

Deterministische Signalverarbeitung

Repräsentation und Approximation in Vektorräumen

Univ.-Prof., Dr.-Ing. Markus Rupp
SS 04

Letzte Änderung: 11.04.2004

Lernziele

- Repräsentation und Approximation in Vektorräumen
 - Approximationsproblem im Hilbert Raum
 - Orthogonalitätsprinzip
 - Minimierung mit Gradientenverfahren
 - Least Squares Filterung,
 - lineare Regression, parametrische Schätzung, Iteratives LS Problem
 - Signaltransformation und verallgemeinerte Fourierreihe
 - Beispiele für orthogonale Funktionen, Wavelets
 - Mehrraten Systeme

Motivation

- Der Erfolg der modernen Nachrichtentechnik besteht darin, Informationen unter optimaler Ausnutzung der vorhandenen begrenzten Bandbreiten und in gestörter Umgebung ohne wesentliche Verluste zu übertragen.
- Dieser Erfolg beruht vor allem auf einer erfolgreichen Querkodierung, die es erlaubt nicht benötigte Redundanz aus den Signalen zu nehmen, um damit die zu übertragene Datenrate gering zu halten.
- Damit verbunden ist eine Approximation der Signale.

Approximation im Hilbertraum

- Problemstellung: Sei $(S, ||\cdot||)$ ein linearer, genormter Vektorraum und $T = \{p_1, p_2, \dots, p_m\}$ eine Teilmenge von linear unabhängigen Vektoren aus S und $V = \text{span}(T)$. Gegeben ein Vektor $\underline{x}$ aus S , finde die Koeffizienten c_m derart, dass

$$\hat{\underline{x}} = c_1 \underline{p}_1 + c_2 \underline{p}_2 + \dots + c_m \underline{p}_m$$

- $\underline{x}$ durch diese Linearkombination bestmöglich angenähert wird, also der Fehler $\underline{e}$

$$\underline{e} = \underline{x} - \hat{\underline{x}}$$

- minimal wird.

Approximation im Hilbertraum

- Um $\underline{e}$ zu minimieren, ist es vorteilhaft eine Norm einzuführen.
- Würde man die l_1 -oder l_{∞} -Norm einführen, so wäre das Problem mathematisch sehr schwierig zu lösen.
- Nutzt man jedoch die induzierte l_2 -Norm, so erhält man typischerweise quadratische Gleichungen, die durch Ableitung lösbar sind.
- Wir werden später noch iterative LS Verfahren vorstellen, die auch Lösungen in anderen Normen zulassen.

Approximation im Hilbertraum

- Beachte, dass wenn $\underline{x}$ in V liegt der Fehler natürlich Null werden kann.
- Liegt aber $\underline{x}$ nicht in V , so ist es nur möglich einen möglichst kleinen Wert für $\|\underline{e}\|_2$ zu finden.
- Anwendung: Sei $\underline{x}$ ein zu übertragendes Signal. Durch die Vektoren in T wird eine approximative Lösung bestimmt mit sehr kleinem Fehler $\|\underline{e}\|_2$. Am Empfänger ist T bekannt. Somit müssen nur m Koeffizienten c_m übertragen werden. Ist die Anzahl m der Koeffizienten wesentlich geringer als die Anzahl der Abtastwerte in $\underline{x}$, so erhält man eine erhebliche Datenreduktion.

