

**Mehrkanalige adaptive Filter für die  
Unterdrückung von multiplen Reflexionen in  
Verbindung mit der freien Oberfläche in  
marinen Seismogrammen.**

**Multichannel adaptive filters for the suppression  
of surface related multiple reflections in  
marine seismograms**

---

**Andreas Rosenberger**

Andreas Rosenberger

Alfred-Wegener-Institut für Polar- und Meeresforschung  
Postfach 120161  
Columbusstraße  
D-2850 Bremerhaven

Die vorliegende Arbeit ist die inhaltlich unveränderte Fassung einer  
Dissertation, die 1991 dem Fachbereich Geowissenschaften der  
Universität Bremen vorgelegt wurde.

Zusammenfassung	
Summary	
Vorwort.....	1
1. Notation.....	4
2. Einführung.....	5
3.Theorie.....	9
3.1 Autoregressives Modell, Systemidentifikation und Vorhersageoperatoren.....	10
3.1.1 Formulierung des stationären 1-D Vorhersageproblems.....	11
3.1.2 Autokorrelations-, Covarianz- und "Prewindowed"- Formulierung.....	14
3.1.3 Lewinson-Durbin Rekursion und Block- Verarbeitings Techniken.....	16
3.2 Eigenschaften der Least-Squares Lösung.....	17
3.2.1 Projektionsoperatoren für Vorhersage- und Vorhersagefehler-Operatoren.....	19
3.2.2 Aufdatierungen für Projektionsoperatoren .....	23
3.3 Skizze des Lattice-Algorithmus.....	27
3.4 Adaptive Formulierung für den instationären Fall .....	33
3.5 Joint-Process Estimation.....	35
3.6 Mehrdimensionale Formulierung des Lattice- Algorithmus.....	36
3.6.1 Direkte Lösung : Rang-1 Aufdatierungen.....	39
3.6.2 Konditionsproblem, Rang-Entscheidungs- Problem.....	39
3.7 QR-Zerlegung: der Algorithmus von Lewis.....	49
3.7.1 QR-Zerlegung und Kondition.....	49
3.7.2 Rangentscheidung mit Hilfe der KL- Transformation.....	55
3.7.3 Rangentscheidung mit Hilfe der Diskreten Cosinus-Transformation.....	60
4. Anwendungen.....	67
4.1 Multiple Reflexionen in Verbindung mit der freien Oberfläche in Common-Shot-Point (CSP) Seismogrammen.....	68

4.2 Filterdesign.....	7 1
4.3 Eindimensionale adaptive Filter für die Unterdrückung multipler Reflexionen.....	7 3
4.3.1 Beispiel: Synthetische CSP-Daten.....	7 6
4.3.2 Beispiel: Reale CSP Seismogramme.....	8 0
4.4 Adaptive Mehrkanal-Filter für die Unterdrückung multipler Reflexionen.....	8 6
4.4.1 Beispiel: Synthetische CSP-Daten.....	8 6
4.4.2 Beispiel: Reale CSP-Seismogramme.....	9 1
5. Ungelöste Probleme.....	9 7
6. Zusammenfassung, Einschätzung.....	9 9
7. Literatur.....	1 0 5

## ZUSAMMENFASSUNG

Multiple Reflexionen im Wellenleiter des Wasserkörpers sind ein häufig beobachtetes Phänomen in marinen seismischen Daten insbesondere aus den polaren Regionen. Es ist offensichtlich, daß der Prozess, der die mehrfach reflektierten, auf verschiedenen Wegen und mit verschiedenen Laufzeiten beim Empfänger eintreffenden Wellen generiert, mit einem eindimensionalen Modell nicht beschrieben werden kann. Aus diesem Grund sind die Ergebnisse, die bei der Filterung der Daten mit eindimensionalen Vorhersage-Filtern erzielt werden können oft unbefriedigend. Darüberhinaus zeigen Multipleneinsätze ein nichtstationäres Verhalten. Derzeit verwendete Block-Verarbeitungsverfahren, die auf dem Wiener-Levinson-Algorithmus basieren, hängen von der Töplitzstruktur der Datenmatrix ab und sind ungenau hinsichtlich der Behandlung von Daten außerhalb des Beobachtungsfensters oder numerisch ineffizient, wenn auf die Töplitzstruktur verzichtet wird und stattdessen Kovarianz-Matrizen invertiert werden müssen. In dieser Arbeit wird für die Behandlung des Multiplenproblems ein mehrkanaliges adaptives Lattice-Filter vorgeschlagen. Die numerischen Schwierigkeiten aufgrund schlecht konditionierter Projektionsoperatoren werden diskutiert und es wird gezeigt, wie das Rang-Entscheidungs-Problem, das mit bestimmten Eigenschaften von seismischen Mehrkanaldaten verbunden ist, durch Einbindung der Diskreten Cosinus-Transformation gelöst werden kann. Synthetische Daten und ein komplettes seismisches Profil werden mit einem Einkanal- und einem Merkanalalgorithmus bearbeitet. Die Resultate werden vorgestellt und diskutiert.

## SUMMARY

Multiple reflections in the water-layer waveguide are often encountered in marine seismic surveys especially in polar regions. The multipathing effect is clearly not one-dimensional so one-dimensional prediction-filters applied to multichannel recordings often show unsatisfactory results. Multiple reflections are a non-stationary phenomenon and current block-processing methods based on the Wiener-Lewinson-Algorithm suffer from either inaccurate treatment of data outside the observation-interval when autocorrelation matrices are computed or prohibitive computational costs when non-toeplitz covariance matrices have to be inverted. In this paper the approach via a multichannel adaptive lattice algorithm is taken. The numerical difficulties due to illconditioned projection-operators are discussed and it is shown how the remaining rank-decision-problem arising from certain properties of multichannel seismic recordings can be solved by incorporation of the discrete-cosine-transform into the algorithm. Processing results for synthetic data and a complete seismic profile are presented for both a single- and a multichannel adaptive lattice.

## Vorwort

In dieser Arbeit wird ein neues Verfahren zur Behandlung des Multiplenproblems in marinen seismischen Daten vorgeschlagen. Neu ist dieses Verfahren in Bezug auf die algorithmische Umsetzung, ihm liegt jedoch ein Modell zugrunde, das seit den Anfängen der Analyse von diskreten Signalen bekannt ist. Die Rede ist vom sogenannten Autoregressive-Moving-Average Modell (ARMA), dessen Erfolgsgeschichte im 2. Weltkrieg mit Aufkommen der militärischen Radartechnik beginnt und dessen Theorie eng mit dem Namen Norbert Wiener verknüpft ist (Wiener 1949).

Die Entwicklung dieser Theorie vollzog sich in den frühen fünfziger Jahren einigermaßen stürmisch, zumal unmittelbar nach dem Krieg alle diejenigen Publikationen erschienen, die auf Arbeiten basierten, die noch während des Krieges als "classified" galten. Als Markstein für den Einzug dieser Theorie in die geophysikalische Signalanalyse gilt die Dissertation von E.A. Robinson: "Predictive decomposition of time series with application to seismic exploration" (Robinson 1954).

Die Schwächen und Unzulänglichkeiten dieser Theorie und der auf ihr basierenden inversen Verfahren oder Modellparametrisierungen sind ebensolange bekannt, wie die Theorie selbst. Reale physikalische Prozesse variieren in der Zeit oder wesentliche Parameter dieser Prozesse sind unbekannt. In den allermeisten Fällen dürfte beides der Fall sein, ganz abgesehen davon, daß das ARMA-Modell selbst nur ein recht grobes Modell eines Ausschnittes aus der "realen", nichtlinearen Welt" darstellt.

Dem in der Regel nichtstationären Verhalten realer Signale versuchten so frühe Arbeiten wie beispielsweise die von Bonton (Bonton, 1952) gerecht zu werden. Hierbei und in folgenden Arbeiten (z.B. Wang 1969) handelte es sich zumeist um Untersuchungen, inwieweit ein nichtstationärer Prozess durch ein auf Teilintervallen (Datenfenstern) stückweise stationäres (lineares) Modell approximiert werden kann oder, wenn es wie bei Wang (Wang, R.J. 1969) um die Bestimmung von optimalen Fensterlängen

ging, um die Beantwortung der Frage: "wie instationär sind seismische Daten ?"

Der Übergang von diesen Block-Verarbeitungsverfahren zu den adaptiven Filtern ist historisch schwer zu lokalisieren, zumal die Forschung an den entsprechenden Algorithmen über eine ständig wachsende Zahl verschiedener wissenschaftlicher Disziplinen im Spektrum von Operations Research bis hin zur Meß- und Regelungstechnik verteilt war, die zum Teil eine jeweils eigene Terminologie entwickelten. Irgendwann in den späten 60'er Jahren, scheint es, wurden Gradientenverfahren zum ersten Mal erfolgreich eingesetzt, um instationäre Prozesse mit einem AR-Modell mit zeitabhängigen Koeffizienten zu parametrisieren (Widrow et al. 1967, Sondhi, M. 1967). Die früheste mir bekannte Arbeit auf dem Gebiet der Geophysik, die sich bezeichnender Weise mit dem instationären Multiplenproblem auseinandersetzt und ein eindimensionales Gradientenverfahren für einen adaptiven Filter vorschlägt, ist die von Griffiths (Griffiths, 1977).

Eindimensionale adaptive Filter, die im Gegensatz zu Gradientenverfahren das Minimierungsproblem exakt lösen und darüberhinaus sehr viel bessere Konvergenzeigenschaften zeigen, sind von Beginn der 80'er Jahre an inzwischen zur "Lehrbuchreife" entwickelt worden (Alexander 1986, Bellanger 1987).

Diese recht komplexen aber auch sehr recheneffizienten Algorithmen haben in die geophysikalische Literatur kaum Eingang gefunden, möglicherweise weil das Interesse der Mehrzahl der mit inversen Problemen beschäftigten Geophysiker sich in dieser Zeit auf die deterministischen Algorithmen verlagert hat und dort beeindruckende Fortschritte beispielsweise mit Wellengleichungsverfahren erzielt wurden.

Es scheint mir daher gerechtfertigt in der vorliegenden Arbeit auch den eindimensionalen Vertreter der sogenannten Lattice-Algorithmen auf seine Anwendbarkeit auf das Multiplenproblem in marinen seismischen Reflexionsdaten zu untersuchen. Für die Darstellung der wichtigsten theoretischen Zusammenhänge ist der

Weg über den eindimensionalen Algorithmus ohnehin notwendig und bezogen auf die praktische Anwendung ist ein Vergleich der beiden Verfahren sinnvoll.

Mehrdimensionale oder genauer mehrkanalige adaptive Filter sind Gegenstand aktueller Forschung auf dem Gebiet der Signal- und Bildverarbeitung (Lewis, 1990). Die numerischen Probleme, die diesen Algorithmen anhaften und die Suche nach Möglichkeiten, diese Probleme zu lösen bilden einen Schwerpunkt dieser Arbeit.

Fachsprache in der Signaltheorie ist, wie in anderen Gebieten auch, Englisch. In den meisten Fällen wurden Fachausdrücke in die deutsche Sprache übersetzt. Um dem Leser die Orientierung in der englischsprachigen Sekundärliteratur zu erleichtern, sind die englischen Termini jedoch in den meisten Fällen mit angegeben. Wo eine Übersetzung nicht gelang oder nicht sinnvoll erschien, sind die Originalbegriffe, dort wo sie eingeführt werden, im Text erläutert.

## 1. Notation

In dieser Arbeit werden die folgenden Symbole verwendet:

$\bar{x}$	reeller Vektor
$x_k$	k-te Komponente von $\bar{x}$
$D$	reelle Matrix
$D^T$	transponierte Matrix
$\text{tr}(D)$	Spur der Matrix $D$
$\{D\}$	Spalten-Vektorraum der Matrix $D$
$\text{diag}(\lambda_1, \lambda_2 \dots \lambda_n)$	Diagonalmatrix mit Elementen $\lambda_1, \lambda_2 \dots$
$\lambda_n$	
$\bar{d}_k$	k-ter Zeilenvektor der Matrix $D$
$\bar{d}^k$	k-ter Spaltenvektor der Matrix $D$
$\langle \bar{x}, \bar{x} \rangle$	Skalarprodukt $\bar{x}^T \bar{x}$

## 2. Einführung

Multiple Reflexionen sind zunächst ein allgemeines Problem der marinen Seismik. Die seismische Energie, die in der Nähe der Wasseroberfläche freigesetzt wird, breitet sich in der Wasserschicht aus und dringt nur zum Teil in das Gestein des Meeresbodens ein.

Die Grenzfläche Wasser/Meeresboden zusammen mit der Grenzfläche Wasser/Luft bilden einen Wellenleiter, in dem die seismische Energie hin und her reflektiert wird und nur langsam abnimmt.

Speziell in den polaren Breiten ist der Anteil, der vom Meeresboden unmittelbar reflektiert wird, besonders groß und so wird das Problem der Wasserschicht-Multiplen oft mit der marinen Seismik in hohen Breiten assoziiert (vergl. Clearbout, 1985).

Die Frage, wieso in diesen Regionen der Meeresboden die akustische Energie so stark reflektiert, ist nicht abschließend geklärt und das Spektrum der Hypothesen entsprechend breit.

Dort, wo Gletscher im Bereich der Schelfe existiert haben, werden jüngere Sedimente durch die Auflast zum Teil überkompaktiert, zum Teil auch abgeschliffen und dadurch älteres, stärker verfestigtes Sediment freigelegt. Beides würde zu höheren Dichte- und Geschwindigkeitskontrasten und damit zu stärkeren Reflexionen an der Grenzfläche Wasser/Meeresboden führen als in Gebieten, in denen junge unverfestigte Sedimente mit hohem Wassergehalt den Meeresboden bilden.

Clearbout (Clearbout, 1985) führt an, daß in Wasser gelöstes Methan den Gefrierpunkt von Wasser unter hohen Drücken anhebt und es möglich wäre, daß methan-gesättigtes Eis in die Lockersedimente eingebettet ist, was ebenfalls zu hohen Reflexionskoeffizienten führen würde.

Die im Kapitel 4 dieser Arbeit vorgestellten reflexionsseismischen Daten wurden im Verlauf der Arktisreise ARK VII-3 des Forschungseisbrechers POLARSTERN im Jahr 1990 von der Seismik-Arbeitsgruppe des Alfred-Wegener Instituts im Scoreby-Sund aufgezeichnet.

Der Scoreby-Sund an der Ostküste Grönlands ist eines der größten Fjordsysteme der Erde und bildet die Nahtstelle zwischen drei völlig unterschiedlichen geologischen Provinzen. Die Natur dieser Nahtstelle sollte mit unterschiedlich auflösenden reflexionsseismischen Methoden und am Fjordsystem orientierten refraktionsseismischen Messungen aufgeklärt werden.

Das für diese Arbeit ausgewählte Profil 90550 wurde in Hall Bredning, dem nördlichen Hauptbecken des Scorsby Sunds aufgezeichnet. Die Wassertiefen längs des Profils liegen zwischen 300 und 500 m.

Die seismische Sektion in Abbildung 4.3.2 zeigt stark ausgeprägte Multiple, deren Energie von Einsatz zu Einsatz nur langsam abnimmt, und aufgrund der geringen Wassertiefen fallen die ersten Einsätze in den Zeitbereich, in dem Einsätze von Primärreflexionen noch erwartet werden können. Da offensichtlich nur wenig Energie in die tieferen Schichten des Meeresbodens eindringt, wären diese Primärreflexionen jedoch nur schwach und möglicherweise unter den Multipleneinsätzen nicht zu erkennen.

Dieses Beispiel zeigt auch, daß die Meeresbodenmultiple, zumal wenn sie derart ausgeprägt ist, in einem seismischen Profil leicht zu identifizieren ist. Ein weitaus größeres Problem stellen die Multiplen von Reflexionen aus dem Untergrund dar, die sogenannten Peg-Leg Multiplen. Die filigrane Struktur der Primärreflexionen der Sedimente unterhalb des Meeresbodens im Profil 90550 machen die Zuordnung von primärer und multipler Reflexion zu einer schwierigen Aufgabe, zumal aufeinanderfolgende Multiple einen Phasensprung von 180 Grad aufweisen.

Die seismischen Verfahren sind, wie beispielsweise auch Radar, Laufzeitverfahren. Multiple Reflexionen gehören zu den sogenannten "Multipathing"-Effekten, die von einem Objekt reflektierte Welle trifft auf vielen verschiedenen Wegen wieder beim Empfänger ein. Der Anteil des Wellenfeldes, der den Empfänger auf direktem Weg erreicht, muß vom mehrfach reflektierten Wellenfeld getrennt werden. Dies wird in der Radartechnik u.a. dadurch erreicht, daß Sende- und Empfangsantennen mit extrem guter Richtcharakteristik verwendet werden. Die Richtung, aus der ein reflektiertes Signal relativ zur Richtung des abgestrahlten Signals empfangen wird, ist der Schlüssel zur Diskriminierung von Einfach- und Mehrfachreflexionen. In der marinen Reflexionsseismik ist der Prozess, der die akustischen Echos generiert, recht gut bekannt. Die vorliegende Arbeit entwickelt ein neues Verfahren zur Unterdrückung multipler Reflexionen in marinen seismischen Daten, für das die a-priori bekannten Perioden der Multipleneinsätze den Schlüssel zur Diskriminierung darstellen.

In Kapitel 3 wird die Theorie dieses Verfahrens schrittweise entwickelt, soweit sie zum Verständnis insbesondere der numerischen Probleme des adaptiven Mehrkanal-Filterverfahrens notwendig ist.

Der Abschnitt 3.1 führt in das autoregressive Modell zur allgemeinen Beschreibung eines Signals ein und erläutert das auf diesem Modell basierende inverse Problem der linearen Vorhersage. Verschiedene Möglichkeiten der Formulierung dieses Problems werden vorgestellt und die Konsequenzen für eine algorithmische Lösung im Fall nichtstationärer Daten diskutiert. Hier wird die sogenannte "Prewindowed"-Form eingeführt, die einer rekursiven Lösung zugänglich ist. Projektionsoperatoren bilden den Kern eines effizienten rekursiven Algorithmus, und in Abschnitt 3.2 werden deren Eigenschaften beschrieben. Die spezielle Struktur der Datenmatrix des Vorhersageproblems führt auf den sogenannten Lattice Algorithmus, der mit Hilfe des Projektions-Operator-Formalismus in Abschnitt 3.3 skizziert wird. An dieser Stelle sei darauf hingewiesen, daß ein interessanter

Zusammenhang zwischen Lattice-Struktur und Verfahren zur Berechnung der Impulsantwort eines einfachen, geschichteten Mediums besteht (Bruckstein 1987, Robinson, 1980), auf den in dieser Arbeit jedoch nicht weiter eingegangen wird. In den Abschnitten 3.4 und 3.5 wird der Lattice Algorithmus zum eindimensionalen adaptiven Multiplenfilter erweitert.

Im Abschnitt 3.6 schließlich wird ein adaptiver Mehrkanal-Lattice Algorithmus beschrieben. Die Implementation dieses Algorithmus förderte erhebliche numerische Probleme insbesondere bei der Anwendung auf seismische Daten zutage, die hier analysiert werden, während Abschnitt 3.7 den Lösungsmöglichkeiten dieser Probleme gewidmet ist. Der Algorithmus von Lewis (Lewis, 1990) liefert hierzu einen entscheidenden Beitrag ebenso wie die Diskrete Cosinus-Transformation, die in den Algorithmus aufgenommen wurde.

In Kapitel 4 werden bestimmte Eigenschaften multipler Reflexionen in Verbindung mit der freien Wasseroberfläche zur argumentativen Untermauerung des in dieser Arbeit favorisierten Mehrkanalverfahrens beleuchtet. Einkanal- und Mehrkanalfilter werden auf synthetische und reale Daten angewandt, die Ergebnisse verglichen und diskutiert. Die Probleme, die in dieser Arbeit nicht gelöst werden konnten, werden im Kapitel 5 geschildert, und im Kapitel 7 sind die Ergebnisse der Arbeit abschließend zusammengefaßt.

### 3.Theorie

Das Autoregressive-Moving Average Modell (ARMA) ist das elementarste Modell der linearen Signaltheorie zur Beschreibung von diskret vorliegenden Signalen, die hier als Folgen von Abtastwerten  $x(t)$  und  $y(t)$  eingeführt werden. Dieses Modell bildet die Basis für inverse Probleme wie beispielsweise das Dekonvolutionsproblem, das Problem der System-Identifikation oder das Problem der linearen Vorhersage.

$$x(t) = \sum_{j=0}^{N-1} h_j y(t-j) \quad (3.1)$$

$$y(t) = \sum_{k=1}^P a_k y(t-k) + e(t) \quad (3.2)$$

Die Gleichungen 3.1 und 3.2 beschreiben das ARMA-Modell, Gleichung 3.2 den autoregressiven Teil. Angetrieben von einer in der Regel unbekanntem, stochastischen Zeitreihe  $e(t)$  bestehen die Ausgangsdaten des Systems  $y(t)$  aus einer Linearkombination vergangener Werte und bildet die Eingangsdaten für das Moving-Average System in Gleichung 3.1. Am Ausgang des gesamten Systems erscheint die Zeitreihe  $x(t)$ .

Schon im Hinblick auf die Anwendung, die in dieser Arbeit im Vordergrund steht, ist Abschnitt 3.1 auf die Darstellung des autoregressiven (AR) Modells und die Formulierung von zwei Beispielen für inverse Probleme beschränkt.

### 3.1 Autoregressives Modell, Systemidentifikation und Vorhersageoperatoren

Ausgehend von Gleichung (3.2) lassen sich zwei grundlegende inverse Probleme formulieren. Die vom System produzierten Daten seien bekannt, ebenso die Eigenschaften der stochastischen Zeitreihe  $e(t)$ . Die Suche nach Koeffizienten  $\hat{a}_k$  mit

$$\hat{y}(t) = \sum_{k=1}^M \hat{a}_k y(t-k),$$

die bezüglich eines zunächst beliebigen Kriteriums die wahren Ausgangsdaten des Systems optimal approximieren, also

$\|\hat{y}(t) - y(t)\|_p$  bezüglich irgendeiner Norm  $p$  minimieren, bezeichnet man als Problem der Systemidentifikation oder der AR-Modellierung. Von Interesse ist dieses Problem unter anderem bei der Berechnung hochauflösender Spektren (high resolution spectral estimation).

Bei bekannten Parametern  $a_k$  kann die Transferfunktion des Systems sofort angegeben werden:

$$H(z) = \frac{1}{1 - A(z)}$$
$$A(z) = \sum_{k=1}^M a_k z^k \quad (3.1.1)$$

Gleichung (3.1.1) beschreibt einen sogenannten All-Pol Prozess. Es ist klar, daß das von (3.1.1) beschriebene Spektrum extrem schmalbandig und von großen Magnituden sein kann. Die Auswertung von (3.1.1) für  $z = e^{j\omega}$  kann nun mit beliebig guter Auflösung in  $\omega$  erfolgen, wohingegen Fourier-Spektren in ihrer Auflösung von vornherein beschränkt sind.

Ein zweites Beispiel für ein auf (3.2) basierendes Verfahren ist das lineare Vorhersageproblem. Aufgrund der Kenntnis der Daten  $y(t-k)$ ,  $k>0$ , soll ein zukünftiger Wert  $\hat{y}(t)$  möglichst gut vorhergesagt werden. Möglichst gut wiederum im Sinne eines Optimalitätskriteriums, sodaß Vorhersage  $\hat{y}(t)$  und der sich tatsächlich einstellende Wert  $y(t)$  möglichst wenig voneinander abweichen.

Das Vorhersageproblem unterscheidet sich auf den ersten Blick nicht wesentlich vom Identifikationsproblem. Formuliert man die Optimierungsaufgabe als Least-Squares-Problem, minimiert man also die Euklidische- oder  $L_2$ -Norm, kann für das Vorhersageproblem jedoch auf die explizite Berechnung der Koeffizienten  $a_k$  verzichtet werden, was erheblich zur Effizienz der verwendeten Algorithmen beiträgt (siehe Abschnitt 3.2.1).

### 3.1.1 Formulierung des stationären 1-D Vorhersageproblems

Wie schon in der Einführung und im vorangegangenen Abschnitt wird auch hier wieder vorausgesetzt, daß Daten diskret vorliegen. Ein Datenvektor  $\bar{y}$  besteht demnach aus Komponenten  $y_t = y(n \times \delta t)$ .

Das lineare, stationäre Vorhersageproblem in Anlehnung an Gleichung (3.2) läßt sich nun folgendermaßen formulieren:

$$e_t = y_t - \sum_{i=1}^P a_i y_{t-i}, \quad t=0, \dots, T+p \quad (3.1.1.1)$$

mit  $\|\bar{e}\|_2 = \langle \bar{e}, \bar{e} \rangle^{\frac{1}{2}} = \min \quad (3.1.1.2).$

Gleichung (3.1.1.1) beschreibt ein i.d.R. überbestimmtes, lineares Gleichungssystem. Aus der Menge aller Lösungen greift (3.1.1.2) diejenige mit der kleinsten Euklidischen Norm heraus.

Die Wahl der  $L_2$ - oder Euklidischen Norm für die Restriktion läßt sich statistisch begründen, die Argumente sind u.a. in Chui (Chui 1987, Kap. 1.3) nachzulesen. Vom Standpunkt der linearen Algebra aus argumentiert, ist die Wahl von  $L_2$  wohl hauptsächlich begründet in der Handhabbarkeit der Optimierungsaufgabe und der Stabilität der Lösung.

Es soll an dieser Stelle darauf hingewiesen werden, daß die  $L_2$ -Norm trotz ihrer Fundierung in der Signal-Statistik nicht immer die geeignete Wahl für das Optimierungsproblem darstellt. Ein Beispiel dafür, wie robuste und sinnvolle Lösungen des

Dekonvolutionsproblems mit einer  $L_1$  Restriktion,  $\|\tilde{e}\|_1 = \sum_{k=1}^N e_k$ ,

erreicht werden können, findet sich in Dorau (Dorau 1991).

Um die Problematik diskreter, endlicher Signale genauer aufzuzeigen, wird Gleichung (3.1.1.1) in eine explizite Matrixschreibweise übertragen:

$$\begin{aligned} & \begin{bmatrix} e_0 & \dots & e_{T+p-1} & 0 \end{bmatrix} = \\ & = \begin{bmatrix} y_0 & \dots & y_T & 0 & \dots & 0 \end{bmatrix} - \begin{bmatrix} a_1 & \dots & a_p \end{bmatrix} \mathbf{Y}^T \quad (3.1.1.3) \end{aligned}$$

mit

$$\mathbf{Y}^T := \begin{bmatrix} 0 & y_0 & \dots & y_{p-1} & \dots & y_{T-1} & 0 & \dots & 0 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 0 & \dots & 0 & y_0 & \dots & y_{T-p} & \dots & y_{T-1} & \dots \end{bmatrix}$$

$\mathbf{Y}$  ist die Datenmatrix, in der die einzelnen Elemente des Datenvektors  $\tilde{y}$  entsprechend der Konvolutionsgleichung eingetragen sind.

Die Bezeichnungen *Datenmatrix* für  $\mathbf{Y}$  und *Datenvektor* für  $\bar{\mathbf{y}}$  werden im weiteren beibehalten.

Hierbei ist berücksichtigt, daß das Signal  $\bar{\mathbf{y}}$  kausal und in seiner Aufzeichnung von endlicher Dauer ist. In der Datenmatrix  $\mathbf{Y}$  wird dem Rechnung getragen, indem Werte  $y_t$  für  $t < 0$  und  $t > T$  gleich Null gesetzt sind. Diese Vorgehensweise liegt auf der Hand, hat aber nichttriviale Konsequenzen für den Gültigkeitsbereich der Lösung und die Übereinstimmung von algebraischen und statistischen Formulierungen.

Entsprechend der Restriktion (3.1.1.2) führt die Bildung der partiellen Ableitung und Nullsetzen für die Suche nach dem Extremum

$$\frac{\partial}{\partial \epsilon_i} \|\mathbf{Y}\hat{\mathbf{a}} - \bar{\mathbf{y}}\|_2^2 = 0$$

auf die sogenannten Yule-Walker Gleichungen

$$\mathbf{Y}^T \mathbf{Y} \hat{\mathbf{a}} - \mathbf{Y}^T \bar{\mathbf{y}} = 0 \quad (3.1.1.4).$$

Unter der Voraussetzung  $\mathbf{Y} \in \mathbb{R}^{n,m}$ ,  $n > m$ ,  $\text{Rg}(\mathbf{Y}) = m$  ist  $\mathbf{Y}^T \mathbf{Y}$  positiv definit und invertierbar, das Gleichungssystem (3.1.1.4) hat also eine eindeutige Lösung  $\hat{\mathbf{a}}$ .

