Estimation of Tracking Errors Using Array Heads

L. C. Barbosa, C. E. MacDonald and J. K. Wolf
Center for Magnetic Recording Research, University of California at San Diego, La Jolla, CA 92093

Abstract

The present trend of using narrow tracks in magnetic recording is severely limited by difficulties due to track misregistration. Reference [1] suggests that the use of array heads can reduce these difficulties. Here we will examine in more detail such application of array heads. Two classes of procedures for acquiring misregistration information will be presented here: the first class is decision oriented, that is, the data sequence is assumed known, which requires an independent detection mechanism. The second class is independent of the knowledge of the written sequence: the amount of misregistration is estimated by comparing the read back signals from the heads in the array.

1: Introduction

In the past, there has been a concern in digital magnetic recording to keep neighbor tracks well apart to avoid inter-track interference and to accommodate for deficiencies in tracking procedures. Therefore, density increases have been aggressively pursued by increasing linear density to the point of allowing intersymbol interference (ISI), and moderately done by increasing track density to the point of using "guardbands" to avoid ISI. Current attempts to increase linear density have been limited by severe non-linearities. However, the disciplines of Communications and Signal Processing offer great opportunities for pushing track densities much more aggressively than in the past.

Array heads for optical digital recording have been built and reported in the literature [3,4]. Elliptical spots and laser diode arrays have also been reported for optical recording [5]. Techniques to simultaneously detect the data recorded in several tracks and the use of the array for servo purposes were discussed in [1,2]. In [3] an array was used to cancel interference from adjacent tracks on the central track, and [4] used the same system to correct the skewing of the heads over the tracks.

Array heads are also common in systems using magnetic tapes. In both cases of optical recording and tapes the heads are kept well separated to avoid "cross-talk" [6]. The present fear of crosstalk can be relieved if the techniques suggested in [1,2] are used. Here we will explore yet another clear advantage of using array heads: they allow for efficient estimation of track misregistration from the read-back set of voltages. The estimates can be used either to take into account the tracking error in the detection process, or to provide an error signal for a fine servo system.

2: A simple example

As an example, consider the case of two tracks being read simultaneously by three identical heads. In addition, assume that the total width of the head array is slightly smaller than the total width of the two tracks as shown in Fig. 1. The theory to be presented covers a much broader spectrum than this simple example indicates; nevertheless, the example exposes the main features of the theory in a very clear way.

Following Fig. 1, the read back voltages in the three heads are:

\[ y_1 = H_1 a_1 + n_1 \]
\[ y_2 = H_2 a_2 + n_2 \]
\[ y_3 = (0.5 + \epsilon)H_1 a_1 + (0.5 - \epsilon)H_2 a_2 + n_3. \]

Here, \( \epsilon \) is the misregistration in fraction of the head width, \( a_i \) is the bit sequence recorded in the i-th track, \( n_j(t) \) is the noise associated with the j-th head and the \( H_i's \) are mappings from the data sequence spaces to the observation spaces [1,2]. For this example, they are defined as:

\[ H_i a_i = \sum_n a_i(n) h_i(t - nT_i) \]
where $a_i(n)$ is the bit at position $n$ in track $i$, $h_i(\cdot)$ is the channel response from track $i$ to head $i$, and $T_i$ is the clock interval for track $i$.

Fig. 1. Tracks and head configuration for the simple example. Here, $\varepsilon = \Delta/W_H$

The noise is zero mean and jointly normally distributed, necessarily white and not necessarily uncorrelated across tracks. The central head reads $y_3$, a combination of the signals from both tracks; the amount of the combination depends on $E$. Equations (1) can be written in a compact form using vectorial notation as in [1, 2, 7], $y = H a + n$, where $y(t)$ is the tuple $[y_1(t)\ y_2(t)\ y_3(t)]$; a similar notation is used for the noise and for the data sequence $a = [a_1\ a_2]$. The operator $H$, which depends on $\varepsilon$, is a mapping from $S = S_1 \times S_2$ to $Y = Y_1 \times Y_2 \times Y_3$. $H$ is further decomposed into two parts, as:

$$H = H_0 + \varepsilon H_\varepsilon = \begin{bmatrix} H_1 & 0 \\ 0 & H_2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ .5H_1 & .5H_2 \end{bmatrix} \varepsilon \begin{bmatrix} H_1 & -H_2 \\ 0 & 0 \end{bmatrix}.$$

Therefore, the readback tuple can be written as:

$$y = H_0 a + \varepsilon H_\varepsilon a + n.$$

$H_0$ is the operator when the heads are perfectly positioned over the tracks and $H_\varepsilon$ is the operator that takes into account the misregistration $\varepsilon$. The structures of $H_0$ and $H_\varepsilon$ are given as above.