Approximation im Hilbertraum

- Um das Problem und seine Lösung möglichst anschaulich zu gestalten, betrachten wir zunächst nur einen Vektor $T=\{\underline{p}_1\}$.
- Somit gilt: $\underline{x} = c_1 \underline{p}_1 + \underline{e}$.
- Zu minimieren ist:

$$\begin{aligned} \min \|\underline{e}\|_2^2 &= \min_{c_1} \|\underline{x} - c_1 \underline{p}_1\|_2^2 = \min_{c_1} (\underline{x} - c_1 \underline{p}_1)^T (\underline{x} - c_1 \underline{p}_1) \\ \frac{\partial}{\partial c_1} (\underline{x} - c_1 \underline{p}_1)^T (\underline{x} - c_1 \underline{p}_1) &= \frac{\partial}{\partial c_1} (\underline{x}^T \underline{x} - c_1 \underline{p}_1^T \underline{x} - c_1 \underline{x}^T \underline{p}_1 + c_1^2 \underline{p}_1^T \underline{p}_1) \\ &= -\underline{p}_1^T \underline{x} - \underline{x}^T \underline{p}_1 + 2c_1 \underline{p}_1^T \underline{p}_1 = 0 \\ c_1 &= \frac{\underline{x}^T \underline{p}_1}{\underline{p}_1^T \underline{p}_1} = \frac{\langle \underline{x}, \underline{p}_1 \rangle}{\|\underline{p}_1\|_2^2} \end{aligned}$$

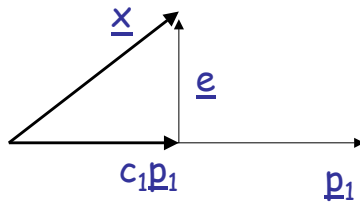
Approximation im Hilbertraum

- Beachte, dass dieses Problem auch durch Ableiten nach einem Vektor erreicht werden kann:

$$\begin{aligned} \min \|\underline{e}\|_2^2 &= \min_{c_1} \|\underline{x} - c_1 \underline{p}_1\|_2^2 = \min_{c_1} (\underline{x} - c_1 \underline{p}_1)^T (\underline{x} - c_1 \underline{p}_1) \\ \frac{\partial}{\partial c_1} (\underline{x} - c_1 \underline{p}_1)^T (\underline{x} - c_1 \underline{p}_1) &= -\underline{p}_1^T (\underline{x} - c_1 \underline{p}_1) - (\underline{x} - c_1 \underline{p}_1)^T \underline{p}_1 \\ &= -\underline{p}_1^T \underline{x} - \underline{x}^T \underline{p}_1 + 2c_1 \underline{p}_1^T \underline{p}_1 = 0 \\ c_1 &= \frac{\underline{x}^T \underline{p}_1}{\underline{p}_1^T \underline{p}_1} = \frac{\langle \underline{x}, \underline{p}_1 \rangle}{\|\underline{p}_1\|_2^2} \end{aligned}$$

Approximation im Hilbertraum

- Die Vorgehensweise lässt sich anschaulich geometrisch interpretieren:



- Es hat den Anschein, als ob der Fehler $\underline{e}$ und $\underline{p}_1$ senkrecht aufeinander stehen.

Approximation im Hilbertraum

- Genauere Analyse zeigt:

$$\underline{e}^T \underline{p}_1 = (\underline{x} - c_1 \underline{p}_1)^T \underline{p}_1 = \left(\underline{x} - \frac{\langle \underline{x}, \underline{p}_1 \rangle}{\|\underline{p}_1\|_2^2} \underline{p}_1 \right)^T \underline{p}_1$$

$$= \underline{x}^T \underline{p}_1 - \langle \underline{x}, \underline{p}_1 \rangle = 0$$

- In der Tat ist der Koeffizient so gewählt, dass der Fehler $\underline{e}$ senkrecht auf $\underline{p}_1$ steht.
- Beachte auch:

$$\underline{e}^T \underline{e} = \underline{e}^T (\underline{x} - c_1 \underline{p}_1) = \underline{e}^T \underline{x} - c_1 \underline{e}^T \underline{p}_1 = \underline{e}^T \underline{x}$$

Approximation im Hilbertraum

- Um die Vorgehensweise für mehrere Vektoren zu zeigen, müssen wir systematisch vorgehen. Wir wollen $\underline{x}$ durch die Linearkombination

$$\hat{\underline{x}} = c_1 \underline{p}_1 + c_2 \underline{p}_2 + \dots + c_m \underline{p}_m$$

