

**Heinrich-Hertz-Institut
für Nachrichtentechnik
Berlin GmbH**

Technischer Bericht Nr. 211

VERGLEICH VERSCHIEDENER VERFAHREN

ZUR LINEAREN PRÄDIKTION

von

Dieter Wörner

Zusammenfassung

Im ersten Abschnitt wird kurz erklärt, was hier unter linearer Prädiktion verstanden wird. Es werden die Bestimmungsgleichungen für die Prädiktorkoeffizienten aufgestellt, wobei grundsätzlich zwei Methoden unterschieden werden: die Kovarianz und die Autokorrelation Methode. Aus der Levinson-Rekursion wird dann die Lattice-Filterstruktur abgeleitet.

Im zweiten Abschnitt werden von der Lattice-Filterstruktur ausgehend, einem Aufsatz von John Makhoul folgend, verschiedene Möglichkeiten angegeben, die PARCOR-Koeffizienten zu definieren und zu bestimmen, die die Stabilität des (inversen) Filters garantieren. Diese Bestimmungsgleichungen können sowohl rekursiv (über die Prädiktorkoeffizienten a_i) als auch nichtrekursiv (aus den Zustandsgrößen des Lattice-Modells) gelöst werden.

Im dritten Abschnitt werden für die nichtrekursiven Methoden aus dem zweiten Abschnitt zu deren Lösung Signalflußgraphen angegeben, die die entsprechenden Gleichungen exakt repräsentieren. Daraus werden dann aus ingenieurmäßigen Gründen Näherungen abgeleitet.

Inhaltsverzeichnis

| | |
|---|----|
| 1. Lineare Prädiktion | 1 |
| 1.1 Einleitende Bemerkung | 1 |
| 1.2 Aufstellen der Bestimmungsgleichungen für die Prädiktorkoeffizienten | 4 |
| 1.3 Die LEVINSON-Rekursion | 5 |
| 1.4 Die Lattice-Filterstruktur | 6 |
| | |
| 2. Lattice Methoden | 11 |
| 2.1 Vorbetrachtung | 11 |
| 2.2 Stable and Efficient Lattice Methods for Linear Prediction | 13 |
| 2.2.1 The Covariance-Lattice Method | 16 |
| | |
| 3. Umsetzung der Algorithmen in Signalflußgraphen | 18 |
| 3.1 Der Algorithmus von ITAKURA und SAITO (Autokorrelation) | 18 |
| 3.2 Bestimmung der PARCOR-Koeffizienten nach BURG | 22 |
| 3.2.1 Verringerung des Hardwareaufwands durch ein rekursives Filter 1. Ordnung | 26 |
| 3.3 Eine Begründung für die getroffene Auswahl | 29 |
| | |
| Literatur | 30 |

1. Lineare Prädiktion

1.1 Einleitende Bemerkung

Bei einem Sprachsignal $s(\cdot)$ ist der aktuelle Wert des Signals $s(t)$ korreliert mit Werten zu anderen Zeitpunkten t' . Eine Möglichkeit, diese Redundanz zu beseitigen, besteht darin, einen Fehler einzuführen zwischen den Werten $s_k := s(kT)$ der Funktion und Schätzwerten \hat{s}_k

$$(1.1) \quad e_k := s_k - \hat{s}_k$$

Im allgemeinen sind die Schätzwerte \hat{s}_k eine Linearkombination eines Teils der (Eingangs) Folge. In dieser Zusammenstellung werden die Schätzwerte durch lineare Prädiktion /3/ gewonnen,

$$(1.2) \quad \hat{s}_k := - \sum_{i=1}^p a_i s_{k-i}$$

also durch Linearkombination vergangener Signalwerte, und die Koeffizienten a_i des Prädiktors H_p , die Prädiktorkoeffizienten, werden durch die Statistik der Eingangsfolge festgelegt, indem eine geeignete Größe minimiert wird. Ist die Eingangsfolge nicht stationär, so müssen die Prädiktorkoeffizienten an das Signal angepaßt werden.

Im Bildbereich ($\{v_i\} \xrightarrow{z} V(z)$) erhält man

$$(1.3) \quad \begin{aligned} \hat{S}(z) &= H_p(z) S(z) \\ E(z) &= S(z) - \hat{S}(z) = (1-H_p(z))S(z) \text{ mit } H_p(z) = - \sum_{i=1}^p a_i z^{-i} \end{aligned}$$

Die Prädiktorkoeffizienten ergeben sich aus der Forderung

$$(1.4) \quad \alpha = \sum_k e_k^2 \stackrel{!}{=} \text{Min.}$$

Wird der Fehler e quantisiert, so kann das (vereinfachte) Strukturbild angegeben werden; Bild 1.1.

Bezeichnet man als Quantisierungsrauschen q_k die Differenz $q_k := e_k - \tilde{e}_k$, so ergibt sich

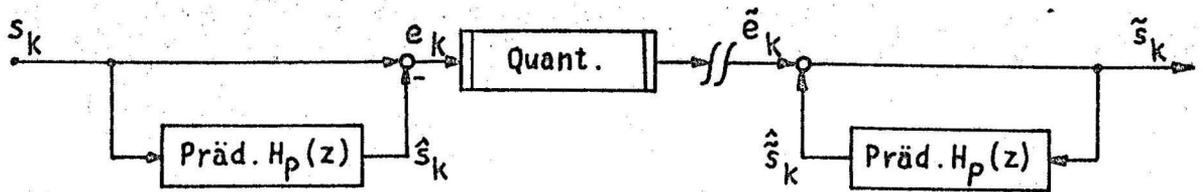


Bild 1.1 Quantisierung des Fehlers e

$$s_k - \tilde{s}_k = e_k + \hat{s}_k - (\tilde{e}_k + \hat{\tilde{s}}_k) = q_k + (\hat{s}_k - \hat{\tilde{s}}_k) \text{ mit}$$

$$\hat{\tilde{s}}_k = - \sum_{i=1}^p a_i \tilde{s}_{k-i}$$

Das Quantisierungsrauschen im empfangenen Signal \tilde{s}_k besteht also nicht nur aus dem Quantisierungsrauschen des Fehlersignals, sondern zusätzlich noch aus dem Quantisierungsfehler der Schätzung. Dieser Nachteil kann vermieden werden, wenn der Quantisierer in einen Regelkreis gelegt wird /4/:

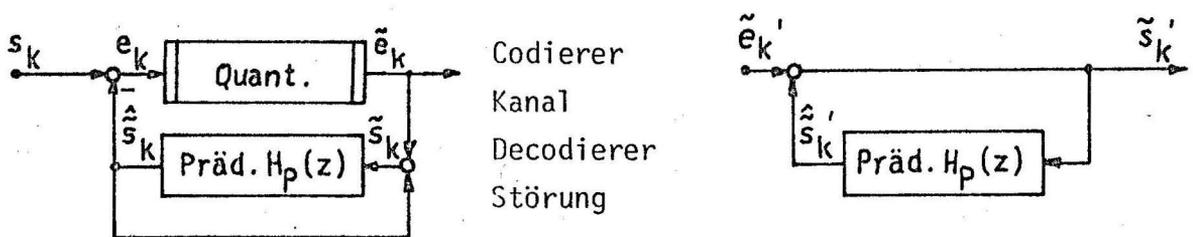


Bild 1.2 Allgemeines Schema einer DPCM

Das Quantisierungsrauschen im rekonstruierten Signal \tilde{s}_k ist dasselbe wie im Fehlersignal

$$q_k = e_k - \tilde{e}_k = s_k - \hat{s}_k - (\tilde{s}_k - \hat{\tilde{s}}_k) = s_k - \tilde{s}_k$$

Um die Prädiktorkoeffizienten zu bestimmen, gelte nun $e_k = \tilde{e}_k = \tilde{e}_k'$; somit kann folgender Signalflußplan angegeben werden.

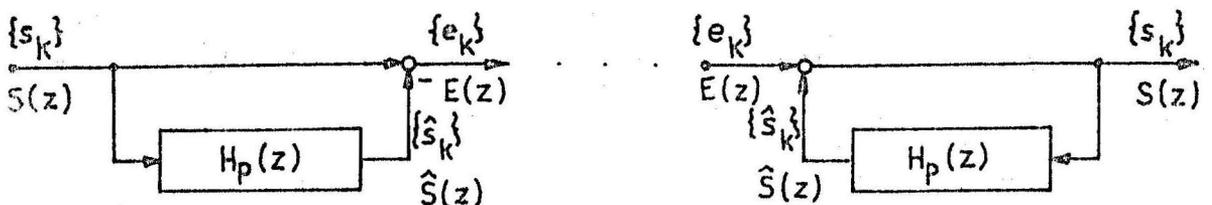


Bild 1.3

Signalflußgraph einer DPCM zur Bestimmung der Prädiktorkoeffizienten.

Die Methode der linearen Prädiktion wurde von Saito und Itakura /1/ und von Atal und Schroeder /2/ zum ersten Mal auf die Analyse bzw. Synthese von Sprache angewendet.

Bei der ADPCM werden die Prädiktorkoeffizienten an die (nicht stationäre) Statistik des Signals angepaßt. Diese Werte müssen quantisiert und übertragen werden. Die Stabilitätsprüfung des Filters $(1-H_p(z))^{-1}$ auf der Syntheseseite z.B. mittels des ZYPKIN-Verfahrens ist sehr aufwendig; weiter sind die Nullstellen von $(1-H_p(z))$ gegen Parametervariation der Prädiktorkoeffizienten nach Gl. (1.3) empfindlich, so daß, selbst wenn auf der Analyseseite alle Wurzeln von $(1-H_p(z))$ im Einheitskreis liegen, dies auf der Syntheseseite nicht der Fall zu sein braucht (quantisieren, co-decodieren, Kanalstörungen). Diese Überlegungen, u.a., führen dazu, eine andere Filterstruktur zu verwenden, die auch eine einfachere Stabilitätsprüfung ermöglicht. Ein wesentlicher Teil dieser Zusammenstellung besteht in dem Auflisten einiger aus der Literatur bekannter Lösungsverfahren zur Bestimmung der Koeffizienten dieses Filters.