### 3.1.2 Autokorrelations-, Kovarianz- und "Prewindowed"-Formulierung

In der Gleichung (3.1.1.3) hat die Datenmatrix die sogenannte Autokorrelations-Form, sodaß die Elemente von  $\mathbf{Y}^T \mathbf{Y}$  der Vorschrift

$$R_\tau = \sum_{t=0}^{T-1} y_t y_{t+\tau} \quad \text{und} \\ y_{t+\tau} = 0 \quad \text{für } t+\tau > T-1 \quad (3.1.2.1)$$

für die Berechnung der Koeffizienten der diskreten Autokorrelation entsprechen. Hier ist

$$\mathbf{Y}^T \mathbf{Y} =: \mathbf{R} = \begin{bmatrix} R_0 & R_1 & \dots & R_{p-1} \\ R_{-1} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ R_{-p-1} & \cdot & \cdot & R_0 \end{bmatrix} \quad (3.1.2.2)$$

aus der Yule-Walker Gleichung (3.1.1.4) die Autokorrelationsmatrix des Datenvektors  $\bar{y}$ .

Diese Matrix hat Töplitzstruktur mit  $R_{ij} = R_{i-j}$  und ist entsprechend (3.1.2.1) symmetrisch  $R_{ij} = R_{ji}$ .

Die Töplitzstruktur ist unmittelbar davon abhängig, daß Datenwerte außerhalb des Beobachtungsintervalls  $[0, T]$  identisch Null sind. Dies kann eine, den physikalischen Eigenschaften des Signals oder der Anordnung eines Experiments entsprechend, absolut vernünftige Annahme sein, ist gelegentlich jedoch unsinnig und Quelle von Fehlern.

Abweichend vom ursprünglichen Problem in Gl.(3.1.1.3) existieren zwei weitere Formulierungen für das Vorhersageproblem, wobei die Modifikation in der Datenmatrix verantwortlich ist für den jeweiligen Namen.

Die Prewindowed-Formulierung zieht Datenwerte  $y_t$ ,  $t > T-1$  nicht in Betracht, sie entspricht

$$e_{p,t} = y_t - \sum_{k=1}^p a_k y_{t-k}, \quad t=0, \dots, T \quad (3.1.2.3),$$

in expliziter Matrix-Form:

$$\begin{bmatrix} e_0 & \dots & e_T \end{bmatrix} = \begin{bmatrix} y_0 & \dots & y_T \end{bmatrix} - \begin{bmatrix} a_1 & \dots & a_p \end{bmatrix} \begin{bmatrix} 0 & y_0 & \dots & y_{p-1} & \dots & y_{T-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \dots & 0 & y_0 & \dots & y_{T-p} \end{bmatrix} \quad (3.1.2.4)$$

Die Bezeichnung "Prewindowed" illustriert den Fall, daß ein kausales Signal  $y(t)$  auf einem "uneigentlichen" Intervall  $[0,t]$ ,  $t=0, \dots, T-1$  betrachtet wird. In die Berechnung der Koeffizienten gehen im Gegensatz zum Autokorrelations-Fall nur Werte  $0, y_0, \dots, y_{T-1}$  ein, bezüglich der Werte  $y_t$   $t > T$  werden keine Annahmen getroffen. Analog dazu existiert auch die sog. "Postwindowed"-Form, deren Datenmatrix mit Elementen des Datenvektors über dem Intervall  $[t, T-1]$ ,  $t = 0, \dots, T-1$  bestückt ist, Nullelemente stehen in der Postwindowed-Datenmatrix anstelle der Datenwerte für  $t > T-1$ .

Die Covarianz-Form berechnet

$$e_{p,t} = y_t - \sum_{k=1}^p a_k y_{t-k}, \quad t=p, \dots, T,$$

entsprechend

$$\begin{bmatrix} e_p & \dots & e_T \end{bmatrix} = \begin{bmatrix} y_0 & \dots & y_T \end{bmatrix} - \begin{bmatrix} a_1 & \dots & a_p \end{bmatrix} \begin{bmatrix} y_{p-1} & \dots & y_{T-1} \\ \vdots & & \vdots \\ y_0 & \dots & y_{T-p} \end{bmatrix}.$$

Diese Formulierung schränkt die Daten, die Eingang in die Lösung finden, auf ein abgeschlossenes Intervall  $[0, T]$  ein, sodaß keine Annahmen über nicht beobachtete Datenwerte getroffen werden müssen.

Der Lösungsvektor  $\bar{a}$  ist für jede der drei Formulierungen zunächst verschieden, man kann jedoch zeigen, daß sie im Grenzfall  $T \rightarrow \infty$  gegeneinander konvergieren (Friedlander 1982). Entscheidend ist jedoch, daß die Korrelationsmatrix  $\mathbf{Y}^T \mathbf{Y}$  nur in der Autokorrelations-Form Töplitzstruktur hat.

### 3.1.3 Lewinson-Durbin Rekursion und Block-Verarbeitungen Techniken

Mit der Lewinson-Durbin Rekursion existiert ein sehr effizienter Algorithmus für die Lösung der Yule-Walker Gleichungen für den Autokorrelations-Fall. Der Algorithmus nutzt die Töplitzstruktur der Autokorrelationsmatrix aus, deren Elemente längs der Haupt-

und Nebendiagonalen konstant sind. Versucht man, ein instationäres Signal stückweise stationär zu approximieren, führt dies entweder dazu, daß auf den Datenfenstern der Teilprobleme die Töplitzstruktur der Matrix  $\mathbf{Y}^T\mathbf{Y}$  durch die Prewindowed-, Postwindowed- oder Covarianz-Formulierung verloren geht, oder will man die Töplitzstruktur des Autokorrelations-Falls erzwingen- Datenwerte außerhalb des Teilintervalls willkürlich Null gesetzt werden müssen, die tatsächlich von Null verschieden sind.

Im erstgenannten Fall müßte zur Lösung des Gleichungssystems ein allgemeiner und sehr rechenaufwendiger Algorithmus für die Inversion symmetrischer Matrizen eingesetzt werden. Die Anzahl der Datensamples im Beobachtungsintervall muß zudem groß sein im Verhältnis zur Operatorlänge  $P$ , damit die Approximation der Autokorrelationsmatrix hinreichend genau ist.

### 3.2 Eigenschaften der Least-Squares Lösung

Zur Vereinfachung der Darstellung sei hier zunächst das allgemeine lineare Ausgleichsproblem

$$\mathbf{A}\bar{\mathbf{x}} = \bar{\mathbf{b}} + \bar{\mathbf{e}} = \hat{\mathbf{b}} \quad (3.2.1)$$

$$\langle \bar{\mathbf{e}}, \bar{\mathbf{e}} \rangle = \|\mathbf{A}\bar{\mathbf{x}} - \bar{\mathbf{b}}\|_2^2 \stackrel{!}{=} \min,$$

$$\mathbf{A} \in \mathbb{R}^{m,n}, \text{ pos. def. } \bar{\mathbf{b}} \in \mathbb{R}^m, m > n$$

betrachtet, mit  $\hat{\mathbf{x}}$  als Lösung von

$$D_{\hat{\mathbf{x}}}(\|\mathbf{A}\hat{\mathbf{x}} - \bar{\mathbf{b}}\|_2^2) = 2\mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} - 2\mathbf{A}^T\bar{\mathbf{b}} = \bar{\mathbf{0}}, \quad (3.2.2)$$

entsprechend Gl.(3.1.1.4).

Sei ferner  $U := \{ \mathbf{A}\bar{\mathbf{x}} = \sum_{i=1}^n \bar{a}_i \bar{x}_i : \bar{\mathbf{x}} \in \mathbb{R}^n \}$

der Spaltenraum von  $\mathbf{A}$  und

$U_{\perp} := \{ \bar{\mathbf{r}} \in \mathbb{R}^m, \bar{\mathbf{r}}^T \mathbf{A}\bar{\mathbf{x}} = \bar{0}, \forall \bar{\mathbf{x}} \in U \}$  der zu  $U$

orthogonale Raum.

Der Vektor  $\hat{\mathbf{x}}$  löst sowohl (3.2.1) als auch (3.2.2),

$$\mathbf{A}\hat{\mathbf{x}} - \bar{\mathbf{b}} = \bar{\mathbf{e}}$$

$$\mathbf{A}^T \mathbf{A}\hat{\mathbf{x}} - \mathbf{A}^T \bar{\mathbf{b}} = \mathbf{0}$$

also

$$\mathbf{A}^T (\mathbf{A}\hat{\mathbf{x}} - \bar{\mathbf{b}}) = \mathbf{0}$$

$$\mathbf{A}^T \bar{\mathbf{e}} = \mathbf{0} \quad (3.2.3)$$

demzufolge  $\bar{\mathbf{e}} \in U_{\perp}$ .

Für den Vektor  $\bar{\mathbf{b}}$  im ursprünglichen Problem gilt:

$$\mathbf{A}\hat{\mathbf{x}} = \bar{\mathbf{b}} + \bar{\mathbf{e}} = \hat{\mathbf{b}}$$

mit (3.2.2)

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \bar{\mathbf{b}}$$

$$\hat{\mathbf{b}} = \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \bar{\mathbf{b}},$$

$$\hat{\mathbf{b}} = \mathbf{P}\bar{\mathbf{b}}, \quad \mathbf{P} := \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \quad (3.2.4)$$

$$\bar{\mathbf{e}} = \bar{\mathbf{b}} - \hat{\mathbf{b}} = (\mathbf{1} - \mathbf{P})\bar{\mathbf{b}} = \mathbf{P}_\perp \bar{\mathbf{b}}, \quad \mathbf{P}_\perp := (\mathbf{1} - \mathbf{P}). \quad (3.2.5)$$

Bezogen auf das Problem der linearen Vorhersage in Gl(3.1.1.3)

$$\mathbf{Y}\bar{\mathbf{a}} - \bar{\mathbf{y}} = \bar{\mathbf{e}}, \quad \langle \bar{\mathbf{e}}, \bar{\mathbf{e}} \rangle \stackrel{!}{=} \min,$$

erkennt man bei Identifikation der geeigneten Größen in (3.2.4) und (3.2.5), daß das Vorhersageproblem ohne explizite Berechnung der Koeffizienten des Operators  $a_k$  behandelt werden kann. Allein aus der Kenntnis der Datenmatrix  $\mathbf{Y}$  und des Datenvektors  $\bar{\mathbf{y}}$  können der Vektor  $\hat{\mathbf{y}}$  der Vorhersage und der Fehlervektor  $\bar{\mathbf{e}}$  berechnet werden.

In der auf dem Konzept von Vektorräumen basierenden Behandlung des AR-Modells spielt die Orthogonalitätseigenschaft (3.2.3) zusammen mit den Gleichungen (3.2.4) und (3.2.5) für die Vorhersage- und Vorhersagefehler-Projektionen eine zentrale Rolle. Zusammen mit einer abgewandelten Lewinson-Durbin Rekursion bilden sie die Basis für einen effizienten, adaptiven Filteralgorithmus.

### 3.2.1 Projektionsoperatoren für Vorhersage- und Vorhersagefehler-Operatoren

Man kann leicht nachrechnen, daß die Matrizen  $\mathbf{P}$  und  $\mathbf{P}_\perp$  aus (3.2.4) und (3.2.5) folgende Eigenschaften besitzen

$$\text{i) } \mathbf{P}^T = \mathbf{P}, \quad \mathbf{P}_\perp^T \mathbf{P}_\perp = \mathbf{P}_\perp$$

$$\text{ii) } \mathbf{P}\mathbf{P}^T = \mathbf{P}^T \mathbf{P} = \mathbf{P}\mathbf{P} = \mathbf{P}, \quad \mathbf{P}_\perp \mathbf{P}_\perp^T = \mathbf{P}_\perp^T \mathbf{P}_\perp = \mathbf{P}_\perp \mathbf{P}_\perp = \mathbf{P}_\perp$$

mit  $\mathbf{P}_\perp = (\mathbf{1} - \mathbf{P})$ , ( $\mathbf{1}$  : Einheitsmatrix),

$$\text{iii) } \mathbf{P}_\perp \mathbf{P} = \mathbf{P} \mathbf{P}_\perp = \mathbf{0} \quad (3.2.1.1).$$

Die Eigenschaften i)-iii) sind die Eigenschaften eines Projektionsoperators.

Unabhängig von der Form des Ausgangsproblems, also unabhängig davon, ob Autokorrelations-, Kovarianz- oder Prewindowed-Formulierung für die Beschreibung des zu untersuchenden Prozesses angemessen sind, liegen die Projektionsoperatoren für jede Ordnung  $p$ ,  $p=0, \dots, P$  und für jeden diskreten Zeitpunkt  $t$ ,  $t=0, \dots, T-1$ , fest. Die hochgestellten Indizes werden im folgenden zur Kennzeichnung dieses Zusammenhangs benutzt.

Der Projektionsoperator  $\mathbf{P}^{p,t}$  hängt demnach von der Datenmatrix  $\mathbf{Y}^{p,t}$  der Ordnung  $p$  zum Zeitpunkt  $t$  ab mit

$$\mathbf{P}^{p,t} = \mathbf{Y}^{p,t} (\mathbf{Y}^{p,tT} \mathbf{Y}^{p,t})^{-1} \mathbf{Y}^{p,tT}.$$

Die Form der Datenmatrix  $\mathbf{Y}^{p,t}$  ist selbstverständlich abhängig von der gewählten Formulierung. Diskutiert werden soll hier nur die Prewindowed-Formulierung. Diese Form wurde für die hier in dieser Arbeit verwendeten Algorithmen benutzt, die Kovarianz-Formulierung läßt sich ebenfalls zu einem adaptiven Verfahren entwickeln, Friedlander (Friedlander, 1982) gibt jedoch an, daß für die Autokorrelationsform keine effiziente, adaptive, also in der Zeit rekursive Lösung angegeben werden kann. Ein Grund hierfür wird im folgenden, wenn auch eher intuitiv, beschrieben.

Beispielhaft soll hier die Veränderung der Prewindowed-Datenmatrix bei einer Aufdatierung in der Ordnung und einer Aufdatierung in der Zeit explizit dargestellt werden. Der Übersichtlichkeit der Darstellung kommt es entgegen, wenn hier wie bspw. schon in (3.1.1.3) die Transponierten betrachtet werden.

1) Aufdatierung in der Ordnung

Sei

$$\mathbf{Y}^{p,t^T} := \begin{bmatrix} 0 & y_0 & \dots & y_{p-1} & \dots & y_{t-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \dots & 0 & y_0 & \dots & y_{t-p} \end{bmatrix}$$

die Datenmatrix der Ordnung  $p$  zur Zeit  $t$ , die Matrix

$$\mathbf{Y}^{p+1,t^T} := \begin{bmatrix} 0 & y_0 & \dots & y_{p-1} & \dots & y_{t-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & y_0 & \dots & y_{t-p} \\ 0 & \dots & 0 & y_0 & \dots & y_{t-p-1} \end{bmatrix}$$

entsteht aus  $\mathbf{Y}^{p,t^T}$  durch Anfügen des Zeilenvektors

$$\bar{y}^{p+1,t^T} = [0 \dots 0 y_0 \dots y_{t-p-1}]$$

$$\mathbf{Y}^{p+1,t^T} = \begin{bmatrix} \mathbf{Y}^{p,t^T} \\ \bar{y}^{p+1,t^T} \end{bmatrix},$$

wobei  $\bar{y}^{p+1,t}$  eine um  $p+1$  Zeitschritte zeitverzögerte Version des ursprünglichen Datenvektors  $\bar{y}^{0,t}$  darstellt, dessen letzte  $T-p-1$  Komponenten im Prewindowed-Fall abgeschnitten sind.

2) Aufdatierung in der Zeit

$$\mathbf{Y}^{p,tT} = \begin{bmatrix} 0 & y_0 & \dots & y_{p-1} & \dots & y_{t-2} & y_{t-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \dots & 0 & y_0 & \dots & y_{t-p-2} & y_{t-p} \end{bmatrix}$$

entsteht durch Anfügen des Spaltenvektors

$$\bar{y}^{p,tT} = [y_{t-1} \dots y_{t-p}]^T$$

an die ursprüngliche Datenmatrix

$$\mathbf{Y}^{p,tT} = \left[ \mathbf{Y}^{p,t-1T} \bar{y}^{p,t} \right] .$$

Die Datenvektoren  $\bar{e}_t$  und  $\bar{y}_t$  wachsen um eine Komponente.

### 3.2.2 Aufdatierungen für Projektionsoperatoren

Im Sinne der am Eingang dieses Kapitels vorgestellten Darstellung der linearen Vorhersage eines Signals als Projektion in bestimmte Vektorräume und deren orthogonale Komplementäräume, die jeweils von der 'aktuellen' Datenmatrix aufgespannt werden, soll hier nun dargestellt werden, wie sich die Projektionen ändern über dem zugrundeliegenden, sich mit der Datenmatrix ändernden Vektorraum. Die im vorangegangenen Abschnitt vorgestellten Projektionsoperatoren eignen sich im Prinzip zur Konstruktion eines Einfachalgorithmus, mit dem ein zeit- und ordnungsrekursiver Vorhersage-Filter realisiert werden könnte. Die für jede Aufdatierung der Datenmatrix erforderliche Neuberechnung entsprechend (3.2.4) wäre allerdings ungeheuer aufwendig und ineffizient.

Die Frage ist daher, wie sich der Projektionsoperator  $P_{\{U\}}$  über einem Vektorraum  $U$  ändert, wenn ein zusätzlicher Vektor  $\vec{v}$  in diesen Vektorraum aufgenommen wird. Der Vektorraum  $U$  kann im Vorhersageproblem beispielsweise durch die Datenmatrix  $Y^{p,t}$  repräsentiert sein.

Seien  $\vec{u}_i$  die Basisvektoren des Vektorraumes  $U$ . Die Basis  $\vec{u}_i$  kann zu einer Basis für den erweiterten Vektorraum  $\{U, \vec{v}\}$  ergänzt werden, indem die zu  $U$  orthogonale Komponente  $\vec{w}$  von  $\vec{v}$  in die Basis aufgenommen wird.

Für die Projektionsoperatoren bedeutet dies einfach

$$P_{\{U, \vec{v}\}} = P_U + P_{U\vec{w}} = P_U + P_{U\perp\vec{v}}$$

oder mit der Definition des Projektionsoperators

$$P_{\{U, \vec{v}\}} = P_U + P_{\perp U \vec{v}} \langle P_{\perp U \vec{v}}, P_{\perp U \vec{v}} \rangle^{-1} \vec{v}^T P_{\perp U} \quad (3.2.2.1a),$$

für den orthogonalen Operator gilt entsprechend (3.2.1.1):

$$\mathbf{P}_{\perp\{U,\bar{v}\}} = \mathbf{1} - \mathbf{P}_{\{U,\bar{v}\}} = \mathbf{P}_{\perp U} - \mathbf{P}_{\perp U \bar{v}} \langle \mathbf{P}_{\perp U \bar{v}}, \mathbf{P}_{\perp U \bar{v}} \rangle^{-1} \bar{v}^T \mathbf{P}_{\perp U}$$

(3.2.2.1b).

Die Gleichungen (3.2.2.1) behalten ihre Gültigkeit, wenn sie mit einem beliebigen, in der Dimension kompatiblen Vektor oder einer kompatiblen Matrix von links bzw. von rechts multipliziert werden.

$$\mathbf{W}^T \mathbf{P}_{\{U,\bar{v}\}} \mathbf{Z} = \mathbf{W}^T \mathbf{P}_U \mathbf{Z} + \mathbf{W}^T \mathbf{P}_{\perp U \bar{v}} \langle \mathbf{P}_{\perp U \bar{v}}, \mathbf{P}_{\perp U \bar{v}} \rangle^{-1} \bar{v}^T \mathbf{P}_{\perp U} \mathbf{Z}$$

(3.2.2.2a)

$$\begin{aligned} \mathbf{W}^T \mathbf{P}_{\perp\{U,\bar{v}\}} \mathbf{Z} &= \mathbf{W}^T (\mathbf{1} - \mathbf{P}_{\{U,\bar{v}\}}) \mathbf{Z} = \\ &= \mathbf{W}^T \mathbf{P}_{\perp U} \mathbf{Z} - \mathbf{W}^T \mathbf{P}_{\perp U \bar{v}} \langle \mathbf{P}_{\perp U \bar{v}}, \mathbf{P}_{\perp U \bar{v}} \rangle^{-1} \bar{v}^T \mathbf{P}_{\perp U} \mathbf{Z} \end{aligned}$$

(3.2.2.2b) .

Auf diese Weise können nunmehr nicht nur die Aufdatierungen der Projektionsoperatoren berechnet werden, sondern gleichzeitig auch die Aufdatierungen der zu projizierenden Variablen.  $\mathbf{W}$  und  $\mathbf{Z}$  können in (3.2.2.2) beliebig durch Matrizen oder Vektoren ersetzt werden. Die Aufdatierungs-Relationen für skalare Größen erhält man, indem man an geeigneter Stelle den sogenannten "pinning-vector"  $\bar{\pi}_n := [0 \dots 0 \ 1 \ 0 \dots 0]^T$  einsetzt, dessen n-te Komponente gleich Eins ist während die anderen verschwinden.

Die Aufdatierungs-Relationen für das Vorhersageproblem, also geeignete Substitutionen für die Variablen  $U$  und  $\bar{v}$ , sollen im folgenden vorgestellt werden.

Aufdatierungen des Projektionsoperators für eine wachsende Ordnung des Problems liegen auf der Hand: Der neu in den Vektorraum eintretende Vektor  $\bar{y}^{p+1,t}$  ist ein Spaltenvektor der Datenmatrix (ein Zeilenvektor in der Transponierten), der resultierende neue Vektorraum ist einfach  $\{Y^{p,t}, \bar{y}^{p+1,t}\}$ . Etwas komplizierter ist der Formalismus der Aufdatierungen in der Zeit, hier ändert sich die Dimension des ganzen Gleichungssystems. Ein Zeilenvektor wird in die Datenmatrix aufgenommen, oder besser gesagt die Spalten der Datenmatrix werden jeweils um eine Komponente ergänzt, ebenso die Datenvektoren  $\bar{e}^{p,t}$  und  $\bar{y}^{p,t}$ .

Ein für die Aufdatierung des Projektionsoperators, der ebenfalls in der Dimension wachsen muß, geeigneter Kandidat ergibt sich aus den oben schon angeführten Überlegungen. Ein Vektor, der die Dimension des Vektorraums  $U$  um Eins erhöht und zudem orthogonal zur bisherigen Basis ist, ist der Einheitsvektor in der entsprechenden "Richtung"  $\bar{\pi}_t = [0, \dots, 0, 1]^T$ , dessen Komponenten alle bis auf die an der Stelle  $t$  verschwinden. Mit den folgenden Definitionen

$$Y^{p,t-1T} := \begin{bmatrix} 0 & y_0 & \dots & y_{p-1} & \dots & y_{T-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \dots & 0 & y_0 & \dots & y_{T-p-1} \end{bmatrix},$$

$$Y^{p,tT} := \begin{bmatrix} 0 & y_0 & \dots & y_{p-1} & \dots & y_{T-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \dots & 0 & y_0 & \dots & y_{T-p} \end{bmatrix},$$

$$\bar{\pi}_t^T := [0, \dots, 0, 1]$$

und aufgrund der Tatsache, daß die Vektorräume  $\{Y^{pt}, \bar{\pi}_t\}$  und  $\{Y^{pt-1}, \bar{\pi}_t\}$  äquivalent sind, zeigt Friedlander (Friedlander 1982), daß der folgende Zusammenhang zwischen den Projektionsoperatoren für aufeinander folgende Zeitschritte besteht:

$$\mathbf{P}_{\{\mathbf{Y}^{p,t}, \bar{\pi}_t\}} = \begin{bmatrix} \mathbf{P}_{\{\mathbf{Y}^{p,t-1}\}} & \bar{\mathbf{0}} \\ \bar{\mathbf{0}}^T & 1 \end{bmatrix} \quad (3.2.2.3a)$$

und entsprechend für den orthogonalen Operator

$$\mathbf{P}^\perp_{\{\mathbf{Y}^{p,t}, \bar{\pi}_t\}} = \mathbf{1} - \begin{bmatrix} \mathbf{P}_{\{\mathbf{Y}^{p,t-1}\}} & \bar{\mathbf{0}} \\ \bar{\mathbf{0}}^T & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{P}^\perp_{\{\mathbf{Y}^{p,t-1}\}} & \bar{\mathbf{0}} \\ \bar{\mathbf{0}}^T & 0 \end{bmatrix} \quad (3.2.2.3b).$$

Mit  $\bar{\mathbf{v}} = \bar{\pi}_t$  und den Gleichungen 3.2.2.1 lauten die expliziten Aufdatierungs-Relationen

$$\mathbf{P}_{\{\mathbf{Y}^{p,t}\}} = \begin{bmatrix} \mathbf{P}_{\{\mathbf{Y}^{p,t-1}\}} & \bar{\mathbf{0}} \\ \bar{\mathbf{0}}^T & 1 \end{bmatrix} - \frac{\mathbf{P}^\perp_{\{\mathbf{Y}^{p,t-1}\}} \bar{\pi}_t \bar{\pi}_t^T \mathbf{P}^\perp_{\{\mathbf{Y}^{p,t-1}\}}}{\bar{\pi}_t^T \mathbf{P}^\perp_{\{\mathbf{Y}^{p,t-1}\}} \bar{\pi}_t} \quad (3.2.2.3a)$$

$$\mathbf{P}^\perp_{\{\mathbf{Y}^{p,t}\}} = \begin{bmatrix} \mathbf{P}^\perp_{\{\mathbf{Y}^{p,t-1}\}} & \bar{\mathbf{0}} \\ \bar{\mathbf{0}}^T & 0 \end{bmatrix} - \frac{\mathbf{P}^\perp_{\{\mathbf{Y}^{p,t-1}\}} \bar{\pi}_t \bar{\pi}_t^T \mathbf{P}^\perp_{\{\mathbf{Y}^{p,t-1}\}}}{\bar{\pi}_t^T \mathbf{P}^\perp_{\{\mathbf{Y}^{p,t-1}\}} \bar{\pi}_t} \quad (3.2.2.3b)$$

Die wesentlichen Ingredienzien für einen sowohl in der Ordnung als auch in der Zeit rekursiven Vorhersagefilter oder Vorhersagefehler-Filter stehen damit bereit.

In der Tat ließe sich ein solcher Filter auf der Basis der Relationen (3.2.2.2) und (3.2.2.3) realisieren. Diese lassen sich darüberhinaus jedoch maßgeschneidert an die spezielle Struktur der Datenmatrix im Prewindowed AR-Modell anpassen, deren Spalten sich jeweils nur durch eine Verschiebung auf der diskreten Zeitachse und verschiedene "Reichweiten" in der Zeit unterscheiden.

Die Ausnutzung dieses Umstands folgt der grundlegenden Idee des Levinson-Durbin Algorithmus über die Definition eines Hilfsproblems der Rückwärtsvorhersage (Backward-Prediction) um wiederholte Berechnungen auf den gleichen Daten zu vermeiden. Am oben dargestellten Beispiel der Aufdatierungen in der Ordnung der Datenmatrix wird klar, daß hier nicht ein gänzlich neuer Vektor in die Datenmatrix eintritt. Nur eine Komponente dieses Vektors, ein Datensample, ist neu, alle anderen sind bereits bekannt und besetzen lediglich andere, gegenüber dem vorhergehenden Spaltenvektor um einen Zeitschritt verschobene, Komponenten. Damit ist zugleich auch deutlich geworden, daß für die Autokorrelationsform der Datenmatrix vergleichbar einfache Aufdatierungen in der Zeit weder für die Datenmatrix noch für die dazugehörigen Projektionsoperatoren existieren. In der Datenmatrix würde sich jeweils ein ganzer Block ändern und die Projektionsoperatoren müßten vollständig neu berechnet werden.

Der Lattice-Algorithmus, der im nächsten Abschnitt grob skizziert wird, sollte nicht mit dem Levinson-Durbin Algorithmus verwechselt werden, der Töplitz-Systeme löst und damit von der Autokorrelationsform der Datenmatrix abhängig ist.

### **3.3 Skizze des Lattice-Algorithmus**

Wie im letzten Abschnitt bereits angedeutet wurde, besteht eine Möglichkeit, die Effizienz des Projektionsalgorithmus und insbesondere der Aufdatierungen in der Ordnung für die Operatoren durch die Ausnutzung der strukturellen Redundanz der Datenmatrix zu steigern. Hierfür erweist sich die Formulierung eines Rückwärts-Vorhersage-Problems als hilfreich.

$$\bar{e}_{p,t} = \bar{y}_t - \sum_{k=1}^p \hat{a}_k y_{t-k}, \quad t=0, \dots, T$$

(3.3.1)

$$\bar{r}_{p,t-1} = \bar{y}_{t-p-1} - \sum_{k=1}^p \hat{b}_{p+1-k} y_{t-k} \quad t=0, \dots, T$$

(3.3.2a)

$$[0 \ r_{p,0} \ \dots \ r_{p,T-1}] = [0 \ \dots \ 0 \ y_0 \ \dots \ y_{T-p-1}] - [b_p \ \dots \ b_1] \mathbf{Y}^{p,T^T}$$

(3.3.2b)

In den Gleichungen (3.3.1) und (3.3.2) sind Vorwärts- und Rückwärts-Vorhersage-Problem einander gegenübergestellt, (3.3.2b) zeigt zum Vergleich auch die explizite Matrixdarstellung für die Prewindowed- Formulierung. Der Fehlervektor der Rückwärts-Vorhersage wird in (3.2.2) zur Zeit t-1 ausgewertet, dadurch wird erreicht, daß die Datenmatrix  $\mathbf{Y}^{p,T}$  für beide Probleme dieselbe ist.