- approximieren, so dass die Norm des Fehlers minimal wird, also

$$\frac{\partial}{\partial c_k} \|\underline{e}\|_2^2 = \left(\frac{\partial}{\partial c_k} \underline{e} \right)^T \underline{e} + \underline{e}^T \underbrace{\frac{\partial}{\partial c_k} \underline{e}}_{-\underline{p}_k} = 0; k = 1, 2, \dots, m$$

Approximation im Hilbertraum

- Somit erhalten wir also m Gleichungen, die wir in einer Matrixform zusammenfassen können:

$$\begin{bmatrix} \langle \underline{p}_1, \underline{p}_1 \rangle & \langle \underline{p}_2, \underline{p}_1 \rangle & \dots & \langle \underline{p}_m, \underline{p}_1 \rangle \\ \langle \underline{p}_1, \underline{p}_2 \rangle & \langle \underline{p}_2, \underline{p}_2 \rangle & & \langle \underline{p}_m, \underline{p}_2 \rangle \\ \vdots & & \ddots & \vdots \\ \langle \underline{p}_1, \underline{p}_m \rangle & \langle \underline{p}_2, \underline{p}_m \rangle & \dots & \langle \underline{p}_m, \underline{p}_m \rangle \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} = \begin{bmatrix} \langle \underline{x}, \underline{p}_1 \rangle \\ \langle \underline{x}, \underline{p}_2 \rangle \\ \vdots \\ \langle \underline{x}, \underline{p}_m \rangle \end{bmatrix}$$

$$R \underline{c} = \underline{p}$$

- Die Lösung dieses Gleichungssystems wird (lineare) Least-Square Lösung genannt.

Approximation im Hilbertraum

- Ob solch ein Gleichungssystem eine eindeutige Lösung hat, hängt entscheidend von der Beschaffenheit der Matrix R ab.
- Definition 3.1:** Eine durch innere Vektorprodukte $\underline{p}_i$ $i=1,2,\dots,m$ aus T gebildete $m \times m$ Matrix R heißt Grammian der Menge T .

$$R = \begin{bmatrix} \langle \underline{p}_1, \underline{p}_1 \rangle & \langle \underline{p}_2, \underline{p}_1 \rangle & \dots & \langle \underline{p}_m, \underline{p}_1 \rangle \\ \langle \underline{p}_1, \underline{p}_2 \rangle & \langle \underline{p}_2, \underline{p}_2 \rangle & & \langle \underline{p}_m, \underline{p}_2 \rangle \\ \vdots & & \ddots & \vdots \\ \langle \underline{p}_1, \underline{p}_m \rangle & \langle \underline{p}_2, \underline{p}_m \rangle & \dots & \langle \underline{p}_m, \underline{p}_m \rangle \end{bmatrix}$$

- Es gilt: $R_{ij} = \langle \underline{p}_j, \underline{p}_i \rangle$

$$R^H = R$$

- Zur eindeutigen Lösung muss R positiv-definit sein.
- Definition 3.2:** Eine Matrix heißt positiv-definit, wenn für beliebige Vektoren $\underline{q}$ ungleich Null gilt:

$$\underline{q}^H R \underline{q} > 0$$

Approximation im Hilbertraum

- Satz 3.1:** Die Grammian Matrix R ist positiv semidefinit. Sie ist dann und nur dann positiv-definit, wenn die Vektoren $\underline{p}_1, \underline{p}_2, \dots, \underline{p}_m$ linear unabhängig sind.
- Beweis:** Sei $\underline{q} = [q_1, q_2, \dots, q_m]$ ein beliebiger Vektor:

$$\begin{aligned} \underline{q}^H R \underline{q} &= \sum_{i=1}^m \sum_{j=1}^m q_i^* q_j R_{ij} = \sum_{i=1}^m \sum_{j=1}^m q_i^* q_j \langle \underline{p}_j, \underline{p}_i \rangle \\ &= \sum_{i=1}^m \sum_{j=1}^m \langle q_j \underline{p}_j, q_i \underline{p}_i \rangle = \left\langle \sum_{j=1}^m q_j \underline{p}_j, \sum_{i=1}^m q_i \underline{p}_i \right\rangle \\ &= \left\| \sum_{i=1}^m q_i \underline{p}_i \right\|_2^2 \geq 0 \end{aligned}$$