Zunächst wird die Bestimmungsgleichung für die Prädiktorkoeffizienten angegeben, wobei grundsätzlich zwei Methoden unterschieden werden: die Kovarianz (nichtstationäres Signal) - und die Autokorrelation (stationäres Signal) - Methode.

1.2 Aufstellen der Bestimmungsgleichungen für die Prädiktorkoeffizienten

Literatur /5/

Gegeben sei eine Folge von N Abtastwerten $\{s(n)\}_{n=0}^{N-1}$. Es soll zu $s(n)$ ein Schätzwert $\hat{s}(n)$ als Linearkombination p vergangener Abtastwerte

$$\hat{s}(n) := - \sum_{i=1}^p a_i s(n-i)$$

derart bestimmt werden, daß die Summe über die Fehlerquadrate

$$\alpha := \sum_n e^2(n) \quad \text{mit dem Fehler}$$

$$e(n) = s(n) - \hat{s}(n)$$

das Infimum annimmt.

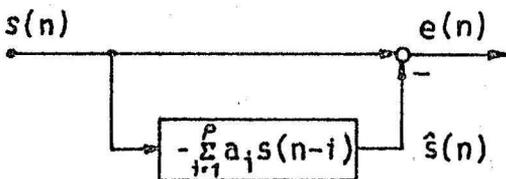


Bild 1.4 Signalflußgraph zur linearen Prädiktion

Der Freiheitsgrad besteht somit in der Wahl der Koeffizienten a_i , $i=1,2,\dots,p$ des Prädiktors H_p .

Bei der Covarianz-Methode erstreckt sich die Summation in Gleichung (1.4) von $k = p$ bis $N-1$, so daß α nur im Intervall $[p, N-1]$ minimisiert wird und alle N Abtastwerte zur Bestimmung der Elemente c_{ij} der Covarianzmatrix \underline{C} verwendet werden,

$$(1.5) \quad c_{ij} := \sum_{n=p}^{N-1} s(n-i)s(n-j); \quad i,j = 0,1,\dots,p$$

Die Prädiktorkoeffizienten ergeben sich als Lösung des Gleichungssystems.

$$(1.6) \quad \sum_{i=1}^p a_i c_{ik} = -c_{0k}; \quad k = 1,2, \dots, p$$

Der minimale Wert von α in Gl. (1.4) ist gegeben durch

$$\alpha_{\min} = \inf_{a_i} \{ \alpha : i = 1, 2, \dots, p \}$$

$$(1.7) \quad \alpha_{\min} = c_{00} + \sum_{i=1}^p a_i c_{i0}; \quad a_i \text{ nach Gl. (1.6)}$$

Die Summe $\sum_{i=1}^p a_i c_{i0}$ kann nicht positiv sein, da sonst durch die Wahl von $a_i = 0$ für alle i α_{\min} verkleinert würde (was ein Widerspruch ist). Andererseits ist α als Summe von quadratischen Gliedern immer positiv

$$0 \leq \alpha_{\min} \leq c_{00}$$

Bei der Autokorrelation-Methode erstreckt sich die Summation in Gl. (1.4) von $k = -\infty$ bis $+\infty$. Für s_k wird definitionsgemäß der Wert Null eingesetzt für k außerhalb des Intervalls $[0, N-1]$. Die Koeffizienten der Kovarianzmatrix vereinfachen sich zu

$$c_{ij} = \sum_{n=0}^{N-1-|i-j|} s(n) s(n + |i-j|)$$

$$(1.8) \quad r_{|i-j|} := c_{ij}$$

Somit folgt für den minimalen Wert des mittleren quadratischen Fehlers

$$(1.9) \quad \alpha_{\min} = r_0 + \sum_{i=1}^p a_i r_i$$

wobei die a_i sich ergeben als Lösung der Gleichung

$$(1.10) \quad \sum_{i=1}^p a_i r_{|i-k|} = -r_k \quad ; \quad k = 1, 2, \dots, p$$

1.3 Die LEVINSON-Rekursion zur Lösung des Gleichungssystems (1.10)

Literatur /6/. (Levinson hat ein allgemeineres Gleichungssystem untersucht) Die Gleichung (1.10) wird rekursiv gelöst, die Filterkoeffizienten doppelt indiziert, m sei der Rekursionsschritt (erster Index), zu lösen ist Gl.(1.10)

$$\sum_{i=1}^m a_{m,i} + r^{|k-i|} = -r_k \quad ; \quad k = 1, 2, \dots, m \quad ; \quad m = 1, 2, \dots, p$$

Dazu wird eine Hilfsfolge eingeführt

$$b_{1,1} = -\frac{r_1}{r_0}$$

$$b_{m,1} = -\frac{\sum_{i=1}^{m-1} b_{m-1,i} r_i + r_m}{\sum_{i=1}^{m-1} b_{m-1,m-i} r_i + r_0} \quad ; \quad m = 2, 3, \dots, p$$

$$b_{m,k} = b_{m-1,k-1} + b_{m,1} \cdot b_{m-1, m+1-k} \quad ; \quad k = 2, 3, \dots, m$$

$$a_{1,1} = -\frac{r_1}{r_0} \quad \text{folgt aus (1.10)}$$

$$a_{m,k} = b_{m, m+1-k} \quad ; \quad k = 1, 2, \dots, m \quad ; \quad m = 2, 3, \dots, p$$

LEVINSON - Rekursion

1.4 Die Lattice-Filterstruktur

Ausgehend von der Levinson-Rekursion kann die Lattice-Filterstruktur abgeleitet werden. Es ist (Bild 1.4)

$$e(n) = s(n) - \hat{s}(n) = \sum_{i=0}^m a_{m,i} s(n-i) \quad \text{mit } a_{m,0} = 1$$

die \mathcal{Z} -Transformation liefert

$$E(z) = \sum_{i=0}^m a_{m,i} z^{-i} S(z)$$

$$A_m(z) := \frac{E(z)}{S(z)} = \sum_{i=0}^m a_{m,i} z^{-i} \quad ; \quad a_{m,0} = 1$$

wird ein Polynom definiert

$$B_m(z) := \sum_{i=1}^{m+1} b_{m,i} z^{-i} \quad ; \quad b_{m,m+1} = 1$$

so ergibt sich eine Verknüpfung zwischen den beiden Polynomen durch die Rekursion

$$A_m(z) = A_{m-1}(z) + k_m B_{m-1}(z)$$

$$B_m(z) = z^{-1} [k_m A_{m-1}(z) + B_{m-1}(z)] \quad \text{mit}$$

$$k_m := b_{m,1} = a_{m,m}$$

$$B_m(z) = z^{-(m+1)} A_m(z^{-1})$$

wird weiter definiert

$$X_m^+(z) := A_m(z) S(z)$$

(1.11)

$$X_m^-(z) := B_m(z) S(z)$$

und werden diese neuen Größen in die darüberstehenden Gleichungen eingesetzt, so liefert die inverse \mathcal{Z} -Transformation

$$x_m^+(n) = x_{m-1}^+(n) + k_m x_{m-1}^-(n) \quad ; \quad x_0^+(n) = s(n)$$

(1.12)

$$x_m^-(n+1) = k_m x_{m-1}^+(n) + x_{m-1}^-(n) \quad ; \quad x_0^-(n+1) = s(n) \quad m = 1, 2, \dots, p$$

Aus Gl. (1.12) läßt sich ein Signalflußgraph konstruieren

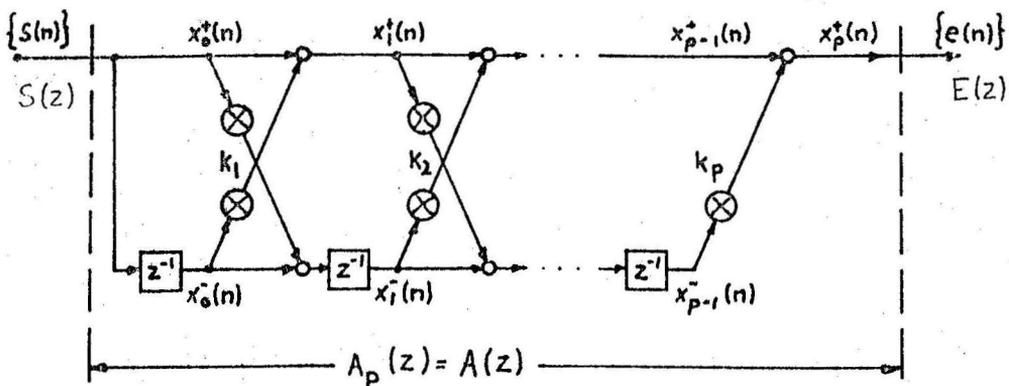


Bild 1.5 Lattice-Filterstruktur zur linearen Prädiktion

Bild 1.5 wurde von Wakita /7/ angegeben und die Parameter k_i von Itakura und Saito /8,9,10/ als partielle Correlationskoeffizienten, PARCOR-Koeffizienten, definiert.

Eliminiert man in der Levinson-Rekursion die Koeffizienten b_i durch die entsprechenden a_i und schreibt den Nenner von $b_{m,1}$ rekursiv, so erhält man den Rekursionsalgorithmus von DURBIN (zitiert in /11/)

$$\begin{aligned}
 E_0 &= r_0 \\
 k_1 &= a_{1,1} = - \frac{r_1}{E_0} \\
 k_m &= - \frac{\sum_{i=1}^{m-1} a_{m-1,i} r_{m-i} + r_m}{E_{m-1}} \\
 E_{m-1} &= (1 - k_{m-1}^2) E_{m-2} \\
 a_{m,i} &= a_{m-1,i} + k_m a_{m-1,m-i} \\
 a_{m,m} &= k_m \quad ; \quad m = 2, 3, \dots, p \quad ; \quad i = 1, 2, \dots, m-1
 \end{aligned}$$

DURBIN - Rekursion

Der Algorithmus setzt bei der Berechnung von k_m Optimalität des $(m-1)$ -ten Rekursionsschrittes voraus, k_{m-1} kann also nicht innerhalb der Rekursion quantisiert werden!

