Some Mathematics Behind Multi-Channel Prediction

Heiko Purnhagen
c/o Hendrik Fuchs, Institut für Theoretische Nachrichtentechnik
und Informationsverarbeitung, Uni Hannover
(purnhage@tnt.uni-hannover.de / fuchs@tnt.uni-hannover.de)

30 September 1994

1 Introduction

Multi-channel prediction is a technique to exploit redundancy between different channels of a multi-channel audio signal. Thus, the coding of these signals can be improved [1, 2].

The basic idea is to calculate an estimate $\hat{z}(n)$ for the signal $z(n)$ in a channel z based on the signals $x(n)$ and $y(n)$ in the channels x and y. If desired, the prediction can also be based on only a single signal $x(n)$.

When using prediction, instead of the signal $z(n)$ only the predictor error signal

$$e(n) = z(n) - \hat{z}(n) \tag{1}$$

has to be transmitted. The parameters describing the applied predictor (predictor coefficients etc.) have to be transmitted as side-information.

In the decoder, $x(n)$, $y(n)$ and the transmitted predictor parameters are used to calculate $\hat{z}(n)$ as in the encoder. Together with the transmitted predictor error signal $e(n)$, the original signal $z(n)$ can be reconstructed.

Prediction can be applied independently for several channels $z_1, z_2, \ldots$ if desired. If a subband-based coding scheme is used, prediction can (and should) be applied independently in the different subbands.

2 Multi-Channel Prediction

The estimate $\hat{z}(n)$ of the actual sample $z(n)$ is calculated as a linear combination of samples from $x(n)$ and $y(n)$

$$\hat{z}(n) = \sum_{k=0}^{K} a_k \, x(n - d_x - k) + \sum_{l=0}^{L} b_l \, y(n - d_y - l), \tag{2}$$

where a_k and b_l are the predictor coefficients, d_x and d_y are used for time delay compensation and K and L are the predictor orders.

Figure 1 shows a block diagram of a multi-channel predictor. Figure 2 illustrates the realisation of the "delay" and "FIR predictor" blocks for a single channel in such a predictor.

Compared to the power σ_z^2 of the original signal $z(n)$, the prediction error signal $e(n)$ only has the power σ_e^2 and thus the prediction gain

$$G = \frac{\sigma_z^2}{\sigma_e^2} = \frac{E[z^2(n)]}{E[e^2(n)]} \tag{3}$$

is obtained.

