(\gamma)}] \quad (30)$$

where $\bar{\beta}$ is the boolean negation of β ; this property is known as *polar symmetry*.

E. Noise Regressive Model

Following [2], this section reviews the assumptions which make the magnetic channel noise a *non-stationary regressive model* like this one:

$$n_k = \sum_{i=0}^{\varpi-1} c_i^{(k)} \xi_{k-i} \quad (31)$$

where $c_i^{(k)} \in \mathbb{R}$ depends on the instant $t = kT$ and $\{\xi_k\}$ are random variables I.I.D. $\sim \mathcal{N}(0, 1)$; according to this hypothesis, the stochastic process \mathbf{n} can be conceived as the output of a *non-stationary* ϖ -taps FIR digital filter excited by white gaussian noise.

The unknown quantities $c_i^{(k)}$ can be found by matching the correlation coefficients of the two different noise expressions (13-31) and by solving the following iterative non-linear system:

$$\begin{cases} R(0)^{(k)} = \sum_{i=0}^{\varpi-1} c_i^{(k)2} \\ R(1)^{(k)} = \sum_{i=1}^{\varpi-1} c_i^{(k)} c_{i-1}^{(k-1)} \\ \dots \\ R(\gamma)^{(k)} = \sum_{i=\gamma}^{\varpi-1} c_i^{(k)} c_{i-\gamma}^{(k-\gamma)} \quad \gamma \leq \varpi - 1 \end{cases} \quad (32)$$

Since the binary sequence \mathbf{b} has a finite length N , the first $\varpi - 1$ samples of noise sequence can be expressed as:

$$n_k = \sum_{i=0}^{k-1} c_i^{(k)} \xi_{k-i} \quad \text{for } k \in [1; \varpi - 1] \quad (33)$$

As a consequence, we may re-formulate the system (32) into the matricial form:

$$\mathbf{R}_{(\mathbf{b})} = \mathbf{C}_{(\mathbf{b})}^T \mathbf{C}_{(\mathbf{b})} \quad (34)$$

where $\mathbf{R}_{(\mathbf{b})}$ is the $(2\varpi - 1)$ -diagonal autocorrelation matrix of \mathbf{n} and $\mathbf{C}_{(\mathbf{b})}$ is the upper triangular matrix of coefficients $\{c_i^{(k)}\}$:

$$\mathbf{R}_{(\mathbf{b})} = E[\mathbf{n}^T \mathbf{n}] \quad (35)$$

$$\begin{aligned} \mathbf{C}_{(\mathbf{b})}(i, i+j) &= c_j^{(i+j)} \quad j \in [0; \varpi - 1], i \in [1; N] \\ \mathbf{C}_{(\mathbf{b})}(i, i+j) &= 0 \quad j < 0, j \geq \varpi \end{aligned} \quad (36)$$

Commonly, the matrix product $\mathbf{C}_{(\mathbf{b})}^T \mathbf{C}_{(\mathbf{b})}$ is called the *Cholesky decomposition* of $\mathbf{R}_{(\mathbf{b})}$.

It is remarkable that the noise expression in (31) is admissible if and only if noise is *zero mean*. Under the first-order approximation of noisy read-back signal, noise is zero-mean (see section III-C).

F. Linear Dependent Noise

The Cholesky decomposition of a square matrix \mathbf{A} is possible if and only if \mathbf{A} is positive definite; in general, the autocorrelation matrix \mathbf{R} is semi-positive definite.

In other words, for any $\mathbf{a} = [a_1 \dots a_N] \in \mathbb{R}^N$:

$$\mathbf{a} \mathbf{R} \mathbf{a} = \mathbf{a} E[\mathbf{n}^T \mathbf{n}] \mathbf{a}^T \quad (37)$$

$$= E[(\mathbf{a} \mathbf{n}^T)(\mathbf{n} \mathbf{a}^T)] \quad (38)$$

$$= E[(\mathbf{a} \mathbf{n}^T)^2] \geq 0 \quad (39)$$

As a consequence, the only circumstance where \mathbf{R} can not be factorized as in (34) is when:

$$\exists \mathbf{a} \in \mathbf{R}^N : E[(\mathbf{a} \mathbf{n}^T)^2] = 0 \iff p(\mathbf{a} \mathbf{n}^T = 0) = 1 \quad (40)$$

that is the condition for which noise samples are linearly dependent.