Bemerkung zu den eingeführten Polynomen $B_m(z)$.

Wird Gl. (1.11) durch die entsprechenden Größen im Folgenbereich ausgedrückt und definiert man

$$\hat{s}_m(n) = \sum_{i=1}^m a_{m,i} s(n-i)$$

$$\hat{s}_m(n-(m+1)) := - \sum_{i=1}^m b_{m,i} s(n-i)$$

so kann $x_m^+(n) = s(n) - \hat{s}_m(n)$ Vorwärtsschätzfehler eines Prädiktors m -ter

Ordnung genannt werden; für $m = p$ gilt dann $x_p^+(n) = e(n)$. Analog kann $x_m^-(n) = s(n-(m+1)) - \hat{s}_m(n-(m+1))$ als Rückwärtsschätzfehler eines Prädiktors m -ter Ordnung bezeichnet werden, da aus m -vergangenen Werten von $s(n)$ der $(m+1)$ -te vergangene geschätzt wird. Bild 1.6

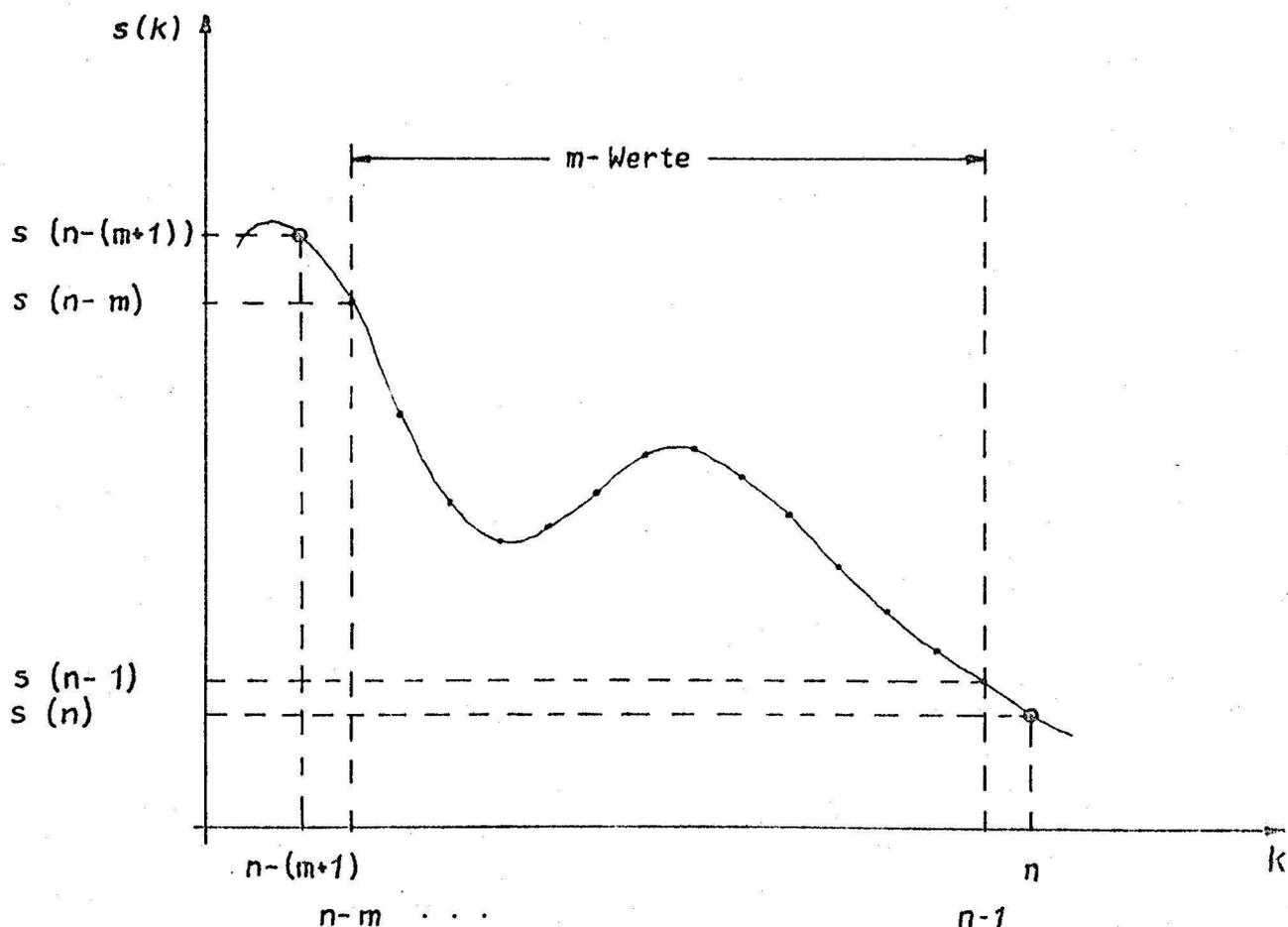


Bild 1.6

Es ist der eingekreiste Wert $s(n)$ [$s(n-(m+1))$] durch Vorwärtsprädiktion [Rückwärtsprädiktion] mittels m vergangener Werte $s(n-1), s(n-2), \dots, s(n-m)$ zu schätzen.

Es sei nochmals darauf hingewiesen, daß das in den Abschnitten 1.3 und 1.4 gesagte nur für die Autokorrelation-Methode gültig ist.

Eliminiert man in der angegebenen Polynomrekursion das Polynom $B_m(z)$, so folgt

$$(1.13) A_m(z) = A_{m-1}(z) + k_m z^{-m} A_{m-1}(z^{-1}) ; m = 1, 2, \dots, p ; A_0(z) = 1$$

Sind also das Polynom $A_{m-1}(z)$ und k_m bekannt, so liegt $A_m(z)$ fest. Es ist in der Regel nur bei der Autokorrelation-Methode möglich, bei vorgelegtem optimalen $A_{m-1}(z)$ ein k_m so zu bestimmen, daß auch $A_m(z)$ optimal ist (optimal im Sinne von Gl. (1.6) bzw. Gl. (1.10)). In einem Aufsatz von John Makhoul /12/ wird daher ein anderer Weg eingeschlagen.

2. Lattice-Methoden

2.1 Vorbetrachtung

Fordert man vom "Lattice-Filter"

$$x_m^+(n) = x_{m-1}^+(n) + k_m x_{m-1}^-(n) \quad ; \quad x_0^+(n) = s(n)$$

(2.1)

$$x_m^-(n+1) = k_m x_{m-1}^+(n) + x_{m-1}^-(n) \quad ; \quad x_0^-(n+1) = s(n)$$
$$m = 1, 2, \dots, p$$

dasselbe Eingangs-Ausgangsverhalten wie vom "Prädiktorfilter"

$$(2.2) \quad e(n) = s(n) + \sum_{i=1}^p a_i s(n-i) \quad ; \quad p \text{ fest}$$

also $e(n) = x_p^+(n)$, so ergibt sich daraus ein nichtlinearer Zusammenhang zwischen den Koeffizienten der beiden Filter

$$a_i = f_i(k_1, k_2, \dots, k_i) \quad ; \quad i = 1, 2, \dots, p$$

Bestimmt man die PARCOR-Koeffizienten nun so, daß gilt

$$\sum_n [x_p^+(n)]^2 \stackrel{!}{=} \text{Min}$$

(durch partielles differenzieren nach den k_m), so führt das auf ein nicht-lineares Gleichungssystem für die PARCOR-Koeffizienten k_m . Berücksichtigt man die aufgrund des gleichen Eingangs-Ausgangsverhaltens aufgestellte Beziehung $a_i = f_i(k_1, k_2, \dots, k_i)$, so folgt

$$\sum_n [e(n)]^2 = \text{Min}$$

Ist also die Übertragungsfunktion (Eingangsfolge s , Ausgangsfolge e) optimal bezüglich der PARCOR-Koeffizienten, so ist sie es auch bezüglich der Prädiktorkoeffizienten und umgekehrt.

Wegen der erwähnten Nichtlinearitäten wird jedoch anders verfahren. Ausge-

gangen wird von einem Prädiktorfilter der Ordnung m

$$(2.3) \quad e_m(n) = s(n) + \sum_{i=1}^m a_{m,i} s(n-i) \quad ; \quad m = 1, 2, \dots, p$$

wobei gilt $e_p(n) = e(n)$ nach Gl. (2.2)

Gefordert wird nun gleiches Eingangs-Ausgangsverhalten in der m -ten Stufe, also

$$e_m(n) = x_m^+(n) \quad ; \quad m = 1, 2, \dots, p$$

Das führt auf ein lineares rekursives Gleichungssystem (Gl. (2.4) s.u.). Bestimmt man nun die PARCOR-Koeffizienten so, daß gilt

$$\sum_n [x_m^+(n)]^2 \stackrel{!}{=} \text{Min} \quad ; \quad m = 1, 2, \dots, p$$

betrachtet also $x_m^+(n)$ in Abhängigkeit von nur k_m und berücksichtigt Gl. (2.4), s.u., so folgt nicht

$$\sum_n [e_m(n)]^2 = \text{Min}$$

(bei der Autokorrelation-Methode gilt es). Die analoge Aussage gilt umgekehrt: Sei nämlich das Prädiktorfilter $(m-1)$ -ter Ordnung optimal, d.h. die Koeffizienten $a_{m-1,j}$; $j = 1, 2, \dots, m-1$ sind so bestimmt, daß gilt $\sum_n [e_{m-1}(n)]^2 = \text{Min}$. Dann legt Gl. (2.4), s.u., nahe, k_m so zu bestimmen, daß auch das Prädiktorfilter m -ter Ordnung optimal ist. Das ist jedoch allgemein nicht möglich (bei der Autokorrelationsmethode wiederum gilt es). Lattice-Methoden liefern also in der Regel nur suboptimale Lösungen.