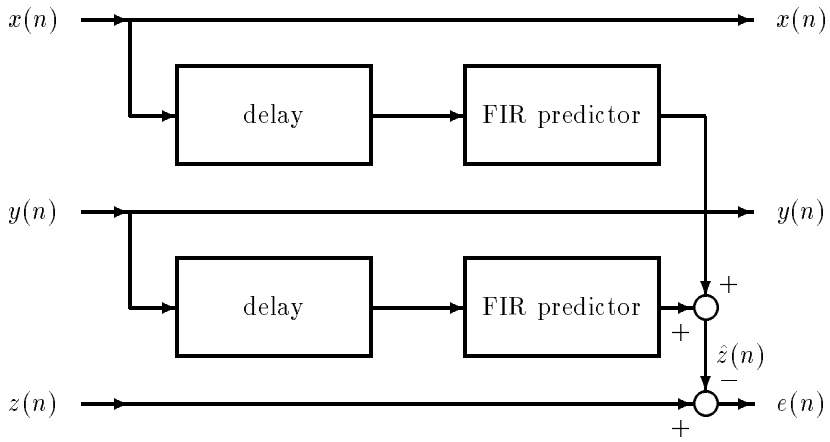

Figure 1: Block diagram of a multi-channel predictor

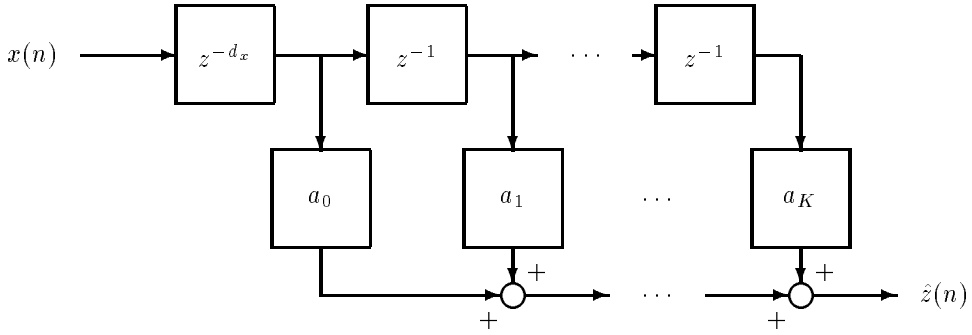

Figure 2: Realisation of delay and FIR predictor for a single channel x

2.1 Calculating The Predictor Coefficients

Assuming given values for the delay compensation d_x and d_y and given predictor orders K and L, the optimum predictor coefficients $a_0, \ldots, a_K, b_0, \ldots, b_L$ can be found by minimising the power of the predictiction error signal σ_e^2.

$$\sigma_e^2 = E[(z(n) - \hat{z}(n))^2] \tag{4}$$

$$= E\left[\left(z(n) - \sum_{k=0}^{K} a_k\, x(n - d_x - k) - \sum_{l=0}^{L} b_l\, x(n - d_y - l)\right)^2\right] \tag{5}$$

$$= E[z^2(n)] - 2\sum_{k=0}^{K} a_k\, E[z(n)\, x(n - d_x - k)] - 2\sum_{l=0}^{L} b_l\, E[z(n)\, y(n - d_y - l)] +$$

$$+ \sum_{k=0}^{K} a_k \sum_{i=0}^{K} a_i\, E[x(n - d_x - k)\, x(n - d_x - i)] + \sum_{l=0}^{L} b_l \sum_{j=0}^{L} b_j\, E[y(n - d_y - l)\, y(n - d_y - j)] +$$

$$+ 2\sum_{k=0}^{K} a_k \sum_{l=0}^{L} b_l\, E[x(n - d_x - k)\, y(n - d_y - l)] \tag{6}$$

To minimise σ_e^2, its partial derivatives with respect to the predictor coefficients $a_0, \ldots, a_K, b_0, \ldots, b_L$ are set to zero. For $a_0, \ldots, a_K$, this leads to a set of $K+1$ equations:

$$0 \stackrel{!}{=} \frac{dE[e^2(n)]}{da_i}, \quad 0 \le i \le K \tag{7}$$

$$= -2E[c(n)\,a(n-d_x-i)] + 2\sum_{k=0}^{K} a_k\, E[x(n-d_x-i)\,x(n-d_x-k)] +$$

$$+2\sum_{l=0}^{L} b_l\, E[x(n-d_x-i)\,y(n-d_y-l)], \quad 0 \le i \le K, \tag{8}$$

For $b_0, \ldots, b_L$, an equivalent set of $L+1$ equations is obtained.

These $K+L+2$ equations can be summarised in a linear equation system. To simplify its notation, the signal powers

$$\sigma_{x_k}^2 = E[x^2(n-d_x-k)], \tag{9}$$

$$\sigma_{y_l}^2 = E[y^2(n-d_y-l)], \tag{10}$$

$$\sigma_{z^2} = E[z^2(n)], \tag{11}$$

the auto-correlation coefficients

$$r_{x_{k,i}} = \frac{E[x(n-d_x-k)\,x(n-d_x-i)]}{\sigma_{x_k}\sigma_{x_i}}, \tag{12}$$

$$r_{y_{l,j}} = \frac{E[y(n-d_y-l)\,y(n-d_y-j)]}{\sigma_{y_l}\sigma_{y_j}}, \tag{13}$$

