Minimum mean square error linear predictor with rounding

Fu Yu Tsai and Hui Peng
Department of Electrical Engineering
National Tsin Hua University
Hsinchu, Taiwan, R.O.C

ABSTRACT

Many digital signal processing and image coding systems implement the linear predictor with rounding. Usually, people will obtain the linear predictors by solving the Yule-Walker equations or doing something equivalent. The predictors obtained in this way will not necessarily be the true minimum mean square error predictor considering the effect of rounding. In this paper, we address the issue of finding the true optimum mean square error rounded linear predictor. Experiment results show that when the prediction results are rounded, this true MMSE linear predictor could outperform the conventional one without considering the effect of rounding very significantly for data of low prediction errors.

Keywords: linear prediction, rounding, optimum prediction, coding, data compression.

1. INTRODUCTION

Linear predictors are commonly used in data compression, modeling, parameter estimation, and filter design, etc. Obtaining the best mean square error linear predictor is desirable because it implies best error-free compression (for Gaussian Source), prediction, modeling, and estimation with minimum error energy. In linear predictive coding of error-free data compression, the prediction results will be rounded because all data are bounded integers and only predictors of integer values will be allowed. Due to this rounding, linear predictors that yield minimum mean square error, obtained by solving the Yule-Walker equations conventionally, will not necessarily be the true "minimum mean square error" linear predictor. This paper addresses the issue of finding the optimum mean square error linear predictor considering the prediction results are rounded and both the original and the error signals are bounded integers.

This paper is partitioned into five parts. In Part II, we review the background of the linear predictors and the Orthogonality Principle which is used in deriving the conventional linear predictors without rounding. In Part III, we develop a method to find the first order minimum mean square error causal linear predictor with rounding and discuss its extension to high order cases. In Part IV, we show some experiment results we have obtained on the minimum mean square error rounded linear predictor. Finally, the summary is stated in Part V.

2. BACKGROUND

Fig. 1. Error signal generation in a predictive coding system.
Fig. 1 is the block diagram of a typical predictive coding scheme. In Fig. 2(a), we show a one dimensional kth order causal linear predictor without rounding of which the output of the predictor is a real number. If the linear predictor in Fig. 1 is as the predictor in Fig. 2(a),

\[ k \sum_{j=1}^{k} \alpha_j Y_{i-j} \]

(a)

Fig. 2. Linear Predictors without and with rounding
(a) Linear Predictor without rounding
(b) Linear Predictor with rounding

\[ e_i = Y_i - \sum_{j=1}^{N} \alpha_j Y_{i-j}. \]  

(1)

Denote

\[ \hat{Y}_i = \sum_{j=1}^{N} \alpha_j Y_j, \]

\[ a = (\alpha_1, \alpha_2, \ldots, \alpha_k) \]

\[ X = (Y_{i-1}, Y_{i-2}, \ldots, Y_{i-N})^T, \]

then,

\[ e_i = Y_i - \hat{Y}_i = Y_i - a^T X = Y_i - X^T a. \]  

(2)

To obtain the optimal linear predictor without rounding, we need to find \( a^* \) such that \( E[(Y_i - a^T X)^2] \) is minimized. The Orthogonality Principle\(^1\) states that the optimal \( a^T \) is the one that causes the error vector \( e_i \) to be orthogonal to (have zero correlation with) the vector \( X \). It can be shown that the optimal \( a^* \) is a solution of

\[ a^T R_s = E[YX^T], \]  

(3)

where \( R_s = E[XX^T] \). If \( R_s \) is invertible, then the optimal \( a^* \) is given by

\[ a^* = E[YX^T]R_s^{-1}. \]  

(4)

For example, in the case of the one dimensional 1st order linear predictor,

\[ a^* = \alpha, \]

\[ X = Y_{i-1}, \]

thus

\[ \alpha = E[Y_{i-1}](E[Y_{i-1}])^{-1}. \]  

(5)