Zusammenstellung der wichtigsten Beziehungen:

| | |
|-------|--|
| (2.1) | $x_m^+(n) = x_{m-1}^+(n) + k_m x_{m-1}^-(n) \quad ; \quad x_0^+(n) = s(n)$ $x_m^-(n+1) = k_m x_{m-1}^+(n) + x_{m-1}^-(n) \quad ; \quad x_0^-(n+1) = s(n)$ <p style="text-align: right; margin-right: 100px;">$m = 1, 2, \dots, p$</p> |
| (2.3) | $e_m(n) = \sum_{i=0}^m a_{m,i} s(n-i) \quad ; \quad a_{m,0} = 1 \quad ; \quad m = 1, 2, \dots, p$ |

Forderung $e_m(n) = x_m^+(n)$; $m = 1, 2, \dots, p \Rightarrow$ Gl. (2.4), (2.4b)

$$(2.4) \quad a_{m,m} = k_m$$

$$a_{m,j} = a_{m-1,j} + k_m a_{m-1,m-j} ; \quad j = 1, 2, \dots, m-1$$

$$(2.4a) \quad x_m^+(n) = \sum_{i=0}^m a_{m,i} s(n-i) ; \quad a_{m,0} = 1 ; \quad m = 1, 2, \dots, p$$

$$(2.4b) \quad x_m^-(n+1) = \sum_{i=0}^m a_{m,m-i} s(n-i)$$

2.2 Stable and Efficient Lattice Methods for Linear Prediction

nach John Makhoul /12/

Es werden Lattice-Methoden für die lineare Prädiktion untersucht, die die Stabilität des Filters auch beim Rechnen mit "fester Wortlänge" (FWL, finite word length) garantieren. Für Zwecke der Datenreduktion können in der Rekursion die Reflexionskoeffizienten (PARCOR) quantisiert werden, ohne die Stabilität des Filters zu beeinflussen.

Ausgegangen wird von Gl. (2.1) und Gl. (2.3), deren Zusammenhang durch Gl. (2.4) gegeben ist. Bevor die verschiedenen Lattice-Methoden abgeleitet werden, werden noch drei wichtige Größen definiert:

$$(2.5a) \quad F_m(n) := \sum_n [x_m^+(n)]^2$$

$$(2.5b) \quad B_m(n) := \sum_n [x_m^-(n+1)]^2$$

$$(2.5c) \quad C_m(n) := \sum_n [x_m^+(n) \cdot x_m^-(n)]$$

a. forward-method

k_m wird so bestimmt, daß der mittlere quadratische Fehler bei der "Vorwärts-

prädiktion" minimal wird

$$\inf_{k_m} \{F_m(n)\} \Rightarrow (k_m)_f = - \frac{C_{m-1}(n)}{B_{m-1}(n-1)}$$

(der Index f steht für forward)

b. backward-method

k_m wird so bestimmt, daß der mittlere quadratische Fehler bei der "Rückwärtsprädiktion" minimal wird

$$\inf_{k_m} \{B_m(n)\} \Rightarrow (k_m)_b = - \frac{C_{m-1}(n)}{F_{m-1}(n)}$$

(der Index b steht für backward)

c. geometric-mean-method (Itakura)

Die beiden vorherigen k_m sind nicht notwendig betragsmäßig kleiner eins. Daher hat Itakura die Reflexionskoeffizienten wie folgt definiert (Index I steht für Itakura)

$$(k_m)_I := - \frac{C_{m-1}(n)}{\sqrt{F_{m-1}(n) \cdot B_{m-1}(n-1)}}$$

definiert deshalb, weil $(k_m)_I$ nicht durch Minimierung eines Fehlerkriteriums gewonnen werden kann; der Betrag von $(k_m)_I$ ist jedoch immer kleiner als 1 (Cauchy-Schwarzsche Ungleichung. Streng genommen natürlich kleiner gleich 1. Gleichheit genau dann, wenn $x_{m-1}^+(n)$ und $x_{m-1}^-(n)$ linear abhängig sind). $(k_m)_I$ ist bis auf das Vorzeichen das geometrische Mittel von $(k_m)_f$ und $(k_m)_b$ (daher geometric-mean-method) und deshalb gilt

$$\min \{ |(k_m)_f|, |(k_m)_b| \} \leq |(k_m)_I| \leq \max \{ |(k_m)_f|, |(k_m)_b| \}$$

aus $|(k_m)_I| < 1$ folgt dann

$$\text{sei } |(k_m)_f| > 1 \Rightarrow |(k_m)_b| < 1$$

$$\text{sei } |(k_m)_b| > 1 \Rightarrow |(k_m)_f| < 1$$

Daraus läßt sich eine andere mögliche Definition der Reflexionskoeffizienten ableiten, so daß Stabilität des Filters gewährleistet ist.

d. minimum-method

$$(k_m)_M := S \min \{ |(k_m)_f|, |(k_m)_b| \} ;$$

$$\text{mit } S = \text{sign } (k_m)_f = \text{sign } (k_m)_b$$

e. general-method

Da alle Reflexionskoeffizienten, die "zwischen" $(k_m)_M$ und $(k_m)_I$ liegen die Stabilität des Filters gewährleisten, kann allgemein definiert werden

$$(k_m)_r := S \left[\frac{1}{2} (|(k_m)_f|^r + |(k_m)_b|^r) \right]^{\frac{1}{r}} ; r \leq 0$$

und es gilt

$$|(k_m)_{r-1}| \leq |(k_m)_r|$$

(HÖLDERSche Ungleichung). Für $r=0$ und $r=-\infty$ erhält man die unter c. und d. angegebenen Werte,

$$(k_m)_0 = (k_m)$$

$$(k_m)_{-\infty} = (k_m)_M$$

Im stationären Fall gilt $(k_m)_f = (k_m)_b = (k_m)_r$ für alle $r \leq 0$.

f. harmonic-mean-method (Burg)

Für $r = -1$ hat $(k_m)_r$ einige interessante Eigenschaften.

$$(2.6) \quad (k_m)_B = (k_m)_{-1} = 2 \frac{(k_m)_f \cdot (k_m)_b}{(k_m)_f + (k_m)_b} = - \frac{2 C_{m-1}(n)}{F_{m-1}(n) + B_{m-1}(n-1)}$$

(der Index B steht für Burg) und $(k_m)_B$ ist das harmonische Mittel von $(k_m)_f$ und $(k_m)_b$. Im Gegensatz zu $(k_m)_I$ und $(k_m)_M$ kann $(k_m)_B$ abgeleitet werden aus

einem vernünftig erscheinenden Fehlerkriterium:

$$\inf_{k_m} \{ F_m(n) + B_m(n) \} \Rightarrow (k_m)_B$$

2.2.1 The Covariance-Lattice Method

Drückt man die für die Berechnung von $(k_m)_B$ nach Gl. (2.6) benötigten Größen durch die Koeffizienten $a_{m,i}$ des Prädiktorfilters aus mittels der Rekursion nach Gl. (2.4), so erhält man mit Gl. (2.7)

$$(2.7) \quad \Phi(i,j) := \sum_n s(n-i) s(n-j)$$

$$(2.8a) \quad F_{m-1}(n) = \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} a_{m-1,i} \Phi(i,j) a_{m-1,j}$$

$$(2.8b) \quad B_{m-1}(n-1) = \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} a_{m-1,i} \Phi(m-i, m-j) a_{m-1,j}$$

$$(2.8c) \quad C_{m-1}(n) = \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} a_{m-1,i} \Phi(i, m-j) a_{m-1,j}$$

$$a_{m-1,0} = 1$$

Der Algorithmus zur Berechnung der Reflexionskoeffizienten, genannt "Covariance-Lattice Method" kann nun in einem Flußdiagramm prinzipiell wie folgt dargestellt werden

Berechne die Covarianz nach Gl. (2.7)

$$\Phi(i,j) = \sum_n s(n-i) s(n-j) ; i,j = 0,1, \dots, p$$

m := 1

Berechne nach den Gl.n. (2.8)

$$F_{m-1}(n) = \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} a_{m-1,i} \Phi(i,j) a_{m-1,j}$$

$$B_{m-1}(n-1) = \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} a_{m-1,i} \Phi(m-i,m-j) a_{m-1,j}$$

$$C_{m-1}(n) = \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} a_{m-1,i} \Phi(i,m-j) a_{m-1,j}$$

mit $a_{m-1,0} = 1$

Berechne nach Gl. (2.6)

$$(k_m)_B = -2 \frac{C_{m-1}(n)}{F_{m-1}(n) + B_{m-1}(n-1)}$$

Quantisiere $(k_m)_B$ wenn erwünscht

Berechne nach Gl. (2.4)

$$a_{m,m} = (k_m)_B$$

$$a_{m,j} = a_{m-1,j} + (k_m)_B a_{m-1,m-j} ; j = 1,2, \dots, m-1$$

m := m+1

n $\left\{ \begin{array}{l} m > p \\ j \end{array} \right.$ Stop

3. Umsetzung der Algorithmen in Signalflußgraphen

3.1 Der Algorithmus von ITAKURA und SAITO (Autokorrelation)

Bei der Autokorrelation-Methode wird definitionsgemäß gesetzt

$$s(n) = 0 \text{ für } n \notin [0, N-1]$$

Daraus folgt aus den Gl'n. (2.5a), (2.5b), (2.5c)

$$F_{m-1}(n) = \sum_{n=0}^{N-2+m} [x_{m-1}^+(n)]^2$$

$$B_{m-1}(n-1) = \sum_{n=1}^{N-1+m} [x_{m-1}^-(n)]^2 = F_{m-1}(n)$$

$$C_{m-1}(n) = \sum_{n=1}^{N-2+m} x_{m-1}^+(n) x_{m-1}^-(n)$$

Für den unter Abschnitt 2.2c nach der geometric-mean-method bestimmten Reflexionskoeffizienten $(k_m)_I$ ergibt sich

$$(k_m)_I = - \frac{C_{m-1}(n)}{F_{m-1}(n)} = - \frac{C_{m-1}(n)}{B_{m-1}(n-1)} ; m = 1, 2, \dots, p$$

Gemäß Gl. (2.1) wird, nachdem $(k_m)_I$ bekannt ist, $x_m^+(n)$ und $x_m^-(n+1)$ berechnet, m um eins erhöht, usw. Eine Realisierung in Form eines Signalflußgraphen für den m -ten Schritt $m=1, 2, \dots, p$ ist in Bild 3.1 dargestellt. Beim Anlegen der Eingangsfolgen wird sich nach $(N-2+m)$ -Takten der eingezeichnete Zustand einstellen. Danach kann $(k_m)_I$ berechnet und wenn gewünscht quantisiert werden. Mit diesem (quantisierten) Wert werden die Ausgangsfolgen bestimmt. Die Eingangs- bzw. Ausgangsfolgen sind außerhalb des angegebenen Intervalls per definitionem Null.

