OPTIMAL DATA RETRIEVAL FOR HIGH DENSITY STORAGE

Hans Burkhardt
Technische Informatik I
Technische Universität Hamburg-Harburg, Postfach 90 14 03
D-2100 Hamburg 90, West-Germany

Abstract

Going to extreme high densities in digital storage media - either magnetic or optical - the readback signals are distorted by intersymbol interference and by additive noise. Due to the dispersion of the media as well as the write- and readback-channel neighbouring bits affect each other and cause amplitude distortions and peak shifts. For storage techniques with a low track density this intersymbol interference affects only one dimension (linear density). However, for increased track densities or in optical storage media (high areal density) we observe a crosstalk between the tracks and hence a two-dimensional dispersion. The contribution discusses maximum-a-posteriori (MAP) data detection techniques for signals corrupted by one- and two-dimensional intersymbol interference and additive noise. The distortions are modelled by Markov processes and dynamic programming in form of the Viterbi algorithm is applied to the problem of optimal data retrieval. Various aspects like detection performance and implementation considerations will be discussed.

1 Introduction

The write/read characteristic of storage technology (either magnetic or optical) may be modelled by a communication channel (Fig. 1). It is in general characterised by a linear or nonlinear dispersion function H and additive noise R. 

\[ Z = H(X) + R = Y + R. \]  

(1)

The finite dimension and the stray field of magnetic heads or non-focused laser beams in optical recording cause intersymbol interference and hence limit the channel capacity.

For low track densities intersymbol interference is given only along the tracks and we have a one-dimensional problem like in communication theory. Going to increased areal densities we get crosstalk between the tracks and the dispersion must be modelled by two-dimensional spatial functions like in image processing.

In digital recording the source information belongs to a constrained finite signal alphabet

\[ X \in X, \]  

(2)

Figure 1: Degradation model of a digital recording channel

and we talk about a digital communication channel. In the case of pulse-amplitude-modulation (PAM), for example, discrete-time and discrete-valued sequences are sent through analog channels.

Optimal data detection schemes may be applied to increase data density without raising the bit error rates. The signal detector has to cope with increased intersymbol interference, closure of eye-patterns and ill-conditioned or singular transfer functions in the presence of additive noise [1].

The application of optimal linear filters (linear equalizer, Wiener or Kalman filter) results in a poor performance improvement for channels with severe distortions. Linear equalization shows a clear trade-off between pulse slimming and noise enhancement [2]. Linear filters are non-optimal for digital data detection because the a-priori knowledge of binary valued source signals may not be incorporated into the solution.

A nonlinear technique, called decision feedback [1], gives further improvements by compensating intersymbol interference by feeding back tentative decisions. This solution, however, may suffer from error propagation.

Finally the Viterbi algorithm (VA) [3] utilizes the principle of dynamic programming to achieve maximum-a-posteriori (MAP) detection of a finite signal alphabet passing through a channel with known transfer function. The resulting nonlinear recursive filter provides the optimal solution with a superior performance at the expense of an increased detector complexity.

Dynamic programming is a recursive optimization technique for the solution of locally coupled deterministic and stochastic decision processes based on the optimality prin-
cipe of Bellman [4]. Based on dynamic programming Viterbi developed an algorithm [5] to decode convolutional codes. Forney [3] applied the algorithm to the maximum-a-posteriori analysis of discrete Markov processes and solved the problem of optimal data detection in dispersive digital communication channels. The transition probabilities of a finite-state Markov process and hence the trellis with a linear complexity in the number of decision stages in contrast to an exponentially growing complexity of analyzing all possible paths in a brute force technique. The metric for the analysis is determined by the distribution density of the given Markov process.

Meanwhile the Viterbi algorithm plays a dominant role in digital communication [1], in speech and image processing [6]. It is a nonlinear optimal filter with a remarkable improvement over linear equalizing filters, which are not able to consider a-priori available signal or structural constraints. Based on the one-dimensional theory of digital communication, the VA was proposed to increase the linear density in magnetic recording [7-9].

It was possible to extend the Viterbi algorithm to two-dimensional image restoration ([10]-[12]) on the basis of a maximum-a-posteriori criterion for problems with a non-convex constraint in form of a finite discrete set of amplitudes of the original image (up to the extreme of a black and white original). The image distortions are modelled with a two-dimensional Markov model in analogy to a communication channel. This model is able to consider nonlinear and space variant dispersions and is not restricted to normally distributed noise.