Approximation im Hilbertraum

- Ist R nicht positiv-definit, so muss es einen Vektor $\underline{q}$ geben (ungleich dem Nullvektor) für den gilt:

$$\underline{q}^H R \underline{q} = 0$$

- Also auch:

$$\left\| \sum_{i=1}^m q_i \underline{p}_i \right\|_2^2 = 0$$

$$\sum_{i=1}^m q_i \underline{p}_i = \underline{0}$$

- Was ja bedeutet, dass die Vektoren $\underline{p}_1, \underline{p}_2, \dots, \underline{p}_m$ linear abhängig sind.

Approximation im Hilbertraum

- Beachte, dass diese Methode durch die Matrixinversion der Grammian Matrix einen erheblichen Rechenaufwand erfordert.
- Dies kann man dadurch erheblich reduzieren, wenn man die Vektoren $\underline{p}_1, \underline{p}_2, \dots, \underline{p}_m$ orthonormal wählt.
- In diesem Falle wird die Grammian zur Einheitsmatrix.
- Wir werden uns daher im wesentlichen auf die Suche nach orthonormalen Basen begeben.

Orthogonalitätsprinzip

- **Satz 3.2:** Sei $(S, ||\cdot||)$ ein linearer genormter Vektorraum und $T=\{\underline{p}_1, \underline{p}_2, \dots, \underline{p}_m\}$ eine Teilmenge von linear unabhängigen Vektoren aus S und $V=\text{span}(T)$. Gegeben sei ein Vektor $\underline{x}$ aus S . Die Koeffizienten c_m , die durch die Linearkombination

$$\hat{\underline{x}} = c_1 \underline{p}_1 + c_2 \underline{p}_2 + \dots + c_m \underline{p}_m$$

den Fehler $\underline{e}$ in Form der induzierten l_2 -Norm minimieren, verursachen, dass der Fehlervektor $\underline{e}$ orthogonal zu allen Vektoren in T liegt.

Orthogonalitätsprinzip

- **Beweis:** zu zeigen (durch Substitution)

$$\begin{aligned} \langle \underline{e}, \underline{p}_j \rangle &= 0 \\ &= \left\langle \underline{x} - \sum_{i=1}^m c_i \underline{p}_i, \underline{p}_j \right\rangle \xrightarrow{\text{für alle } j} \underline{p} - R\underline{c} = \underline{0} \end{aligned}$$

- Beachte: dadurch, dass $\underline{e}$ auf jedem Vektor $\underline{p}_j$ orthogonal steht, muss $\underline{e}$ auch dem Schätzwert orthogonal stehen:

$$\langle \underline{e}, \hat{\underline{x}} \rangle = 0$$

Orthogonalitätsprinzip

- **Beispiel 3.1:** Ein nichtlineares System $f(x)$ werde harmonisch angeregt. Welche Amplituden haben die Oberschwingungen?
- Eine Möglichkeit besteht darin, das nichtlineare System in Form von Polynomen anzugeben (exakt oder approximiert). Bei hohem Polynomgrad kann dies sehr aufwendig zu berechnen sein.
- Eine andere Möglichkeit besteht darin, das Ausgangssignal als gegeben anzunehmen:

$$\begin{aligned} f(\sin(x)) &= a_0 + a_1 \sin(x) + a_2 \sin(2x) + \dots + b_1 \cos(x) + b_2 \cos(2x) + \dots \\ e(x) &= f(\sin(x)) - (a_0 + a_1 \sin(x) + a_2 \sin(2x) + \dots + b_1 \cos(x) + b_2 \cos(2x) + \dots) \end{aligned}$$

- Da die Funktionen eine orthogonale Basis bilden, lassen sich die Ergebnisse leicht berechnen.