Die Signalverzögerung in der m -ten Stufe beträgt $(N-2+m)$ Takteinheiten; bei einem Prädiktorfilter der Ordnung p also insgesamt $[p(N-2) + \frac{1}{2} p(p+1)]$ Takteinheiten.

Geht man davon aus, daß sich die Statistik der Sprache zeitlich nur "langsam" ändert, so kann bei kontinuierlicher Verarbeitung von Sprache auch eine "schon etwas ältere" Schätzung verwendet werden. Bild 3.2

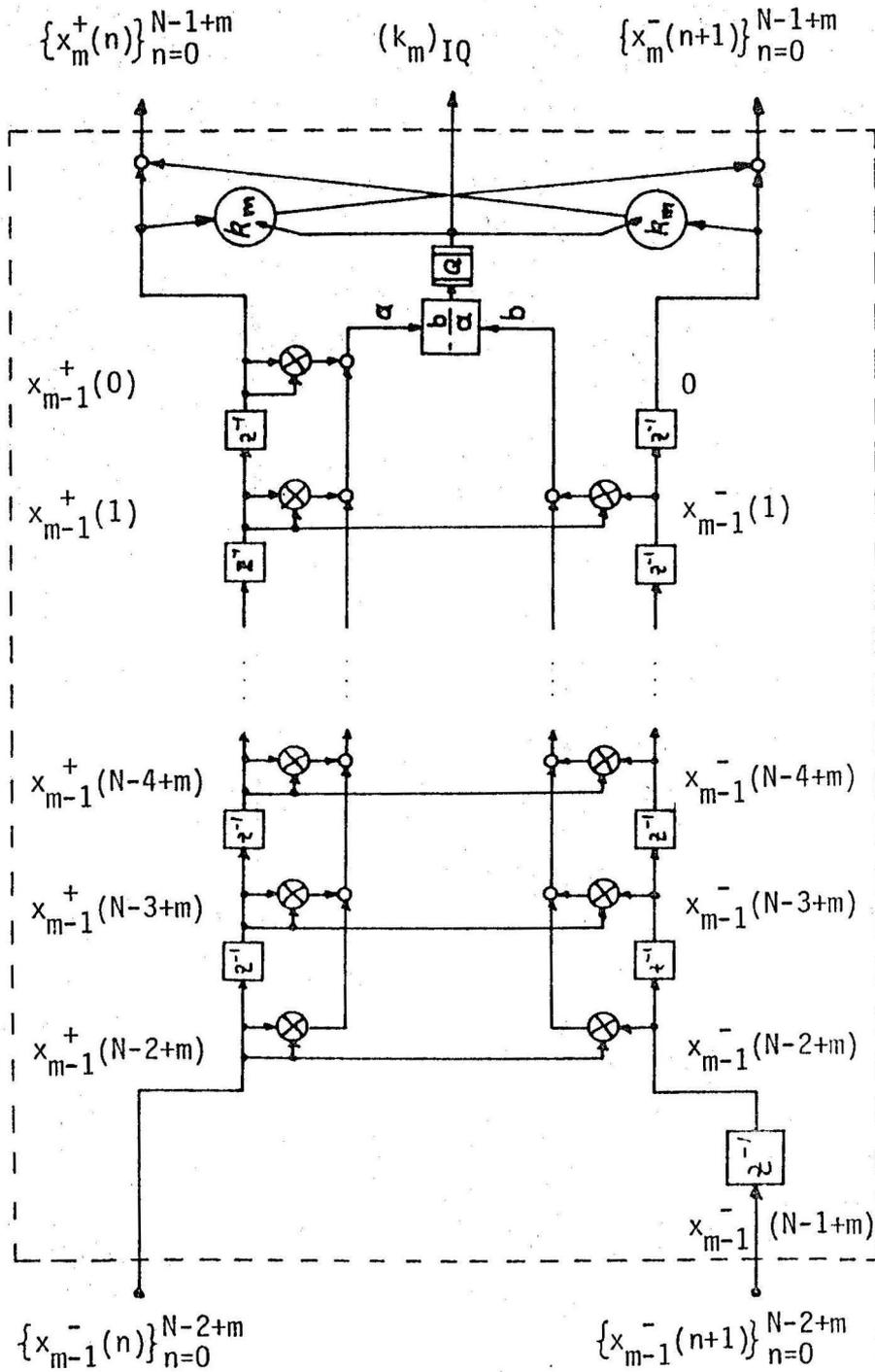


Bild 3.1 m-te Stufe eines Signalflußgraphen zur linearen Prädiktion, $m = 1, 2, \dots, p$

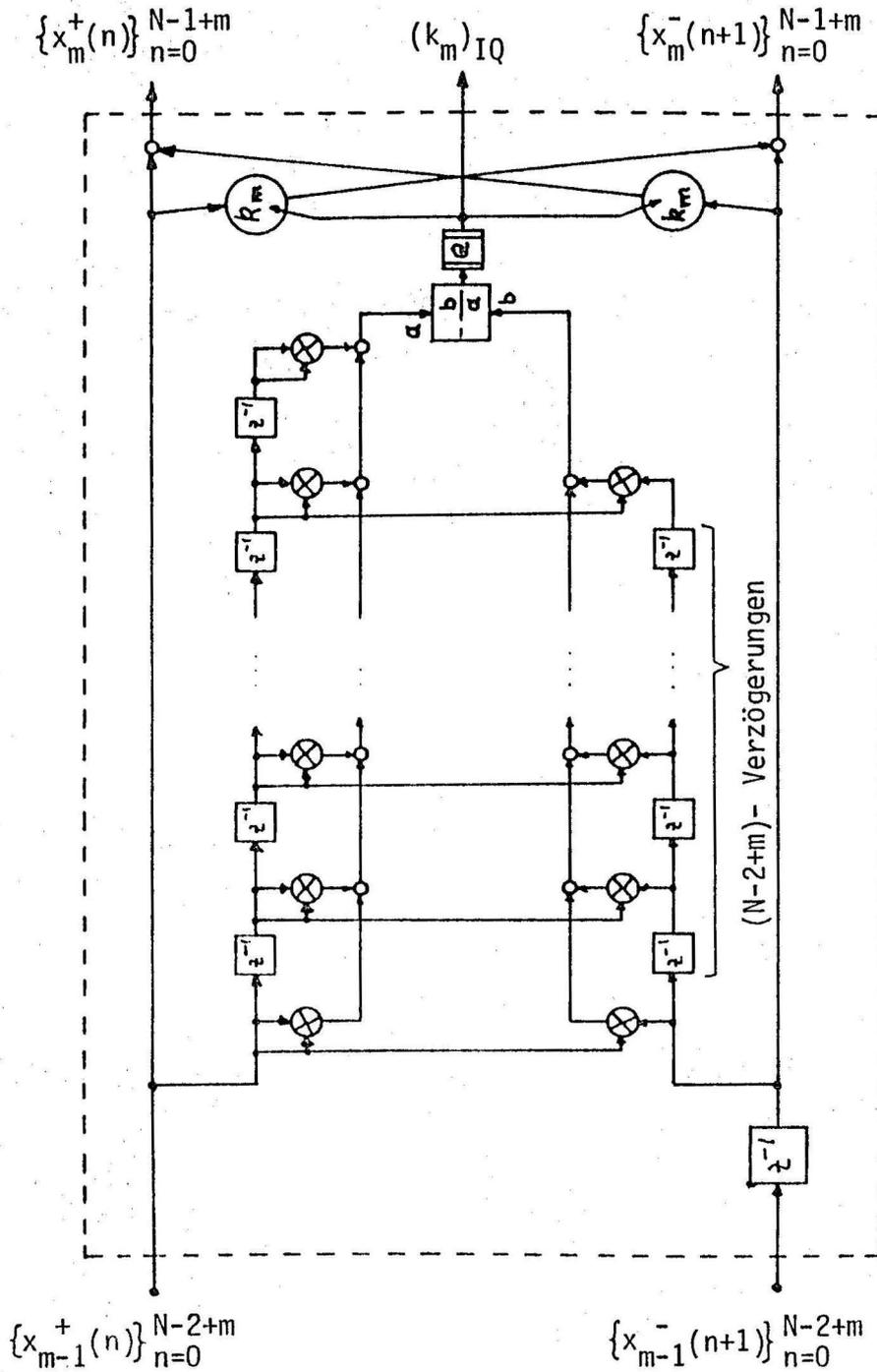


Bild 3.2 m-te Stufe eines Signalflußgraphen zur linearen Prädiktion, $m = 1, 2, \dots, p$. k_m wird nicht aus dem aktuell verarbeiteten, sondern aus dem vorherigen Sprachstück bestimmt.

Bemerkung: Daß die Ausgangsfolge $\{e(n)\}_{n=0}^{N-1+p} \equiv \{x_p^+(n)\}_{n=0}^{N-1+p}$ um p Glieder länger ist als die Eingangsfolge $\{s(n)\}_{n=0}^{N-1}$, ist auf die in der Lattice-Filterstruktur enthaltenen p Verzögerungseinheiten zurückzuführen und nur scheinbar, da ja eine "gefensterte" Eingangsfolge vorausgesetzt wurde, die die Eigenschaft hat $s(N) = s(N+1) = \dots = s(N-1+p) = 0$ und diese "Nullen" müssen am Eingang des Filters auch tatsächlich vorhanden sein. Liegt stattdessen das nächste Sprachstück (N -Abtastwerte) an, so kommt es zu einer Überlagerung der Fehler (Ausgangsfolge), die bei der Lattice-Struktur und dem zugrundegelegten Algorithmus, aufgrund der Rekursivität der auftretenden Gleichungen aus (zunächst) zeitlichen Gründen nicht abgeschätzt werden kann. (Um diesen Fehler zu vermeiden, könnte man z.B. zwischen zwei aufeinanderfolgende Sprachstücke p Nullen einschieben.)