Sei  $\mathbf{P}^{p,t}$  der zur aktuellen Datenmatrix gehörende Projektionsoperator, dann gilt nach (3.2.5)

$$\bar{e}_{p,t} = \mathbf{P}_{\perp}^{p,t} \begin{bmatrix} y_0 \\ \vdots \\ y_t \end{bmatrix}$$

für eine Aufdatierung in der Ordnung entsprechend:

$$\bar{e}_{p+1,t} = \mathbf{P}_{\perp}^{p+1,t} \begin{bmatrix} y_0 \\ \cdot \\ \cdot \\ y_t \end{bmatrix}, \text{ wobei hier für die Erweiterung des}$$

Vektorraums mit

$$\bar{v}^T = [0 \dots 0 y_0 \dots y_{t-p-1}]$$

die Aufdatierungsgleichung für den Projektionsoperator entsprechend (3.2.2.1)

$$\mathbf{P}_{\perp}^{p+1,t} = \mathbf{P}_{\perp}^{p,t} - \mathbf{P}_{\perp}^{p,t} \bar{v} \langle \mathbf{P}_{\perp}^{p,t} \bar{v}, \mathbf{P}_{\perp}^{p,t} \bar{v} \rangle^{-1} \bar{v}^T \mathbf{P}_{\perp}^{p,t} \quad \text{gilt.}$$

Der Vergleich mit (3.2.2b) zeigt, daß  $\bar{v}$  dort gleichzeitig dem Datenvektor  $\bar{y}$  des Rückwärtsproblems entspricht, die Gleichung oben läßt sich daher durch Substitution von

$$\mathbf{P}^{p,t} \bar{v} = \begin{bmatrix} \hat{y}_0 \\ \cdot \\ \cdot \\ \hat{y}_t \end{bmatrix}$$

$$\mathbf{P}_{\perp}^{p,t} \bar{v} = \mathbf{P}_{\perp}^{p,t} \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ 0 \\ y_0 \\ \cdot \\ \cdot \\ y_{t-p-1} \end{bmatrix} = \bar{r}_{p,t-1}$$

und Einsetzen nach (3.2.2.2b) mit  $\mathbf{W}=\mathbf{1}$ ,  $\mathbf{Z} = \begin{bmatrix} y_0 \\ \cdot \\ \cdot \\ y_t \end{bmatrix}$

zu

$$\bar{e}_{p+1,t} = \bar{e}_{p,t} - \frac{\bar{r}_{p,t-1} \bar{r}_{p,t-1}^T}{\bar{r}_{p,t-1}^T \bar{r}_{p,t-1}} \begin{bmatrix} y_0 \\ \cdot \\ \cdot \\ y_t \end{bmatrix}$$

und mit den Beziehungen

$$\begin{bmatrix} y_0 \\ \cdot \\ \cdot \\ y_t \end{bmatrix} = \bar{e}_{p,t} + \hat{y}_{p,t} = \bar{e}_{p,t} + \mathbf{P}^{p,t} \begin{bmatrix} y_0 \\ \cdot \\ \cdot \\ y_t \end{bmatrix},$$

$$\bar{r}_{p,t-1} (\bar{e}_{p,t} + \mathbf{P}^{p,t} \begin{bmatrix} y_0 \\ \cdot \\ \cdot \\ y_t \end{bmatrix}) = \mathbf{P}_\perp^{p,t} \bar{e}_{p,t} + \mathbf{P}_\perp^{p,t} \mathbf{P}^{p,t} \begin{bmatrix} y_0 \\ \cdot \\ \cdot \\ y_t \end{bmatrix}$$

mit  $\mathbf{P}_\perp^{p,t} \cdot \mathbf{P}^{p,t} = \mathbf{0}$

zu 
$$\bar{e}_{p+1,t} = \bar{e}_{p,t} - \frac{\bar{r}_{p,t-1} \bar{r}_{p,t-1}^T}{\bar{r}_{p,t-1}^T \bar{r}_{p,t-1}} \cdot \bar{e}_{p,t} \quad (3.3.3a)$$

vereinfachen.

Einen ganz ähnlichen Ausdruck findet man für den Fehlervektor der Rückwärts-Vorhersage, wenn man Aufdatierungen in Zeit und Ordnung zusammenfaßt. Für die Datenmatrix gilt:

$$\mathbf{Y}^{p+1,t+1} = \begin{bmatrix} y_0 \dots y_t \\ \mathbf{Y}^{p,t} \end{bmatrix},$$

Substitution von  $\bar{\mathbf{v}}^T = [y_0 \dots y_t]$  in (3.2.2.1) mit  $\mathbf{P}_\perp^{p,t} \bar{\mathbf{v}} = \bar{\mathbf{e}}^{p,t}$  führt analog zur Herleitung von (3.3.3a) schließlich auf

$$\begin{aligned} \bar{\mathbf{r}}_{p+1,t} &= \bar{\mathbf{r}}_{p,t-1} - \frac{\bar{\mathbf{e}}_{p,t} \bar{\mathbf{e}}_{p,t}^T}{\bar{\mathbf{e}}_{p,t}^T \bar{\mathbf{e}}_{p,t}} \cdot \begin{bmatrix} 0 \\ \vdots \\ 0 \\ y_0 \\ \vdots \\ y_{t-p-1} \end{bmatrix} = \\ &= \bar{\mathbf{r}}_{p,t-1} - \frac{\bar{\mathbf{e}}_{p,t} \bar{\mathbf{e}}_{p,t}^T}{\bar{\mathbf{e}}_{p,t}^T \bar{\mathbf{e}}_{p,t}} \cdot \bar{\mathbf{r}}_{p,t-1} \end{aligned} \quad (3.3.3b).$$

Die Gleichungen (3.3.3) bilden das Gerüst des sogenannten Lattice-Algorithmus, die Lattice- (Gitter-) Struktur illustriert den Datenfluß im Algorithmus, der gleichsam kreuzweise Vorwärts- und Rückwärts-Vorhersage-Fehler miteinander verknüpft. Diese Verknüpfung ist das Resultat der Anpassung des Projektionsformalismus an die spezielle Struktur der Datenmatrix. Sie bewirkt, daß wechselweise der zum bestehenden Vektorraum hinzutretende Vektor  $\bar{\mathbf{v}}$  im Vorwärtsproblem mit dem zu projizierenden Vektor  $\bar{\mathbf{y}}$  im Rückwärtsproblem identisch ist, und umgekehrt, der hinzutretende Vektor im Rückwärtsproblem dem zu projizierenden Vektor des Vorwärtsproblems gleicht.

Es sei darauf hingewiesen, daß mit (3.3.3) für jedes  $p$  jeweils ein kompletter  $t$ -Komponenten-Vektor  $\bar{\mathbf{e}}_{p+1,t}$  bzw.  $\bar{\mathbf{r}}_{p+1,t}$  berechnet

wird. Diese Darstellung eignet sich sicher nicht für die Rechnerimplementation, erlaubt aber eine interessante Interpretation des Algorithmus:

Mit den Initialisierungen  $\bar{e}_{0,t} = \bar{r}_{0,t} = \bar{y}_t$  berechnet der Algorithmus die sukzessiven orthogonalen Komplemente zu den ursprünglichen Daten  $\bar{y}_t$ . Mit wachsender Ordnung  $p$  konvergiert der Algorithmus in Bezug auf die Projektionsoperatoren. Diese ändern sich nicht mehr, wenn keine orthogonalen, außerhalb des vom Operator aufgespannten Vektorraumes liegenden, Komponenten hinzutreten, das Matrix-Vektorprodukt  $\mathbf{P}_{\perp} \mathbf{U} \bar{v}$  in (3.2.2.1a) also gleich dem Nullvektor ist. In (3.3.3) finden keine Aufdatierungen mehr statt und die optimale Anpassung eines AR-Modells in Bezug auf die Ordnung  $P$  des Problems ist dann theoretisch erreicht. Beim Rechnen mit begrenzter Genauigkeit wird dies jedoch nicht zu realisieren sein, sodaß bei Überparametrisierung die Vektorräume immer noch künstlich mit Rundungsfehlern erweitert werden. Die eindimensionalen Lattice-Algorithmen haben sich jedoch als ausgesprochen stabil gegenüber diesem Phänomen erwiesen, wozu der im nächsten Abschnitt eingeführte Wichtungsfaktor (Forgetting- oder Limited-Memory-Faktor) beizuträgt.

Die Gleichungen (3.3.3) lassen sich zu Update Relationen für die skalaren Größen mit Hilfe des Pinning-Vektors  $\bar{\pi}_t^T = [0 \dots 0 1]$  umformen

$$e_{p+1,t} = e_{p,t} - \frac{\bar{\pi}_t \bar{r}_{p,t-1} \bar{r}_{p,t-1}^T}{\bar{r}_{p,t-1}^T \bar{r}_{p,t-1}} \cdot \bar{e}_{p,t} \quad (3.3.4a)$$

$$r_{p+1,t} = r_{p,t-1} - \frac{\bar{\pi}_t \bar{e}_{p,t} \bar{e}_{p,t}^T}{\bar{e}_{p,t}^T \bar{e}_{p,t}} \cdot \bar{r}_{p,t-1} \quad (3.3.4b)$$

und stellen nunmehr ein System von Gleichungen dar, das relativ einfach auf einem Rechner implementiert werden kann, wobei die entsprechenden Relationen für die Aufdatierung in der Zeit, die die jeweils notwendige Neuberechnung der Quotienten in (3.3.4) überflüssig machen würden, noch herzuleiten wären. Darüberhinaus sind weitere Vereinfachungen von (3.3.4) möglich. Auch diese sollen in dieser Arbeit nicht diskutiert werden. Hier trennen sich die Wege und die weitere Entwicklung führt auf eine Vielzahl von möglichen Algorithmen, abhängig von der jeweils interessierenden Anwendung. Beispiele für derartige Algorithmen finden sich in Alexander (Alexander 1986), Bellanger (Bellanger, 1987), Lewis (Lewis, 1990), Lee (Lee, 1981), Carayannis (Carayannis, 1986), Bellanger (Bellanger, 1989), Friedlander (Friedlander, 1982).

### 3.4 Adaptive Formulierung für den instationären Fall

In der diskreten Signaltheorie kann die Bezeichnung "adaptiver" Filter dahingehend verstanden werden, daß dem Filterverfahren ein Algorithmus zugrunde liegt, der die relevanten Variablen rekursiv in der Zeit berechnet. Rekursiv in dem Sinne, daß eine Variable  $v_t$  unmittelbar aus  $v_{t-1}$  hervorgeht. Dies bedeutet nichts anderes, als daß Berechnungen mit der größtmöglichen zeitlichen Auflösung der auf dem diskreten Zeitgitter vorliegenden Daten durchgeführt werden können. Hier liegt abgesehen von Schwierigkeiten in der Formulierung, wie sie in Abschnitt 3.1.3 angesprochen wurden, der wesentliche Unterschied zu Block-Verarbeitungs Techniken, deren zeitliche Auflösung i.d.R. nur die Größenordnung von mehreren zehn oder gar mehreren hundert Datenwerten erreicht.

Bislang sind keinerlei Voraussetzungen über die Eigenschaften des diskreten Signals  $y_t$  in die Überlegungen eingeflossen. Auf der Basis statistischer Argumente kann man zeigen, daß die Gleichungen (3.3.4) sich im Fall eines stationären Signals weiter vereinfachen lassen. Es stellt sich heraus, daß die Quotienten in beiden Gleichungen bei stationären Signalen identisch sind. Die Diskussion

des stationären Falls findet sich in Alexander (Alexander, 1986) und Friedlander (Friedlander, 1982).

Im Zusammenhang mit dieser Arbeit interessiert speziell der Fall instationärer Daten. Mit den Initialisierungsbedingungen für die skalaren Aufdatierungs-Relationen  $e_{0,t} = r_{0,t} = y_t$ ,  $t=0, \dots, T-1$  stellen die Gleichungen (3.3.4) ein zeit- und ordnungsrekursives Verfahren zur Lösung von (3.1.2.4) mit der Bedingung  $\|e\|_2^2 \stackrel{!}{=} \min$  dar. Zur Zeit  $\tau$  basiert die Lösung auf allen Datensamples  $y_t$ ,  $t=0, \dots, \tau-1$ , zur Zeit  $t=T-1$  schließlich ist sie identisch der, durch direkte Inversion von  $Y^T Y$  gewonnenen, aus den Yule-Walker-Gleichungen (3.1.1.4).

Im Falle von Daten, die im begründeten Verdacht stehen, instationäres Verhalten zu zeigen, ist speziell dies unerwünscht. Die Lösung für  $t = \tau$  repräsentierte auf diese Weise nicht die Eigenschaften der Daten in der zeitlichen Umgebung von  $\tau$ , sondern eher ein mittleres, durchschnittliches Verhalten bis zu diesem Zeitpunkt, also im Intervall  $[0, \tau)$ .

Eine sehr einfache Methode, die Lösung auf Datenwerte in der zeitlichen Nähe von  $\tau$  zu fokussieren, ist die Einführung einer Gewichtsfunktion  $w(\tau)$ , die weiter in der Vergangenheit zurückliegende Daten mit einem abnehmenden Gewicht versieht und ihnen damit ebenso abnehmende Signifikanz für die aktuelle Lösung zuordnet. Die Gewichtsfunktion muß sich der rekursiven Struktur des Algorithmus anpassen und so liegt es auf der Hand,

$$w(\tau) = \lambda^{t-\tau} \text{ zu wählen, mit}$$

$$0 \ll \lambda < 1.$$

Multiplikation von (3.1.2.3) mit  $w(\tau)$  zeigt, daß sich durch Einführung der Gewichtsfunktion die Minimierungsaufgabe ändert, die Nebenbedingung lautet nun

$$\sum_{t=0}^{T-1} \lambda^{T-t} e_{p,t}^2 \stackrel{!}{=} \min. \quad (3.4.1).$$

Der Algorithmus sieht die Daten in der Vergangenheit gleichsam durch ein abfallendes Exponentialfenster, was dem Faktor Lambda den Namen Forgetting- oder Limited-Memory-Factor eingetragen hat. Eine Konvention zur Ermittlung der effektiven Länge dieses Exponentialfensters findet sich in Friedlander (Friedlander, 1982), der diesen Wert mit  $L = 1/\ln(\lambda)$  angibt.

### 3.5 Joint-Process Estimation

Alle bisherigen Formulierungen ergaben sich in enger Anlehnung an das Modell des autoregressiven Prozesses, bei dem ein Datenwert  $y_t$  linear von Datenwerten  $y_{t-k}$ ,  $k = 1, \dots, p$  ein und desselben Datensatzes abhing. Dieses ursprüngliche Modell läßt sich in der Richtung erweitern, daß die Daten  $y_t$  zusätzlich einen weiteren, sog. Joint-Process steuern:

$$x_t = \sum_{i=1}^R c_i y_{t-i} + e_t^x \quad (3.5.1).$$

Der dazugehörige Vorhersagefehler Filter  $\hat{c}$  löst dementsprechend:

$$e_t^x = x_t - \sum_{i=1}^R \hat{c}_i y_{t-i}.$$

Man sieht sofort, daß die Datenmatrix dieses Problems mit der Datenmatrix des bisher diskutierten Problems der Vorhersage über einen einzelnen Zeitschritt (One Step Ahead Prediction) überein-

stimmt, demzufolge stimmen auch die jeweiligen Projektionsoperatoren überein, nur die zu projizierenden Vektoren entstammen dem neuen Problem. Auf eine Herleitung, die im wesentlichen eine Wiederholung wäre, wird daher verzichtet. Die Erweiterung des Lattice-Algorithmus für den Joint-Process ergibt:

$$e^x_{p+1,t} = e^x_{p,t} - \frac{\bar{\pi}_t \bar{r}_{p,t-1} \bar{r}_{p,t-1}^T}{\bar{r}_{p,t-1}^T \bar{r}_{p,t-1}} \cdot \bar{e}^x_{p,t} \quad (3.5.2).$$

Diese Erweiterung ist von beachtlicher praktischer Bedeutung. Sie ermöglicht u.a. die Behandlung des Multiplenproblems in seismischen Daten. Während der einfache Lattice-Algorithmus nur ein Lösungsverfahren für das Vorhersage-Problem über einen einzelnen Zeitschritt darstellt, ist es möglich, mit der Joint-Process Erweiterung durch Substitution von  $x_t = y_{t+\tau}$ , auch das "Gapped-Prediction" Problem zu lösen, bei dem die Vorhersage über eine Distanz mehrerer Zeitschritte versucht wird. In (3.5.1) wird der, das Echo oder die Multiple generierende Prozess, als unabhängig vom Prozess des Nutzsymbols im Hinblick auf die Parameter interpretiert, der Echoprozess wird jedoch von den Daten des Nutzsymbols  $y(t)$  angetrieben.

Es existieren neben der Anwendung für die "Echo"-Unterdrückung eine Vielzahl von Anwendungen, bei denen  $x(t)$  und  $y(t)$  voneinander verschiedene Signalgenese-Prozesse darstellen, die sich jedoch jeweils gegenseitig beeinflussen. Einen Überblick über mögliche und bestehende Anwendungen von der Antennentechnik bis hin zur Elektro-Kardiographie vermittelt Widrow et.al. (Widrow 1975).

### 3.6 Mehrdimensionale Formulierung des Lattice-Algorithmus

Die in den Abschnitten 3.3 bis 3.5 skizzierte Herleitung des Lattice-Algorithmus für ein eindimensionales, besser einkanaliges, Signal

ist formal einfach erweiterbar auf ein mehrkanaliges oder mehrdimensionales Signal. Um die Betrachtungen nicht zu sehr zu verallgemeinern, soll im folgenden die typische Situation der Akquisition mariner seismischer Daten als Modell dienen. Die Daten eines Experimentes, eines Schusses, werden an verschiedenen räumlichen Positionen (Offsets) mehrkanalig aufgezeichnet. Diese Aufzeichnung liefert eine 2-dimensionale Darstellung in Raum und Zeit. Die an die Mehrkanalsituation angepasste Formulierung des Ausgangsproblems gleicht der Prewindowed Form in (3.1.2.4)

$$\begin{aligned}
 [\bar{e}_0 \dots \bar{e}_T] &= [\bar{y}_0 \dots \bar{y}_T] - \\
 & \cdot [A_1 \dots A_p] \begin{bmatrix} \bar{0} & \bar{y}_0 & \dots & \bar{y}_{p-1} & \dots & \bar{y}_{T-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \bar{0} & \dots & \bar{0} & \bar{y}_0 & \dots & \bar{y}_{T-p} \end{bmatrix} \quad (3.6.1)
 \end{aligned}$$

mit der Nebenbedingung

$$\text{tr} \left( \sum_{t=0}^T \bar{e}_t \bar{e}_t^T \right) \stackrel{!}{=} \min.$$

Die Elemente der Datenmatrix und die Komponenten des Datenvektors sind hier ihrerseits wiederum Vektoren, deren Komponenten von den Datenwerten der einzelnen Kanäle  $m = 0 \dots K-1$  zur Zeit  $t = \text{const.}$  gebildet werden. Gleiches gilt für die Komponenten des Vorhersage-Fehler-Vektors  $\bar{e}$ . Die skalaren Koeffizienten  $a_i$  des AR-Modells in (3.1.2.4) sind zu  $p \times p$ -Matritzen  $A_i$  gewachsen.

Die Lattice-Herleitung in 3.3 bis 3.5 ist völlig unabhängig von der "Natur" der Daten und mit den Definitionen

$$\mathbf{R}^{p,t-1} := \begin{bmatrix} 0 \\ \bar{\mathbf{r}}_{p,0}^T \\ \bar{\mathbf{r}}_{p,1}^T \\ \cdot \\ \cdot \\ \bar{\mathbf{r}}_{p,t-1}^T \end{bmatrix}$$

$$\mathbf{E}^{p,t} := \begin{bmatrix} \bar{\mathbf{e}}_{p,0}^T \\ \bar{\mathbf{e}}_{p,1}^T \\ \cdot \\ \cdot \\ \bar{\mathbf{e}}_{p,t}^T \end{bmatrix}$$

$$\mathbf{E}_x^{p,t} := \begin{bmatrix} \bar{\mathbf{e}}_{p,0}^x{}^T \\ \bar{\mathbf{e}}_{p,1}^x{}^T \\ \cdot \\ \cdot \\ \bar{\mathbf{e}}_{p,t}^x{}^T \end{bmatrix}$$

können die Lattice-Gleichungen analog zu (3.3.4) und (3.5.2) sofort angegeben werden.

$$\bar{\mathbf{e}}_{p+1,t}^T = \bar{\mathbf{e}}_{p,t}^T - \frac{\bar{\pi}_t \mathbf{R}^{p,t-1} \mathbf{R}^{p,t-1T}}{\mathbf{R}^{p,t-1T} \mathbf{R}^{p,t-1}} \cdot \mathbf{E}^{p,t} \quad (3.6.2a)$$

$$\bar{\mathbf{r}}_{p+1,t}^T = \bar{\mathbf{r}}_{p,t-1}^T - \frac{\bar{\pi}_t \mathbf{E}^{p,t} \mathbf{E}^{p,tT}}{\mathbf{E}^{p,tT} \mathbf{E}^{p,t}} \cdot \mathbf{R}^{p,t} \quad (3.6.2b)$$

$$\bar{\mathbf{e}}_{p+1,t}^{xT} = \bar{\mathbf{e}}_{p,t}^{xT} - \frac{\bar{\pi}_t \mathbf{R}^{p,t-1} \mathbf{R}^{p,t-1T}}{\mathbf{R}^{p,t-1T} \mathbf{R}^{p,t-1}} \cdot \mathbf{E}_x^{p,t} \quad (3.6.2c)$$

Der Skalar im Nenner der Quotienten in (3.3.4) und (3.5.2) ist in (3.6.2) zur inversen Matrix des Typs  $\mathbf{C}^T \mathbf{C}$  mutiert. Man kann leicht zeigen, daß unter der Voraussetzung, daß  $\mathbf{C}$  vollen Rang (bei  $K$  Kanälen also  $\text{Rg}(\mathbf{C}) = K$ ) hat,  $\mathbf{C}^T \mathbf{C}$  positiv definit, also invertierbar ist.

### 3.6.1 Direkte Lösung : Rang-1 Aufdatierungen

Die rekursive Berechnung der Inversen vom Typ  $(\mathbf{C}^T \mathbf{C})^{-1}$  aus (3.6.2) ist bei vollem Rang und hinreichend guter Kondition mit Hilfe des Sherman-Morrison-Lemmas (Hager, 1988)

$$(\mathbf{C} + \bar{\mathbf{v}} \bar{\mathbf{v}}^T)^{-1} = \left[ \frac{\mathbf{C}^{-1} - \mathbf{C}^{-1} \bar{\mathbf{v}} \bar{\mathbf{v}}^T \mathbf{C}^{-1}}{\bar{\mathbf{v}}^T \mathbf{C}^{-1} \bar{\mathbf{v}}} \right]$$

sehr effizient möglich. Die Sherman-Morrison-Formel baut die Inverse rekursiv aus Rang-1 Matrizen auf. Die Inverse einer  $p \times p$  Matrix könnte nach  $p$  Schritten theoretisch selbst vollen Rang haben. Ein schneller Algorithmus für den Mehrkanal-Fall scheint auf den ersten Blick also grundsätzlich machbar, solange für die ersten  $p-1$ -Schritte oder, im praktischen Fall, bis eine befriedigend stabile Inverse entstanden ist, eine generalisierte Inverse berechnet wird.

### 3.6.2 Konditionsproblem, Rang-Entscheidungs-Problem

Zu Beginn dieser Arbeit wurde tatsächlich die oben skizzierte Strategie verfolgt. Im Bewußtsein der Tatsache, daß Matrizen des Typs  $C^T C$  in vielen Fällen schlechte Kondition haben und demzufolge die mit endlicher Genauigkeit berechneten Inversen mit beliebig großen Fehlern behaftet sind, standen am Anfang numerische Experimente mit dem Ziel, die Entwicklung der Kondition auf verschiedenen synthetischen Datensätzen zu beobachten, um so zu einer Regularisierungsstrategie zu gelangen.

Die zur euklidischen Vektornorm kompatible Matrixnorm (Spektralnorm)

$A \in \mathbb{R}^{n,n}$ ,  $\|A\|_2 := (\max_{j=1, \dots, n} (\sigma_j))$  mit

$\sigma_j :=$  Singulärwerte von  $A$

und die darauf definierte Kondition

$$\text{cond}_2(A) = \|A\|_2 * \|A^{-1}\|_2 = \sigma_{\max} / \sigma_{\min},$$

wobei  $\text{cond}_2(A) \geq 1$ ,

bietet u.a. ein Maß für die Abschätzung des relativen Fehlers  $\|\Delta \bar{x}\| / \|\bar{x}\|$  in der Lösung des Systems  $A \bar{x} = \bar{b}$  für eine gestörte Matrix  $(A + \Delta A)$  (Voß, 1987)

$$\frac{\|\Delta \bar{x}\|}{\|\bar{x}\|} \leq \frac{\text{cond}(A)}{1 - \text{cond}(A) \cdot \frac{\|\Delta A\|}{\|A\|}} \cdot \frac{\|\Delta A\|}{\|A\|}.$$

Man erkennt, daß der relative Fehler  $\|\Delta \bar{x}\| / \|\bar{x}\|$  in der Lösung für  $\text{cond}(A) \gg 1$  selbst für kleine relative Fehler in der Matrix  $A$  beliebig groß werden kann. In der numerischen Praxis heißt dies, daß die Lösung  $\bar{x}$  bei schlechter Kondition der Matrix  $A$ , die in jedem Fall mit dem Fehler der Maschinengenauigkeit ( dem

Rundungsfehler beim Rechnen mit endlicher Genauigkeit) behaftet ist, absolut unbrauchbar sein kann.

Eine Möglichkeit, die Entwicklung der Kondition der Matrizen  $E^{p,t}$ ,  $R^{p,t-1}$ ,  $E_x^{p,t}$  in (3.6.2) zu beobachten und gleichzeitig durch Regularisierung die Lösung zu stabilisieren, besteht darin, die Singulärwert-Zerlegung der betreffenden Matrix zu berechnen (Baumeister, 1987).

Sei  $A$  positiv definit und

$$V\Sigma V^T = A, \quad A \in \mathbb{R}^{n,n}$$

die Singulärwert-Zerlegung der Matrix  $A$

mit  $V^T V = V V^T = \mathbf{1}$ , orthogonal

und  $\Sigma = \text{diag}(s_1, \dots, s_n)$

die Diagonalmatrix der Singulären Werte von  $A$ .

Für die Berechnung einer effektiven Pseudoinversen  $A^+$  stehen zwei Verfahren zur Verfügung:

1) Rangreduzierung (Baumeister, 1987, Voß, 1987)

$$A^+ = V^T \Sigma^+ V$$

mit

$$\Sigma^+ = \begin{cases} \text{diag}(1/s_i) & \text{für } s_i > d \\ 0 & \text{für } s_i \leq d \end{cases}$$

Hier werden kleine Singuläre Werte unterhalb einer willkürlichen Schwelle  $d$  abgeschnitten, der Rang der Pseudoinversen reduziert.

Wenn  $s_m$  der kleinste Singuläre Wert größer  $d$  ist, verbessert sich die Kondition der Matrix  $\mathbf{A}$  auf

$$s_{\max}/s_m.$$

2) Regularisierung nach Tichonow (Baumeister, 1987, Voß, 1987)

$$\Sigma^+ = \text{diag} (1/\sqrt{s_i^2 + \alpha}) \text{ mit } \alpha > 0.$$

Bei der Regularisierung nach Tichonow wird der Einfluß kleiner Singulärer Werte bedämpft, die Kondition verbessert sich entsprechend

$$\sqrt{\frac{s_{\max}^2 + \alpha}{s_{\min}^2 + \alpha}}.$$

In beiden Fällen wird ein Verfahrensfehler in Kauf genommen zugunsten einer stabilen Lösung, die weniger empfindlich gegenüber Datenfehlern in  $\mathbf{A}$  ist.