Gradientenverfahren

- Liegen die Vektoren $\underline{p}_1, \underline{p}_2, \dots, \underline{p}_m$ nicht als orthonormale (orthogonale) Menge vor, so kann die Matrixinversion beschwerlich sein.
- Eine Abhilfe können iterative Verfahren nach dem Gradientenprinzip (Verfahren des stärksten Abstiegs, engl.: steepest descent method) sein.
- Hierbei löst man nach $\underline{c}$ stückweise:

$$\underline{c}_k = \underline{c}_k + \mu(\underline{p} - R\underline{c}_k)$$

- Statt einer Matrixinversion wird also mehrfach eine Matrixmultiplikation angewendet.

Gradientenverfahren

- Der Wahl der Schrittweite μ kommt hierbei große Bedeutung zu.
 - Ist μ zu klein, so müssen viele Iterationen durchgeführt werden, um die Lösung zu finden.
 - Ist μ zu groß, so konvergiert das Verfahren nicht.
- Genauer erfährt man in
 - Vorlesung stochastische Signalverarbeitung
 - Vorlesung adaptive Filterverfahren

Least Squares Filterung

- Schreiben wir das bisherige Problem in Matrix-Form durch Verwendung einer Matrix $A=[p_1, p_2, \dots, p_m]$:

$$\hat{\underline{x}} = A\underline{c}$$

$$\langle \underline{x} - A\underline{c}, A \rangle = 0; \text{ Orthogonalität zu jedem } \underline{p}_j$$

$$\langle \underline{x} - A\underline{c}, A \rangle = \langle A^H, \underline{x} - A\underline{c} \rangle = A^H \underline{x} - A^H A\underline{c} = 0$$

$$A^H A\underline{c} = A^H \underline{x}$$

$$\underline{c} = (A^H A)^{-1} A^H \underline{x} = B\underline{x}$$

Least Squares Filterung

- **Definition 3.3:** Die Matrix $B=(A^H A)^{-1} A^H$ heißt Pseudoinverse von A .
- Beachte, dass $\underline{c}$ aus einer linearen Transformation B der Beobachtung $\underline{x}$ gewonnen wird.
- Ebenso erhält man den Schätzwert aus einer linearen Transformation der Beobachtung $\underline{x}$:

$$\hat{\underline{x}} = A\underline{c} = A(A^H A)^{-1} A^H \underline{x}$$

- Die Matrix $P_A = A(A^H A)^{-1} A^H$ heißt Projektionsmatrix und verdient nähere Betrachtung.

Least Squares Filterung

- Sei S ein linearer Vektorraum, der sich aus den zwei disjunkten (engl. disjoint) Teilräumen W und V zusammensetzt: $S=W+V$. Damit lässt sich ein Vektor $\underline{x}$ aus S eindeutig zusammensetzen aus einem Vektor $\underline{w}$ aus W und einem Vektor $\underline{v}$ aus V .
- Wenn doch diese Zusammensetzung eindeutig ist, dann müsste doch auch $\underline{w}$ und $\underline{v}$ aus $\underline{x}$ wieder gewonnen werden können.
- Der Operator P_v der $\underline{x}$ auf $\underline{v} = P_v \underline{x}$ abbildet, heißt Projektionsoperator, oder kurz: Projektion.