3.2 Bestimmung der PARCOR-Koeffizienten nach BURG

Für die im Abschnitt 2.2 abgeleiteten Verfahren zur Bestimmung der PARCOR-Koeffizienten lassen sich analog zu Abschnitt 3.1 Signalflußgraphen angeben, die ebenfalls auf die Rekursion nach Gl. (2.4) verzichten. Da die Bestimmung der PARCOR-Koeffizienten nach BURG (Abschnitt 2.2, f. harmonic-mean-method) in der Literatur bevorzugt behandelt wird (/12/, /5/, pp 248), sind nachstehend zwei Realisierungen mittels Signalflußgraphen angegeben.

Die zur Berechnung von $(k_m)_B$ nach Gl. (2.6) benötigten Größen erhält man nach Gl. (2.5) und es folgt:

$$(3.1) \quad (k_m)_B = - \frac{\sum_n [x_{m-1}^+(n) + x_{m-1}^-(n)]^2 - [x_{m-1}^+(n) - x_{m-1}^-(n)]^2}{\sum_n [x_{m-1}^+(n) + x_{m-1}^-(n)]^2 + [x_{m-1}^+(n) - x_{m-1}^-(n)]^2}$$

Die für jedes n benötigten Teilsummanden werden aus $x_{m-1}^+(n)$ und $x_{m-1}^-(n)$ berechnet, siehe Bild 3.3

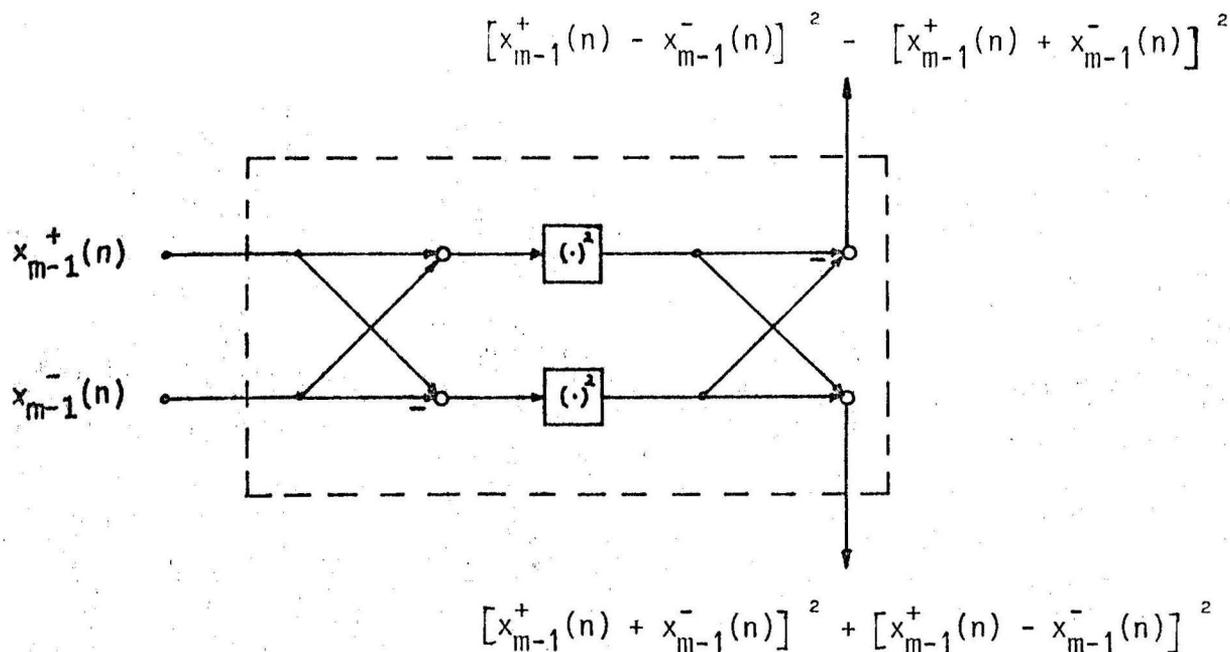


Bild 3.3

Die Grenzen von n in Gl.(3.1) definieren das Intervall, auf dem die PARCOR-Koeffizienten "optimal" sind. Bei der Covarianz-Methode erstreckt sich die Summation (z.B.) von $n=p$ bis $n=N-1$. (In /12/ werden noch andere Möglichkeiten für die Wahl der Grenzen erwähnt.)

In Bild 3.4 ist die exakte Bestimmung des m -ten PARCOR-Koeffizienten zu dem entsprechenden Sprachstück mittels eines Signalflußgraphen dargestellt (exakt im Sinne von Gl. (3.1)). Sobald nach Anlegen der Eingangsfolgen sich der eingezeichnete Zustand eingestellt hat, kann k_m berechnet und wenn erwünscht quantisiert werden. Mit dem so bestimmten k_m werden dann die Ausgangsfolgen berechnet. Die Verzögerungszeit durch sämtliche p -Stufen beträgt $p \cdot (N-1)$ - Takteinheiten. Geht man davon aus, daß sich die Statistik der Sprache zeitlich nur "langsam" verändert, so kann bei kontinuierlicher Verarbeitung von Sprache auch eine "schon etwas ältere" Schätzung verwendet werden; Bild 3.5 (analog zu Abschnitt 3.1).

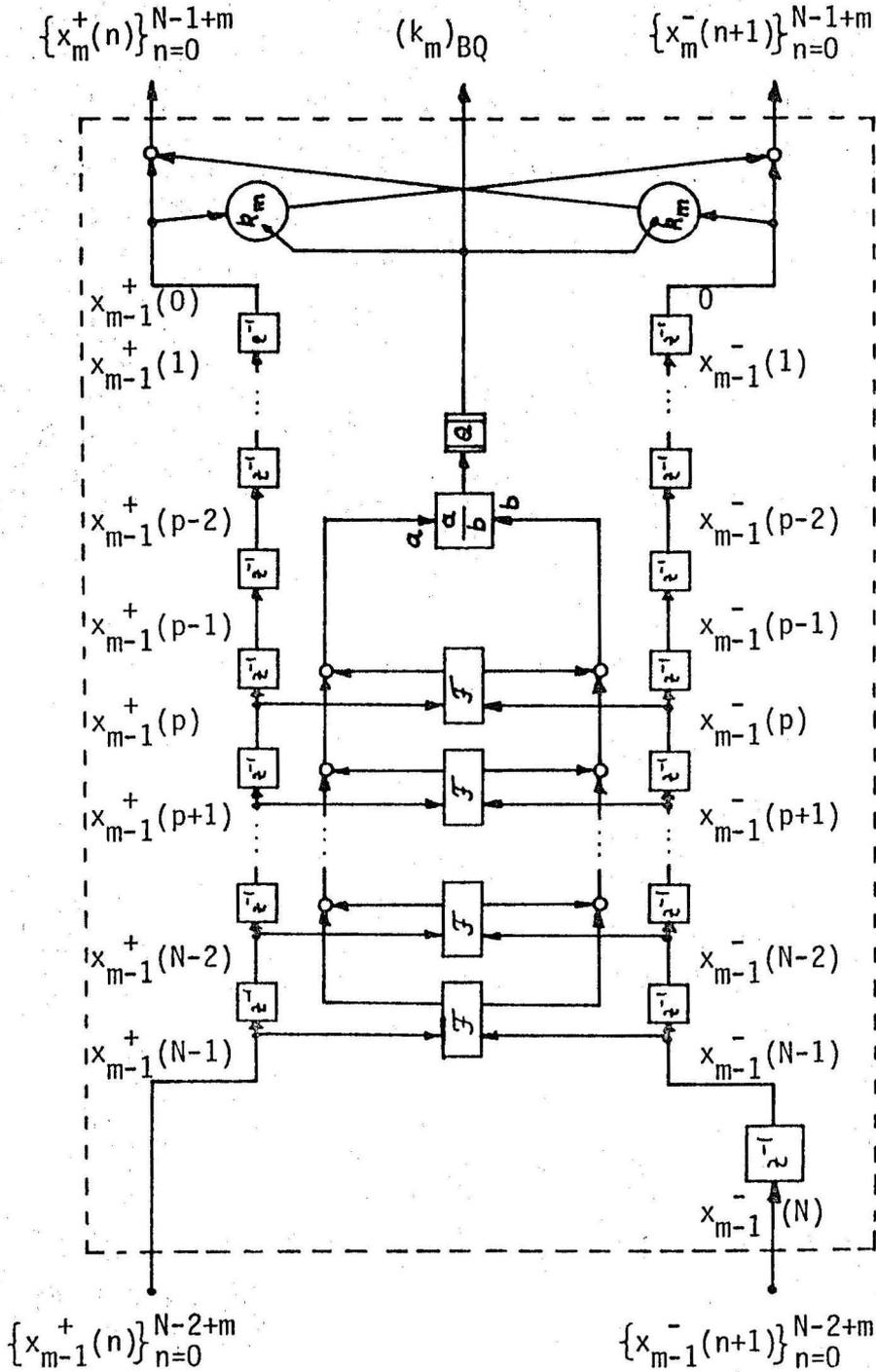


Bild 3.4 m -te Stufe eines Signalflußgraphen zur linearen Prädiktion, $m = 1, 2, \dots, p$.