and the cross-correlation coefficients

$$c_{x_k,y_l} = \frac{E[x(n-d_x-k)\,y(n-d_y-l)]}{\sigma_{x_k}\sigma_{y_l}}, \tag{14}$$

$$c_{x_k,z} = \frac{E[x(n-d_x-k)\,z(n)]}{\sigma_{x_k}\sigma_z}, \tag{15}$$

$$c_{y_l,z} = \frac{E[y(n-d_y-l)\,z(n)]}{\sigma_{y_l}\sigma_z}, \tag{16}$$

are introduced. It should be noted that in this text, the symbols σ, r and c and the term "correlation coefficient" are defined in a way which is slightly different from their normal usage since constant signal components are *not* subtracted here.

Using these symbols, the linear equation system can be written as

$$\begin{pmatrix} r_{x_0,0} & \cdots & r_{x_0,K} & c_{x_0,y_0} & \cdots & c_{x_0,y_L} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ r_{x_K,0} & \cdots & r_{x_K,K} & c_{x_K,y_0} & \cdots & c_{x_K,y_L} \\ c_{y_0,x_0} & \cdots & c_{y_0,x_K} & r_{y_0,0} & \cdots & r_{y_0,L} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ c_{y_L,x_0} & \cdots & c_{y_L,x_K} & r_{y_L,0} & \cdots & r_{y_L,L} \end{pmatrix} \begin{pmatrix} \sigma_{x_0} a_0 \\ \vdots \\ \sigma_{x_K} a_K \\ \sigma_{y_0} b_0 \\ \vdots \\ \sigma_{y_L} b_L \end{pmatrix} = \sigma_z \begin{pmatrix} c_{x_0,z} \\ \vdots \\ c_{x_K,z} \\ c_{y_0,z} \\ \vdots \\ c_{y_L,z} \end{pmatrix}. \tag{17}$$

The matrix in this equation system is symmetric with the diagonal elements $r_{x_k,k} = r_{y_l,l} = 1$. To solve this equation system and thus find the optimum predictor coefficients, the *Cholesky decomposition* (see Section 3) can be used. The Colesky decomposition is possible since the matrix in Equation 17 in practice

is usually positive definite (see Section 4). After the optimum predictor coefficients are found, the power of the prediction error signal $e(n)$ can be calculated as

$$\sigma_e^2 = \sigma_z^2 - 2\sum_{k=0}^{K} a_k \sigma_{x_k} \sigma_z c_{x_k,z} - 2\sum_{l=0}^{L} b_l \sigma_{y_l} \sigma_z c_{y_l,z} + \sum_{k=0}^{K} a_k \sigma_{x_k} \sum_{i=0}^{K} a_i \sigma_{x_i} r_{x_k,i} + \sum_{l=0}^{L} b_l \sigma_{y_l} \sum_{j=0}^{L} b_j \sigma_{y_j} r_{y_l,j} +$$
$$+2\sum_{k=0}^{K} a_k \sigma_{x_k} \sum_{l=0}^{L} b_l \sigma_{y_l} c_{x_k,y_l}. \tag{18}$$

Using the relations in the linear equation system (Equation 17), Equation 18 can be written as

$$\sigma_e^2 = \sigma_z^2 - \sigma_z \left(\sum_{k=0}^{K} a_k \sigma_{x_k} c_{x_k,z} + \sum_{l=0}^{L} b_l \sigma_{y_l} c_{y_l,z} \right) \tag{19}$$

and thus the prediction gain G (Equation 3) can be calculated:

$$G = \frac{1}{1 - \frac{1}{\sigma_z} \left(\sum_{k=0}^{K} a_k \sigma_{x_k} c_{x_k,z} + \sum_{l=0}^{L} b_l \sigma_{y_l} c_{y_l,z} \right)} \tag{20}$$