The properties of autocorrelation matrix can be usefully considered in two different instances:

- if $\sigma_e = 0$, there exist some particular binary patterns which imply linear dependence among noise samples; two examples will be shown.

1) Let

$$\boldsymbol{\vartheta}_1 = [b_{k-\lambda_p-2} \dots b_{k+\lambda_f+2}] = \underbrace{[0 \dots 0]}_{\lambda_p+2} \underbrace{[1 \dots 1]}_{\lambda_f+3}$$

In this case, noise samples n_k and n_{k+1} can be written as:

$$\begin{cases} n_k = a_k \tau_k \dot{h}_0^{eq} \\ n_{k+1} = a_k \tau_k \dot{h}_1^{eq} \end{cases} \implies n_k = [\dot{h}_0^{eq} / \dot{h}_1^{eq}] n_{k+1}$$

2) Let

$$\boldsymbol{\vartheta}_2 = [b_{k-\lambda_p-2} \dots b_{k+\lambda_f+3}] = \underbrace{[0 \dots 0]}_{\lambda_p+2} \underbrace{[1 \ 0 \dots 0]}_{\lambda_f+3}$$

so that:

$$\begin{cases} n_k &= a_k \tau_k \dot{h}_0^{eq} + a_{k+1} \tau_{k+1} \dot{h}_{-1}^{eq} \\ n_{k+1} &= a_k \tau_k \dot{h}_1^{eq} + a_{k+1} \tau_{k+1} \dot{h}_0^{eq} \\ n_{k+2} &= a_k \tau_k \dot{h}_2^{eq} + a_{k+1} \tau_{k+1} \dot{h}_1^{eq} \end{cases}$$

The aim is to demonstrate that exist $(A, B) \in \mathbb{R}^2$ such that:

$$n_k = An_{k+1} + Bn_{k+2} \quad (41)$$

The unknown quantities A, B satisfy the following linear system:

$$\begin{bmatrix} \dot{h}_1^{eq} & \dot{h}_2^{eq} \\ \dot{h}_0^{eq} & \dot{h}_1^{eq} \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} \dot{h}_0^{eq} \\ \dot{h}_{-1}^{eq} \end{bmatrix} \quad (42)$$

which is solvable if $\dot{h}_1^{eq2} - \dot{h}_0 \dot{h}_2^{eq} \neq 0$.

As a *general rule*, it is possible to assert that, under the first-order approximation, the stochastic process \mathbf{n} is linearly dependent if and only if it exists a subset Λ of noise samples generated by the linear combination of the same set Ω of random variables: in the first case, $\Lambda = \{n_k, n_{k+1}\}, \Omega = \{\tau_k\}$; in the second $\Lambda = \{n_k, n_{k+1}, n_{k+2}\}, \Omega = \{\tau_k, \tau_{k+1}\}$. This is possible only if there are within the whole sequence \mathbf{b} at least two subsequences containing no transitions respectively of length $\psi_1 \geq \lambda_p + 2, \psi_2 \geq \lambda_f + 3$; consequently, the probability that \mathbf{n} is linearly dependent rises with N .

- As a consequence of the former rule, if $\sigma_e \neq 0$ the matrix \mathbf{R} is positive definite, because the presence of AWG equalized noise makes it impossible to find a subset of noise samples generated by the same random variables.

So, it has been mathematically demonstrated under which channel conditions ($\sigma_e \neq 0$) the regressive noise model (31) is well-founded.

Remarkably enough, the former considerations are valid even if the read-back signal is not equalized.