Die Abbildungen 3.6.1 und 3.6.2 zeigen die Entwicklung der Kondition der Matrix  $\mathbf{E}^{p,tT} \mathbf{E}^{p,t}$  für verschiedene drei- und vierkanalige Signale.

Die Abbildung 3.6.1.a zeigt ein dreikanaliges AR-Signal, das gemäß

$$\bar{y}_t = \sum_{k=1}^4 \mathbf{A}_k \bar{y}_{t-k} + \bar{e}_t, \quad e_{i,t} \neq e_{j,t} \text{ f. } \#i$$

erzeugt wurde, wobei die Elemente der Koeffizientenmatrizen alle identisch 0.082 gesetzt wurden, und das den Prozess antreibende Signal  $\tilde{e}_t$  aus inkohärentem, d.h. für alle Kanäle verschiedenem, gaussverteilterm Rauschen mit Mittelwert Null und Varianz Eins besteht.

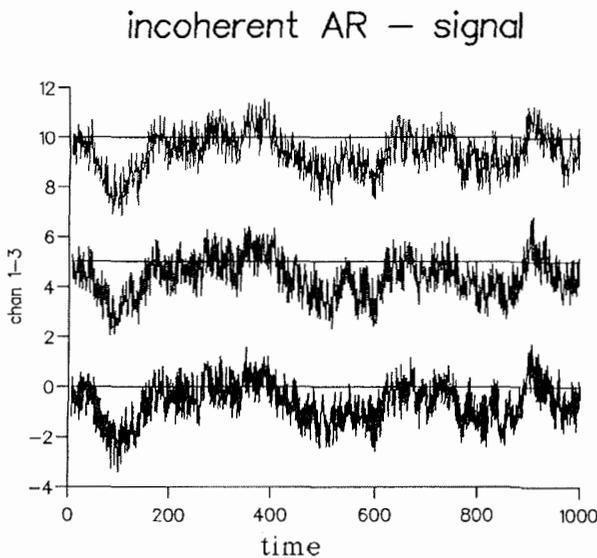


Abb. 3.6.1a

Die Abbildungen 3.6.1b - f zeigen die Entwicklung des Logarithmus der Kondition im Lattice-Algorithmus für Filterordnungen 3 bis 6. Bis hin zur Ordnung 5 ist allen Darstellungen gemeinsam, daß die Kondition zum Beginn der Rekursion ( $t=0$ ) schlechter als  $10^{10}$  ist, dann aber exponentiell besser wird, und solange der Filter nicht überparametrisiert ist, also bis zur Ordnung Vier, gegen einen Wert nahe Eins konvergiert.

### Condition Number Order 2

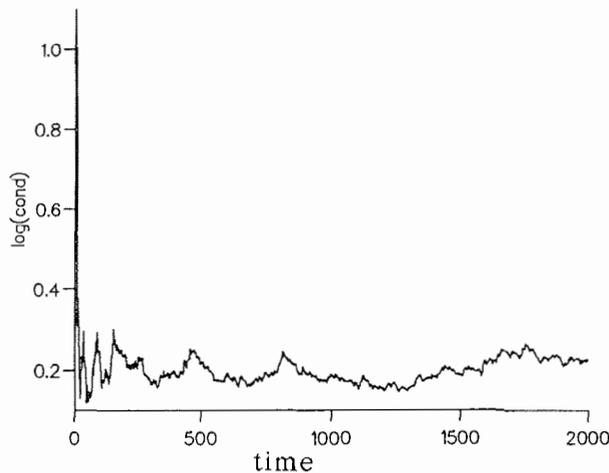


Abb. 3.6.1b

Ab Filterordnung Fünf macht sich die Überparametrisierung bemerkbar, ab der Ordnung Sechs nimmt die Kondition nicht mehr ab, sondern bleibt schlechter als  $10^{12}$ .

Die schlechte  
Kondition für Werte  
nahe  $t=0$  kennzeichnet  
die Einschwingphase  
des Filters. Aufgrund  
der Tatsache, daß die  
Matrix  $\mathbf{E}^{p,tT} \mathbf{E}^{p,t}$  re-  
kursiv aus Rang-1-  
Matritzen aufgebaut  
wird, besitzt sie zu-  
nächst nicht vollen  
Rang. Die kleinen  
Singulären Werte  
werden durch Run-  
dungsfehler hervorge-  
rufen.

### Condition Number Order 3

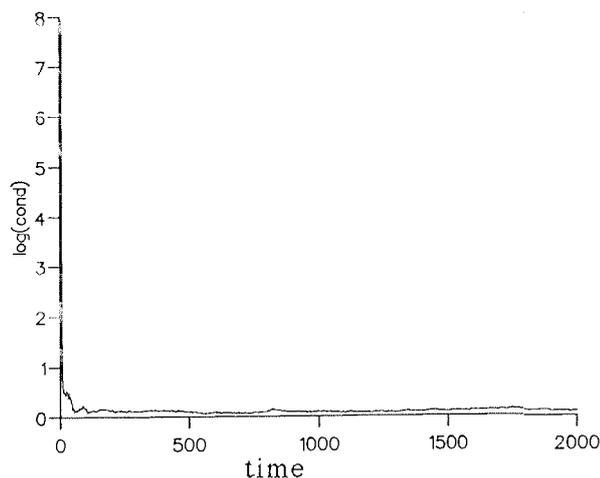


Abb. 3.6.1c

Zu diesem Beispiel ist  
zu bemerken, daß das  
synthetische Signal  
stationär ist und  
dementsprechend der  
Filter im stationären  
Modus, also mit einem  
Gewichts-Faktor  $\lambda=1$   
betrieben wurde. Dies  
hat zur Folge, daß im  
Bereich der Über-  
parametrisierung, für  
Ordnungen größer  
Vier, in der nächst hö-  
heren Stufe des Filters nur noch Rundungsfehler projiziert werden,  
die auch hier für die kleinen Singulären Werte der Fehler-Ko-  
varianz-Matritzen  $\mathbf{E}^{p,tT} \mathbf{E}^{p,t}$  verantwortlich sind.

### Condition Number Order 4

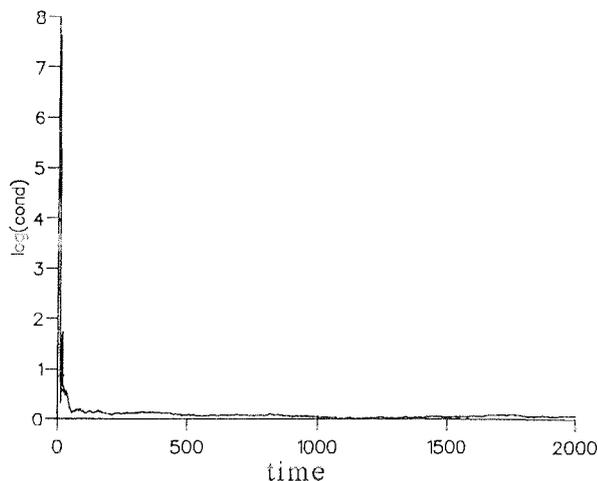


Abb. 3.6.1d

Generell beobachtet man eine Abnahme bzw. eine Verbesserung der Kondition im eingeschwungenen Zustand bis zur optimalen Filterordnung.

Betrachtet man die Norm des Vorhersagefehlers für die Filterordnungen Eins bis Sechs, erkennt man, daß selbst für die Filterordnung Sechs

der sogenannte Innovations-Fehler gegen ein Minimum konvergiert. Dies ist eine Folge der hier durch Rangreduzierung vorgenommenen Regularisierung, mit der die Kondition von  $E^{p,tT}E^{p,t}$  auf  $10^6$  beschränkt wurde. Die Beschränkung auf  $10^6$  ist vollkommen willkürlich und durch keine a-priori Information gestützt, sie kann nur durch den praktischen Erfolg gerechtfertigt werden.

Condition Number Order 5

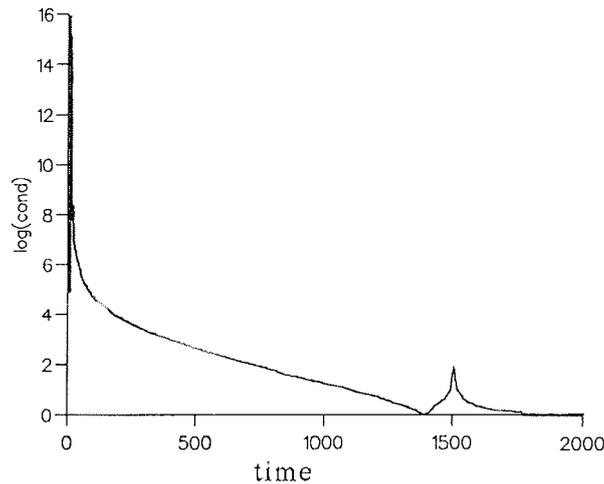


Abb.3.6.1e

Condition Number Order 6

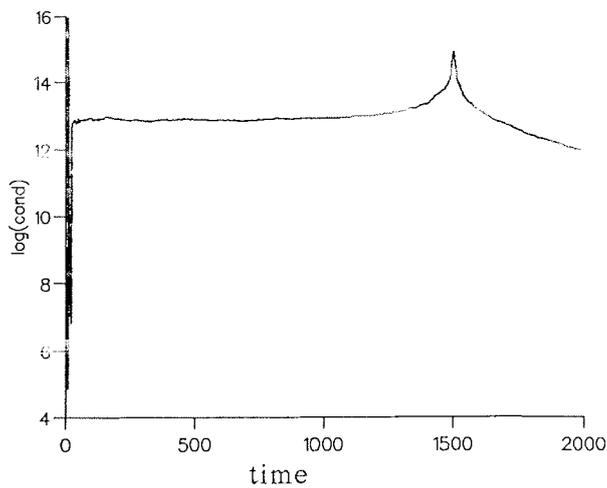


Abb. 3.6.1f

### 4-channel seismogram

Abbildung 3.6.2a zeigt ein vierkanaliges, synthetisches Seismogramm, das mit einem Finite-Differenzen-Algorithmus von einem einfachen Zweischichten-Modell erzeugt wurde. Die Ränder des Modells genügen der

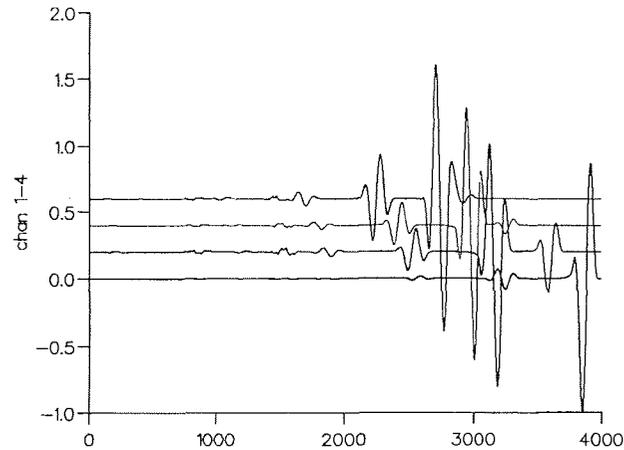


Abb. 3.6.2a

Sommerfeld-Ausstrahlungsbedingung für hyperbolische Differentialgleichungen (Miller, 1981, Orlanski, 1976). Erkennbar sind direkt gelaufene Welle, eine Primärreflexion und ein Zug von Multiplen. Die Ränder des Modells sind mit der Orlanski-Ausstrahlungsbedingung nicht völlig reflexionsfrei, daher ist eine Reflexion vom Boden des Modells bei  $t = 1500$  sichtbar. Die Zeitachse des Seismogramms ist umgedreht, der Lattice-Algorithmus läuft, aus Gründen, die später erläutert werden sollen, rückwärts in der Zeit.

Die Abbildungen 3.6.2b - d zeigen die Entwicklung der Kondition für die Filterordnungen Drei bis Fünf.

Zunächst ist festzustellen, daß hier ein völlig anderes Verhalten als auf dem synthetischen AR-Modell zu beobachten ist. Die Kondition von  $E^{p,tT}E^{p,t}$  konvergiert weder für konstante Ordnung in der Zeit, noch ist eine Abnahme der Kondition mit zunehmender Ordnung zu verzeichnen. Insgesamt ist ein eher erratic Verhalten zu konstatieren.

### Condition Number Order 3

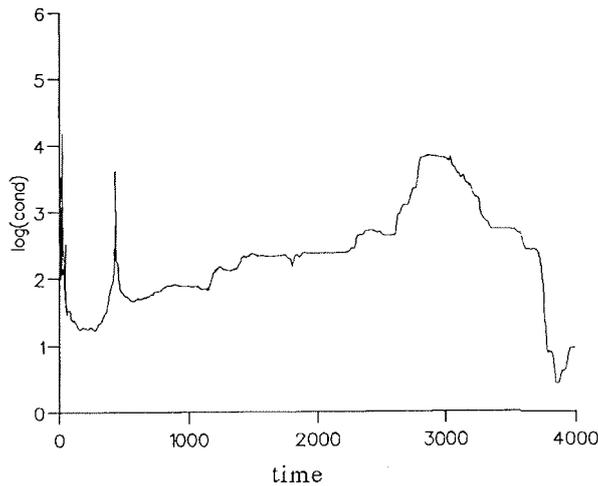


Abb. 3.6.2b

Alle Versuche, dieses Verhalten mit anderen synthetischen Signalen zu modellieren, die beispielsweise den transienten Charakter seismischer Daten nachahmen oder in einem anderen Experiment die periodischen Anteile hervorheben bzw. den quasi eindimensionalen (kollinearen) Charakter für späte Zeiten, wo Common-Shot-Point-

### Condition Number Order 4

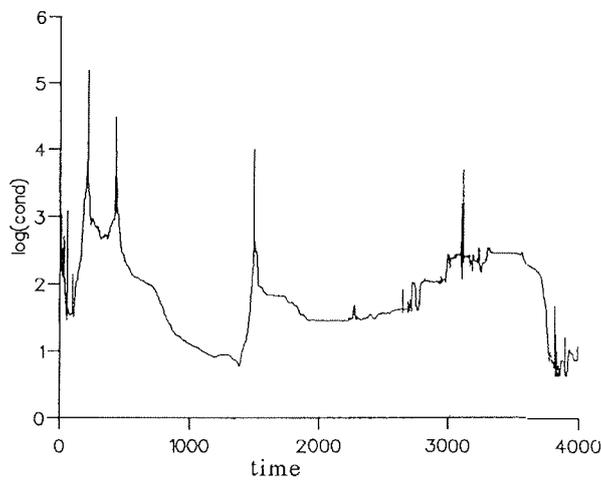


Abb. 3.6.2c

(CSP) Ensembles praktisch keinen Move-Out mehr aufweisen, haben zu keinem Ergebnis geführt, das eine Klassifizierung des Problems erlaubt hätte.

Die Regularisierung sowohl durch Rangreduzierung mit verschiedenen Schwellwerten als auch, nach Tichonow, mit verschiedenen  $\alpha$  zeigt keinen nachhaltigen Erfolg. Im Gegenteil, die oben angesprochenen Verfahrensfehler, die durch die Regularisierung entstehen, scheinen zur Divergenz noch beizutragen.

### Condition Number Order 5

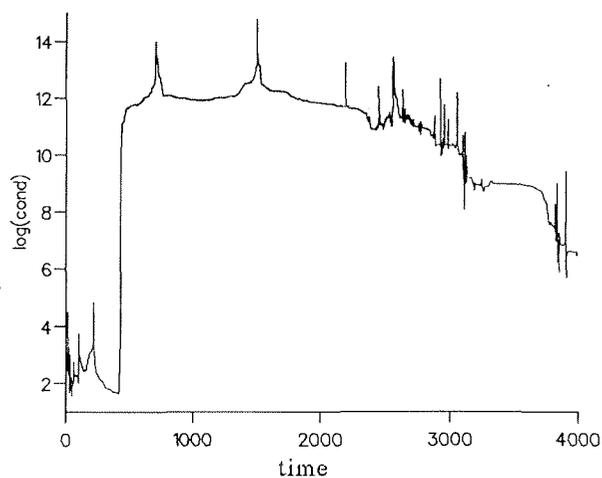


Abb. 3.6.2d

Alle Erfahrungen deuten jedoch darauf hin, daß seismische CSP Daten im Zusammenhang mit dem beschriebenen Mehrkanal-Algorithmus auf folgendes Dilemma führen:

Neben dem Problem schlechter Kondition existiert ein Rang-Entscheidungs-Problem. Aufgrund der geometrischen Ausbreitungsbedingungen einer Kugelwelle zeigen Einsätze zu späteren Zeiten im x-t CSP Seismogramm in Fall söhlicher Schichten kaum noch Move-Out. In einem solchen Zeitbereich enthält die seismische Datenmatrix von Kanal zu Kanal im Wesentlichen nur redundante Information, die Spaltenvektoren sind hochgradig kollinear. Das gilt entsprechend für die Fehler-Matrizen  $\mathbf{E}$  und  $\mathbf{R}$  jeder Ordnung. In diesem Zeitbereich ist es sinnvoll, durch Rangreduzierung zu regularisieren. Ohne a-priori Information über den "wahren" Rang ist dies jedoch praktisch nicht möglich (vgl. Baumeister 1987), zumal dieser sich in der Zeit ändert. In der Folge werden "falsche" Projektionsoperatoren berechnet, die auf unter- oder überdimensionierten Vektorräumen basieren. Im letzteren Fall passiert phänomenologisch dasselbe wie im Fall der Überparametrisierung: Rundungsfehler bilden die Basis der Projektionen. Im ersteren Fall

können die Verfahrensfehler die Projektion verfälschen und, wenn sie hinreichend groß sind, die Konvergenz verhindern.

Die oben geführte Diskussion dieses Phänomens wird im Abschnitt 3.7 noch einmal aufgenommen mit einer mehr mathematischen Argumentation.

### 3.7 QR-Zerlegung: der Algorithmus von Lewis

Im vorangegangenen Abschnitt wurden die Probleme diskutiert, die sich aus der direkten Erweiterung des eindimensionalen zum mehrdimensionalen adaptiven Lattice-Algorithmus ergeben. Sowohl das Konditionsproblem als auch das Rang-Entscheidungsproblem sind numerischer Natur, d.h. eine Folge des Rechnens mit begrenzter Genauigkeit, und nicht etwa die Folge schlechter, sprich veräuschter Daten. Am Beispiel des synthetischen Seismogramms im letzten Abschnitt ist dies deutlich geworden.

Einen entscheidenden Beitrag zur Lösung des Konditionsproblems liefert der Algorithmus von Lewis (Lewis 1990). Lewis zeigt einen Weg auf, die direkte Inversion der schlecht konditionierten Matrix vom Typ  $C^T C$  zu vermeiden.

#### 3.7.1 QR-Zerlegung und Kondition

Das Verfahren beruht auf der sog. QR-Zerlegung (vergl. Ciarlet 1989, Hager 1988) einer beliebigen reellen Matrix  $C \in \mathbb{R}^{m,n}$ ,  $m > n$

$$C = Q \begin{bmatrix} R \\ 0 \end{bmatrix}$$

mit  $Q^T Q = 1$ , orthonormal

und  $R \in \mathbb{R}^{n,n}$ : obere Dreiecksmatrix.

Unter der zusätzlichen Voraussetzung, daß  $C$  vollen Rang hat, ist  $R$  invertierbar.

Definiert man eine Matrix  $M \in \mathbb{R}^{n,n}$ , die die ersten  $n$  Zeilen der Matrix  $Q$  enthält

$$M = [ \mathbf{1} \ \mathbf{0} ] Q ,$$

so ist

$$C = M R.$$

Für den zur Matrix  $C$  gehörenden Projektionsoperator gilt:

$$\begin{aligned} P_{\{C\}} &= C(C^T C)^{-1} C^T \\ &= M R (R^T M^T M R)^{-1} R^T M^T \\ &= M R (R^T R)^{-1} R^T M^T \\ &= M R R^{-1} R^{-1 T} R^T M^T \\ &= M M^T. \end{aligned}$$

Hierbei wurde von der Voraussetzung  $R$  invertierbar, also  $\text{Rg}(C) = n$  Gebrauch gemacht. Wie man schon an

$$M = C R^{-1}$$

sehen kann, spannt die orthonormale Matrix  $M$  denselben Vektorraum auf wie  $C$  und der Ausdruck für den Projektionsoperator über  $\{C\}$  vereinfacht sich zu

$$P_{\{C\}} = M M^T.$$

Die Kondition der Matrix  $C$  wird durch die orthonormale Transformation  $M^T C = R$  nicht verschlechtert.

Mit bekannter QR- bzw. MR-Zerlegung der Matrizen  $\mathbf{R}^{p,t-1}$ ,  $\mathbf{E}^{p,t}$  und  $\mathbf{E}_x^{p,t}$  lauten die Lattice-Gleichungen (3.6.2)

$$\bar{\mathbf{e}}_{p+1,t}^T = \bar{\mathbf{e}}_{p,t}^T - \pi_t^T \mathbf{M}_R^{p,t} \mathbf{M}_R^{p,t^T} \mathbf{E}^{p,t} \quad (3.7.1a)$$

$$\bar{\mathbf{r}}_{p+1,t}^T = \bar{\mathbf{r}}_{p,t-1}^T - \pi_t^T \mathbf{M}_E^{p,t} \mathbf{M}_E^{p,t^T} \mathbf{R}^{p,t} \quad (3.7.1b)$$

$$\bar{\mathbf{e}}_x^{p+1,t} = \bar{\mathbf{e}}_x^{p,t} - \pi_t^T \mathbf{M}_{E_x}^{p,t} \mathbf{M}_{E_x}^{p,t^T} \mathbf{E}_x^{p,t} \quad (3.7.1c)$$

Die rekursive Berechnung der Transformations-Matrizen  $\mathbf{M}$  und der oberen Dreiecksmatrizen  $\mathbf{R}$  ist mit Givens-Rotationen (vergl. bspw. Ciarlet 1989, Hager 1988) möglich. Die Details zweier sehr effizienter adaptiver Lattice-Algorithmen, die aus diesem Prinzip abgeleitet sind und auf Givens- bzw. Fast-Givens-Rotationen basieren, sind in Lewis (Lewis 1990) entwickelt.

Durch das oben angedeutete Verfahren wird bereits die explizite Berechnung der Kovarianzmatrix  $\mathbf{C}^T \mathbf{C}$  vermieden, die i.d.R. von noch erheblich schlechterer Kondition ist als die Matrix  $\mathbf{C}$  selbst. Durch die Transformation wird die Kondition nicht verschlechtert. In der Praxis ist der QR-basierte Algorithmus von Lewis erheblich stabiler als die direkte Lösung.

Voraussetzung für ein solches Verfahren ist die Invertierbarkeit von  $\mathbf{R}$ , die letztlich davon abhängig ist, daß  $\mathbf{C}$  vollen Rang hat. Im vorangegangenen Abschnitt wurde dargestellt, daß dies bei seismischen Daten nicht vorausgesetzt werden kann.

Auch hier hat man allerdings zu differenzieren. Reale seismische Daten sind mit Meßfehlern behaftet, die von statistischen Rauschmodellen recht gut beschrieben werden. Dieses Rauschen ist auf den einzelnen Kanälen der seismischen Aufzeichnung inkohä-

rent verteilt und erzeugt eine Datenmatrix, die in der Tat vollen Rang hat. Das Spektrum einer solchen Datenmatrix enthält allerdings dann auch (relativ) sehr kleine Singuläre Werte, mit anderen Worten, schon die ursprüngliche Datenmatrix ist u.U. von schlechter Kondition und wenn der QR-Formalismus diese auch nicht zusätzlich verschlechtert, können auch hier unbrauchbare Lösungen das Resultat sein.

Dies soll zur argumentativen Vorbereitung des nächsten Abschnitts hier noch einer etwas formaleren Betrachtung unterzogen werden.

Sei  $\mathbf{C} \in \mathbb{R}^{m,n}$ ,  $m > n$ ,  $\text{Rg}(\mathbf{C}) < n$ , die Matrix der idealen, rauschfreien, exakten Daten,  $\mathbf{F} \in \mathbb{R}^{m,n}$ ,  $\text{Rg}(\mathbf{F}) = n$ , die Matrix der inkohärenten Meßfehler und entsprechend  $\mathbf{D} = \mathbf{C} + \mathbf{F}$  die mit den Meßfehlern behaftete Matrix der Daten. Zusätzliche Voraussetzung sei hier, daß die Meßfehler und Daten unabhängig voneinander bzw. unkorreliert sein sollen. Übertragen in die Terminologie der linearen Algebra heißt dies, daß das Rauschen und die exakten Daten jeweils in Vektorräumen leben, die zueinander orthogonale Basen haben.

Mit dieser Voraussetzung verschwinden die gemischten Terme in

$$\begin{aligned} \mathbf{D}^T \mathbf{D} &= \mathbf{C}^T \mathbf{C} + \mathbf{C}^T \mathbf{F} + (\mathbf{C}^T \mathbf{F})^T + \mathbf{F}^T \mathbf{F} \\ &= \mathbf{C}^T \mathbf{C} + \mathbf{F}^T \mathbf{F} \end{aligned}$$

und mit der Voraussetzung, daß das Rauschen selbst inkohärent ist, sind entsprechend die Spaltenvektoren von  $\mathbf{F}$  einander orthogonal, die Matrix  $\mathbf{F}^T \mathbf{F}$  ist eine Diagonalmatrix. Eine weitere Vereinfachung ergibt sich wenn das Rauschen eine konstante Leistung aufweist. Die Skalarprodukte der Spaltenvektoren von  $\mathbf{F}$  sind dann näherungsweise gleich,  $\langle \mathbf{f}^i, \mathbf{f}^i \rangle \approx \alpha^2$ , und damit ist

$$\mathbf{F}^T \mathbf{F} \approx \alpha^2 \mathbf{1}, \quad (\mathbf{1} : \text{Einheitsmatrix})$$

und mit dieser Näherung

$$\mathbf{D}^T \mathbf{D} \approx \mathbf{C}^T \mathbf{C} + \alpha^2 \mathbf{1}.$$

Die reelle Matrix  $\mathbf{D}$  hat nach Voraussetzung vollen Rang, somit ist  $\mathbf{D}^T\mathbf{D}$  pos.def. und diagonalisierbar.

Sei  $\mathbf{X}$  die Matrix der Eigenvektoren von  $\mathbf{D}^T\mathbf{D}$  mit

$$\mathbf{X}^T\mathbf{D}^T\mathbf{D}\mathbf{X} = \text{diag}(\lambda_i), i=1,\dots,n,$$

dann gilt

$$\begin{aligned} \text{diag}(\lambda_i) &= \mathbf{X}^T[\mathbf{C}^T\mathbf{C} + \alpha^2 \mathbf{1}]\mathbf{X} \\ &= \mathbf{X}^T\mathbf{C}^T\mathbf{C}\mathbf{X} + \alpha^2 \mathbf{X}^T\mathbf{X} \\ &= \mathbf{X}^T\mathbf{C}^T\mathbf{C}\mathbf{X} + \alpha^2 \mathbf{1}. \end{aligned}$$

Da beide Summanden Diagonalmatrizen sein müssen, ist auch  $\mathbf{X}^T\mathbf{C}^T\mathbf{C}\mathbf{X}$  Diagonalmatrix und mit  $\text{Rg}(\mathbf{C}) = p < n$

$$\text{diag}(\lambda_i) = \text{diag}(g_i) + \alpha^2 \mathbf{1}, \quad g_i \begin{cases} > 0, & i = 1, \dots, p < n \\ = 0, & i = p+1, \dots, n \end{cases}$$

(3.7.1.1).

Die Eigenwerte von  $\mathbf{D}^T\mathbf{D}$  sind die Singulären Werte von  $\mathbf{D}$ . Die Matrix  $\mathbf{C}^T\mathbf{C}$  ist positiv, dementsprechend gilt  $g_k \geq 0$  und dementsprechend  $\lambda_i \geq \alpha^2$ ,  $i=1,\dots,n$ .

Die untere Schranke für den kleinsten Singulären Wert wird also von der Leistung des Rauschens gebildet. Mit der Definition der Kondition,  $\text{cond}_2(\mathbf{D}) = \lambda_{\max}/\lambda_{\min}$  macht man sich sofort klar, daß die Gutartigkeit der Daten gegenüber dem Algorithmus, bei nicht vollem Rang der Datenmatrix, vom Rauschen in den Daten abhängig ist. Rauschen stabilisiert den Algorithmus in diesem Sinn, und die

Ähnlichkeit der Form von (3.7.1.1) mit der der Tichonow-Regularisierung aus Abschnitt 3.6.1 ist natürlich nicht zufällig.

In Abschnitt 4 werden Beispiele vorgestellt, die das Verhalten des Algorithmus auf synthetischen, rauschfreien Daten illustrieren. Ein Grund, weshalb der Algorithmus auf synthetischen Daten weniger eindrucksvolle Ergebnisse liefert, ist oben genannt.