2.2 Prediction Based On A Single Channel

When only a single signal $x(n)$ is used as the basis for estimating $z(n)$, the linear equation system (Equation 17) is reduced to

$$\begin{pmatrix} r_{x_0,0} & \cdots & r_{x_0,K} \\ \vdots & \ddots & \vdots \\ r_{x_K,0} & \cdots & r_{x_K,K} \end{pmatrix} \begin{pmatrix} \sigma_{x_0} a_0 \\ \vdots \\ \sigma_{x_K} a_K \end{pmatrix} = \sigma_z \begin{pmatrix} c_{x_0,z} \\ \vdots \\ c_{x_K,z} \end{pmatrix}. \tag{21}$$

Instead of using Equation 18, σ_e^2 can now be calculated as

$$\sigma_e^2 = \sigma_z^2 - 2\sum_{k=0}^{K} a_k \sigma_{x_k} \sigma_z c_{x_k,z} + \sum_{k=0}^{K} a_k \sigma_{x_k} \sum_{i=0}^{K} a_i \sigma_{x_i} r_{x_k,i}. \tag{22}$$

Using the relations in the linear equation system (Equation 21), Equation 22 can be written as

$$\sigma_e^2 = \sigma_z^2 - \sigma_z \sum_{k=0}^{K} a_k \sigma_{x_k} c_{x_k,z} \tag{23}$$

and thus the prediction gain G can be calculated:

$$G = \frac{1}{1 - \frac{1}{\sigma_z} \sum_{k=0}^{K} a_k \sigma_{x_k} c_{x_k,z}} \tag{24}$$

2.3 Some Simple Predictors

To illustrate some of the basic properties of predictors, two very simple examples will be given now:

For the simplest predictor based on a single signal $x(n)$ (predictor order $K = 0$), the optimum predictor coefficient is

$$a_0 = \frac{\sigma_z}{\sigma_{x_0}} c_{x_0,z}. \tag{25}$$

Such a predictor achieves the mimimum predictor error signal power

$$\sigma_e^2 = \sigma_z^2(1 - c_{x_0,z}^2) \tag{26}$$

which corresponds to a prediction gain of

$$G = \frac{1}{1 - c_{x_0,z}^2}. \tag{27}$$

For the simplest predictor based on two signals $x(n)$, $y(n)$ (predictor orders $K = L = 0$), the optimum predictor coefficients are

$$a_0 = \frac{\sigma_z}{\sigma_{x_0}} \frac{c_{x_0,z} - c_{y_0,z}c_{x_0,y_0}}{1 - c_{x_0,y_0}^2} \tag{28}$$

$$b_0 = \frac{\sigma_z}{\sigma_{y_0}} \frac{c_{y_0,z} - c_{x_0,z}c_{x_0,y_0}}{1 - c_{x_0,y_0}^2}. \tag{29}$$

Such a predictor achieves the mimimum predictor error signal power

$$\sigma_e^2 = \sigma_z^2 \left(1 - \frac{c_{x_0,z}^2 + c_{y_0,z}^2 - 2c_{x_0,z}c_{y_0,z}c_{x_0,y_0}}{1 - c_{x_0,y_0}^2}\right) \tag{30}$$

which corresponds to a prediction gain of

$$G = \frac{1}{1 - \frac{c_{x_0,z}^2 + c_{y_0,z}^2 - 2c_{x_0,z}c_{y_0,z}c_{x_0,y_0}}{1 - c_{x_0,y_0}^2}}. \tag{31}$$

2.4 Adaptive Prediction

Since the properties of the signal beeing coded usually vary with the time, the predictor should be automatically adapted to the signal properties to obtain the maximum performance of this prediction technique. If a block-based coding scheme is used, this adaption can be implemented by optimising the predictor parameters individually for each block of N samples by using the short-time averages

$$\frac{1}{N}\sum_{i=0}^{N-1} f(n+i) \quad \text{instead of} \quad E[f(n)]. \tag{32}$$

Thus, the energy of the prediction error signal within each block is minimised.