G. Non-Markovianity of Noise

Referring to equation (31), the iterative substitution:

$$\xi_{k-i} \longrightarrow \frac{1}{c_0^{(k-i)}} \left(n_{k-i} - \sum_{j=1}^{\varpi-1} c_i^{(k-i)} \xi_{k-j-i} \right) \quad \forall i > 0 \quad (43)$$

results in:

$$n_k = c_0^{(k)} \xi_k + \sum_{i=1}^{\infty} a_i^{(k)} n_{k-i} \quad (44)$$

where $a_i^{(k)}$ are real time-dependent coefficients; if the noise sequence length is N , the last equation becomes:

$$n_k = c_0^{(k)} \xi_k + \sum_{i=1}^{k-1} a_i^{(k)} n_{k-i} \quad k \in [1; N] \quad (45)$$

Noise sample n_k results to be the sum of the linear combination of all previous noise samples and the term $c_0^{(k)} \xi_k$ *uncorrelated* with them:

$$E[\xi_k n_{k-i}] = 0 \quad \forall k, i \geq 1 \quad (46)$$

Exploiting the probability theory, it follows that:

$$\begin{aligned} E[\xi_k n_{k-i}] &= \int_{\mathbb{R}^2} \xi_k n_{k-i} p_{df_{\xi_k, n_{k-i}}}(x_1, x_2) dx_1 dx_2 \\ &= E_{n_{k-i}} [n_{k-i} E(\xi_k | n_{k-i})] = 0 \quad \forall k, i \geq 1 \\ \implies E(\xi_k | n_{k-i}) &= 0 \quad \forall k, i \geq 1 \end{aligned} \quad (47)$$

By definition of linear stochastic model, the terms $a_i^{(k)}$ are the *regression coefficients* of the random variable n_k on:

$$\mathbf{n}_{k-1}^1 = [n_{k-1}, \dots, n_1] \quad (48)$$

and then:

$$E[n_k | \mathbf{n}_{k-1}^1] = \sum_{i=1}^{k-1} a_i^{(k)} n_{k-i} \quad (49)$$

Moreover, the variance of n_k assuming to know the value of n_{k-1}, \dots, n_1 is the variance of $c_0^{(k)} \xi_k$:

$$Var[n_k | \mathbf{n}_{k-1}^1] = Var[c_0^{(k)} \xi_k] = c_0^{(k)2} \quad (50)$$

Given that $\xi_k \sim \mathcal{N}(0, 1)$:

$$p(n_k | \mathbf{n}_{k-1}^1, \mathbf{b}) = \mathcal{N}\left(\sum_{i=1}^{k-1} a_i^{(k)} n_{k-i}, c_0^{(k)2}\right) \quad (51)$$

The last equation stresses the fact that the stored binary pattern \mathbf{b} is known.

H. Maximum Likelihood Algorithm

It is easy to show [5] that the expression of the maximum-likelihood binary sequence is equivalent to:

$$\tilde{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmax}} \prod_{k=1}^N p(n_k | \mathbf{n}_{k-1}^1, \mathbf{S}_{k-1}^{\mathbf{b}} \cup \mathbf{S}_k^{\mathbf{b}}) \quad (52)$$

$$= \underset{\mathbf{b}}{\operatorname{argmax}} \prod_{k=1}^N \mathcal{N}\left(\sum_{i=1}^{k-1} a_i^{(k)} n_{k-i}, c_0^{(k)2}\right) \quad (53)$$

where the sequence $\mathbf{S}_k^{\mathbf{b}}$:

$$\mathbf{S}_k^{\mathbf{b}} = [b_{k-\lambda_p} \dots b_{k+\lambda_f}] \quad (54)$$

can be considered the *state* of the Viterbi trellis.

The log-likelihood metric $M_k^{\mathbf{b}}$ associated to the state transition $\mathbf{S}_{k-1}^{(\mathbf{b})} \rightarrow \mathbf{S}_k^{(\mathbf{b})}$ is:

$$M_k^{\mathbf{b}} = \ln [c_0^{(k)}] + \frac{\bar{\xi}_k^2}{2} \quad (55)$$

where $c_0^{(k)}$ and $\bar{\xi}_k^2$ are determined in *every* state and in *every* temporal step of the trellis by the joint solution of the non-linear system (32) and of the equation:

$$\bar{\xi}_k = \frac{1}{c_0^{(k)}} \left(n_k - \sum_{i=1}^{\varpi-1} c_i^{(k)} \bar{\xi}_{k-i} \right) \quad (56)$$

${}^2\xi_k$ is a random variable, $\bar{\xi}_k$ is its realization.