3: Decision directed approaches

3.1: Maximum likelihood estimation

In this well-known approach, we choose $\varepsilon$ to maximize the probability of observing $y$ given $a$ and $\varepsilon$. Such a conditional probability density function can be readily written as:

$$p[y/a,\varepsilon] = p[n = y - Ha/a,\varepsilon] = K \exp(-.5 \| y - Ha \|^2).$$

Assuming that $a$ is known, the maximization of the conditional probability over $\varepsilon$ is straightforward, yielding:

$$\hat{\varepsilon} = \frac{< y - H_0 a, H_\varepsilon a >}{\| H_\varepsilon a \|^2} = \frac{(H_\varepsilon^* y - H_\varepsilon^* H_0 a, a)}{(a, H_\varepsilon^* H_\varepsilon a)}.$$  (4)

Here, the dot product denoted by $<,>$ and the norm $\|,\|$ are the Reproducing Kernel Hilbert Space dot product and norm respectively. The superscript $*$ indicates the adjoint operator and $(\cdot, \cdot)$ is the usual dot product taken in the sequence space $S$.

Using equations (3) and (4), one gets:

$$\hat{\varepsilon} = \varepsilon + \frac{< n, H_\varepsilon a >}{\| H_\varepsilon a \|^2}$$

and denoting the statistical expectation by $E[\cdot]$, we have:

$$E[\hat{\varepsilon}] = \varepsilon.$$  (5)

$$Var \hat{\varepsilon} = E \left[ \frac{< H_\varepsilon^* a, n >}{\| H_\varepsilon a \|^4} \frac{< n, H_\varepsilon a >}{\| H_\varepsilon a \|^4} \right] = \| H_\varepsilon a \|^2.$$  (6)

The above relations show that this estimate is unbiased and of minimum variance (the reader can verify that its variance achieves the Cramer-Rao lower bound).

3.2: Hypothesis testing

This is the second decision oriented method of estimating the misregistration. The written sequences have already been (correctly) detected by some other means. Here we basically tolerate misregistrations having absolute value less than $\varepsilon_0$. This limit $\varepsilon_0$ can be zero, or a chosen positive number derived, for example, from "bathtub" curves. We proceed with a statistical testing of the "null hypothesis" $H_0$ that $|\varepsilon|$ is indeed smaller than $\varepsilon_0$, versus the alternative hypothesis $H_1$ that $|\varepsilon|$ is larger than $\varepsilon_0$.
Notice that this is a case where both hypotheses are composite. The normal procedure to test $H_0$ against $H_1$ in this case is \[ \text{reject } H_0 \text{ if } |\hat{e}| > e_0 \text{ and } 1-e_m - |\hat{e}| > c \]

where $\hat{e}$ is as in (4) and $e_m = e_0 \text{sign}(\hat{e})$.

This region, known in the hypothesis testing literature as the critical region, is shown in Fig. 2. The constant $c$ is calculated so that the probability of rejecting $H_0$ when $H_0$ is true is set to be not greater than, say, .05.

$$\begin{align*}
\text{Reject } H_0 & \quad \text{Accept } H_0 & \quad \text{Reject } H_0 \\
(x < e_0) & \quad (x > e_0) & \quad (x \neq e_0)
\end{align*}$$

$\hat{e}$ is as in (4) and $e_m = e_0 \text{sign}(\hat{e})$.

Fig. 2: The critical region for the hypothesis testing.

Since $c$ is inversely proportional to $\text{sign}H_0$, it will decrease with the amount of data collected, making the region for accepting $H_0$ tighter as time elapses. As an example, for a (1 - D) channel with $\text{SNR} = 20$ dB and with 100 or more observations, selecting $e_0 = .005$ or greater and $\text{Pr}[\text{rejecting } H_0/\text{H}_0] = .05$, yields $c = 1.65 \cdot \text{sign}H_0^{-1}$.

The $\text{Pr}[\text{rejecting } H_0/\text{H}_0]$ is known in the statistics literature as the significance level of the test [8], and in the communications literature as the probability of false alarm (from the fact that hypothesis testing was first applied in communication in radar detection problems).

Of equal importance is the power of the test when $H_1$ is true, i.e., $\text{Pr}[\text{rejecting } H_0/\text{H}_1]$. Since we want to reject $H_0$ when $H_1$ is true, this probability should be as high as possible. The power of the test was estimated for a (1-D) channel with $\text{SNR} = 20$ and $e_0 = 0$ for several values of $\epsilon$ and $N$, as shown in Figs. 3 and 4. In Fig. 4 the estimated values of $c$ are also plotted. Under the above conditions, Fig. 3 shows that an $\epsilon = .01$ can be detected with reliability of about 90% when using 1000 clock cycles whereas for $\epsilon = .035$ the same reliability is achieved with only 100 clock cycles. It also shows the significance level of the test at the point $\epsilon = 0$. In Fig. 4 the power of the test is plotted against the number of clock cycles. Also plotted there is the estimated value $c$ of the size of the critical region.