3. OPTIMUM LINEAR PREDICTOR WITH ROUNING

Fig. 2(b) is the one dimensional kth order causal linear predictor with rounding of which the output is an integer, where "\( \lceil \cdot \rceil \)" means rounding. With the linear predictor in Fig. 1 having the structure as in Fig. 2(b), the optimum mean square error predictor should be the one which will make the integer error signal having the smallest energy among all possible linear predictors of the same structure as in Fig. 2(b).
We formulate the problem of finding the optimum mean square error rounded linear predictor as finding the coefficient vector $\mathbf{a}^*$ among all possible $\mathbf{a}$'s, such that the mean square error $E[(Y - \hat{Y}(\mathbf{a}))^2]$ is minimized, where $\hat{Y}(\mathbf{a})$ is the predictor which is a linear function of $\mathbf{a}$ and the neighboring pixels of $Y$. The Orthogonality principle which can be used in finding the MMSE linear predictor without rounding can not be used to find the above $\mathbf{a}^*$.

Fig. 3 shows the relation between $X$ and $Z$, where $Z = \text{round}(X)$. Now, let us consider the one dimensional 1st order causal linear predictor with rounding. What we need to do is to find the coefficient $\alpha$ which is the solution of the following optimization problem:

$$\text{Min } E[(Y - \alpha Y_{i-1})^2].$$

(6)

Since rounding is a nonlinear function, we can not use the Yule-Walker equations to solve it. However, we can find the solution of (6) employing the property of rounding. Let's see the following:

For given $\alpha$ and $\hat{Y}$, there exists a $\Delta \alpha$ such that

$$[(\alpha + \Delta \alpha)Y_{i-1}] = \lceil \alpha Y_{i-1} \rceil, \quad \forall \Delta \alpha \in (0, \Delta \alpha)$$

(7)

and

$$[(\alpha + \Delta \alpha)Y_{i-1}] = \lceil \alpha Y_{i-1} \rceil + 1.$$  

(8)

From (8), we have

$$\Delta \alpha Y_{i-1} = \lceil \alpha Y_{i-1} \rceil + 0.5 - \alpha Y_{i-1}, \quad \text{and}$$

$$\Delta \alpha = \frac{\lceil \alpha Y_{i-1} \rceil + 0.5 - \alpha Y_{i-1}}{Y_{i-1}}.$$  

(9)

$\Delta \alpha$ in (9) is a function of $\alpha$ and $Y_{i-1}$. Consider all possible values of $Y_{i-1}$, $Y_{i-1} \in \{I_1, I_2, \ldots, I_n\}$, for a fixed value of $\alpha$'s, we have different $\Delta \alpha$, i.e., $\Delta \alpha(\alpha, Y_{i-1} = I_1)$, $\Delta \alpha(\alpha, Y_{i-1} = I_2)$, ..., and $\Delta \alpha(\alpha, Y_{i-1} = I_n)$. Denoting $\Delta \alpha(\alpha) \equiv \Delta \alpha(\alpha, Y_{i-1} = I_j)$, and letting
\[ \Delta \alpha_{\text{min}}(\alpha) = \min_{j=1, \ldots, n} (\Delta \alpha_j(\alpha)). \]  

we have

\[ \{(\alpha + \hat{\Delta} \hat{\alpha}) \hat{Y}_{t-1} = \hat{\alpha} \hat{Y}_{t-1} \} \quad \forall \Delta \hat{\alpha} \in [0, \Delta \alpha_{\text{min}}(\alpha)] \]  

for all possible values of \( Y_{t-1} \).