In correspondence of every node of the trellis, it is necessary to keep in memory the value of $0.5\varpi^2 + 1.5\varpi - 1$ coefficients:

$$\left\{ \bar{\xi}_{k-i}, i \in [1; \varpi - 1] \right\};$$

$$\left\{ c_{i-\gamma}^{k-\gamma}, \gamma \in [0; \varpi - 1] i \in [\gamma; \varpi - 1]; \right\} \quad (57)$$

The equation (56) can be interpreted as the whitening of the noise sequence in correspondence of every trellis path through an IIR digital filter, which is the exact inverse of the FIR filter generating the noise sequence.

I. Detector Calibration

In correspondence of every Viterbi trellis state, knowledge of the noise sample n_k is necessary (see eq.56) and, under the first-order approximation, the following expression is valid:

$$n_k = r_k - E[r_k | \alpha_k] \quad (58)$$

where $E[r_k | \alpha_k]$ can be ergodically estimated as:

$$\widehat{E}[r_k | \alpha_k] = \frac{1}{N_{\bar{\alpha}^{lev}}} \sum_{i: \alpha_i^{lev} \equiv \bar{\alpha}^{lev}} r_i \quad \alpha_k^{lev} = \bar{\alpha}^{lev}, \forall k \quad (59)$$

where:

$$\alpha_k^{lev} = [b_{k-I_c-1} \ b_{k-I_c} \ \dots \ b_{k+I_a}] \quad (60)$$

and I_a, I_c are respectively the pre- and post-cursor of the dibit response \mathbf{p} .

The expected value of the former estimator can be expressed as:

$$E[n_k] = E[r_k] - E \left[\frac{1}{N_{\bar{\alpha}^{lev}}} \sum_{i: \alpha_i^{lev} = \bar{\alpha}^{lev}} r_i^{cal} \right]$$

$$\equiv 0 \iff I_a \geq \lambda_f, I_c \geq \lambda_p \quad (61)$$

It is important to underline that the last equality holds for *any* degree of approximation order of noise because, if $I_a \geq \lambda_f, I_c \geq \lambda_p$, the samples $r_k, \{r_i, i : \alpha_i^{lev} = \bar{\alpha}^{lev}\}$ can all be considered to be realizations of the same random variable.

Hence, the estimator (58) is unbiased *only* under the first order approximation (see section III-C).

It is valuable that, even if σ_j were so high to make unacceptable the first order approximation, the estimated noise samples would remain *zero mean*.

Even the autocorrelation contributes $R(\gamma | \beta^{(\gamma)}) \forall k$, that are used to solve the iterative non-linear system (32) in each node of the trellis, can be estimated in the ergodic way:

$$\widehat{R}(\gamma | \beta^{(\gamma)}) = \frac{1}{N_{\beta^{(\gamma)}}} \sum_{k: \beta_k^{(\gamma)} \equiv \beta^{(\gamma)}} \widehat{n}_k \widehat{n}_{k-\gamma} \quad (62)$$

where $N_{\beta^{(\gamma)}}$ is the number of recurrences of the pattern $\beta^{(\gamma)}$ within the entire sequence \mathbf{b} and the terms \widehat{n}_k are computed through equations (58,59).

It is important to stress that both the estimates (58) and (62) are computed *off-line* for every value of the patterns α and β .

J. Complexity

If the estimator suggested in (59) is used, then the knowledge of the pattern:

$$\alpha'_k = [b_{k-\max(\lambda_p, I_c)-1} \ \dots \ b_{k+\max(\lambda_f, I_a)}] \quad (63)$$

is necessary to estimate the noise sample n_k in correspondence of every state transition of Viterbi trellis; so, let us define the new *extended* state:

$$\mathbf{S}'_k = [b_{k-\max(\lambda_p, I_c)} \ \dots \ b_{k+\max(\lambda_f, I_a)}] \quad (64)$$

Hence, the number of trellis states of the regressive detector becomes:

$$N_{st}^{REG} = 2^{\max(\lambda_p, I_c) + \max(\lambda_f, I_a) + 1} \quad (65)$$

$$\cong 2^{ISI} \quad (66)$$

where $ISI = I_a + I_c + 1$, because typically $I_c > \lambda_p, I_a \div \lambda_f$.