$\boxed{\mathcal{F}}$ symbolisiert den in Bild 3.3 dargestellten Signalflußgraphen

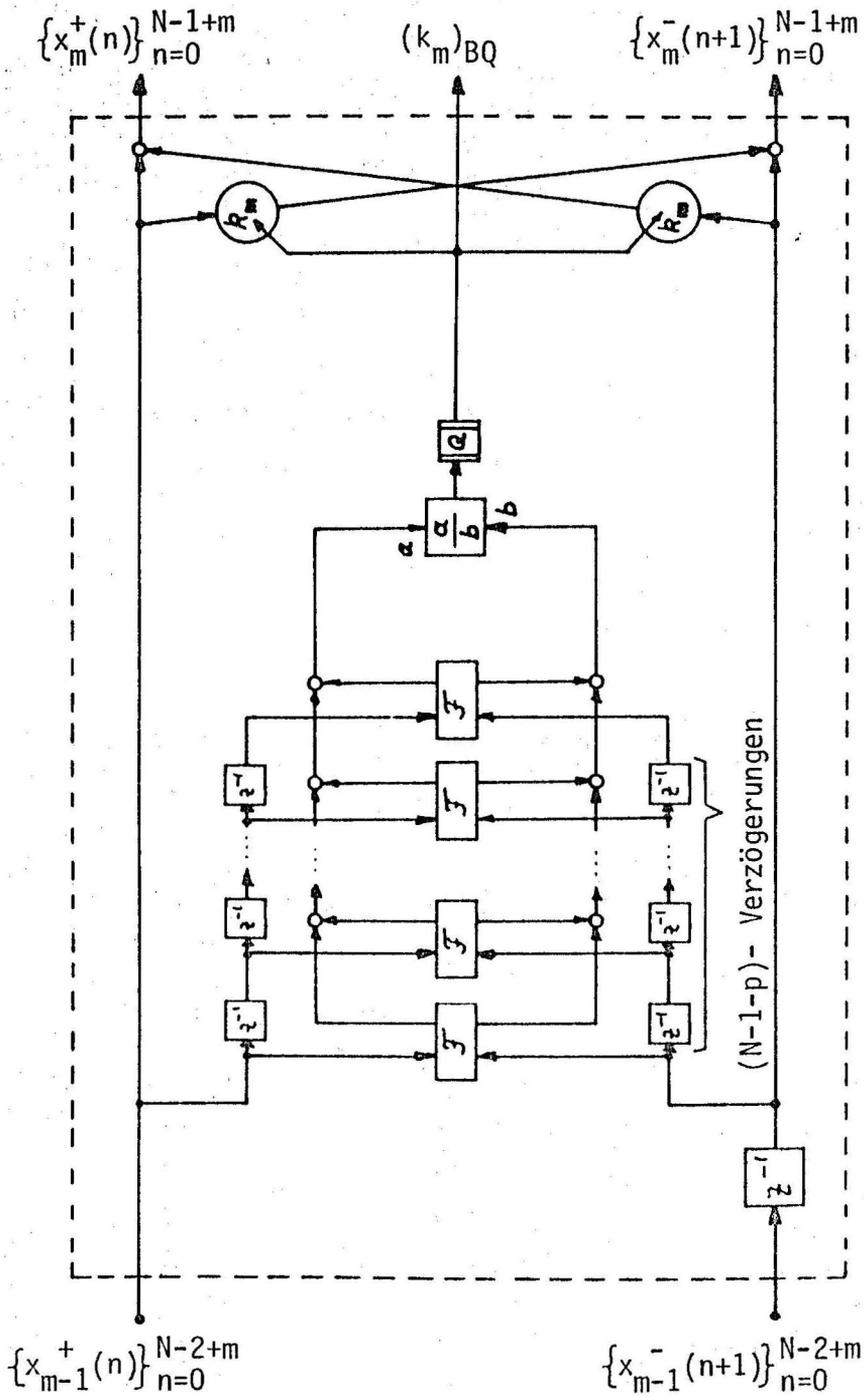


Bild 3.5

m -te Stufe eines Signalflußgraphen zur linearen Prädiktion, $m = 1, 2, \dots, p$. k_m wird nicht aus dem aktuell verarbeiteten, sondern aus dem vorherigen Sprachstück bestimmt.

\mathcal{F} symbolisiert den in Bild 3.3 dargestellten Signalflußgraphen

3.2.1 Verringerung des Hardwareaufwandes durch ein rekursives Filter erster Ordnung

Die in den Gleichungen auftretenden Summen über n der Form $\sum_{n=n_0}^{n_1} a_n \cdot x_n$

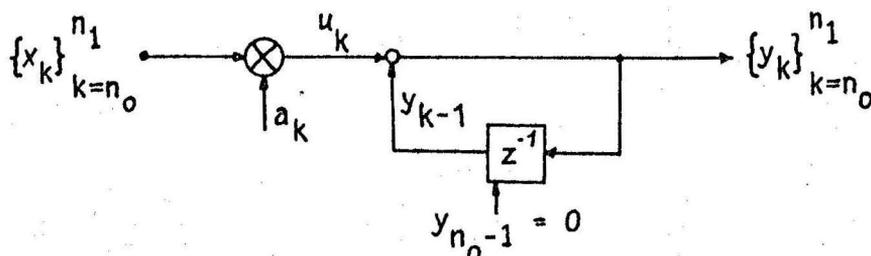
wurden bislang so berechnet, daß die Summen genau dann bestimmt wurden, wenn das Filter nach Anlegen der Eingangsfolge einen definierten Zustand erreicht hatte (parallele Berechnung); Bild 3.1, 3.4.

Im Folgenden soll diese Summation über n mittels eines rekursiven Filters geschehen. Mit

$$u_k := a_k \cdot x_k$$

$$y_k := y_{k-1} + u_k \quad ; \quad k = n_0, n_0+1, \dots, n_1 \quad ; \quad y_{n_0-1} = 0$$

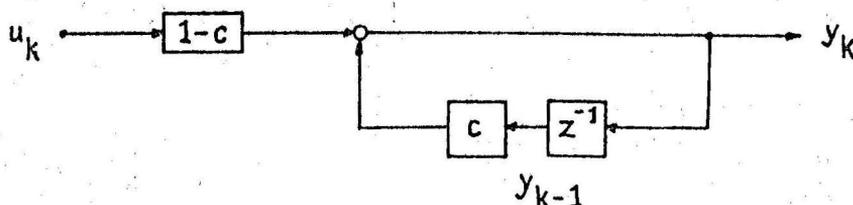
ist dann das rekursive Filter gegeben,



und die zu berechnende Summe ist gegeben durch y_{n_1} .

Da dieses Filter nicht asymptotisch stabil ist, haben Itakura und Saito (zitiert in /5/ pp 248) zur Summation nachstehendes Filter verwendet (linear, zeitinvariant), das kurz diskutiert werden soll:

$$y_k = c y_{k-1} + (1-c) u_k \quad ; \quad 0 \leq c \leq 1 \quad ; \quad y_{-1} = 0$$



Die allgemeine Lösung ist

$$y_k = (1-c) \sum_{i=0}^k c^{k-i} u_i ; k = 0, 1, \dots$$

und die Sprungantwort

$$y_k = 1 - c^{k+1} \quad \text{für } u_i = 1 ; i = 0, 1, \dots$$

sowie die Impulsantwort (Gewichtsfolge)

$$y_k = (1-c) c^k \quad \text{für } u_i = \begin{cases} 1 & \text{für } i = 0 \\ 0 & \text{für } i > 0 \end{cases}$$

Bild 3.6 zeigt die Impulsantwort mit c als Parameter

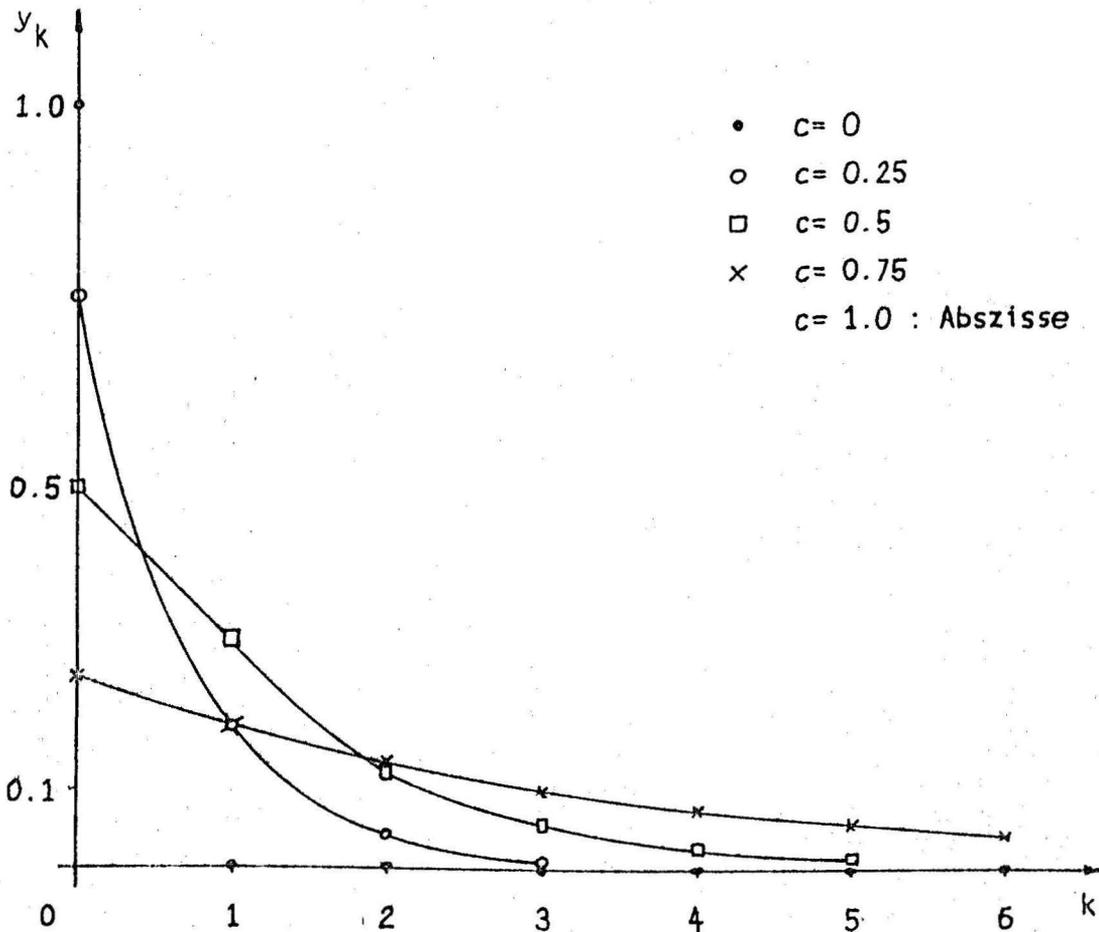


Bild 3.6 Impulsantwort des rekursiven Filters

Mit dem rekursiven Filter kann k_m , entsprechend Gl. 3.1, wie folgt aus den Folgen $\{x_{m-1}^+(n)\}$ und $\{x_{m-1}^-(n)\}$ bestimmt werden, Bild 3.7. Diese Realisierung entspricht dem in Bild 3.5 angegebenen Signalflußgraphen