4: Non-decision directed approaches

In this approach, we take advantage of the relationships among the read voltages without trying to estimate the written data or the clocks. Notice that one can eliminate the unknowns $H_1$ and $H_2$ by substituting their values taken from the first two equations of (1) and plugging them in the third equation, yielding:

$$z \triangleq y_3 - .5(y_1 + y_2) = \epsilon(y_1 y_2) - \epsilon(n_1 n_2) - .5(n_1 + n_2) + n_3.$$
Notice that when \( \varepsilon \) is zero and in absence of noise, \( z \) will be identically zero. By multiplying \( z \) by \( y_1-y_2 \) and taking the expectations (denoted by an upper bar), one finds:

\[
\varepsilon = \frac{z(y_1 - y_2)}{(y_1 - y_2)^2} - \sigma^2 \frac{\rho_{13} - \rho_{23}}{(1 - \rho_{12})^2}.
\] (6)

Now, we can substitute the statistical averages by time averages using the ergodic properties of the process under consideration. Examples of this will be given in the simulation section of this paper.

5: Simulations

For the purposes of simulation, the two track, three head model shown in Figure 1 will be used. The block diagram of the system is as shown in Figure 5.

Figure 5: Block diagram of system

The noise is white Gaussian noise of variance \( \sigma^2 \), uncorrelated from head to head. It should be pointed out that this is simply one particular head configuration, and that many others are possible. The magnetic channel will be modeled as a (1-D) channel. Following (3) we thus have:

\[
y = (H_0 + \varepsilon H_2) a + n
\]

where

\[
a = [a_1(i) \ a_2(i)]^T = [a_1(i) \ a_1(i-1) \ a_2(i) \ a_2(i-1)]^T
\]

(called the Information State Vector), and

\[
H_1 = H_2 = [1 \ -1].
\]

For the decision oriented approach, in this example, we will use maximum likelihood estimates of the information sequences \( a_1 \) and \( a_2 \). The Viterbi detector estimates these sequences using a 4 state trellis (two tracks with a memory of two bit periods each.) The trellis is as shown in Figure 6. \( L_j(i) \) is the path metric to state \( j \) at time \( i \), and \( c_{k,j}(i) \) is the update metric from state \( k \) to state \( j \) at time \( i \). Associated with each node metric at a given time instant is an estimate of \( \varepsilon \) to that node. In the calculation of update metrics from each state to a given state at the succeeding time instant one can use the estimate of \( \varepsilon \) from each state, or the estimate of \( \varepsilon \) associated with the minimum path metric overall at time \( i \). Simulation results showed that the latter approach provided practically the same performance, and was simpler to implement than the former method.

Figure 6: Trellis and update metrics

The update metrics in this case are easily computed as:

\[
c_{k,j}(i) = [a(i)]^2 = [y(i) - H \cdot a_{k,j}]^2
\]
where $a_{k,j}$ = information state vector going from state $k$ to state $j$. Using (4), an update to the estimate $\hat{e}$ is made at each time instant. However, whenever the denominator $\left( H_{5a(i)} \right) = 0$, no update information is provided. This case occurs 6/16 of the time in the decision oriented estimate for this example. This is not the case for the non-decision oriented approach; here $\hat{e}$ is computed using (6), where $\rho_{ij}$ is equal to 0 if $i \neq j$ and equal to 1 when $i = j$. Now the case of no update information occurs only when $y_1 = y_2$, i.e. when 

\[
(Ha(i) + \sigma_i(i)) - (Ha_j(i) + \sigma_j(i)) = (1-D)(a_1(i) - a_2(i)) + (a_1(i) - a_2(i)) = 0.
\]

This rarely happens due to the independent noise terms.

Since the noise is additive white Gaussian and the signal processing is linear, we can bound $\hat{e}$ by confidence intervals. It can be shown that $\sigma^2_{\hat{e}} = \sigma^2 / N$, where $N$ = number of bits used in the estimate. Figures 7 through 10 show the decision and non-decision oriented estimates of $\epsilon$ and the $\sigma_\epsilon$ and $2\sigma_\epsilon$ confidence intervals for several values of SNR, here defined as

$\text{SNR} = 10 \log(1/\sigma^2_N)$. A constant $\epsilon$ of 0.05 was used in all trials. We see that the estimates always fall within the $2\sigma_\epsilon$ (95%) bounds, and that the estimates converge quite rapidly for larger SNR. More importantly, the decision and non-decision oriented approaches have similar convergence properties.
6: Conclusions

The general theory of array heads for data recovery and servoing purposes has been described for both decision oriented and non-decision oriented approaches. Simulation results for both these approaches using a two track, three head model were shown, demonstrating the convergence of the estimate of $E$ as a function of the SNR. The decision and non-decision directed approaches appear to provide similar estimates of $E$ for the 1-D channel. However, the non-decision directed approach can be computed independently of the data detection process, if one has knowledge of the noise correlation, making this method simpler to implement.

7: References