Die Kollinearitäten zu späten Zeiten im CSP-Seismogramm sorgen dafür, daß die Matrix der Daten in diesem Bereich nicht den vollen Rang hat. Der Algorithmus von Lewis in der ursprünglichen Form ist demgegenüber besonders empfindlich, da er von der Invertierbarkeit der Datenmatrix noch genauso abhängig ist wie die direkte Lösung. Der nächste Abschnitt ist geeigneten Desensibilisierungsmaßnahmen gewidmet.

### 3.7.2 Rangentscheidung mit Hilfe der KL-Transformation

Im vorherigen Abschnitt wurde darauf hingewiesen, daß der Algorithmus von Lewis beschränkt ist darauf, daß die aus der QR-Zerlegung resultierende Matrix  $R$  invertierbar ist. Ohne weitere Vorkehrungen ist dies unmittelbar davon abhängig, daß die Matrix seismischer Daten vollen Rang hat und, da der Algorithmus diese Matrix rekursiv aus ihren Zeilenvektoren aufbaut, daß ebenso jede Untermatrix  $S \in R^{p,m}$ ,  $p < m$  diese Eigenschaft hat.

Es ist nun allerdings auch möglich, eine QR-Zerlegung für die Matrix  $C \in R^{m,n}$ ,  $m > n$ ,  $\text{Rg}(C) < n$  zu finden, deren Transformationsmatrix  $Q$  ebenfalls den gesamten Spaltenraum  $\{C\}$  aufspannt, auch wenn  $C$  nicht vollen Rang hat. Erreicht wird dies durch Spaltenpivotierung (vgl. Hager 1988). Wird die Zerlegung beispielsweise durch Housholder-Transformationen  $H_i$  berechnet, gilt

$$R = H_n \dots H_2 H_1 C P_1 P_2 \dots P_{n-1},$$

wobei im  $i$ -ten Schritt  $P_i$  den ersten Spaltenvektor der verbleibenden Submatrix vertauscht mit dem Spaltenvektor, der die größte Norm von allen verbleibenden Spaltenvektoren besitzt.

Sei  $\Pi = P_1 P_2 \dots P_{n-1}$  die (Spalten-) Permutationsmatrix. Für die Zerlegung gilt:

$$\begin{aligned} C\Pi &= [Q_1 \ Q_2] \begin{bmatrix} R_1 & R_2 \\ 0 & 0 \end{bmatrix} \\ &= [Q_1 R_1 \ Q_1 R_2] \end{aligned} \quad (3.7.2.1),$$

wobei hier  $R_1$  eine invertierbare, quadratische obere Dreiecksmatrix ist und mit einfachen Argumenten gezeigt werden kann (Hager 1988), daß  $Q_1 \{C\}$  aufspannt.

Die numerische Berechnung der Zerlegung (3.7.2.1) beseitigt allerdings keine der bis hierher angesprochenen Schwierigkeiten. Der untere Block der Matrizen  $\mathbf{R}_1 \mathbf{R}_2$  in (3.7.2.1) sollte theoretisch zwar Null sein, in der Praxis werden diese Koeffizienten jedoch aufgrund von Rundungsfehlern und Rauschen in den Daten (s.o.) von Null verschieden sein, so daß hier wiederum ohne a-priori-Information ein Schwellwert eingeführt werden muß, unterhalb dessen Koeffizienten (willkürlich) Null gesetzt werden. Dieser Schwellwert beeinflußt natürlich wiederum den (Pseudo-) Rang der Matrix  $\mathbf{C}$  und legt ihn willkürlich fest.

Aus diesem Grund wurde dieser Ansatz in dieser Arbeit nicht weiter verfolgt und stattdessen ein Verfahren gewählt, daß eine Rang-Entscheidung vorverlegt in die Phase des Filterdesigns, wo der Anwender die Möglichkeit hat, nach anderen Kriterien über die Gültigkeit einer Entscheidung zu urteilen.

Das oben angesprochene Verfahren stammt aus der Bildverarbeitung und wird dort als Karhunen-Loevé- (KL-) Transformation bezeichnet (Rosenfeld 1982, Jones 1987, Pratt 1972).

Sei  $\mathbf{C} \in \mathbb{R}^{m,n}$  die Matrix der seismischen Daten. Man löst das Eigenvektor/Eigenwert-Problem

$$\mathbf{X}^T \mathbf{C}^T \mathbf{C} \mathbf{X} = \text{diag}(\lambda_i), i=1, \dots, n.$$

$\mathbf{X}$  ist die orthonormale Matrix der Eigenvektoren von  $\mathbf{C}^T \mathbf{C}$ ,  $\lambda_i$  sind die dazugehörigen Eigenwerte.

Man definiert nun

$$\Phi = \mathbf{C} \mathbf{X}$$

als KL-Transformierte der seismischen Datenmatrix, wobei

$$\phi_{ij} = \langle \bar{c}_i, \bar{x}^j \rangle$$

mit  $\bar{c}_i$  : i-ter Zeilenvektor von  $C$ ,  $\bar{x}^j$ : j-ter Spaltenvektor von  $X$ .

Zunächst sei angenommen,  $C$  habe vollen Rang.  $C^T C$  ist dann pos.def. und besitzt  $n$  verschiedene Eigenwerte  $\lambda_j > 0$ . Rosenfeld (Rosenfeld 1982) zeigt, daß für die Rekonstruktion der Matrix  $C = X^T \Phi^T$  die Eigenvektoren  $\bar{x}^j$  eine dem Betrag der dazugehörigen Eigenwerte entsprechend abgestufte Signifikanz besitzen. Eine gute (und im Sinne der  $L_2$ -Norm sogar optimale) Rekonstruktion gelingt mit einer Transformation

$$\tilde{\phi}_{ij} = \langle \bar{c}_i, \bar{x}^j \rangle \quad \text{für alle } j \text{ für die gilt: } \lambda_j > s.$$

$s$  ist hier wiederum ein vom Anwender vorzugebender Schwellwert. Hier allerdings mit dem Vorzug, daß die Qualität der Approximation schon beim Design des Filters durch Rücktransformation  $\tilde{C} = X^T \tilde{\Phi}^T$  überprüft werden kann. Man beachte, daß der Spaltenraum der Transformierten kleiner geworden ist,  $\tilde{\Phi} \in R^{m,p}$ ,  $p < n$ ,  $\tilde{\Phi}$  spannt nur noch einen Unterraum von  $\{C\}$  auf.

Hat  $C$  nicht vollen Rang, so sind ein oder mehrere Eigenwerte  $\lambda_i$  gleich Null und man bildet

$$\tilde{\phi}_{ij} = \langle \bar{c}_i, \bar{x}^j \rangle \quad \text{für alle } j \text{ für die gilt: } \lambda_j > 0$$

mit dem Resultat, daß  $\tilde{\Phi} \in R^{m,p}$ ,  $p < n$  vollen Rang hat und in diesem Fall sogar den ganzen Vektorraum  $\{C\}$  aufspannt.

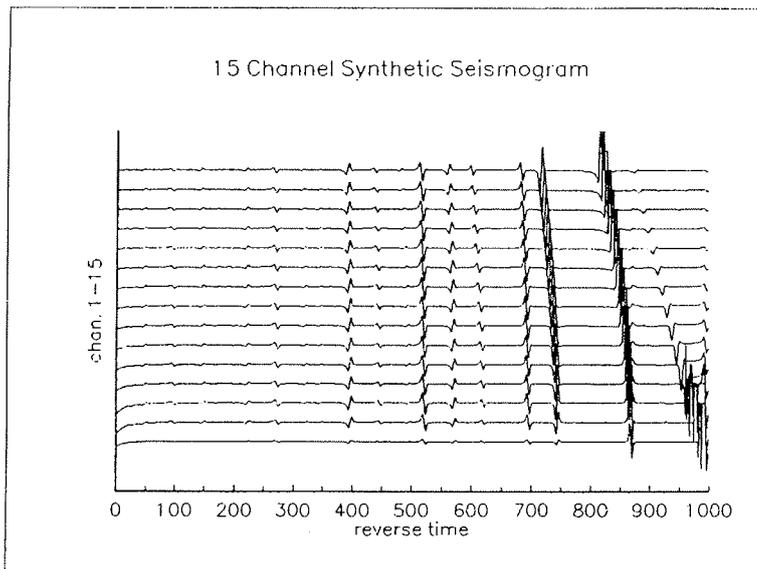


Abb. 3.7.2.1

Die Abbildung 3.7.2.1 zeigt 15 Spuren eines synthetischen Seismogramms. Die Zeitachse ist auch hier wieder umgedreht zur Verdeutlichung der Tatsache, daß der adaptive Lattice-Filter aus praktischen Erwägungen, die in Kapitel 4 erläutert werden, rückwärts in der Zeit läuft.

Im Bereich 0-600 Zeitschritte erkennt man die hohe Redundanz der 2-D seismischen Daten, von der vorher die Rede war. Erst oberhalb 500 Zeitschritte ist nennenswert Move-Out erkennbar, ab hier wird die Information erst wirklich mehrdimensional.

Diese Beobachtungen lassen sich erheblich verfeinern, wenn die KL-Transformierte desselben Seismogramms in Abbildung 3.7.2.2 betrachtet wird:

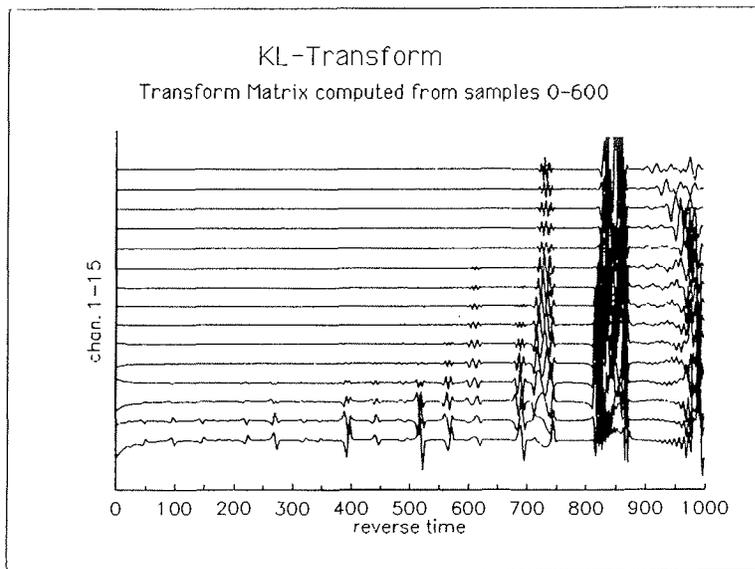


Abb. 3.7.2.2

Im Bereich 0-200  $\Delta t$  ist die gesamte Information in den ersten vier Kanälen unterzubringen. Statt des vollen Ranges von fünfzehn hat die seismische Datenmatrix hier also etwa den Rang Vier. Dieser nimmt im Bereich 200-500  $\Delta t$  auf Fünf zu und springt dann im Bereich 500-650  $\Delta t$  auf ca. Elf. Ab 650  $\Delta t$  werden dann alle fünfzehn Kanäle benötigt, um die Information vollständig abzubilden.

Die Rangentscheidung wird hier wie oben angegeben getroffen. Beispielsweise kann die Schranke  $s$  so gewählt werden, daß mindestens 98 % der Gesamtenergie (vgl. Jones 1987) in der Transformaten repräsentiert ist:

$$s = 0.02 \sum_{i=1}^n \lambda_i.$$

Eine Kontrolle ist stichprobenartig durch Rücktransformation und anschließendem Vergleich mit den Originaldaten einfach möglich.

Um die KL-Transformation im Sinne der Informationsverdichtung im Bereich 0-600  $\Delta t$  möglichst effektiv zu machen, wurde die Transformationsmatrix nur mit Daten aus diesem Bereich berech-

net. Es ist die herausragende Eigenschaft der Karhunen-Loevé-Transformation, daß sie diese Aufgabe von allen denkbaren orthonormalen Transformationen am optimalsten löst (vgl. Rosenfeld 1982).

Ein Nachteil besteht jedoch darin, daß die KL-Transformation sich der rekursiven Struktur des Lattice-Algorithmus nicht anpaßt. Alle Daten aus dem relevanten Zeitfenster müssen vorher verfügbar sein, um die Matrix der Eigenvektoren als Transformationsmatrix berechnen zu können, was zudem zusätzlichen Rechen- und Zeitaufwand erfordert.

Eine ebenfalls orthonormale Transformation, die im Bezug auf eine Informationsverdichtung gegenüber der KL-Transformation nur suboptimal (Rosenfeld 1982, Clarke 1981) arbeitet, deren Transformationsmatrix aber von vornherein festliegt, ist die Diskrete Cosinus-Transformation.

### 3.7.3 Rangentscheidung mit Hilfe der Diskreten Cosinus-Transformation.

Die Diskrete Cosinus-Transformation (DCT) (Rosenfeld 1982)

$$S(u) = \frac{2c(u)}{\sqrt{N}} \sum_{m=0}^{N-1} f(m) \cos \frac{(2m+1)u\pi}{2N},$$

(3.7.2.1)

$$c(u) = \begin{cases} \frac{1}{\sqrt{2}} & \text{für } u = 1, \dots, N \\ 1 & \text{für } u = 0 \end{cases}$$

ist ebenso wie die Karhunen-Loevé-Transformation eine in der Bildverarbeitung verbreitete Methode zur Datenkompression. Die Transformations-Matrix  $S$  mit der o.a Normierung

$$s_{ij} = \frac{2c(j)}{\sqrt{N}} \cos \frac{(2i+1)j\pi}{2N}$$

ist orthonormal, für den Projektionsoperator gilt entsprechend

$$\begin{aligned} P_{\{C\}} &= SC(C^T S^T SC)^{-1} C^T S^T \\ &= SCC^{-1} (S^T S)^{-1} C^{-1 T} C^T S^T \\ &= S (S^T S)^{-1} S^T . \end{aligned}$$

Die Transformation beeinflusst also die Projektion nicht.

Eine theoretische Begründung für die nahezu optimale Informationsverdichtung der DC-Transformation, die im Grenzfall die Leistung der KL-Transformation erreicht, gibt Clarke (Clarke 1981). In der Abbildung 3.7.3.1 ist zum Vergleich die DC-Transformierte des Seismogramms aus Abb. 3.7.2.1 dargestellt,

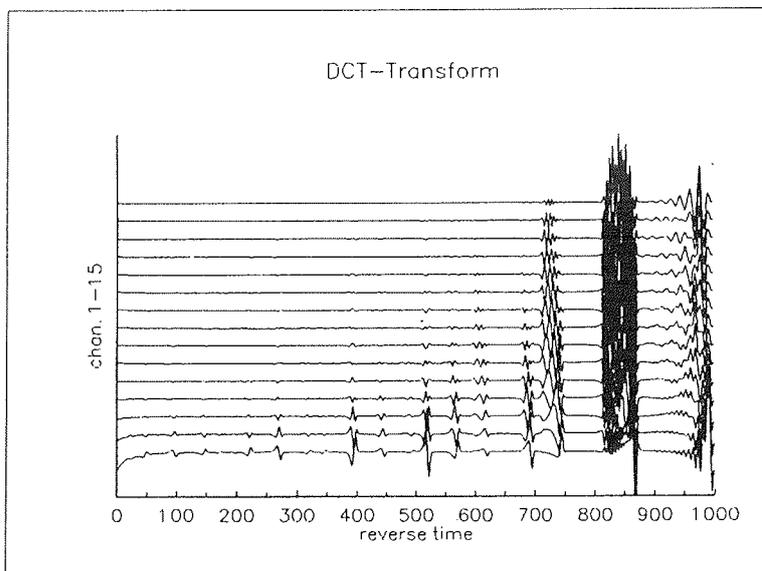


Abb. 3.7.3.1

im hier interessierenden Bereich von 0-600  $\Delta t$  sind kaum Unterschiede zur KL-Transformierten zu erkennen.

Das Kriterium für die Rangentscheidung entspricht dem, das sich in der Bildverarbeitung für die Datenkompression bewährt hat (Rosenfeld 1982). In der Transformierten werden Werte mit Varianzen unterhalb eines Schwellwertes Null gesetzt, auch hier ist eine Kontrolle durch Rücktransformation während der Anpassung des Lattice-Filters möglich.

Die DCT kann rekursiv aus den Daten berechnet werden und die Transformationsmatrix für Daten mit gleicher Anzahl von Kanälen liegt von vornherein fest .

Die Diskrete Cosinus-Transformation ist insofern für die Rangentscheidung in Verbindung mit einem adaptiven Mehrkanal-Lattice-Algorithmus ein sehr gut geeignetes Verfahren.

Die Abbildungen 3.7.3.2 zeigen am Beispiel des synthetischen Seismogramms aus Abb. 3.7.2.1, wie die Rangentscheidung vor der Anwendung des Filters getroffen werden kann.

Abbildung 3.7.3.2a zeigt die Entwicklung des Ranges der seismischen Matrix für die 98%-Rekonstruktion. Werte, die in der DC-Transformierten kleinere Varianzen (Energien) aufweisen als 2% der Gesamtvarianz, werden Null gesetzt.

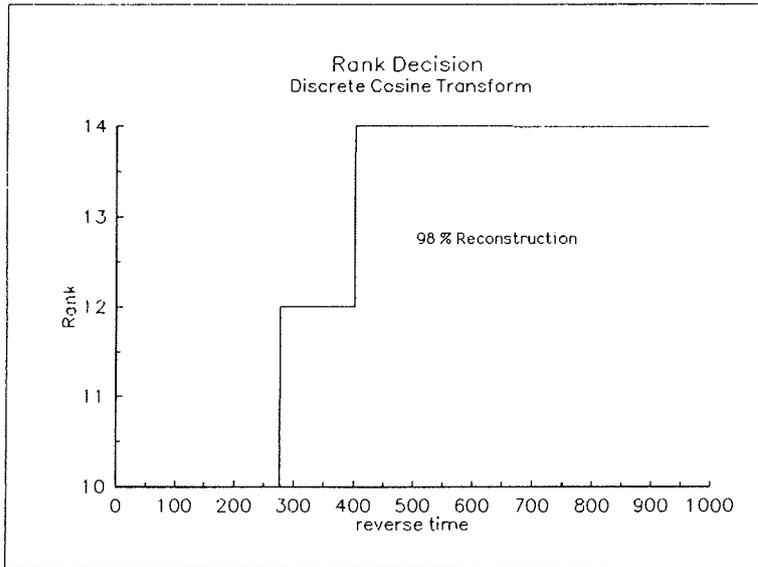


Abb. 3.7.3.2a

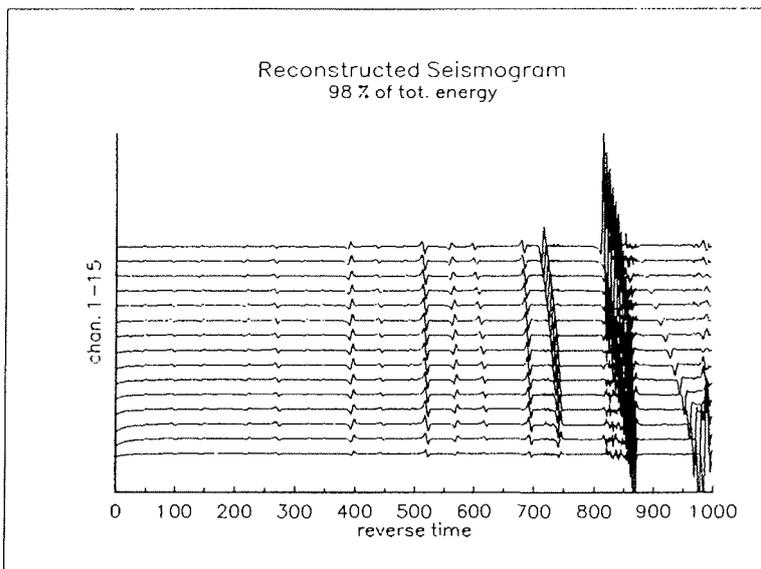


Abb. 3.7.3.2b

Das rekonstruierte Seismogramm in Abb. 3.7.3.2b ist fehlerfrei wiederhergestellt bis zu etwa  $700 \Delta t$ , oberhalb von 700 Zeitschritten macht sich die Fehlanpassung der DCT bemerkbar.

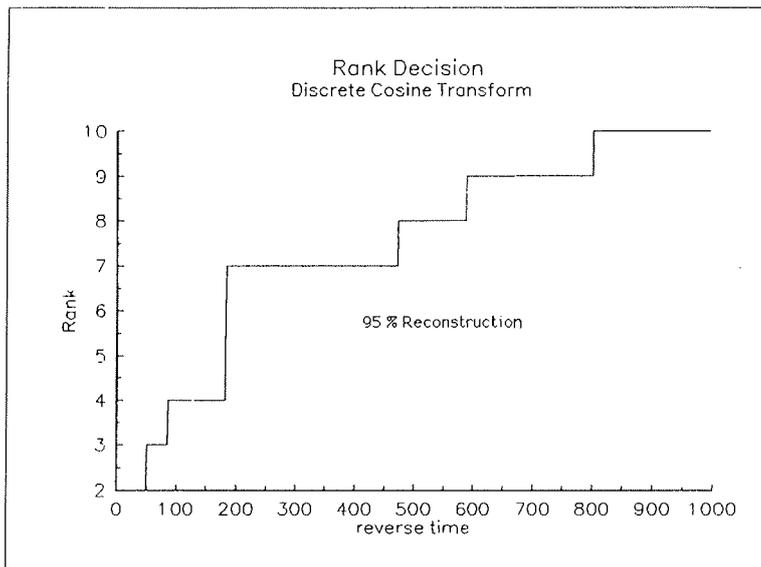


Abb. 3.7.3.2c

Hier müßte die Transformation sozusagen abgestellt werden, damit ab 700  $\Delta t$  mit dem vollen Rang von Fünfzehn rekonstruiert wird.

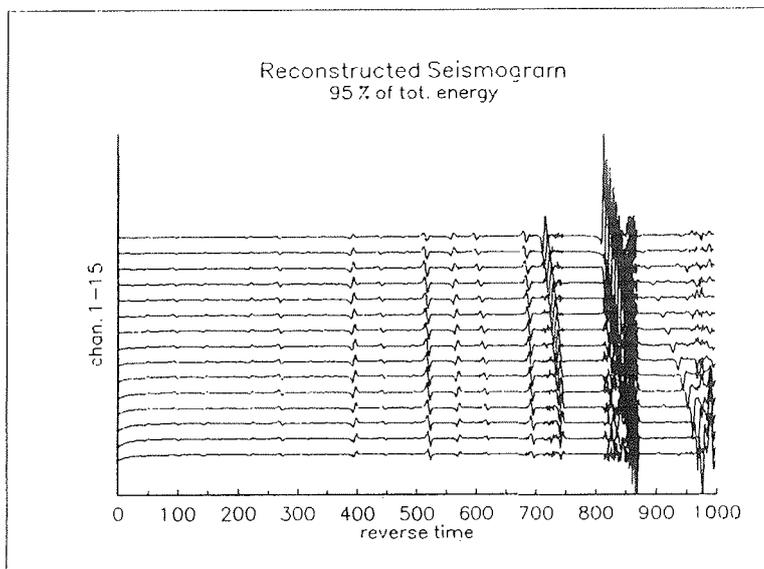


Abb. 7.3.3.2d

Die 98%-Rekonstruktion liefert eine wenig differenzierte Rangentscheidung mit nur drei Werten. Dies wird sehr viel besser bei der 95%-Rekonstruktion, deren Rangentwicklung in Abbildung 3.7.3.2c dargestellt ist.

Die Rekonstruktion (Abb. 3.7.3.2d) gelingt gut bis ca.  $600 \Delta t$ , und nur auf die Daten bis  $600 \Delta t$  würde die Transformation angewendet, wollte man das Seismogramm mit dem adaptiven Mehrkanal-Filter bearbeiten.



#### 4. Anwendungen

Die vorliegende Arbeit beschäftigt sich mit der Anwendbarkeit von adaptiven Filtern für die Unterdrückung von Multiplen Reflexionen, die in Verbindung stehen mit der quasi freien Oberfläche der Wasserschicht. Die Abbildungen A4.1a und 4.1b zeigen typische Ausbreitungswege der seismischen Welle für diesen Typ von Multiplen.

Die Ausbreitung der reinen Wasserschicht-Multiplen ist in Abbildung 4.1.a dargestellt. Ein Teil der seismischen Energie ist im Wellenleiter des Wasserkörpers mit dem gegen Eins strebenden Reflexionskoeffizienten der (fast) freien Wasseroberfläche und dem relativ großen Reflexionskoeffizienten, der den Dichtesprung an der Grenze Wasser-Sediment kennzeichnet, gefangen.

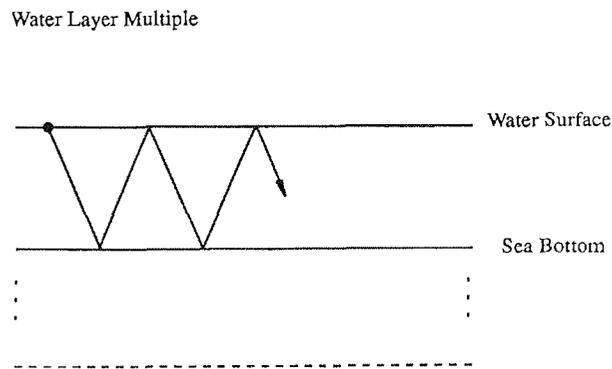


Abb 4.1a

Abbildung 4.1.b zeigt mögliche Wege der Wellenausbreitung der sogenannten Peg-Leg-Multiplen. Die Definition des Begriffes folgt Clearbout (Clearbout 1985). Hier hat ein Teil

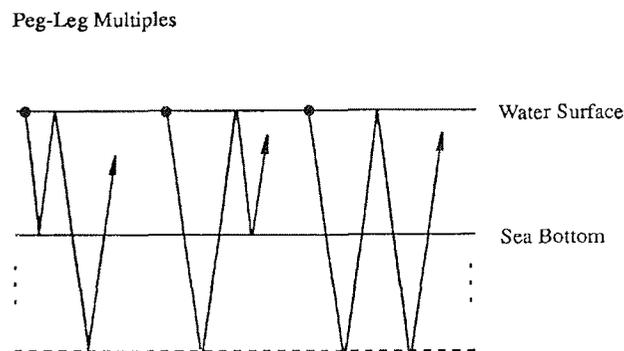


Abb. 4.1b

der seismischen Energie einen tieferen Reflektor erreicht, der Wellenzug wird oder wurde im Verlauf der Ausbreitung jedoch auch mehrfach an Wasseroberfläche und Meeresboden reflektiert.

Die sogenannten Intrabed-Multiplen waren für diese Arbeit nicht von Interesse.

#### **4.1 Multiple Reflexionen in Verbindung mit der freien Oberfläche in Common-Shot-Point (CSP) Seismogrammen**

Taner (Taner 1990) weist darauf hin, daß die konventionelle, Prädiktive Dekonvolution zur Unterdrückung von Wasserschicht-Multiplen im wesentlichen mit zwei systematischen Fehlern behaftet ist. Einer dieser Fehler wurde in der vorliegenden Arbeit bereits behandelt, der sogenannte Fenstereffekt, der sich einstellt, wenn versucht wird, eine Autokorrelations-Matrix auf der Basis einer endlichen Anzahl von Datenwerten zu berechnen. Um die Töplitz-Form der Matrix zu erzwingen, die notwendig ist, um den Wiener-Levinson Algorithmus anwenden zu können, müssen Datenwerte auf Null gezwungen werden, die tatsächlich von Null verschieden sind. Koehler (Koehler, 1985) weist darauf hin, daß Prewindowed-, Postwindowed- und Covarianzform den alles entscheidenden Makel haben, daß die resultierende Korrelationsmatrix erstens nur vergleichsweise aufwendig berechnet werden kann und zweitens, daß diese Matrix, abgesehen von ihrer Symmetrie, keine Struktur aufweist, die von einem "Fast-" Algorithmus ausgebeutet werden könnte. Alle Verfahren, die nicht auf der Töplitz-Struktur der Korrelationsmatrix beruhen, waren also bislang schlicht zu langsam.

Der zweite systematische Fehler beruht auf der vereinfachenden Annahme, der Multiplenprozess sei periodisch, was letztlich nur für Zero-Offset-Seismogramme oder, mehr von akademischem Interesse, im Fall der Anregung mit unendlich ausgedehnten, ebenen Wellen gilt. Selbst eine NMO-Korrektur mit der Wasserschall-Geschwindigkeit simuliert ebene Wellen nur im Wasserkörper, die sog. Peg-Leg-Multiplen besitzen den Move-Out entsprechend der jeweiligen RMS-Geschwindigkeit.

Die Abbildung A4.1.1 aus Taner (Taner 1980) illustriert die typischen Verhältnisse bei der Aufzeichnung mariner seismischer Daten. Die direkt aufeinanderfolgenden Einsätze der Multiplen, die mit einem bestimmten, von Null verschiedenen Winkel im Wasserkörper propagiert, werden jeweils nicht vom gleichen Hydrophon registriert und dementsprechend nicht im gleichen Kanal aufgezeichnet.