Least Squares Filterung

- **Definition 3.4:** Eine lineare Abbildung eines linearen Vektorraums auf sich selbst heißt Projektion, wenn gilt: $P=P^2$. Solch ein Operator wird auch idempotent genannt.
- Offensichtlich handelt es sich bei $P_A=A(A^HA)^{-1}A^H$ um eine Projektionsmatrix.
- **Lemma 3.1:** Wenn P eine Projektionsmatrix ist, dann ist $I-P$ auch eine Projektionsmatrix.
- **Beweis:** $(I-P)^2=I-2P+P^2=I-P$

Least Squares Filterung

- Damit gilt für den LS Fehler $\underline{e}$:

$$\begin{aligned}\underline{e} &= \underline{x} - A\underline{c} = \underline{x} - A(A^HA)^{-1}A^H\underline{x} \\ &= \left(I - A(A^HA)^{-1}A^H\right)\underline{x}\end{aligned}$$

- Auch der Fehler geht somit durch lineare Transformation (Projektion) aus der Beobachtung $\underline{x}$ hervor.
- Da der LS Fehler $\underline{e}$ und der Schätzwert von $\underline{x}$ senkrecht aufeinander stehen, gilt:

$$\|\underline{e}_{LS}\|^2 = \underline{x}^H \underline{e}_{LS} = \underline{x}^H \left[I - A(A^HA)^{-1}A^H\right]\underline{x} \leq \|\underline{x}\|^2$$

Least Squares Filterung

- **Anwendungen:**
 - Data/Curve Fitting
 - Parameter Schätzung
 - Filter Design

Least Squares Filterung

- **Data/Curve fitting**
 - Gegeben seien Beobachtungsdaten in Form von Wertepaaren (x_i, y_i) . Eine bestimmte Kurvenform die den Zusammenhang zwischen x und y Daten beschreibt, wird vermutet. Man möchte diese optimal an die gegebenen Daten anpassen.
 - **Beispiel 3.2:** Polynomannpassung
gegeben sei eine Funktion $f(x)$; diese soll durch ein Polynom der Ordnung m möglichst gut im Intervall $[a, b]$ angepasst werden. Hierzu wird eine quadratische Kostenfunktion gewählt:

Least Squares Filterung

- Also $\min_{\{c_0, c_1, \dots, c_{m-1}\}} \int_a^b (f(x) - c_0 - c_1 x - \dots - c_{m-1} x^{m-1})^2 dx$
- Aufgrund des Orthogonalitätsprinzips wissen wir, dass

$$\langle p(x), p(x) \rangle = \langle f(x), p(x) \rangle$$

$$\begin{bmatrix} \langle 1, 1 \rangle & \langle 1, x \rangle & \dots & \langle 1, x^{m-1} \rangle \\ \langle x, 1 \rangle & \langle x, x \rangle & & \\ \vdots & & \ddots & \\ \langle x^{m-1}, 1 \rangle & \dots & \langle x^{m-1}, x^{m-1} \rangle \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{m-1} \end{bmatrix} = \begin{bmatrix} \langle f, 1 \rangle \\ \langle f, x \rangle \\ \vdots \\ \langle f, x^{m-1} \rangle \end{bmatrix}$$

$$\langle x^i, x^j \rangle = \int_a^b x^{i+j} dx = \frac{b^{i+j+1} - a^{i+j+1}}{i+j+1}$$

Least Squares Filterung

- Normiert man das Intervall auf den Bereich [0,1], so erhält man die so genannte Hilbert Matrix:

$$\begin{bmatrix} 1 & \frac{1}{2} & \dots & \frac{1}{m} \\ \frac{1}{2} & \frac{1}{3} & & \frac{1}{m+1} \\ & \ddots & \ddots & \vdots \\ \frac{1}{m} & \dots & \dots & \frac{1}{2m-1} \end{bmatrix}$$