Furthermore, we can check which prediction coefficient will give the minimum mean square error. Assume the true optimum rounded first order linear predictor has the prediction coefficient lying between \( a \) and \( b \), we define

\[ \Delta \alpha_{\text{min}}^{(k)} = \text{Min}_{j=1, \ldots, n} \Delta \alpha(\alpha_k, Y_{t-1} = I_j). \]  

where

\[ \alpha_0 = a, \]
\[ \alpha_1 = \alpha_0 + \Delta \alpha_{\text{min}}^{(0)}. \]
\[ \vdots \]
\[ \alpha_m = \alpha_{m-1} + \Delta \alpha_{\text{min}}^{(m-1)}. \]
\[ \vdots \]
\[ \alpha_k = \alpha_{k-1} + \Delta \alpha_{\text{min}}^{(k-1)}. \]

and \( k \) is such that \( \alpha_{k-1} < b \leq \alpha_k \), then we will have

\[ |\alpha Y_{t-1}| = |\alpha_j Y_{t-1}|. \]  

for all \( \alpha \in [\alpha_j, \alpha_{j+1}] \), \( j=0, 1, \ldots, \alpha_{j+1} \), and all possible values of \( Y_{t-1} \), i.e., \( I_1, I_2, \ldots, I_n \).

![Graph](image_url)

Fig 4. E(\( \alpha \)) versus \( \alpha \) for \( \alpha \) between (0.993, 0.996).
Therefore,
\[ E\{ (Y_i - [\alpha Y_{i-1}])^2 \} = E\{ (Y_i - [\alpha Y_{i-1}])^2 \}, \quad \forall \alpha \in [\alpha_i, \alpha_{j+1}] \] (15)
and there are only finite number of possible prediction errors each corresponds to \( \alpha = \alpha_i \) in (13) for \( \alpha \in [a, b] \). A typical \( E(\alpha) \) versus \( \alpha \) for a practical signal, e.g., the "LENA" image, is plotted in Fig. 4 for \( \alpha \in (0.993, 0.996) \). The above image data has values between 0 and 255.

One should be interested to know how many steps one, at most, needs to go through in order to find the optimum \( \alpha \). The \( E(\alpha) \)'s that need to be evaluated are the same number as \( \alpha_j \) in (13). If the signal is of integer values between 0 and 255, then between (0.8, 0.9), there are totally 2960 possible \( E(\alpha) \)'s. And in [0.9, 1.0], there are totally 2961 possible \( E(\alpha) \)'s. The number is not too many. Fig. 6 is a plot of the total number of possible \( E \), say \( N(\hat{\alpha}) \), for \( \alpha \) being in [0.8, \( \hat{\alpha} \)]. From it, one can get an idea of the maximum number of \( E(\alpha) \)'s needed to be evaluated for \( \alpha \) lying in \([a, b]\), where \( 0.8 \leq a, b \leq 1 \).

Also there are cases where the number of required searches is much fewer than what is in Fig. 6. If the number of possible value of \( Y_{i-1} \) is few, then the number of possible values of \( E(\alpha) \)'s is also much reduced. Table 1 is the number of \( \alpha \) for \( \alpha \in [0, 1] \) for different number and data value. As the dynamic range of the data is small, the number of \( \alpha \)'s could be as small as 34. For having the same max-min data value, the mid-value of the data, \( X \), will pretty much solely determine the maximum number of searches required. It is interesting to see that the number of required searches is pretty linear to \( X \) as in Fig. 6.

\[ \text{Fig 5. Relation between the number of the intervals of the prediction errors } \Delta \alpha \text{ and the mean } X \text{ of the signal if the data lie in } [X-5, X+5]. \]
The algorithm for finding the coefficient of one dimension first order causal minimum mean square error linear predictor with rounding is stated as the following:

**Initial**: Given \( \alpha = \alpha_n = a \)

\( j = 0 \)

**Loop**: While \( \alpha_j \leq \alpha_{\text{upper}} = b \)

\{ For \( (i = 1 \) to \( n \) ) do \{ \Delta \alpha(\alpha_j, X_j) = \frac{[\alpha_j Y_{i,j} + 0.5 - \alpha_j Y_{i-1}]}{Y_{i-1}} \}