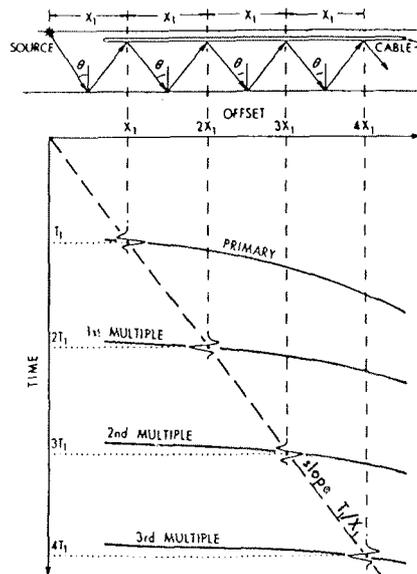


Abb. 4.1.1

In einem  $x$ - $t$ -Koordinatensystem haben demzufolge die Einsätze der Multiplen keine konstante Periode. In einem Kanal werden Multiple aufgezeichnet, die jeweils unter verschiedenen Winkeln reflektiert werden. Daß zudem bei hinreichend großer relativer Apertur, also abhängig von der räumlichen Ausdehnung des Empfänger-Arrays im Verhältnis zur Tiefe des Reflektors, diese, an einer Position konstanten Offsets aufgezeichneten Multiplen in ihren Amplituden ein sehr viel komplexeres Verhältnis zueinander aufweisen können als es die rein geometrische Amplitudenabnahme beschreibt, kann man der Abbildung A4.1.2 entnehmen, die ebenfalls dem Artikel von Taner (Taner 1980) entlehnt ist.

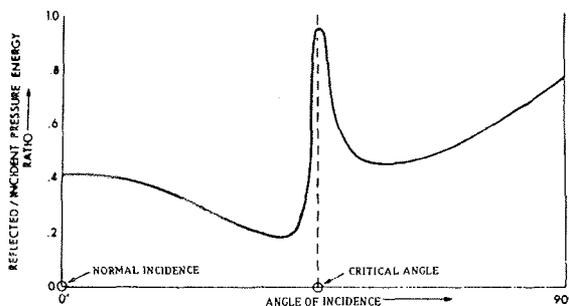


Abb. 4.1.2

Dargestellt ist die Winkelabhängigkeit für eine Druckwelle in einem Einschichtenmodell mit einem Reflexionskoeffizienten von 0,4 für den senkrechten Einfall.

Ein dritter systematischer Fehler besteht darin, daß möglicherweise der Reflexionskoeffizient des Meeresbodens auch in dem von einem einzelnen Experiment, einem Schuß, überdeckten Bereich mehr oder weniger stark variiert, sodaß zur Winkelabhängigkeit noch die Ortsabhängigkeit kommt.

Den hier genannten Eigenschaften von Wasserschicht- und Peg-Leg-Multiplen tragen adaptive Filter Rechnung. Sowohl Einkanal- als auch Mehrkanalfilter sind, wie die Experimente zeigen, in der Lage, die sich langsam ändernden Perioden der Multiplen zu verfolgen. Diese Perioden entsprechen zu späten Zeiten im CSP-Seismogramm der zu erwartenden Zero-Offset Periode und nehmen dann zu früheren Zeiten- für alle Offsets größer Null stetig ab.

Genau dies ist der Grund, warum die Filter rückwärts in der nominalen Zeit laufen. Der Filter wird mit der Zero-Offset-Periode aufgesetzt, schwingt sich in dem Bereich des Seismogramms ein, wo diese Periode für alle Offsets recht gut zutrifft und adaptiert dann kontinuierlich die Koeffizienten, entsprechend den kürzer werdenden Perioden. Die Vorhersage-Richtung bleibt davon unberührt. Sie verläuft in Richtung der positiven Zeitachse, sodaß eine nachfolgende Multiple auf Basis der ihr vorangegangenen vorhergesagt wird. Ein zusätzlicher Vorteil dieses Vorgehens besteht darin, daß die Rechnung "In-Place", d.h. ohne daß eine Kopie des Datensatzes im Speicher gehalten werden müßte, durchgeführt werden kann. Die bearbeiteten Daten werden nicht mehr benötigt und können mit den Ergebnissen überschrieben werden.

Insbesondere der Mehrkanalfilter ist in der Lage, dem mehrdimensionalen Charakter von multiplen Reflexionen und insbesondere der Tatsache, daß Energie gleichsam zwischen den Kanälen der Aufzeichnung propagiert, gerecht zu werden. Dies zahlt sich unter anderem auch ganz besonders im Hinblick auf die benötigte Rechenzeit aus. Der Mehrkanalalgorithmus erledigt seine Aufgabe

um ein Vielfaches schneller als der Einkanalfilter und dürfte speziell in dieser Hinsicht im Vergleich mit anderen Verfahren recht gut abschneiden.

## 4.2 Filterdesign

Aus den Abschnitten zur Theorie der adaptiven Lattice-Filter ist bereits hervorgegangen, daß ein Vorhersage-Fehler-Filter bereits mit drei Parametern vollständig beschrieben ist. Dies sind Vorhersagedistanz (gap), Filterordnung (number of taps) und der Gewichts-Faktor zur Parametrisierung der Anpassungsfähigkeit des Filters an eine sich ändernde Signalstatistik allgemein oder, speziell hier, an eine sich ändernde Vorhersagedistanz und sich ändernde Amplituden und Phasenspektren des Signals.

Die Vorhersagedistanz ist einfach die Zero-Offset-Periode der Wasserschicht-Multiplen, so wie sie aus der Wassertiefe und der Wasser-Schallgeschwindigkeit berechnet werden kann. Wie sich gezeigt hat, ist dieser Wert in der Größenordnung mehrerer Zeitschritte unkritisch, sodaß auch bei fehlerhaften Wassertiefen oder einer schlechten Schätzung der Schallgeschwindigkeit in Wasser der Filter relativ schnell einschwingt. Für die Bearbeitung der synthetischen Daten wurde dieser Wert aus dem Ersteinsatz der Meeresboden-Reflexion abgeschätzt.

Die Festlegung der Filterordnung für den eindimensionalen Filter folgt Überlegungen, die bereits Griffith (Griffith 1977) experimentell bestätigt hat. Zum einen sollte die Filterordnung in der Größenordnung des maximalen Move-Out-Differenz zwischen primärer und multipler Reflexion (gemessen in Zeitschritte) liegen und mindestens der Länge des einfachen Wavelets (ohne Berücksichtigung des Bubbles, im Fall einer Air-Gun-Quelle) entsprechen. Dies garantiert, daß ein gedachter Schwerpunkt des Operators mit der sich tatsächlich ändernden Vorhersagedistanz gleichsam mit genügend Spielraum über die Filterkoeffizienten wandern kann. Auch dieser Wert ist in Grenzen unkritisch, und speziell im eindimensio-

nalen Fall schadet Überparametrisierung offenbar nur im Hinblick auf die drastisch zunehmende Rechenzeit.

Beim Mehrkanalfilter ist die Überparametrisierung kritisch und in den Ausgangsdaten unübersehbar, die bearbeiteten Daten sind i.d.R. nicht wiederzuerkennen. Aufgrund der Tatsache, daß die schußnahen Kanäle auch immer im Datenfenster des Operators liegen, und sich zudem die Multiplenperiode von Kanal zu Kanal nur sehr wenig ändert, muß auf das Driften der Vorhersagedistanz hier keine Rücksicht genommen werden. In allen, im Abschnitt 4.3 vorgestellten Beispielen betrug die Filterordnung nicht mehr als Acht. Überparametrisierungs-Effekte wurden bei den synthetischen Daten bei der Ordnung Zwölf, bei den realen Daten bei der Ordnung Achtzehn beobachtet. Hier liegt ein wesentlicher Grund für die Rechenökonomie des Mehrkanal-Algorithmus, die Anzahl der notwendigen Iterationen pro Zeitschritt ist um eine Größenordnung kleiner als beim eindimensionalen Verfahren.

Der Gewichts-Faktor schließlich, als ein Maß für die zu erwartende Instationarität muß in einem Trial-und-Error Verfahren ermittelt werden. Dasselbe ist im übrigen auch für die Filterordnung zu empfehlen. Die im Verlauf dieser Arbeit durchgeführten Experimente lassen den Schluß zu, daß dieser Parameter irgendwo im Intervall zwischen 0.95 und 0.999 anzusiedeln ist. Eine Faustregel läßt sich dennoch aufstellen. Je mehr man sich der Obergrenze des genannten Intervalls nähert, desto konservativer arbeitet der Filter. Konservativ in dem Sinn, daß weniger aus den Daten herausgefiltert wird, die bearbeiteten Daten also mehr den Rohdaten gleichen. Die Versuche, diesen Parameter so gut wie möglich einzustellen, sollten auf mehreren CSP- (CDP-) Ensembles erfolgen, wobei es sinnvoll ist, auch das Residuum des Filterprozesses, also den Teil der Daten, die herausgefiltert wurden, für die Beurteilung heranzuziehen. Auch hier gilt, daß dieser Parameter für den Einkanal-Algorithmus weniger kritisch ist als für den Mehrkanal-Filter.

In allen Fällen sollte die adaptive Filterung als erster Schritt der Bearbeitung der Rohdaten vor allen anderen Filterprozessen erfolgen. Die Standardfilter in Datenverarbeitungs-Paketen sind oft

Allpol- (IIR-) Filter, dies gilt insbesondere für die sog. Bandsperr-Filter. Den Daten wird mit diesen Filtern ein autoregressives Modell überprägt auf das sich ein inverser adaptiver autoregressiver Filter, wie die hier vorgestellten, unter ungünstigen Umständen leicht einschwingt. Zur Stabilität des Mehrkanalfilters trägt darüberhinaus eine möglichst große Bandbreite bei, sodaß auch die üblicherweise verwendeten Bandpass-Filter erst nach der adaptiven Multiplendekonvolution angewandt werden sollten.

### **4.3 Eindimensionale adaptive Filter für die Unterdrückung multipler Reflexionen**

Beide adaptiven Filter, also der Einkanal- und der Mehrkanal-Algorithmus, wurden im Verlauf der Arbeit an verschiedenen synthetischen und realen Datensätzen getestet. Exemplarisch sind im folgenden die Bearbeitung eines synthetischen CSP-Datensatzes, der mir freundlicherweise von der Fa. PRAKLA SEISMOS, Hannover, zur Verfügung gestellt wurde (Abb. 4.3.1) und eines kompletten seismischen Profils, das während der Arktiskampagne ARK VII-3 des Alfred-Wegener-Instituts, Bremerhaven, im Sommer 1990 in Hall Bredning, dem nördlichen Hauptbecken des Scoreby-Sund Fjordsystems aufgezeichnet wurde (Abb. 4.3.2), dargestellt.

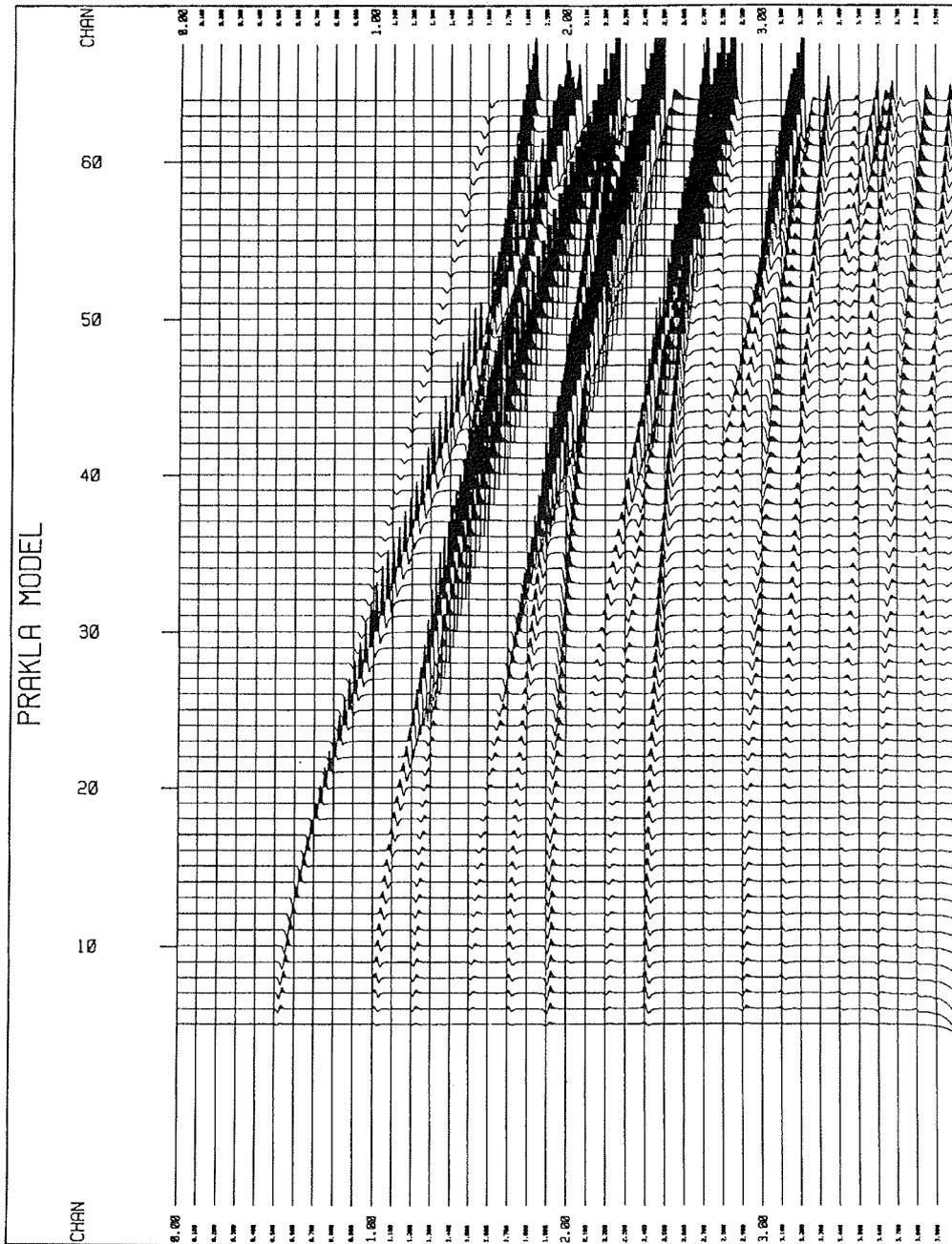


Abb. 4.3.1

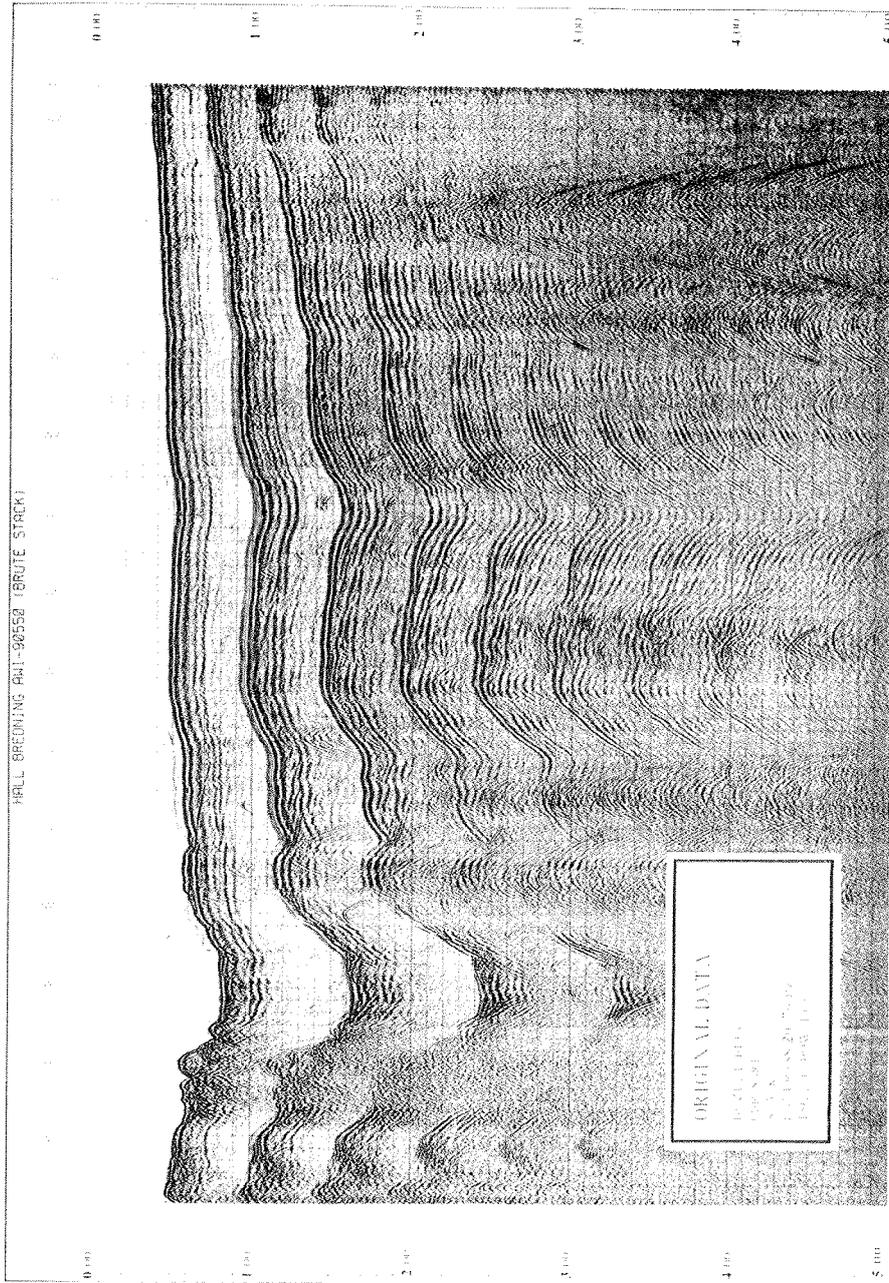


Abb. 4.3.2: Originaldaten des Hall-Breeding Profils AWI-90550

### 4.3.1 Beispiel: Synthetische CSP-Daten

Der synthetische Datensatz in Abbildung 4.3.1 simuliert eine Weitwinkel-Registrierung in sechzig Kanälen in der, bei sorgfältiger Interpretation, neben der Meeresboden-Reflexion bei 500 ms zumindest zwei weitere Primärreflexionen bei 1200 ms und 1900 ms erkennbar sind. Die kritische Distanz liegt etwa bei der Position von Kanal 20, hier ist der Einsatz der am Meeresboden refraktierten Welle bereits vor der Reflexion zu erkennen.

Die Abbildung 4.3.1.1a zeigt das Ergebnis der Einkanalfilterung mit folgenden Parametern: Vorhersagedistanz 480 ms, Ordnung 20,  $\lambda = 0.998$ .

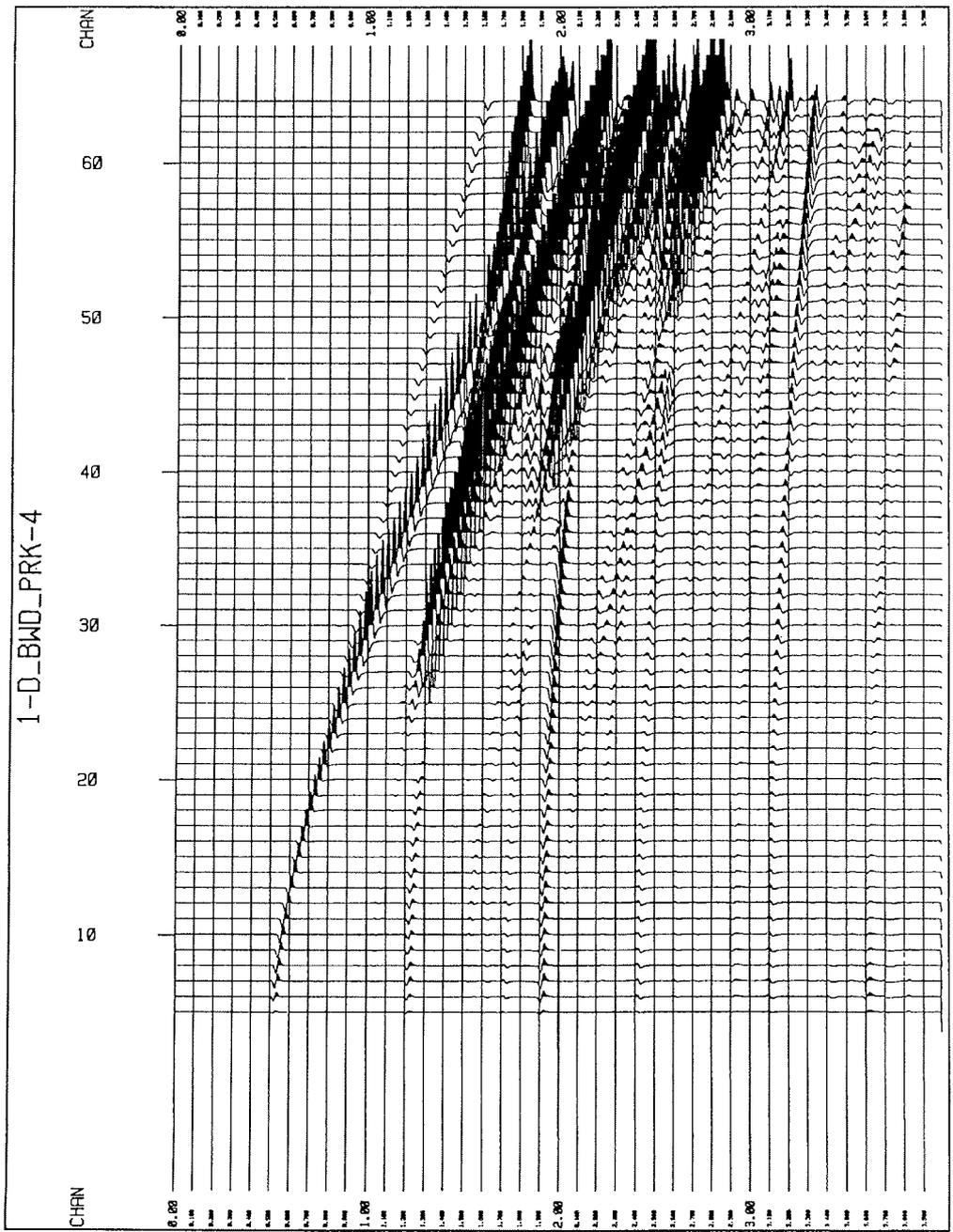


Abb. 4.3.1.1a

Interessant ist hier nur der unterkritische Bereich. Im Bereich überkritischer Einsätze versagt der Filter naturgemäß, da hier auch das AR-Modell keine geeignete Beschreibung für den Multiplenprozess mehr darstellt. Beschränkt man die Betrachtung auf die unterkritischen Kanäle 1-20, so ist zweifelsfrei zu erkennen, daß multiple Einsätze relativ stark gegenüber den primären bedämpft werden. Die Erste Multiple der Meeresbodenreflexion verschwindet sogar fast völlig.

Betrachtet man das Residuum des Filterprozesses, den vom Filter unterdrückten Signalanteil, in Abbildung 4.3.1.1b, findet man diese Beobachtung bestätigt. Im Residuum sind Primäreinsätze nicht abgebildet, der Filter ist offenbar nicht in der Lage, die Multiplen vollkommen zu unterdrücken, läßt aber auf jeden Fall die primären Einsätze unangetastet.

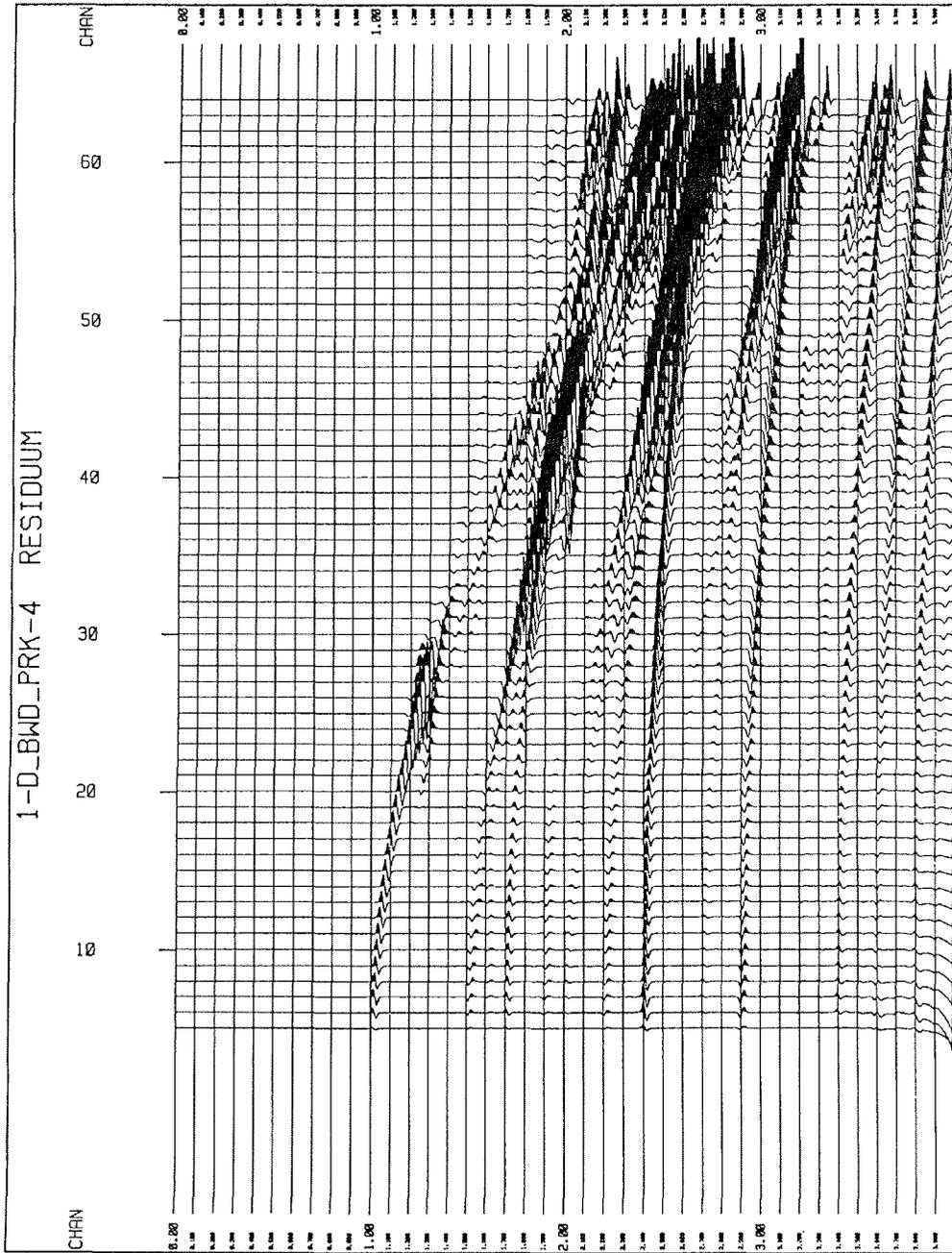


Abb. 4.3.1.1b

### 4.3.2 Beispiel: Reale CSP Seismogramme

Die Abbildung 4.3.2.1 zeigt die Einzelschüsse 20 bis 30 mit je 24 Kanälen des Hall-Bredning-Profiles aus Abb. 4.3.2. Deutlich zu erkennen ist die Meeresbodenmultiple bei 1000 ms, gefolgt von einem Paket von Peg-Legs. Dieses Muster läßt sich mit einer Periode von ca. 500 ms bis zu einer Zeit von 3500 ms im Schuß 30 verfolgen.

Der Filter wurde mit diesen Schüssen getestet. Das Ergebnis der Filterung mit einem Filter der Ordnung 50 und einem Gewichts-Faktor  $\lambda = 0.995$  zeigt Abbildung 4.3.2.2. Die Meeresbodenmultiple ist hier recht erfolgreich unterdrückt, ebenso sind die Amplituden der Peg-Legs reduziert. Die Vorhersagedistanz wurde aus den Werten der Wassertiefe in den Datenheadern und einer, für alle Schüsse konstanten Schallgeschwindigkeit in Wasser von 1480 m/s berechnet.