2.5 Optimising The Delay Compensation

To find the optimum values for the delay compensation d_x, d_y, the prediction gain G_{d_x,d_y} should be calculated for all possible combinations of d_x and d_y within a given range. Thus, the optimum delay compensation

$$(d_x, d_y)_{\text{opt}} = \arg \max_{(d_x,d_y)} G_{d_x,d_y} \tag{33}$$

resulting in a maximum prediction gain G_{opt} for the given predictor orders K, L can be found. If also negative values for d_x, d_y are allowed, the coding delay in the encoder and decoder increases.

2.6 Choosing The Predictor Orders

To determine the optimum values for the predictor orders K, L (which describe the complexity of the predictor), the properties of the actual signals $x(n)$, $y(n)$ and $z(n)$ should be taken into account.

If, for example, the signal $z(n)$ is found to be

$$z(n) = ax(n) + by(n), \quad a, b \text{ constant}, \tag{34}$$

a predictor with $K = L = 0$ is sufficient. For higher values of K or L, the linear equation system (Equation 17) obviously does not any longer have an unique solution for the coefficients a_k, b_l. This means that here, a more complex predictor than one with $K = L = 0$ does not offer any further advantages, but on the other side requires more side-information for transmitting the predictor coefficients.

When determining the optimum predictor orders K, L, the required side-information $S_{K,L}$ for transmitting the predictor parameters has to be considered. The remaining prediction gain

$$R_{K,L} = G_{\text{opt}K,L} - S_{K,L} \tag{35}$$

should be calculated for all combinations of K and L within a given range. Thus, the optimum predictor orders

$$(K, L)_{\text{opt}} = \arg \max_{(K,L)} R_{K,L} \tag{36}$$

allowing the highest remaining prediction gain R_{opt} can be found. If no unique solution to the linear equation system (Equation 17) can be found, the predictor complexity is higher than required.

3 Cholesky Decomposition

The *Cholesky decomposition* is an efficient technique to solve the linear equation system

$$A\vec{x} = \vec{b} \tag{37}$$

if A and $\vec{b}$ are known and if the matrix A is symmetric (i.e. $a_{i,j} = a_{j,i}$) and positive definite [4]. The matrix A is positive definite if

$$\vec{v}^T A \vec{v} > 0 \quad \text{for all} \quad \vec{v} \neq \vec{0}. \tag{38}$$

If A is not positive definite, the Cholesky decompostition of A turns out to be impossible.

In a first step (the decomposition of A), the lower triangular matrix U with

$$A = UU^T \tag{39}$$

is calculated. The $n \times n$ elements $u_{i,j}$ of U are found column by column with $k = 1, 2, \ldots, n$. First, the diagonal element

$$u_{k,k} = +\sqrt{a_{kk} - \sum_{i=1}^{k-1} u_{k,i}^2} \tag{40}$$

in column k and then the remaining elements

$$u_{j,k} = \frac{a_{j,k} - \sum_{i=1}^{k-1} u_{j,i} u_{k,i}}{u_{k,k}}, \quad j = k+1, k+2, \ldots, n \tag{41}$$

in column k are calculated.

In two final steps, first $\vec{c}$ with $U\vec{c} = \vec{b}$ and then $\vec{x}$ with $U^T\vec{x} = \vec{c}$ are calculated. Since

$$A\vec{x} = UU^T\vec{x} = U\vec{c} = \vec{b}, \tag{42}$$

$\vec{x}$ is the solution to the linear equation system $A\vec{x} = \vec{b}$ (Equation 37). First, the elements

$$c_k = \frac{b_k - \sum_{i=1}^{k-1} u_{k,i} c_i}{u_{k,k}} \quad k = 1, 2, \ldots, n. \tag{43}$$

of $\vec{c}$ are calculated. Thereafter, the elements

$$x_k = \frac{c_k - \sum_{i=k+1}^{n} u_{i,k} x_i}{u_{k,k}} \quad k = n, n-1, \ldots, 1. \tag{44}$$

of $\vec{x}$ can be calculated.