Let \( \Delta \alpha_{\text{min}}^{(i)} = \min_{i=1,2,...,n} \Delta \alpha(\alpha_j, X_j) \)

Calculate the Error Signal Energy \, E(j),

Set \( \alpha_{j+1} = \alpha_j + \Delta \alpha_{\text{min}}^{(i)} \)

\( j = j + 1 \)

\}

**End**: Find the Minimum Error Signal Energy \, i.e., \( \min_j E(j) \)
and the corresponding coefficient \( \alpha_j \).
Possible values of data | Number of \( \Delta \alpha \)
---|---
1,3,6,14,31,67,150,200 | 532
0,1,3,6,14,31,67,150 | 296
1,3,6,14,31,67 | 129
1,3,6,14,31,50 | 108
1,3,6,14,31,40 | 101
1,3,6,14,31 | 60
1,3,5,7,9,11 | 34

Table 1.

Second, let's consider the one dimensional higher order causal linear predictor with rounding. The same as the first order case, we need to find the coefficients \( \beta_j, j = 1,2,\ldots, k \), which are the solution of

\[
\min_{\beta_j, j=1,2,\ldots,k} E[(Y - [\sum_{j=1}^{k} \beta_j Y_{j-1}])^2]
\]

For a given \((\beta_1,\beta_2,\ldots,\beta_k)\), there exits a maximum region \(D_{\max}(\beta_1,\beta_2,\ldots,\beta_k)\) surrounding \((\beta_1,\beta_2,\ldots,\beta_k)\), such that

\[
[\sum_{j=1}^{k} (\beta_j + \Delta \beta_j) Y_{j-1}] < [\sum_{j=1}^{k} \beta_j Y_{j-1}]
\]

for \((\beta_1 + \Delta \beta_1,\beta_2 + \Delta \beta_2,\ldots,\beta_k + \Delta \beta_k) \in D_{\max}(\beta_1,\beta_2,\ldots,\beta_k)\) and for all possibile values of \((Y_{i-1},Y_{i-2},\ldots,Y_{i-k})\).

For a given \((Y_{i-1},Y_{i-2},\ldots,Y_{i-k})\), \((\Delta \beta_1,\Delta \beta_2,\ldots,\Delta \beta_k)\) should satisfy, from the property of rounding.

\[
[\sum_{j=1}^{k} \beta_j Y_{j-1}] - 0.5 < \sum_{j=1}^{k} (\beta_j + \Delta \beta_j) Y_{j-1} < [\sum_{j=1}^{k} \beta_j Y_{j-1}] + 0.5.
\]

Equation (17) can be rewritten as

\[
r_{i,m} - 1 \leq \sum_{j=1}^{k} \Delta \beta_j Y_{j-1} < r_{i,m},
\]

where

\[
r_{i,m} = [\sum_{j=1}^{k} \beta_j Y_{j-1}] + 0.5 - \sum_{j=1}^{k} \beta_j Y_{j-1}.
\]

and \(m\) is the index of the element in the set, which contains all possible \((Y_{i-1},Y_{i-2},\ldots,Y_{i-k})\)

\[
C = \{(Y_{i-1},Y_{i-2},\ldots,Y_{i-k}) | \forall Y_i \in \{I_1, I_2, \ldots, I_n\}\}
\]

For different values of \((Y_{i-1},Y_{i-2},\ldots,Y_{i-k})\), we get different inequalities that \(\Delta \beta_1,\Delta \beta_2,\ldots,\Delta \beta_k\) should satisfy. And \(D_{\max}(\beta_1,\beta_2,\ldots,\beta_k)\) is formed by

\[
r_{i,m} - 1 \leq \sum_{j=1}^{k} \Delta \beta_j t_{j,m} < r_{i,m}, \quad m = 1,2,\ldots,n^k,
\]

where \(t_{j,m}\) is the jth component of the mth element in set \(C\). \(C = \{(Y_{i-1},Y_{i-2},\ldots,Y_{i-k}) | \forall Y_i \in \{I_1, I_2, \ldots, I_n\}\}\)