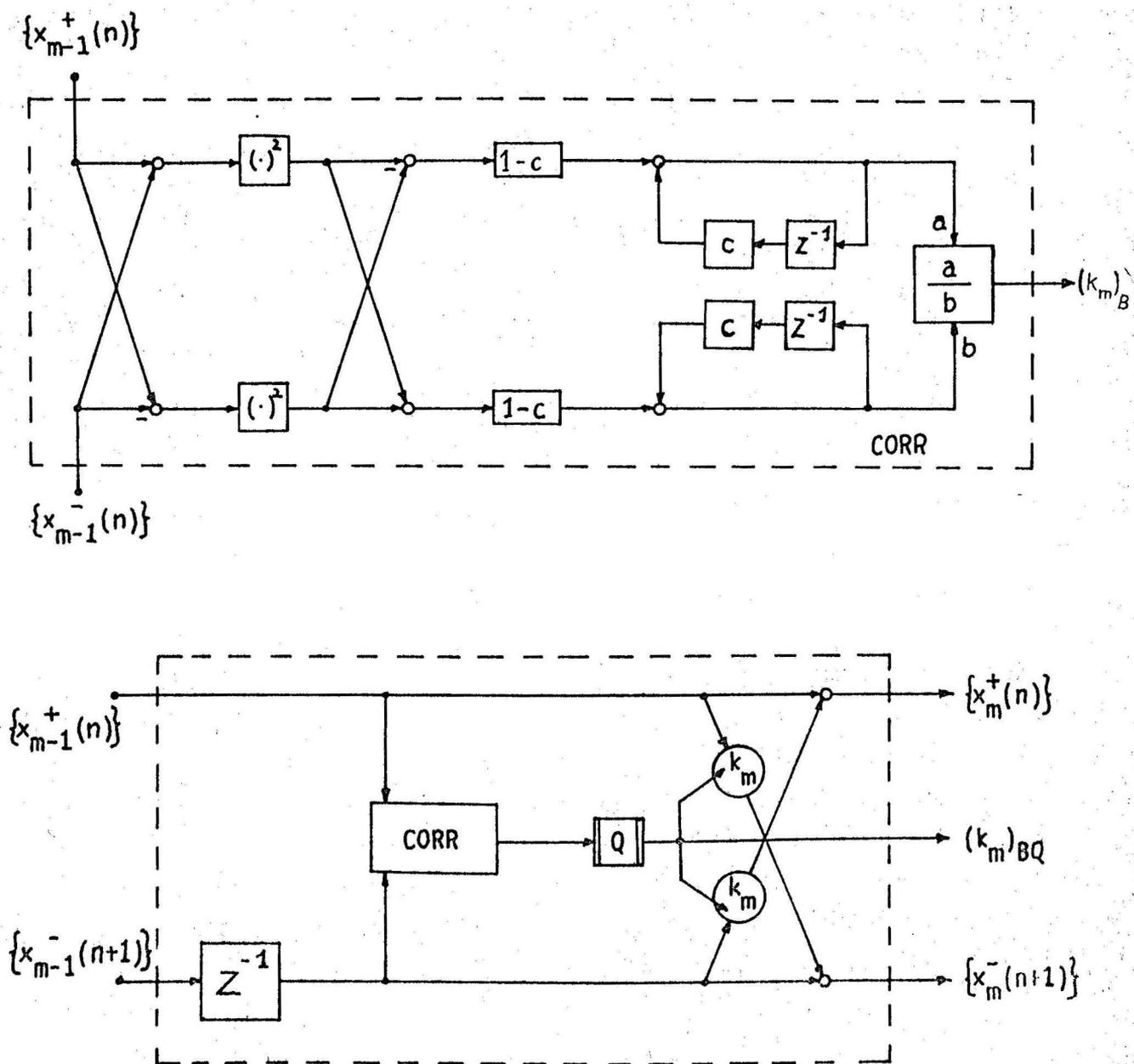


Bild 3.7 m -te Stufe eines Signalflußgraphen zur linearen Prädiktion, $m = 1, 2, \dots, p$. k_m wird nicht aus dem aktuell verarbeiteten, sondern aus dem vorherigen Sprachstück bestimmt.

3.3 Eine Begründung für die getroffene Auswahl

Alle mir bekannten "Nicht-Lattice-Methoden" bestimmen die Prädiktor-Koeffizienten so, daß die Ausgangsgröße (= das Fehlersignal) in Bezug auf die Eingangsgröße (= das zu analysierende Sprachsignal) in einem bestimmten Sinne optimal ist. Werden die aus dieser Optimalitätsbedingung folgenden Gleichungen rekursiv gelöst (z.B. nach Levinson), so ist bei der numerischen Auswertung dieser Gleichungen zu bedenken, daß im $(m+1)$ -ten Rekursionsschritt "numerische Exaktheit" des m -ten Rekursionsschrittes vorausgesetzt wird, was schon bei Maschinen mit "großer Wortlänge" bei hoher Ordnung des Prädiktors ($p > 16$) nicht immer gewährleistet ist und somit zu erheblichen Fehlern führen kann.

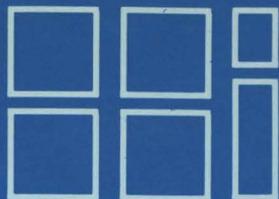
Bei den "Lattice-Methoden" ist dies nicht der Fall, da der Ansatz ein anderer ist; ein Lattice-Filter besteht aus p Stufen, p ist die Ordnung des Prädiktors. Die Ausgangsgröße der $(m+1)$ -ten Stufe wird so bestimmt, daß sie bezüglich der Eingangsgröße derselben Stufe optimal ist, welche durchaus die quantisierte Ausgangsgröße der m -ten Stufe sein kann. Der durch die Quantisierung auftretende Fehler wird also berücksichtigt. Allerdings liefern Lattice-Methoden in der Regel nur suboptimale Lösungen (Abschnitt 2.1).

Die unter Abschnitt 2.2.1 angeführte Lattice-Methode bestimmt die PARCOR-Koeffizienten nicht aus der Statistik der Zustandsgrößen des Lattice-Modells, sondern aus der Statistik des zu analysierenden Sprachsignals und den (rekursiven) Prädiktorkoeffizienten. Der Grund, weshalb ich für diesen Algorithmus keinen Signalflußgraphen angegeben habe, liegt in der Unübersichtlichkeit der dort auftretenden Doppelsummen. Der Signalflußgraph nach Bild 3.4 erzeugt dieselben PARCOR-Koeffizienten und darüber hinaus noch die Fehlergröße (aus den Zustandsgrößen des Lattice-Modells).

Literatur

- /1/ Saito, S.; Itakura, F.: The Theoretical Consideration of Statistically Optimum Methods for Speech Spectral Density.
Report No. 3107, Electrical Communication Laboratory, N.T.T. Tokyo 1966
(in Japanese)
- /2/ Atal, B.S., Schroeder, M.R.: Predictive Coding of Speech Signals.
Proc. 1967 Conf. Commun. and Process., 360- 361 (1967)
- /3/ Elias, P.: Predictive coding
IRE Trans. Information Theory IT-1, 16-33 (1955)
- /4/ McDonald, R.A.: Signal-to-noise and idle channel performance of differential pulse code modulation systems-particular application to voice signals.
Bell System Techn. J. 45, 1123-1151 (1966)
- /5/ Markel, J.D., Gray, A.H.: Linear Prediction of Speech
Springer-Verlag 1976
- /6/ Levinson, N.: The Wiener RMS (Root Mean Square) Error Criterion in Filter Design and Prediction.
Journal of Mathematics and Physics, Vol. 25, No. 4. Jan. 1947, pp. 261-278
(Auch in: Wiener, N.: Extrapolation Interpolation and Smoothing of Stationary Time Series.
MIT Press, Cambridge, Massachusetts 1966)
- /7/ Wakita, H.: Estimation of the Vocal Tract Shape by Optimal Inverse Filtering and Acoustic/Articulatory Conversion Methods.
SCRL Monograph No. 9, Speech Communications Research Laboratory, Santa Barbara, California (1972)
- /8/ Itakura, F., Saito, S.: Speech Analysis-Synthesis System Based on the Partial Autocorrelation Coefficient.
Acoust. Soc. of Japan Meeting (1969)

- /9/ Itakura, F., Saito, S.: On the Optimum Quantization of Feature Parameters in the PARCOR Speech Synthesizer. Conf. Record IEEE 1972 Conf. Speech Commun. and Process., New York, paper L 4, 434-437 (1972)
(Auch in: Flanagan, J.L., Rabiner, L.R., ed.: Speech Synthesis (Dowden, Hutchnison and Ross, Stroudsburg, Pennsylvania, 1973))
- /10/ Itakura, F., Saito, S., Koike, Y., Sawabe, H., Nishikawa, M.:
An Audio Response Unit Based on Partial Correlation.
IEEE Trans. COM-20, pp.792-797 (1972)
- /11/ Makhoul, John: Linear Prediction: A Tutorial Review
Proceedings of the IEEE, Vol. 63, No. 4, April 1975, pp. 561-580
- /12/ Makhoul, John: Stable and Efficient Lattice Methods for Linear Prediction
IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-25,
No. 5, October 1977



**Heinrich-Hertz-Institut
für Nachrichtentechnik
Berlin GmbH**