K. Numerical Instability

It has been demonstrated that, given the binary pattern \mathbf{b}_i , which is associated with the i -th trellis path, the iterative resolution of the non-linear system (32) is equivalent to the on-the-fly Cholesky decomposition of noise autocorrelation matrix $\mathbf{R}_{(\mathbf{b}_i)}$.

Unfortunately, each element of $\mathbf{R}_{(\mathbf{b}_i)}$ can only be statistically estimated as in (62), and the probability that the matrix $\widehat{\mathbf{R}}_{(\mathbf{b}_i)}$ be non-definite (even if $\sigma_e \neq 0$) rises with the decreasing of $N_{\beta^{(\gamma)}}, N_{\alpha^{lev}}$; in this case, the Cholesky decomposition of $\widehat{\mathbf{R}}_{(\mathbf{b}_i)}$ would not be allowed.

Since the following three statements are valid:

- a matrix is positive definite if and only if all its eigenvalues are positive;
- supposing that the eigenvalues of the square matrix $\widehat{\mathbf{R}}_{(\mathbf{b}_i)}$ are $\{\nu_1^{(i)}, \dots, \nu_N^{(i)}\}$, the eigenvalues of $\widehat{\mathbf{R}}_{(\mathbf{b}_i)} + \omega I$ are $\{\nu_1^{(i)} + \omega, \dots, \nu_N^{(i)} + \omega\}$;
- the principal diagonal of the matrices $\mathbf{R}_{(\mathbf{b}_i)}, \forall i$ are constituted by the terms $R(0 | \beta^{(0)})$, while changing $\beta^{(0)}$;

then, if the minimum eigenvalue ν_{min} is *negative*, it is reasonable to carry out the replacement:

$$\widehat{R}(0 | \beta^{(0)}) \longrightarrow \widehat{R}(0 | \beta^{(0)}) + |\nu_{min}| \quad (67)$$

$$\nu_{min} = \min_{r,s} \nu_r^{(s)} \quad (68)$$

after the detector calibration, in order to guarantee the numerical stability of the non-linear system along every trellis path.

Since the terms $R(0 | \beta^{(0)})$ are the only ones depending on the AWGN variance σ_e^2 (see eq.26), so the equation (67) can be interpreted as a fictitious increase of the parameter *mix*.

Remarkably, the physical explanation of this mathematical trick is in accordance with the remarks in section (III-F).

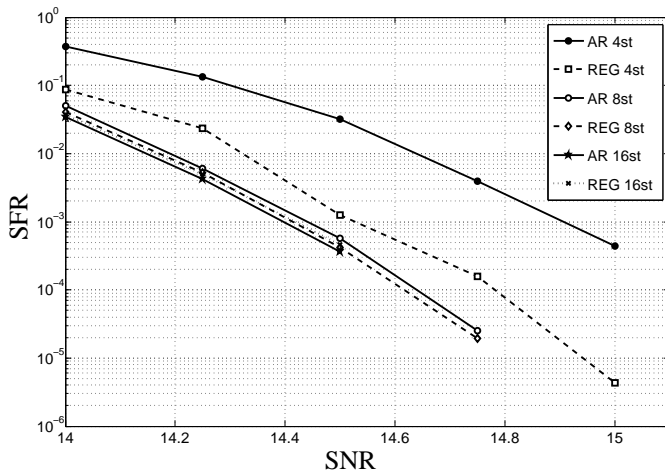


Fig. 1: SFR vs. SNR

IV. COMPARISON BETWEEN THE REGRESSIVE AND THE AR APPROACH

As already mentioned before, the great advantage of regressive approach is expected to be the low number of Viterbi trellis states, which exponentially increases with the ISI only. In fact, the number of predictors does *not* depend on the number of states and the information on all of the past noise samples is contained in the *stored* IIR coefficients (see eq.57).

In the AR-based trellis, the number of predictors L' determines the number of states which grows exponentially with the sum $L' + ISI$.

It stands to reason that the two different methods tend to be closely related when the number L' of AR predictors approaches to the effective channel memory length L . Hence, the regressive detector performances converge faster to the optimum compared to the AR detector when the trellis state number rises. However, it is worth anticipating here that $2^{L'+ISI}$ states may not be strictly necessary to achieve most of the performance gain of the AR detector and several variations of the canonical approach are available. For example, the number of predictors can be left equal to ISI without increasing the number of states; despite suboptimal, this approach generally provides most of the gain. Other combinations in between this approach and the canonical one can be easily envisioned.