It is the objective of this paper to apply the Viterbi algorithm to the optimal detection of data corrupted by spatial intersymbol interference and additive white Gaussian noise. The crosstalk from adjacent tracks is assumed to come from synchronously written data. The proposed solution is derived from results which were given for the optimal restoration of constrained images.

2 Process Modelling

A continuous digital recording channel with a spatially limited extent of the point spread function may be modelled without loss of optimality by a so-called equivalent discrete channel model with white noise, consisting of a cascade of a whitened matched filter followed by a sampler [13]. This discrete channel can then be described by a finite-state discrete-time Markov process observed in memoryless noise.

It is assumed that the original data domain is given by a finite alphabet of signals $X$ with $B$ possible discrete amplitude levels (e.g. $B = 2$ for binary data). The data degradation may be caused by a finite dispersion function $H$ of dimension $m \times n$, corresponding to the point spread function in the case of a linear two-dimensional convolution. Measured data in an observation space $Y \in Y$ of dimension $M \times N$ are uniquely defined by written signals in an area of extension $(M + m - 1) \times (N + n - 1)$. Hence we get: (see also Fig. 1 and Fig. 2)

$$X \in \mathbb{E}^{(M + m - 1) \times (N + n - 1)}$$

$$H \in \mathbb{R}^{m \times n}$$

$$Y, Z \in \mathbb{R}^{M \times N}$$

Therefore the observation space may consist of sampled data of an array of $M$ readback channels (e.g. an array of $M$ magnetic heads) or simply just the sampled data of one channel ($M = 1$, one head).

States of the Markov chain are given by stripes of data of dimension $(M + m - 1) \times (n - 1)$ and transitions are defined by consecutive pairs of states: (see Fig. 4)

$$w^k := (x^{k-1}, x^k)$$

with

$$\dim w^k = (M + m - 1) \times n$$

$$\dim x^k = (M + m - 1) \times (n - 1)$$

and

$$X = \{w^1, w^2, \ldots, w^N\}$$

$$= \{x^0, x^1, \ldots, x^{N-1}\}$$.  

Markov processes form a rather general frame for stochastic modelling. A Markov process is characterized by the property that the probability to be in a state $x^{k+1}$ depends only on the immediate preceding state $x^k$ and is independent of all other previous states $\{x^0, x^1, \ldots, x^{k-1}\}$:

$$P(x^{k+1}|x^k, \ldots, x^1, x^0) = P(x^{k+1}|x^k).$$

These states are characterized by overlapping segments in time domain (Fig. 3) and overlapping areas in the two-dimensional spatial domain (Fig. 4). The transitions in the process may be visualized graphically in a trellis (see Fig. 5).

Markov models may be used to describe linear, nonlinear, time and space variant channels. The only assumption is a dispersion function of finite extension. Markov models get more and more important for modelling deterministic or stochastic context dependencies for example in speech, written text, communication channels, image scenes, image distortions and object contours.

The process is assumed to be observed under the assumption of memoryless noise, namely there is a sequence
Figure 2: Extension of the data in the original space $X \in \mathbb{R}^{(M+m-1)\times(N+n-1)} = \mathbb{R}^{5\times12}$ and the observation space $Y, Z \in \mathbb{R}^{M\times N} = \mathbb{R}^{3\times11}$ for a $3 \times 3$ point spread function $H \in \mathbb{R}^{m\times n} = \mathbb{R}^{3\times3}$.

2. It should be stressed that this criterion tries to optimally restore the whole sequence and not individual data values.

Due to the Markov and memoryless properties the logarithm of this performance criterion may be decomposed into sums and the problem is equivalent to finding the shortest path through a decision tree with weights proportional to:

$$
\lambda(w^k) = -\ln P(x^{k+1}|x^k) - \ln P(z^k|w^k). \quad (10)
$$

Instead of calculating the exponentially growing number of weights of a tree dynamic programming reduces the complexity by tracing a decision trellis, representing $d+\ldots d$ contributions to a total length $L$ recursively. In one node of this trellis, representing the states of the Markov process, we get (see Fig. 3):

$$
L_i^{k+1} = \min((L_i^k + \lambda_x), (L_j^k + \lambda_y), \ldots, (L_r^k + \lambda_r)) \quad (11)
$$

where $\lambda_r$ represents the transition weight for going from state $x_i$ to $x_r$. Thus the VA now tracks the minimal path of accumulated weights along a decision trellis with $B^{(n-1)(M+m-1)}$ states and $B^{(M+m-1)}$ transitions per state and hence a total of the order of $N \cdot B^{(n-1)(M+m-1)}$ probability computations which is of linear arithmetic complexity with respect to the number of samples $N$.