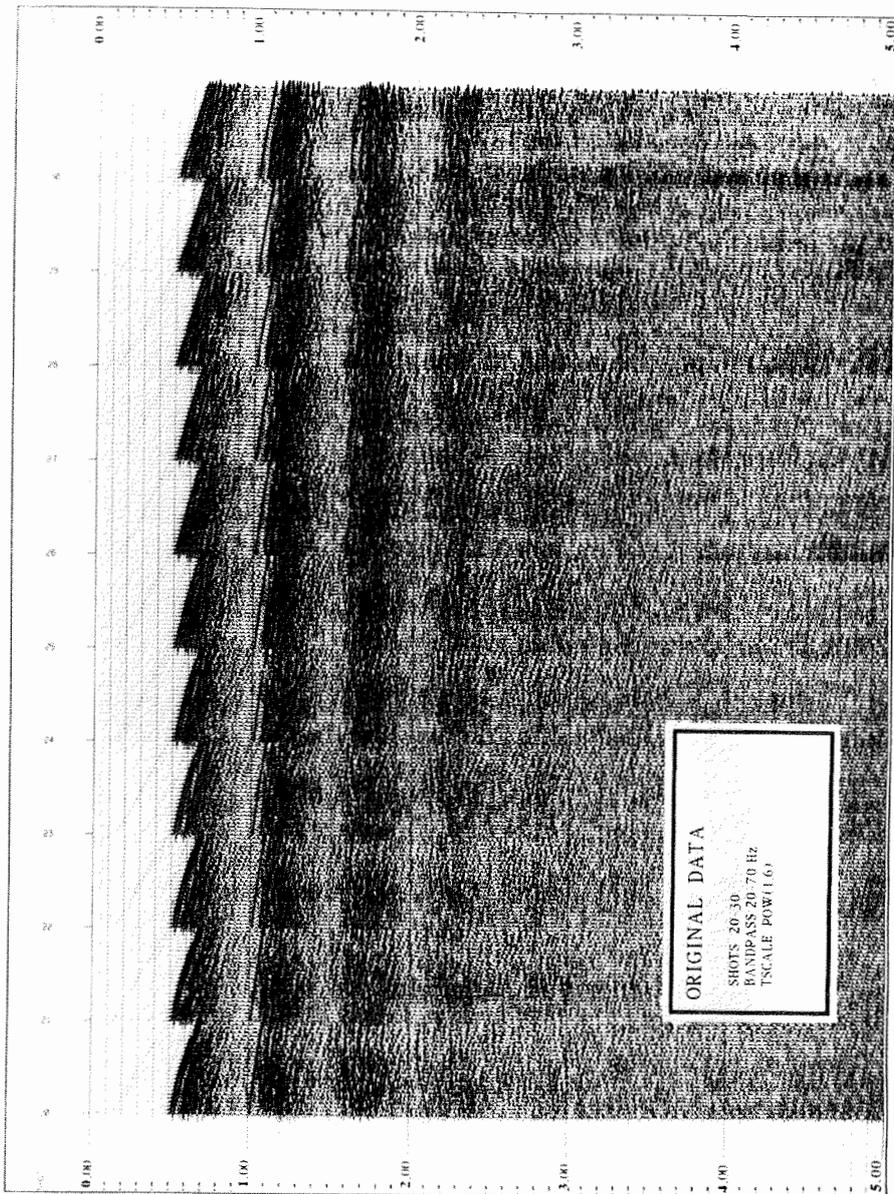


Abb. 4.3.2.1: Originaldaten, Schüsse 20-30, Hall-Bredning AWI-90550

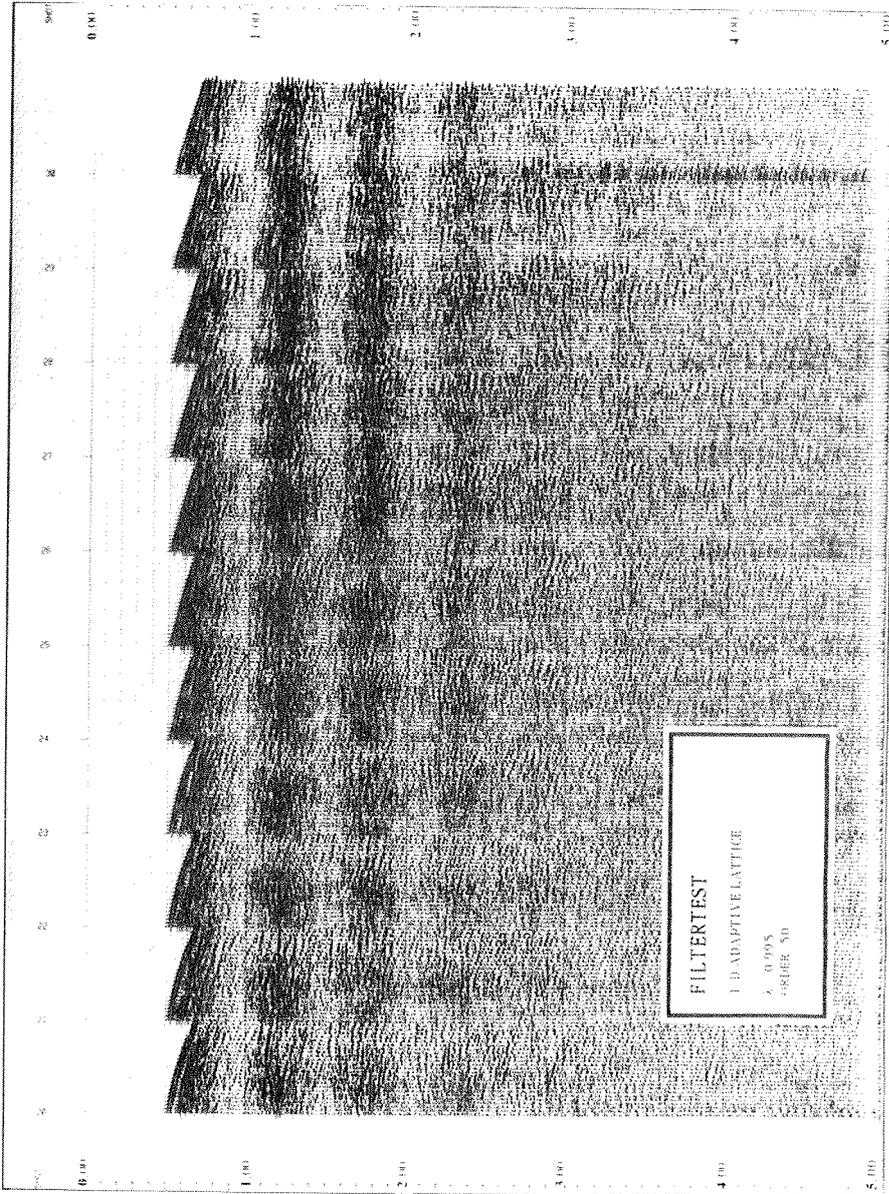


Abb. 4.3.2.2: Einkanalige Filterung der Schüsse 20-30

### 4.3.3 Beispiel: Reale Daten, CDP-Stack, Prestack-Filterung, Prestack-/Poststack-Filterung

Alle Schüsse des aus knapp 1450 CDP's bestehenden Profils wurden mit dem o.g. Filter bearbeitet und anschließend mit einer konstanten NMO-Korrektur von 1500 m/s gestapelt. Dies sollte dafür sorgen, daß die Restenergien der Multiplen in jedem Fall konstruktiv superponiert werden und im Profil gut sichtbar sind.

Das Ergebnis der Prestack-Filterung zeigt die Abbildung 4.3.3.1. Das Multiplenenmuster ist nach wie vor gut zu erkennen, wenn auch die Amplituden der Multiplen reduziert sind. Dieses gestapelte Profil wurde nun noch einmal mit einem Filter der Ordnung 20 und  $\lambda = 0.995$  bearbeitet (Abb. 4.3.3.2). Mit der zusätzlichen Poststack-Bearbeitung kann offensichtlich eine weitere Qualitätsverbesserung erreicht werden.

In beiden Abbildungen ist am Ende der Aufzeichnung bei 5 Sekunden ein Bereich von ca. 150 ms zu erkennen, in dem die Daten zerstört sind. Dies ist charakteristisch für die Ausgangsdaten von adaptiven Filtern. Hier liegt die Einschwingphase des Filters und auch aus diesem Grund ist es vernünftig, den Filter in der nominalen Zeit rückwärts laufen zu lassen. Der Datenverlust ist am Ende der Aufzeichnung am leichtesten zu verschmerzen.

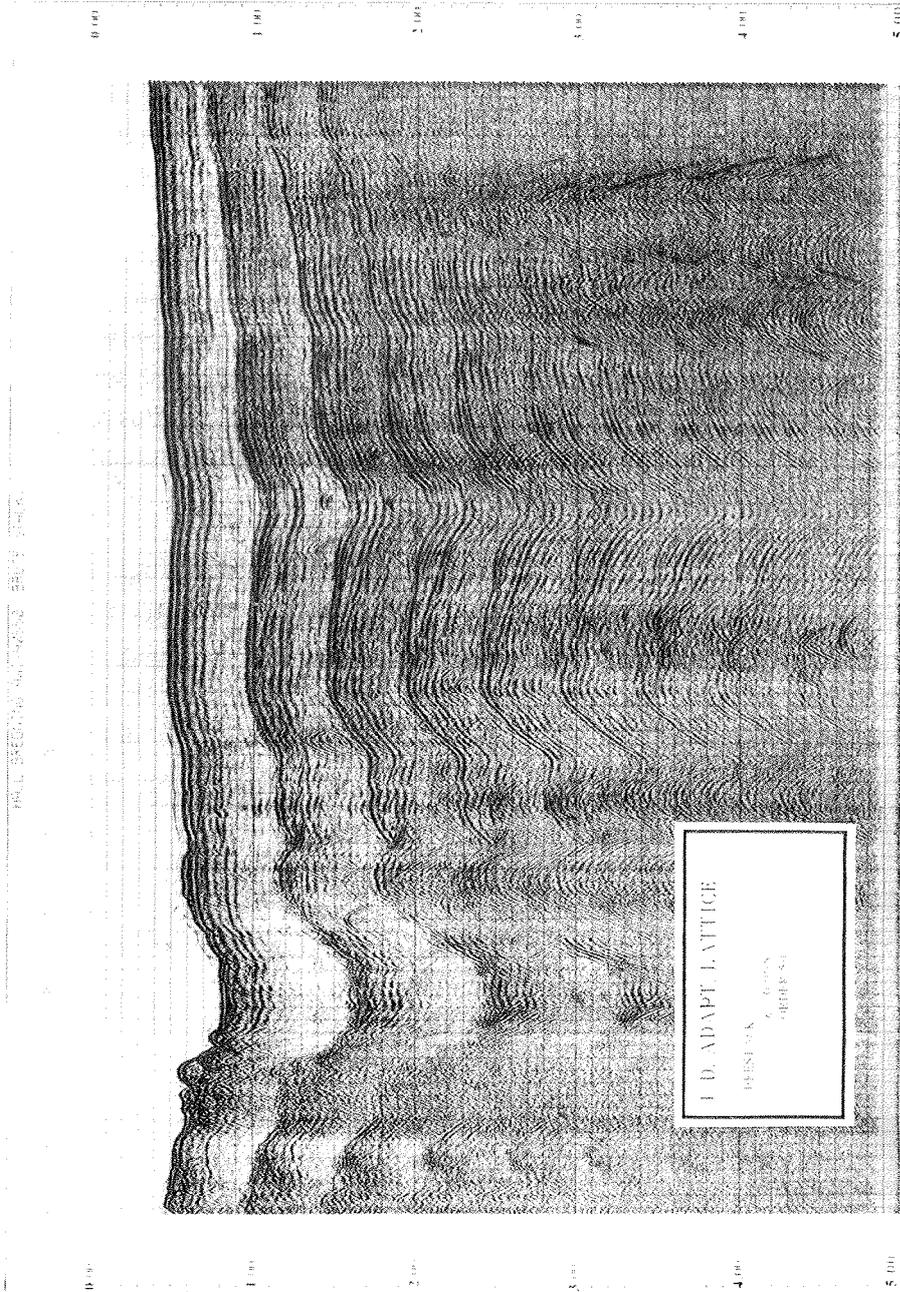


Abb. 4.3.3.1: Einkanalig gefilterte Daten, Filterung vor Stapelung

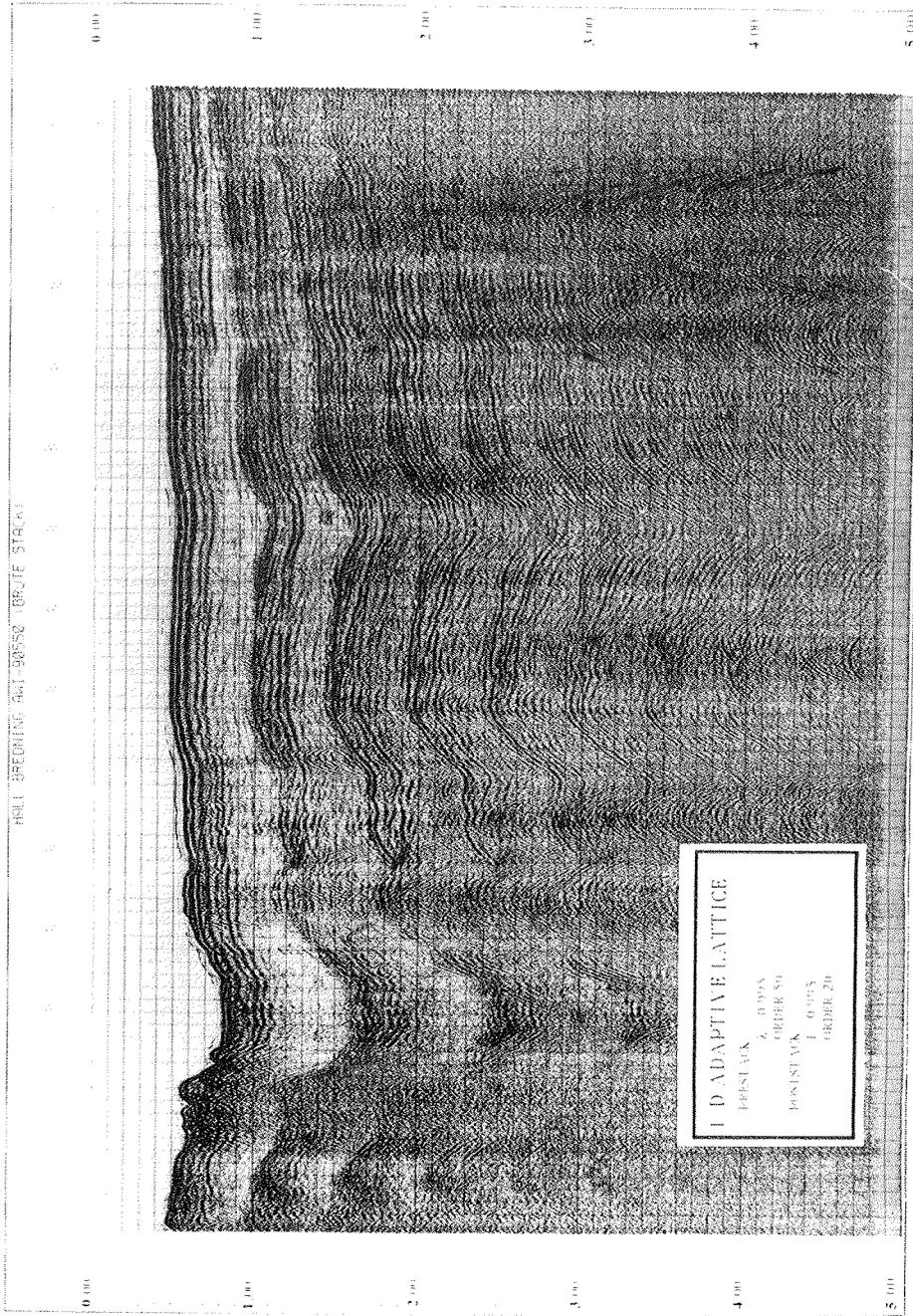


Abb. 4.3.3.2: Einkanalig gefilterte Daten, Filterung vor und nach Stapelung

## **4.4 Adaptive Mehrkanal-Filter für die Unterdrückung multipler Reflexionen**

### **4.4.1 Beispiel: Synthetische CSP-Daten**

Die Abbildungen A4.4.1.1a und A4.4.1.1b zeigen das Ergebnis der Mehrkanal-Filterung der Daten aus Abb. A4.3.1 für zwei verschiedene Werte des Gewichts-Faktors. Der Filter ist von der Ordnung Acht und die vorgegebene Vorhersagedistanz beträgt wie in Fall des Einkanalfilters 480 ms. Bearbeitet wurden hier nur die unterkritischen Spuren.

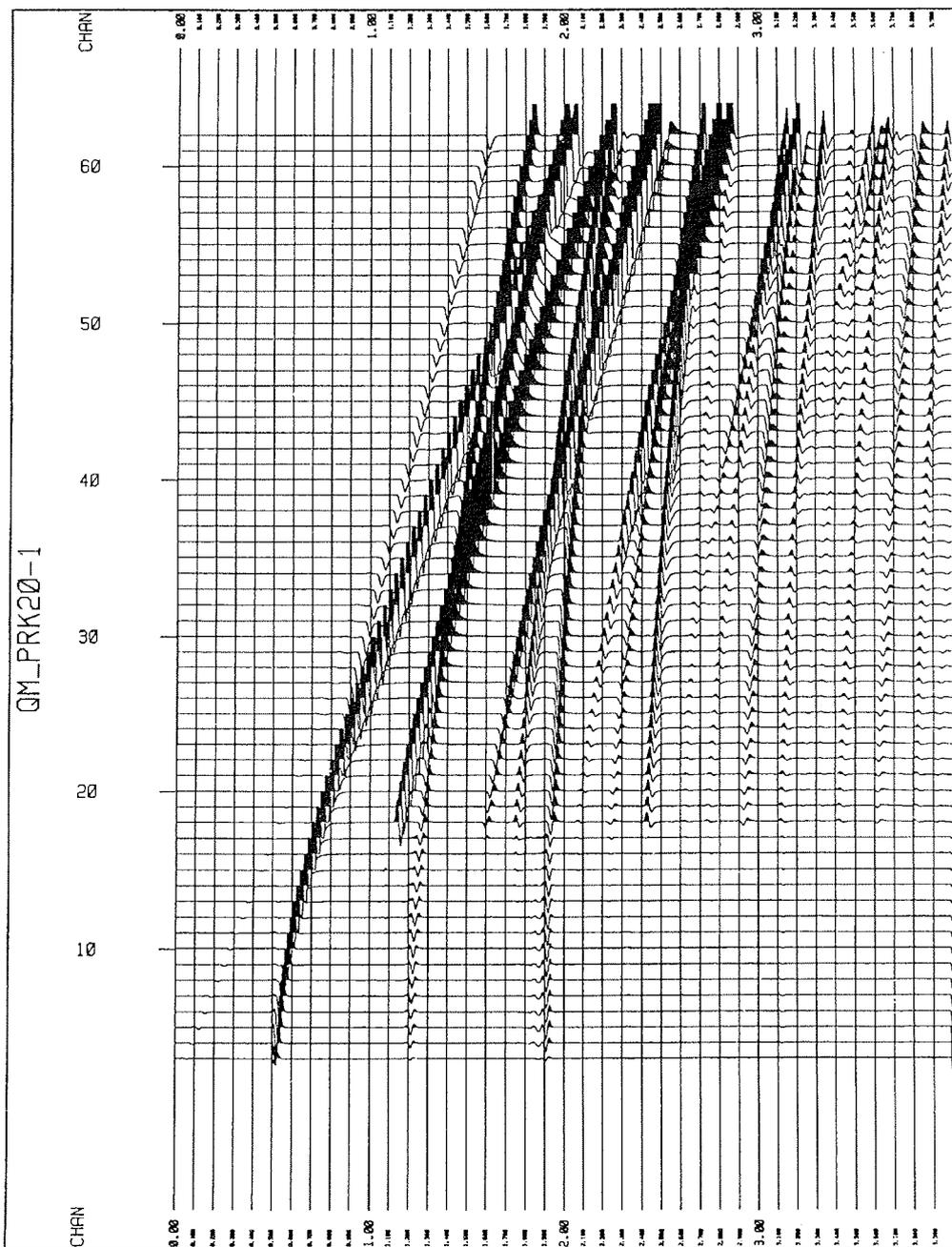


Abb. 4.4.11a

Die Abbildungen zeigen gewissermaßen die Extrema einer sinnvollen Bearbeitung. Im Fall der Abb. 4.4.1.1a wurde der Filter mit  $\lambda = 0.96$  betrieben, mit dem Ergebnis, daß außer den Primärreflexionen bei 1200 und 1900 ms alle Einsätze fast vollständig eliminiert sind. Die primären Einsätze weisen allerdings bereits eine deutliche Phasenverzerrung auf. Das andere Extremum zeigt Abbildung 4.4.1.1b, das Ergebnis der Filterung mit  $\lambda = 0.998$ , also einer extrem konservativen Behandlung der Daten. Die Multiplen unterhalb von 2000 ms sind nach wie vor sehr gut unterdrückt, wenn auch mit lateral leicht unterschiedlicher Qualität, wie der Einsatz bei 1750 ms zeigt, dessen schußferne Spuren noch Multipleneinsätze aufweisen. Unterhalb von 2000 ms sind die Filter-Ausgangsdaten deutlich schlechter, alle dort liegenden Einsätze sind von der Filterung nur wenig unterdrückt. Der Filter hat zudem offenbar eine sehr viel längere Einschwingphase, die vom Ende der Aufzeichnung bis fast 3200 ms reicht und besonders in den schußnahe Spuren mit großen Amplituden deutlich zu erkennen ist.

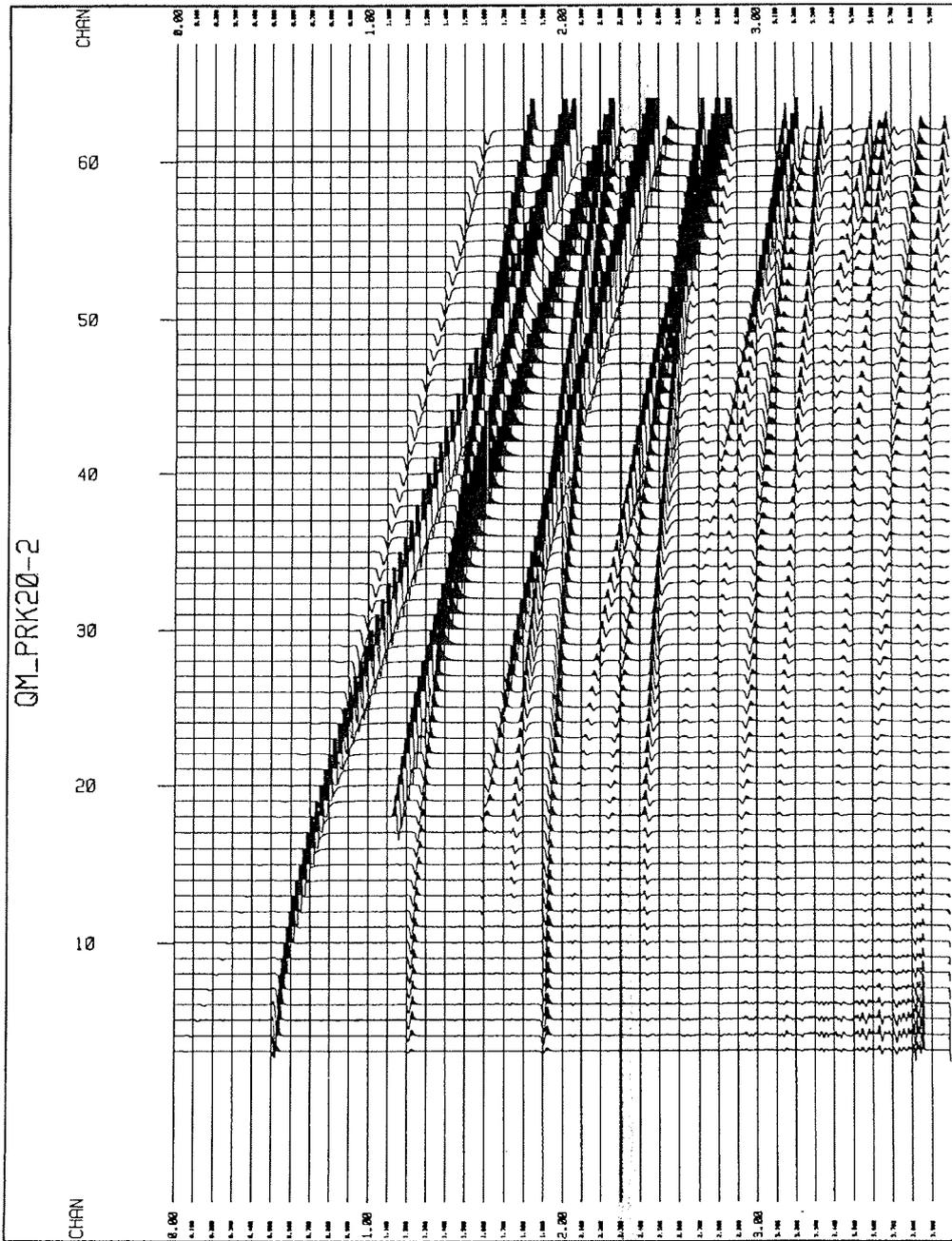


Abb. 4.4.1.1b



#### 4.4.2 Beispiel: Reale CSP-Seismogramme

Die in der Abbildung 4.4.2.1 dargestellten Schüsse 10 bis 20 des Hall-Bredning-Profiles wurden für die experimentelle Bestimmung der Filterparameter des Mehrkanalfilters benutzt. Die Charakteristika der Multiplen und Peg-Legs unterscheiden sich, wie erwartet, nicht von denen der Schüsse in Abb. 4.3.2.1.

Diese Schüsse wurden nun mit einem Mehrkanalfilter der Ordnung Acht und  $\lambda = 0.998$  bearbeitet, die Vorhersagedistanz wurde wieder aus den Wassertiefen auf der Basis von 1480 m/s Wasser-Schallgeschwindigkeit berechnet. Dies entspricht wiederum einer konservativen Einstellung des Filters.

Abbildung 4.4.2.2 zeigt das Ergebnis der Filterung. Bis auf schemenhafte Reste sind die Multiplenpakete völlig aus den Daten verschwunden.