Both detectors have been simulated in a realistic perpendicular channel model (see Section II). For both the implementations 4, 8 and 16 states are compared. The AR detector always adopts the maximum number of predictors (1, 2 and 3 for 4, 8, and 16 states, respectively). The user bit density is 2 which approximately leads to an ISI of 2-3 samples after a Generalized Partial Response equalization. Sectors with 512 bytes user data are protected with a $GF(2^{10})$ Reed-Solomon code with correction power $t = 20$ symbols. The figure 1 demonstrates that the 4-states regressive detector performances are about 0.3 dB better than the 4-states AR

detector in terms of SFR (*Sector Failure Rate*) vs. SNR. When the number of states rises, the two methods tend to be equivalent because the number of predictors L' of the AR detector approach to the effective channel memory length L .

The authors believe that the results presented in [2] are affected by the channel model choice. In fact, the detectors are compared with a channel model developed according to the first order approximation, which matches exactly the assumptions of the regressive detector. For this reason, the performance gains in favour of the regressive detector result to be exaggerated. Actually, both regressive and AR channel assumptions are an approximation of the channel. The approximations become closer each other when the number of states grows. Nevertheless, the present work result confirms that, given a limited number of states, the regressive model captures more accurately the noise characteristics and makes the regressive detector a powerful detection method. When the detection performances are to be maximized both approaches seems to be comparable and other considerations arise. The calibration of the regressive MLSD (*Maximum Likelihood Sequence Detection*) algorithm is very simple because it implies the estimation of ISI (*Inter Symbol Interference*) levels and data-dependent correlation matrices. During the calibration of the AR detector, besides the measurement of ISI levels, it is necessary to estimate and - above all - invert $2^{\varpi-1}$ the square L' - dimensional matrices. However, the Yule-Walker equations can be solved easily with a recursive and cheaper LMS (*Least Mean Square*) approach. Instead, the regressive MLSD algorithm is computationally more onerous (and numerically critical) than the AR one: the first involves the resolution of a non-linear system (see eq.32) with ϖ unknown quantities in *every* node of the trellis, while the second one implies a simple data-dependent FIR filtering of the noise sequence.

V. CONCLUSION

The noise regressive model and the corresponding detector presented in [2] have been reviewed and extended to larger correlation lengths. The channel conditions under which it is valid have been desumed and the numerical problems of this approach have been considered, too. The regressive approach has been compared against the well-known AR detector. The results confirmed that the regressive model is a powerful approach especially when the number of states is limited. However, when detection is to be performed at its best, the two approaches seem to be comparable. This conclusion, which results to be not perfectly in line with the conclusions contained in the original work ([2]), is justified by the adoption of a more realistic channel model which makes it evident that the assumptions of the two approaches are different approximations of the actual channel.

REFERENCES

- [1] A. Kavčić, A. Patapoutian, "The read channel," *Proceedings of the IEEE*, vol. 96, no. 11, pp. 1761 - 1774, Nov. 2008.

- [2] S. Gratrix, R. Jackson, T. Parnell, O. Zaboronski, "Viterbi detector for non-Markov recording channels," *IEEE Transactions on Magnetics*, vol. 44, no. 1, pp. 198-206, Jan. 2008.
- [3] A. Kavčič, A. Patapoutian, "A signal-dependent autoregressive channel model," *IEEE Transactions on Magnetics*, vol. 35, no. 5, pp. 2316-2318, Sept. 1999.
- [4] B. Vasic, E.M.Kurtas (2005), *Coding and Signal Processing for Magnetic Recording Systems*, CRC Press.
- [5] E.M. Kurtas, J. Park, X. Yang, W. Radich, A. Kavčič, "Detection methods for data-dependent noise in storage channels," in B. Vasic, E.M.Kurtas, *Coding and Signal Processing for Magnetic Recording Systems*, CRC Press, 2005.
- [6] A. Kavčič, J.M.F. Moura (2000), "The Viterbi algorithm and Markov noise memory," *IEEE Transactions on Information Theory*, vol. 46, no. 1, pp. 291-301, Jan. 2000.