The \((\beta_1,\beta_2,\ldots,\beta_k)\) space is partitioned by such \(D_{\max}(\beta_1,\beta_2,\ldots,\beta_k)\), and \(E(\beta_1,\beta_2,\ldots,\beta_k)\) will be the same inside each \(D_{\max}(\beta_1,\beta_2,\ldots,\beta_k)\) region. The optimum \(E\) for a bounded area of \((\beta_1,\beta_2,\ldots,\beta_k)\) can be obtained by evaluating and comparing the \(E\)'s for different \(D_{\max}(\beta_1,\beta_2,\ldots,\beta_k)\) regions.
For multi-dimensional linear predictor with rounding, the predictor is identical to the case of one dimensional high order linear predictor with rounding with the

$$\Delta \beta_2$$

Fig. 7(a)

$$\Delta \beta_1$$

Fig. 7(b)

$$Y_{i-1}, j = 1, 2, \ldots, k$$ to be the multi-dimensional neighboring pixels and $$\beta_1, \beta_2, \ldots, \beta_k$$ are the corresponding scaling coefficients. We plot several regions of constant prediction error for the case one dimensional 2nd order linear predictor. They are in Fig. 7. They corresponds to the constant error region around $$(\beta_1, \beta_2) = (0, 0)$$ (Fig. 7(a)), $$(\beta_1, \beta_2) = (0.9, 0.9)$$ (Fig. 7(b)). In the plot, we should note that $$(0, 0)$$ correspond to the above $$(\beta_1, \beta_2)$$.

4. EXPERIMENT RESULT

We have performed two kinds of experiments. The first is on simulated data. The second is on real images. We obtain the true minimum mean square error rounded linear predictor by using the method proposed in section III and compare the results with the minimum mean square error linear predictor without considering the effect of rounding as in section II.

(A). Experiment on first order Markov sequences and decaying sinusoidal

We first perform experiments on the first order Markov sequence, generated according to $$y(n) = 0.9 y(n-1) + e(n)$$, where $$e(n)$$ is zero mean uncorrelated random noise. Table 2 shows the prediction coefficients, prediction errors, and the differences of the MMSE linear predictor considering rounding $$(\alpha^*, E^*)$$ and without considering rounding $$(\alpha, E)$$ for data with different variance of $$e(n)$$ in the generation equation of $$y(n)$$.

We see that, in all cases when the prediction results are rounded, MMSE linear predictors considering the effect of rounding will give smaller prediction error than that of MMSE linear predictors without considering the effect of rounding. Fig. 8 shows the square root of normalized of the MMSE prediction error energies, $$\Delta E_N$$, versus the prediction error energy, which is quite proportional to the energy of $$e(n)$$. We see that $$\Delta E_N$$ generally decreases as the prediction error increases. As the error energy of the MMSE linear prediction without considering rounding is small, say below 1.5, $$\Delta E_N$$ sometimes could be as large as 33% and sometimes are small. When $$E^*$$ (or $$E$$) is small, we see that the number of possible values of
the signal is small and the number of constant prediction error cells is not many. Therefore, depending on the values of α and α*, ΔEN will be zero if both α and α* are located on the same cell and will be big if they are located on different cells. (Since it is normalized by the prediction error and the prediction error is small in these cases). We have also performed experiment to see whether ΔEN has any relationship with the correlation between data samples. The data are first order Markov sequences with different values of β in y(n)=βy(n-1)+e(n) and the variance of e(n) are all 1.7689. The results are summarized in Table 3. We see that the value of ΔEN seems to have nothing to do with how much the data samples are correlated measured by R(1)/R(0).