The Viterbi algorithm may also be interpreted as an efficient technique to classify degraded signals in comparison with the exponentially growing number of all possible $B^{(n-1)(M+m-1)}$ data sequences $X$ in the observation space $Z$. Fig. 5 shows a decision trellis with 64 states and
is also a nonlinear problem. It is possible to show that this distance may efficiently be calculated with the Viterbi algorithm itself by increasing the basis \( B \) [14].

If all original data sequences in \( X \) are equally likely, the MAP criterion is equivalent to the maximum likelihood criterion (MLE).

It is possible to get remarkable improvements in comparison with linear filtering, however, at the cost of a high computational complexity. It is possible to reduce this complexity for areal Markov processes with suboptimal strategies [10].

It is advantageous that the Viterbi algorithm has a parallel implementation thus alleviating the critical timing problem in real time applications at the cost of hardware complexity. In digital communication there are great efforts to provide VLSI implementations.

In contrast to linear filtering it is easy to consider code constraints [14]. This can be done by describing the code with a finite state machine which may easily be incorporated into the analyzing trellis by eliminating transitions which are not allowed by the encoder. In [14] examples are given for run-length limited (RLL) codes [16] like for example MFM. It could be shown that these constraints can straightforwardly be included in the calculation of the minimum distance [14].

It is essential to mention that the VA may easily work on nonlinear channels (see for example [8]) and that it is rather flexible in using different metrics for other noise distributions.

5 Example

The following example was chosen to clarify the method outlined above.

The observations are assumed to come from a single read-head \( (M = 1) \) with 10 samples \( (N = 10) \). The data written to the recording media is assumed to be

\[
\begin{bmatrix}
X_1 \\
X_2 \\
\vdots \\
X_{10}
\end{bmatrix} = \begin{bmatrix}
2 & 5 & 2 & 5 & 6 & 3 & 2 & 1 & 0 & 1 & 7 & 1 \\
0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}
\]

The first notation must be read as an octal representation of the states, for example:

\[
X^1 = (x^0, x^1) = (2, 5, 2, 5, 6, 3, 2, 1, 0, 1, 7, 1)_{\text{Octal}}
\]

\[
X^2 = (x^0, x^2) = (0, 1, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0)_{\text{Octal}}
\]

The write/read characteristic is assumed to be described by a linear convolution with a \( m \times n = 3 \times 3 \) point-spread function, symmetric to the origin, corresponding to a crosstalk from two neighbouring tracks:

\[
H = \begin{bmatrix}
1 & 3 & 1 \\
3 & 7 & 3 \\
1 & 3 & 1
\end{bmatrix}
\]
Figure 6: Eye pattern for the example with a closure of 56%

producing the following readback samples:

\[ Y = [12, 11, 13, 16, 17, 12, 6, 2, 8, 15] \]

The eye pattern (Fig. 6) which characterizes the performance of a detector based on scalar local decisions shows a severe overlap of 56%. This would cause 100% error rates for specific data sequences.

The Viterbi algorithm detected a sequence of

\[ X = \begin{bmatrix} 2 & 5 & 2 & 5 & 6 & 6 & 2 & 4 & 0 & 4 & 7 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \]

for \( R = 0 \) and for

\[ Z = Y + R = [12.21, 11.27, 13.06, 16.07, 16.74, 11.83, 5.64, 2.13, 7.85, 14.89] \]

\( R \) is additive white Gaussian noise with zero mean and standard deviation of \( \sigma(R) = 0.211 \).

Comparing \( X \) with the data written originally, we observe a correctly detected centerline bitstream of

\[ [1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 0] \]

Because of the symmetry of \( H \) the detector may not discriminate for example between

\[ \begin{bmatrix} 0 \\ 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 \\ 0 \end{bmatrix} \]

This means that off-track bits which cause the same amplitudes may have different values, however, they are used only for the complete description of the states and do not belong to the estimated data.

As already mentioned, the detection performance is dominated by the minimum distance \( d_{\text{min}} \) in the observation space, which can be calculated by a slightly modified Viterbi algorithm [14]. Two code sequences are called different if they show deviations in the data bits on the centerline (the offtrack bits do not belong to the detector output).

References