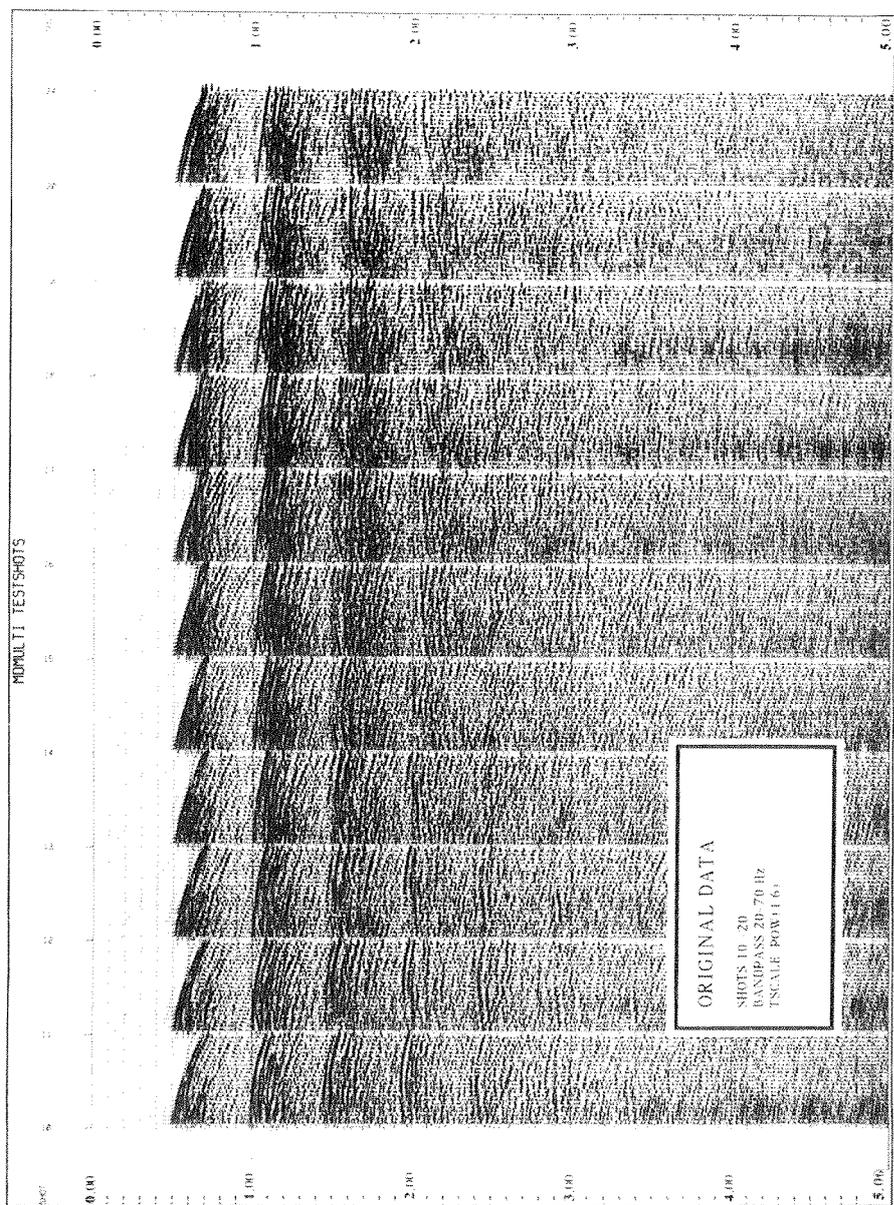
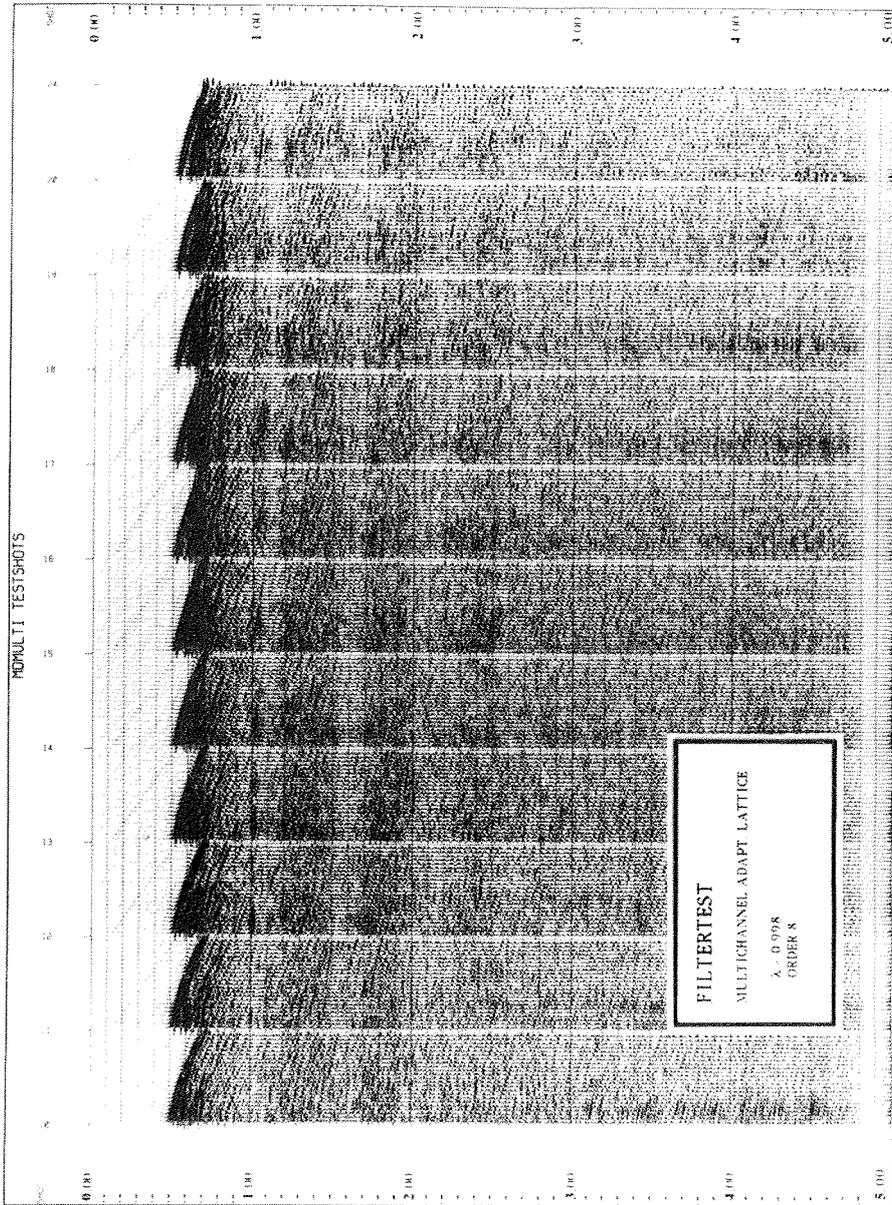


Abb. 4.4.2.1: Originaldaten, Schüsse 10-20, Hall-Bredning AWI-90550



*Abb. 4.4.2.2: Mehrkanalige Filterung der Schüsse 10-20*

#### 4.4.3 Beispiel: Reale Daten CDP-Stack, Prestack-Filterung

Mit dem oben beschriebenen Filter wurden nun alle Schüsse der 1444 CDPs bearbeitet und wie im Experiment mit dem Einkanalfilter, sozusagen "worst case", d.h. mit der Move-Out-Korrektur der Wasser-Schallgeschwindigkeit gestapelt. Abbildung 4.4.3.1 zeigt die gestapelte Sektion. Im Vergleich mit den Originaldaten kann man feststellen, daß die multiplen Reflexionen in hohem Maße unterdrückt sind. Offenbar sind jedoch die obersten Schichten des Meeresbodens aufgrund ihrer hohen Reflektivitäten, die ja auch für extrem starke Ausbildung von multiplen Reflexionen verantwortlich sind, so weitgehend undurchsichtig, daß für größere Laufzeiten keine weiteren Primärreflexionen erkennbar sind.

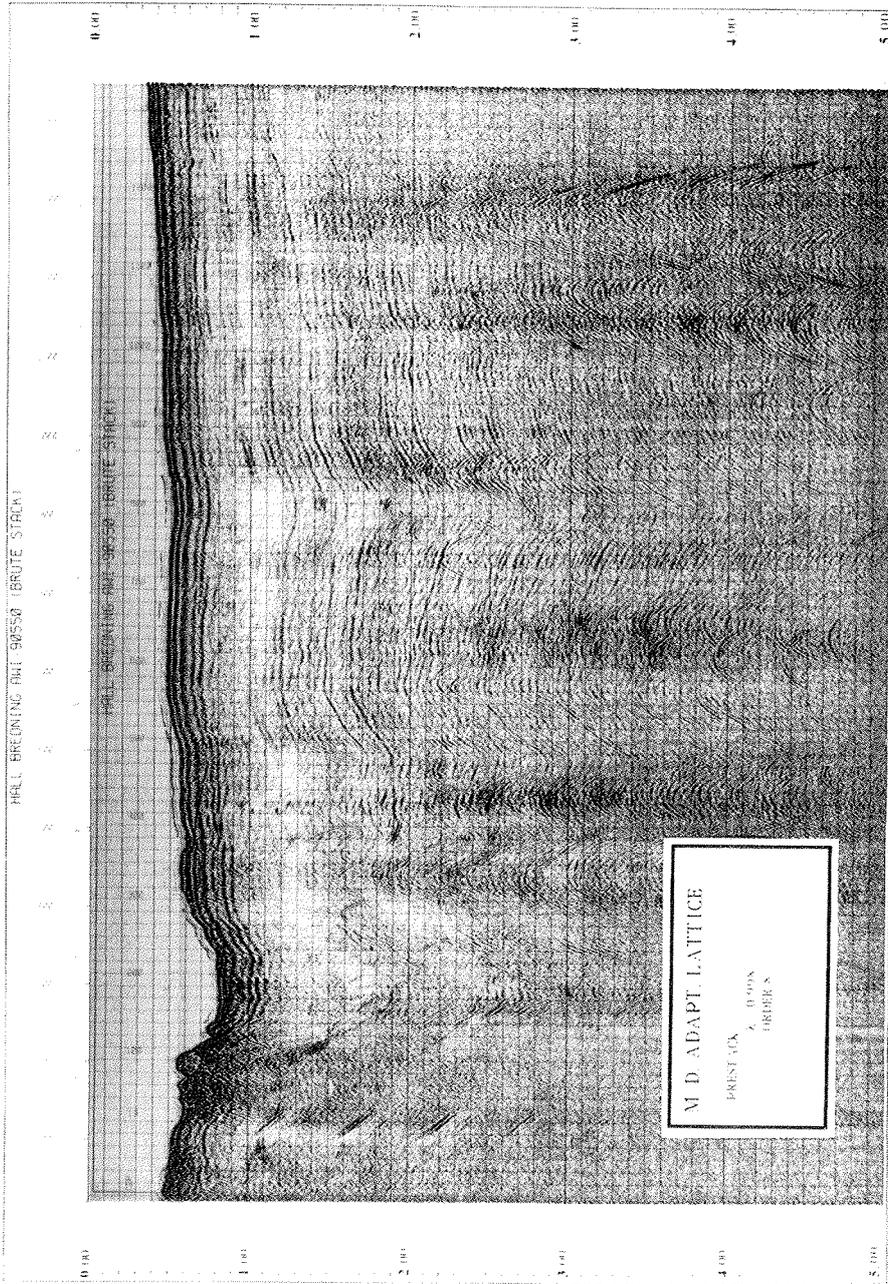


Abb. 4.4.3.1: Mehrkanalig gefilterte Daten, Filterung vor Stapelung

Eine Besonderheit in den Daten sind die, an verschiedenen Stellen isoliert auftauchenden Diffraktionshyperbeln, beispielsweise bei 1600 ms und CDP 400 , 950 ms und CDP 540 oder 1400 ms und CDP 900. Hall Bredning wird unter anderem vom Daugaard-Jensen-Gletscher in Nordwesten des Scoreby-Sund Fjordsystems mit zum Teil riesigen Eisbergen versorgt. Auch zum Zeitpunkt der Aufzeichnung der hier beispielhaft bearbeiteten Daten waren einige dieser Eisberge in der weiteren Umgebung des Schiffes vorhanden. Die Diffraktionen sind also mit großer Wahrscheinlichkeit laterale Reflexionen dieser, in unterschiedlicher Entfernung vom Profil treibenden Eisberge. Alle diese Diffraktionen, auch zu späteren Zeiten (z.B 3500 ms und CDP 750, die möglicherweise nicht von einem Eisberg sondern eher von der Küste stammt), bleiben von der Filterung unbeeinflusst. Wenn also auch in diesem Profil offenbar keine Primärreflexionen unter den Multiplen liegen, so mögen die Diffraktionen der Eisberge, die sich hier wie Primärreflexionen verhalten, einen weiteren Beweis für die korrekte Arbeitsweise des Filters liefern.

Anzumerken bleibt, daß, wie schon oben erwähnt, entgegen allen früheren Erwartungen die Mehrkanalfilterung ein relativ schnelles Verfahren ist. Die vollständige Bearbeitung von 24-Kanal-Daten in 1444 CDPs aus ca. 900 Einzelschüssen benötigte auf dem CONVEX 201-Vektorrechner des Alfred-Wegener-Instituts 724 Minuten CPU-Zeit.

## 5. Ungelöste Probleme

Das im Vergleich zum eindimensionalen Filterverfahren gute Ergebnis der Mehrkanalfilterung beruht nicht zuletzt auf der Tatsache, daß das Mehrkanal-Lattice Energie zwischen den einzelnen Kanälen propagiert.

Single vs. Multichannel Prediction

In der Abbildung 5.1 sind Einkanal- und Mehrkanal-Filter schematisch gegenübergestellt.

Dargestellt sind gedachte Filtermasken, die aus Datensamples, die innerhalb der Maske liegen, die Vorhersagewerte berechnen.

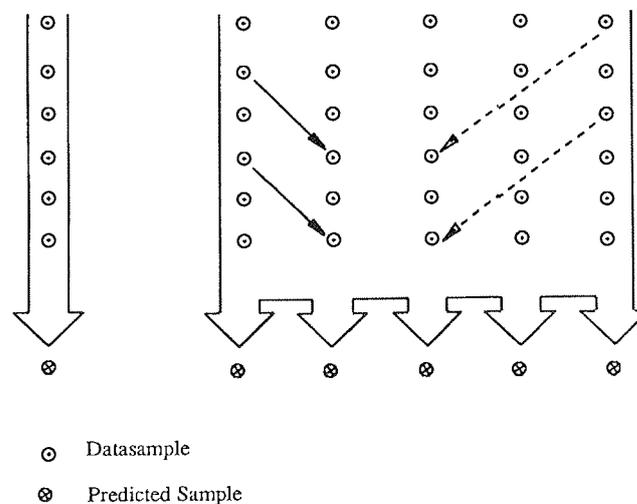


Abb. 5.1

Aufgrund der implizit vorausgesetzten Symmetrie propagiert die Energie nun nicht nur in einer physikalisch sinnvollen Richtung, wie sie beispielsweise die Abbildung 4.1.1 nahelegt und wie sie in Abb. 5.1 mit durchgezogenen Pfeilen angedeutet ist, sondern u.U. auch in der mit gestrichelten Pfeilen angedeuteten Richtung.

Physikalisch bewegt sich die Energie aufeinander folgender Multipleneinsätze vom Ort der Quelle hin zu größeren Offsets. Übertragen in den Formalismus des Mehrkanalfilters hieße dies, daß die Koeffizientenmatrizen  $A_i$  in Gleichung 3.6.1 untere Dreiecksmatrizen sein müssen.

Im Mehrkanal-Lattice-Algorithmus werden diese Matrizen nicht explizit berechnet, und in der hier verwendeten allgemeinen Form besteht keine Möglichkeit, eine Restriktion auf untere Dreiecksmatrizen zu erzwingen.

Dies kann in bestimmten Fällen zu Problemen führen, wenn der Filter in einer "unphysikalischen" Richtung einschwingt, wie sie in Abbildung A5.1 durch die gestrichelten Pfeile dargestellt ist. Diese Probleme werden beispielsweise dann auftauchen, wenn Einsätze mit weiten Offsets und entsprechend großem Move-Out innerhalb der Vorhersagedistanz mit Einsätzen korrelieren, die auf Spuren mit kleinen Offsets liegen. Abhilfe kann dann nur geschaffen werden, wenn ein solcher Datensatz in zwei Datensätze, in denen jeweils schußnahe Spuren und schußferne Spuren zusammengefaßt werden, zerlegt wird.

Die Frage, inwieweit ein ähnlich effizienter Algorithmus mit einer den physikalischen Zusammenhängen entsprechenden Restriktion aus dem adaptiven Mehrkanal-Lattice-Algorithmus abgeleitet werden kann, ist ungeklärt.

## 6. Zusammenfassung, Einschätzung

In dieser Arbeit wurden die theoretischen Grundlagen zweier Vertreter aus der Klasse der adaptiven Filter skizziert. Insbesondere wurde die problematische Numerik der Mehrkanalfilter analysiert und mögliche Lösungen aufgezeigt. Der Algorithmus von Lewis (Lewis 1990) trägt entscheidend zur Lösung des Konditionsproblems bei, das Rangentscheidungsproblem kann mit der Diskreten Cosinus-Transformation gelöst werden.

Während der eindimensionale Algorithmus hier hauptsächlich als Vergleichsverfahren mit behandelt wurde, war das Hauptanliegen dieser Arbeit, die Praktikabilität eines mehrdimensionalen Ansatzes für die Behandlung des Multiplenproblems in marinen seismischen Steilwinkeldaten nachzuweisen. Dieser Nachweis ist letztlich gelungen.

Insbesondere der Vergleich mit dem eindimensionalen Ansatz am Beispiel realer Daten zeigt klar die Überlegenheit des Mehrkanalalgorithmus. Die Resultate lassen es gerechtfertigt erscheinen, den Algorithmus im Sinne einer noch besseren Anpassung an die Charakteristika von multiplen Reflexionen weiterzuentwickeln. Auch hinsichtlich der Rechenzeiten scheint eine weitere Optimierung möglich.

Für die Behandlung des Multiplenproblems in marinen Steilwinkeldaten stehen nur wenige Algorithmen bislang zur Verfügung. In der Praxis sind dies fast ausschließlich die Wiener-Levinson-Algorithmen der Prädiktiven Dekonvolution. Wellengleichungsverfahren, Extrapolationsverfahren oder Dekompositionsverfahren (vgl. Clearbout 1985) haben noch keine weite Verbreitung in die industrielle Praxis gefunden.

Geschwindigkeitsfilter, die auf zweidimensionalen Transformationen beruhen, Radon- ( $\tau$ -p), 2D-Fourier- oder Karhunen-Loevé-Transformation, sind im Steilwinkelbereich oft nur bedingt brauchbar, die Move-Out Differenz zwischen primären

und multiplen Reflexionen im CDP-Seismogramm ist hier meist zu gering, als daß sie durch einen geschwindigkeitsselektiven Filter wirkungsvoll ausgebeutet werden könnte.

Um einen Vergleich zu ermöglichen, wurde das Hall-Bredning Profil mit einem Frequenz-Wellenzahl- ( $f/k$ -) Filter bearbeitet. Das am Alfred-Wegener-Institut verfügbare DISCO-Programmpaket (Cogniseis<sup>TM</sup>) enthält eine Routine, die einen Geschwindigkeits-Filter im 2D-Fourier-Raum realisiert. Hierzu muß zunächst eine Analyse der Stapelgeschwindigkeiten (RMS-Geschwindigkeiten) der CDP-Ensembles durchgeführt werden. Die Stapelgeschwindigkeiten des CDP-Ensembles nehmen mit wachsender Laufzeit monoton zu, der Filter unterdrückt Einsätze, deren Stapelgeschwindigkeiten kleiner sind als es der ermittelten Laufzeit-Geschwindigkeitskurve entspricht.

Schon die Ermittlung der Stapelgeschwindigkeiten ist eine Aufgabe, die umso schwerer zu lösen ist, je weniger ausgeprägt die Move-Out-Differenzen im CDP-Ensemble hervortreten.

In der Abbildung 6.1 ist das Profil nach der  $f/k$ -Filterung dargestellt. Auch hier wurde nach der Filterung wiederum die Stapelung mit der Move-Out-Korrektur der Wasserschallgeschwindigkeit durchgeführt, um die Residuen deutlich hervortreten zu lassen. Zunächst fällt auf, daß die Einsätze nach der Meeresboden-Reflexion zum großen Teil verschwunden sind, ihre Geschwindigkeit unterscheidet sich nicht genügend von den Einsätzen mit Wasserschallgeschwindigkeit, die ja vom Filter unterdrückt werden. Die erste Multiple der Meeresbodenreflexion ist dagegen nur wenig beeinflusst. Peg-Leg-Multiple und alle Multiplen höherer Ordnung sind dagegen recht gut bedämpft.

Es ist nicht ratsam, hier eine Wertung gegenüber dem, in dieser Arbeit vorgeschlagenen, adaptiven Mehrkanal-Filter vorzunehmen. Die Methode der  $f/k$ -Filterung steht und fällt mit einer guten Analyse der Stapelgeschwindigkeiten. Wo diese, wie in diesem Datensatz, schwierig und zum Teil mit großen Unsicherheiten

behaftet ist, werden f/k-Filter keine befriedigenden Ergebnisse zeigen.

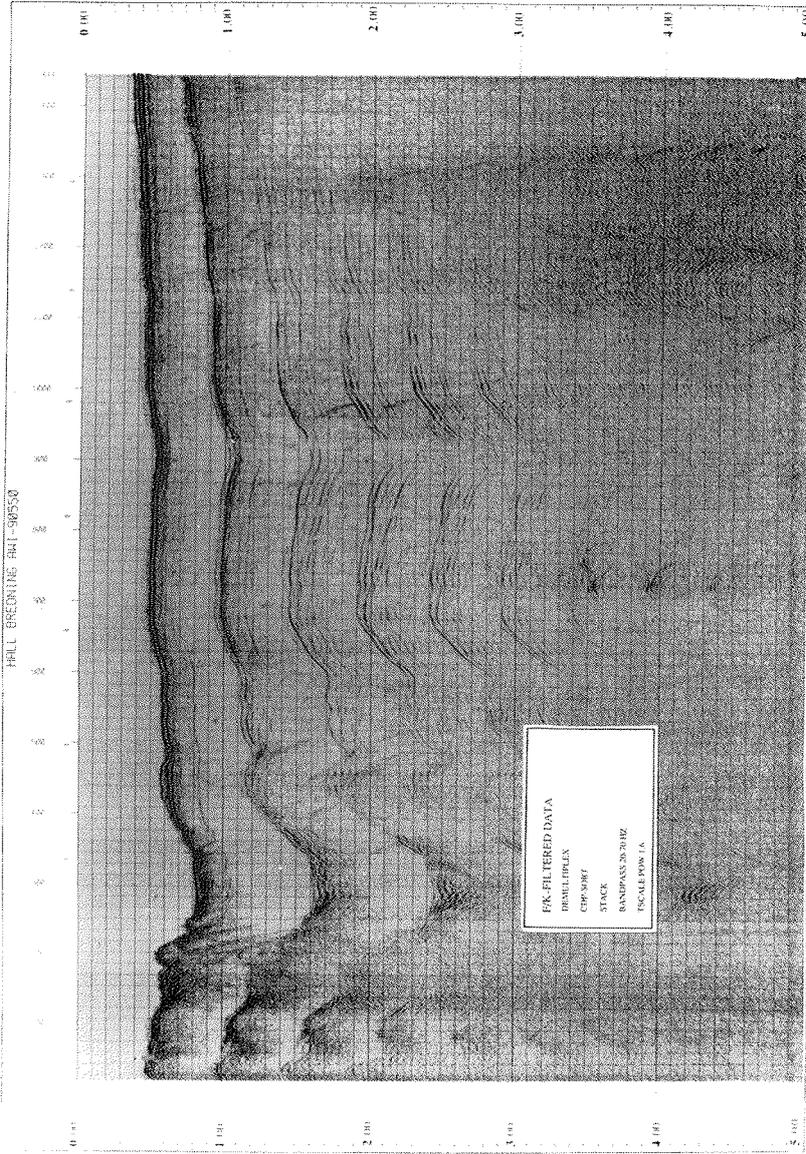


Abb. 6.1: Mit Geschwindigkeitsfilter bearbeitete Daten

Von keiner Methode ist zu erwarten, daß sie auf allen denkbaren Datensätzen gleich gut arbeitet. Das gilt in gleichem Maße auch für das hier vorgeschlagene Verfahren, auch wenn hier noch keine umfangreichen Erfahrungen vorliegen. In diesem Sinne ist der in dieser Arbeit vorgestellte adaptive Mehrkanalfilter jedoch eine sinnvolle und praktikable Ergänzung der für den Steilwinkelbereich bereits existierenden Verfahren.

Die beschriebenen Algorithmen sind prinzipiell auch für die Echtzeit-Verarbeitung von Daten geeignet, mit einer speziellen Hardware ließen sich so möglicherweise multiplenbereinigte Online-Monitor-Aufzeichnungen schon während der Datenakquisition erzeugen, die eine bessere Beurteilung der Datenqualität noch während der Aufzeichnung erlauben.

Adaptive Filter im Joint-Estimation-Modus böten außerdem die Möglichkeit, in Echtzeit deterministisches und auch nicht-deterministisches Rauschen in kontinuierlich abgetasteten Daten zu beseitigen, solange es nur möglich ist, dieses Rauschen isoliert in einem zweiten Kanal aufzuzeichnen. Sehr illustrative Beispiele für Anwendungen, die diesem Prinzip folgen, finden sich in Widrow (Widrow 1975), auch wenn die dort verwendeten Algorithmen nicht dem neuesten Stand entsprechen.



## 7. Literatur

- Alexander, S.T., **1986**, Adaptive signal processing, theory and applications: Springer Verlag, N.Y.
- Baumeister, J. **1987**, Stable solutions of inverse problems: Vieweg advanced lectures in mathematics, Vieweg Verlag, Braunschweig
- Bellanger, M.G. **1987**, Adaptive digital filters and signal analysis: Marcel Dekker Inc., N.Y.
- Bellanger, M.G., **1989**, The fls-qr algorithm for adaptive filtering: Signal Processing, Elsevier Science Publishers, vol 17, pp.291-304
- Boonton,R.C., **1952**,An optimization theory for time varying linear systems with non-stationary statistical inputs: Proc. IRE v.40, p977-981 (zit. nach Griffith et al. 1977).
- Bruckstein,A.,Kailath, T, **1987**, An inverse scattering framework for several problems in signal processing: IEEE Trans. ASSP Magazine, Jan. 1987, pp. 6-20
- Burg, J. **1975**, Maximum entropy spectral analysis, Ph.D. thesis, Dept. of Geophysics, Stanford Univ., CA. , zit. nach Alexander (Alexander, 1986)
- Carayannis, G.,Manolakis, D., Kalouptsidis,N., **1986**, A unified view of parametric processing algorithms for prewindowed signals: Signal Processing, Elsevier Science Publishers, vol 10, pp. 335-368

- Chui,C.K., Chen,G. **1987**, Kalman filtering with real-time applications:  
Springer series in information sciences, Springer-  
Verlag, Berlin Heidelberg
- Clarke, R.J., Tech, B., **1981**, Relation between the Karhunen Loève and  
cosine transforms: IEE Proc., vol. 128, Pt. F, no. 6, pp.  
359-360
- Clearbout, J.F., **1985**, Imaging the earth interior: Blackwell scientific  
publishers, London 1985
- Dorau, F. **1991**, Eindimensionale Inversion bandbegrenzter  
Seismogramme: Dissertation, Institut f. Geophysik,  
Universität Hamburg.
- Griffiths, L.J., Smolka,F.R., Trembly,L.D., **1977**, Adaptive deconvolution:  
a new technique for processing time varying seismic  
data: Geophysics, v.42, no.4, pp. 742-759.
- Hager, W.W. **1988**, Applied numerical algebra: Prentice-Hall,  
London
- Jones, L.F., Levy, S., **1987**, Signal-to-noise ratio enhancement in  
multichannel seismic data via the Karhunen-Loevé  
Transform: Geophysical Prospecting, vol. 35, pp. 12-32
- Koehler, F., Taner, M.T., **1985**, The use of the conjugate gradient  
algorithm in the computation of predictive  
deconvolution operators: Geophysics, v. 50, no. 12, pp.  
2752-2758
- Lee, D.T.L., Morf, M., Friedlander, B., **1981**, Recursive least squares  
ladder estimation algorithms: IEEE Trans. ASSP, vol.  
29, no. 3, pp. 627-641

- Lewis, P.S., **1990**, QR-based algorithms for multichannel adaptive least squares lattice filters: IEEE Trans. ASSP, v. 38, no. 3, pp. 421-432
- Miller, M.J., Thorpe, A.J., **1981**, Radiation conditions for the lateral boundaries of limited-area numeric models: Quart. J. R. Met. Soc, 107, pp.615-628
- Orlanski, J. **1976**, A simple boundary condition for unbounded hyperbolic flows: Journal of Computational Physics, Vol. 21, pp. 251-269
- Peacock, K.L. and Treitel, S., **1969**, Predictive deconvolution: theory and practice: Geophysics, v.34, no.2, pp. 155 -169
- Pratt, W.K., **1972**, Generalized Wiener filtering techniques: IEEE Trans. On Computers, vol. c-21, no. 7, pp. 636-641
- Robinson,E.A., **1954**, Predictive decomposition of time series with applications to seismic exploration: Ph.D. thesis, MIT, Cambridge, Mass., zit. nach Peacock (Peacock, 1969)
- Robinson, E.A. and Treitel, S., **1967**, Principles of digital Wiener filtering: Geophys. Prosp., v. 15,pp. 311-333
- Robinson, E.A. and Treitel, S., **1980**, Maximum entropy and the relationship of the partial autocorrelation to the reflection coefficients of a layered system: IEEE Trans. ASSP, v. 28, no. 2, pp. 224-235
- Rosenfeld, A., Kak, A.C, **1982**, Digital picture processing: Academic Press, N.Y.
- Sondhi, M. **1967**, An adaptive echo canceller: Bell Syst. Tech. J., v. 46, pp.497-511, zit. nach Widrow et al. 1975

- Taner, M.T., **1980**, Long period sea-floor multiples and their suppression: Geophysical Prospecting, v. 28, pp.30-40
- Taner, M.T., O'Doherty, R.F., **1990**, Conjugate gradient x-t-deconvolution: SEG San Francisco 1990
- Voß, H. **1987**, Skriptum zur Vorlesung Numerische Mathematik, Inst.f. Angew. Mathematik, Universität Hamburg, Sommersemester 1987
- Wang, R.J. **1969**, The determination of optimum gate length for time-varying wiener filtering: Geophysics, v.34, no. 5, pp. 683-695.
- Wiener, N., **1949**, Extrapolation, interpolation and smoothing of stationary time series: John Wiley and Sons, Inc., N.Y. (zit. nach Griffith et al. 1977)
- Widrow, B., Mantey, P., Griffith, L., **1967**, Adaptive antenna Systems: Proc. IEEE, v. 55, pp.2143-2159
- Widrow, B., Glover, J.R., McCool, J.M., Kaunitz, J., Williams, C.S., Hearn, R.H., Zeidler, J.R., Dong, E.D., Goodlin, R.C., **1975**, Adaptive noise cancelling: Principles and applications: Proc. IEEE, v. 63, no. 12, pp. 1692-1716