<table>
<thead>
<tr>
<th>Data #</th>
<th>D20</th>
<th>D21</th>
<th>D22</th>
<th>D23</th>
<th>D28</th>
<th>D29</th>
<th>D31</th>
<th>D33</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAR</td>
<td>0.25</td>
<td>1</td>
<td>2.25</td>
<td>4</td>
<td>100</td>
<td>400</td>
<td>1600</td>
<td>3600</td>
</tr>
<tr>
<td>α</td>
<td>0.897355</td>
<td>0.903012</td>
<td>0.892499</td>
<td>0.898314</td>
<td>0.869982</td>
<td>0.846838</td>
<td>0.883811</td>
<td>0.919705</td>
</tr>
<tr>
<td>E</td>
<td>0.42</td>
<td>1.26</td>
<td>2.34</td>
<td>4.19</td>
<td>90.42</td>
<td>258.40</td>
<td>1222.48</td>
<td>2509.36</td>
</tr>
<tr>
<td>α*</td>
<td>0.897511</td>
<td>0.897849</td>
<td>0.897849</td>
<td>0.896226</td>
<td>0.876667</td>
<td>0.861111</td>
<td>0.892857</td>
<td>0.920765</td>
</tr>
<tr>
<td>E*</td>
<td>0.41</td>
<td>1.18</td>
<td>2.31</td>
<td>4.10</td>
<td>89.39</td>
<td>256.19</td>
<td>1215.92</td>
<td>2504.47</td>
</tr>
<tr>
<td>(\frac{\sqrt{E} - \sqrt{E^<em>}}{\sqrt{E^</em>}}) (%)</td>
<td>1.212</td>
<td>3.33</td>
<td>0.674</td>
<td>1.0916</td>
<td>0.574</td>
<td>0.43</td>
<td>0.269</td>
<td>0.0975</td>
</tr>
</tbody>
</table>

Table 2. Experiments on MMSE first order linear predictors with and without considering rounding for first order Markov sequence, y(n)=0.9*y(n-1)+e(n), VAR is the variance of e(n).

Fig 8. The square root of the normalized deviation of the prediction error energy versus the prediction error.
Table 3. Experiment on data with different correlation.

<table>
<thead>
<tr>
<th>α</th>
<th>0.9</th>
<th>0.65</th>
<th>0.45</th>
</tr>
</thead>
<tbody>
<tr>
<td>β</td>
<td>0.895922</td>
<td>0.646985</td>
<td>0.455512</td>
</tr>
<tr>
<td>E</td>
<td>1.95</td>
<td>1.02</td>
<td>0.88</td>
</tr>
<tr>
<td>α*</td>
<td>0.9</td>
<td>0.643333</td>
<td>0.45</td>
</tr>
<tr>
<td>E*</td>
<td>1.82</td>
<td>1.01</td>
<td>0.87</td>
</tr>
<tr>
<td>$\sqrt{E - E^*}$ (%)</td>
<td>26.73</td>
<td>9.95</td>
<td>10.72</td>
</tr>
<tr>
<td>$\sqrt{E - E^*}$ (%)</td>
<td>3.50</td>
<td>0.49</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Fig. 9(a) and Fig 9(b) are two signals which are decaying sinusoids, generated according to $y(n) = a e^{-bn} \cos cn + d$. where (a,b,c,d) are (100,0.01,1,155) and (100,0.01,0.1,155) respectively. Their E's (without considering rounding). E*'s (considering rounding) and the square root of the normalized deviation of the prediction errors ($\Delta E_N$'s) are 230.93, 227.116% and 2.622, 2.385, 23.48%. In these signals, when the sinusoidal is slowly decaying, $\Delta E_N$ is quite big.

(B). Experiments on gray level images

We calculate the first order MMSE linear predictor without considering the effect of rounding and the one considering the effect of rounding for four images: PEPPER, JET, LENA, and HEADMRI. The results are summarized in Table 4. We can see that the difference of the two prediction coefficients are typically smaller than 0.01, and considering the effect of rounding does make the MMSE prediction errors smaller than the ones without considering the effect of rounding. $\Delta E_N$ are all below 4% which is small.