4 Why Correlation Matrices Are Positive Semidefinite

Correlation matrices, like the matrix in Equation 17, are known to be positive semidefinite [3]. This property will be explained here.

A matrix A is positive semidefinite if

$$\vec{v}^T A \vec{v} \geq 0 \quad \text{for all} \quad \vec{v} \neq \vec{0}. \tag{45}$$

To calulate the correlation matrix A, short-time averages over a block of N samples are usually used here (Equation 32). In this case, A can be calculated as

$$A = (N^{-1})^T S^T S N^{-1} \tag{46}$$

where

$$S = \begin{pmatrix} \vec{x}_0 & \cdots & \vec{x}_K & \vec{y}_0 & \cdots & \vec{y}_L \end{pmatrix} \tag{47}$$

consists of vectors

$$\vec{x}_k = \begin{pmatrix} x(n-d_x-k+0) \\ x(n-d_x-k+1) \\ \vdots \\ x(n-d_x-k+N-1) \end{pmatrix} \quad \text{and} \quad \vec{y}_l = \begin{pmatrix} y(n-d_y-l+0) \\ y(n-d_y-l+1) \\ \vdots \\ y(n-d_y-l+N-1) \end{pmatrix} \tag{48}$$

each containing N samples of the signals $x(n)$ and $y(n)$. The matrix

$$N = \begin{pmatrix} |\vec{x}_0| & & & & & \\ & \ddots & & & 0 & \\ & & |\vec{x}_K| & & & \\ & & & |\vec{y}_0| & & \\ & 0 & & & \ddots & \\ & & & & & |\vec{y}_L| \end{pmatrix} \tag{49}$$

is used for normalisation (i.e. calculate correlation *coefficients*) and contains

$$|\vec{x}_k| = \sigma_{x_k} \sqrt{N} \quad \text{and} \quad |\vec{y}_l| = \sigma_{y_l} \sqrt{N} \tag{50}$$

using the definition

$$|\vec{z}| = \sqrt{\vec{z}^T \vec{z}}. \tag{51}$$

By substituting the matrix A in Equation 45 using Equation 46, the equation

$$\vec{v}^T A \vec{v} = \vec{v}^T (N^{-1})^T S^T S N^{-1} \vec{v} \tag{52}$$

is obtained. Using the substitution

$$\vec{w} = S N^{-1} \vec{v}, \tag{53}$$

now results in

$$\vec{v}^T A \vec{v} = \vec{w}^T \vec{w} = |\vec{w}|^2 \geq 0. \tag{54}$$

This shows, that the correlation matrix A is always positive semidefinite.

If $\det(A) \neq 0$ (which means that the equation system has an unique solution) then A is positive definite and thus Choleky decomposition is possible. To show this relation, a matrix A being positive definite is assuned. This matrix has only positive eigenvalues $\lambda > 0$ (see [5, p. 219]) and therefore $\det(A)$ can not be 0 since A otherwise would have at least one eigenvalue $\lambda = 0$. (P.S.: Thanks to Bernd Elder at Uni Hannover, who helped me with these proofs.)

References

[1] Purnhagen, H.: *Implemetation and investigation of a technique for the reduction of stereo redundancy used for digital audio coding* (in German). Thesis (Studienarbeit), Institut für Theoretische Nachrichtentechnik und Informationstheorie, Uni Hannover, December 1992.

[2] Fuchs, H.: *Improving Joint Stereo Audio Coding By Adaptive Inter-Channel Prediction.* Proc. of the 1993 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, October 17-20, 1993, Mohonk Mountain House, New Paltz, New York.

[3] Makhoul, J.: *Linear Prediction: A Tutorial Review.* Proc. IEEE, vol. 63, pp 561–580, Apr. 1975.

[4] Feldmann, D.: *Repetitorium der Ingenieurmathematik, Teil II, Ausgabe H* (in German). Hannover, Summer 1989.

[5] Bronstein, I.N., Semendjajew, K.A.: *Taschenbuch der Mathematik* (in German). Teubner, Leipzig, 1979.