Next we scale down the images to simulate the effect of dimmer illumination in image capturing or having a constant...
scaling down of values for whatever reasons. We see that, for some images (PEPPER, JET, LENA and HEADMRI), though the prediction coefficients are rather different, the prediction errors $E$ and $E^*$ are identical since the prediction error is small, the number of possible data values is small, and $\alpha$ and $\alpha^*$ may likely be located on the same cell of constant prediction error. For the other HEADMRI image, we see that $\Delta E_N$ is about 10%. This is consistent with the high fluctuation of $\Delta E_N$'s for small prediction errors observed before. This results are summarized in Table 5.

<table>
<thead>
<tr>
<th>Image Name</th>
<th>pepper</th>
<th>jet</th>
<th>lena</th>
<th>headmri</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.993392</td>
<td>0.996174</td>
<td>0.9943797</td>
<td>0.9760693</td>
</tr>
<tr>
<td>$E$</td>
<td>185.013140</td>
<td>213.937040</td>
<td>211.678952</td>
<td>71.829029</td>
</tr>
<tr>
<td>$\alpha^*$</td>
<td>0.995098</td>
<td>0.996815</td>
<td>0.996063</td>
<td>0.969565</td>
</tr>
<tr>
<td>$E^*$</td>
<td>184.870178</td>
<td>213.597855</td>
<td>211.471722</td>
<td>71.744730</td>
</tr>
<tr>
<td>$\sqrt{E - E^<em>} / \sqrt{E^</em>}$ (%)</td>
<td>3.55</td>
<td>3.98</td>
<td>3.13</td>
<td>3.43</td>
</tr>
<tr>
<td>$\sqrt{E - E^<em>} / \sqrt{E^</em>}$ (%)</td>
<td>0.063</td>
<td>0.079</td>
<td>0.049</td>
<td>0.059</td>
</tr>
</tbody>
</table>

Table 4. Experiment on gray level images

<table>
<thead>
<tr>
<th>Image Name</th>
<th>pepper</th>
<th>jet</th>
<th>lena</th>
<th>headmri</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.992768858</td>
<td>0.995438176</td>
<td>0.9916523</td>
<td>0.969974289</td>
</tr>
<tr>
<td>$E$</td>
<td>0.580484</td>
<td>0.518290</td>
<td>0.495803</td>
<td>0.279856</td>
</tr>
<tr>
<td>$\alpha^*$</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.937500</td>
</tr>
<tr>
<td>$E^*$</td>
<td>0.580484</td>
<td>0.518290</td>
<td>0.495803</td>
<td>0.271145</td>
</tr>
<tr>
<td>$\sqrt{E - E^<em>} / \sqrt{E^</em>}$ (%)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9.89</td>
</tr>
<tr>
<td>$\sqrt{E - E^<em>} / \sqrt{E^</em>}$ (%)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.488</td>
</tr>
</tbody>
</table>

Table 5. Experiment on scaled down images.

5. SUMMARY

In most digital signal processing and image coding applications, linear prediction results are rounded before the integer-valued prediction error signals are generated. Therefore, the conventional way of finding the MMSE linear prediction without considering this rounding effect does not give the true MMSE linear predictor with rounding.

This paper addresses the issue of finding the true MMSE linear predictor with rounding. We find that the prediction error for integer-valued signal with the prediction result rounded is a piecewise constant function and show that the boundary facets of the constant-prediction-error cells are portions of hyperplanes. For first order linear prediction on integer data with values being between 0 and 255, the prediction error has at most 2961 possible values for $\alpha$ being between 0.9 and 1.
In general, when the number of possible values of the data is small, the number of constant-prediction-error cells is not many. The search of the optimum prediction coefficients can be done very efficiently.

The relative difference of prediction errors between the MMSE predictors with and without considering the effect of rounding could be very significant for signals having small prediction errors.

6. REFERENCES