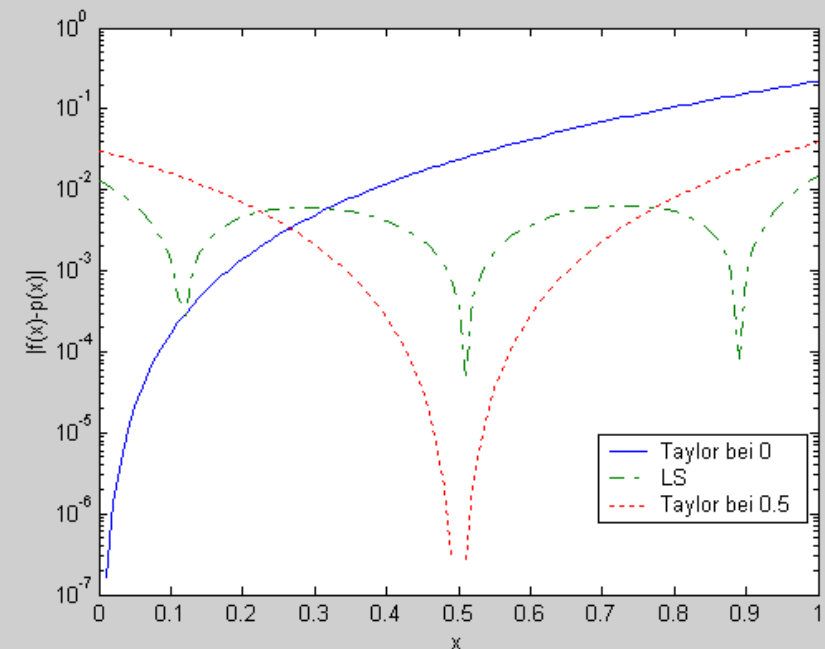
die für wachsende Ordnung m dafür bekannt ist, dass sie schlecht konditioniert ist und somit schwierig zu invertieren ist.

Least Squares Filterung

- Aus diesem Grunde verwendet man im allgemeinen keine Approximation mit (einfachen) Polynomen.
- Bei kleinen Werten von m ist dieser Effekt jedoch nicht so dramatisch.
- Beispiel 3.3: Die Funktion e^x sei zu approximieren. Die Taylorreihenentwicklung ergibt: $e^x = 1 + x + x^2/2 + \dots$. Entwicklung mit LS dagegen liefert: $1.013 + 0.8511x + 0.8392 x^2$

Möchte man, dass der größte Fehler minimal ist, so genügt es nicht die L_2 -Norm zu minimieren, sondern in dem Fall ist die L_∞ -Norm erforderlich:

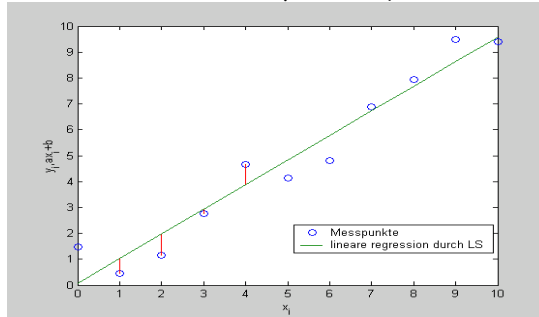
$$\min_{\{c_0, c_1, \dots, c_{m-1}\}} \lim_{p \rightarrow \infty} \left(\int_a^b (f(x) - c_0 - c_1 x - \dots - c_{m-1} x^{m-1})^p dx \right)^{1/p}$$



Least Squares Filterung

■ Beispiel 3.4: Lineare Regression

- Die wohl bekannteste Anwendung des LS Verfahrens ist die lineare Regression. Hierbei wird versucht eine Gerade so zwischen Messpunkte zu legen, dass der Abstand zwischen Gerade und Messpunkte (quadratisch) minimal wird.



Univ.-Prof. Dr.-Ing.
Markus Rupp

Least Squares Filterung

- Ansatz: $y_i = ax_i + b$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} ax_1 + b \\ ax_2 + b \\ \vdots \\ ax_n + b \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \underbrace{\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}}_A \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_c + \underbrace{\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}}_e$$

- Somit ergibt sich die LS Lösung zu:

$$\underline{c} = (A^H A)^{-1} A^H \underline{y}$$

Least Squares Filterung

- Natürlich gibt es nicht nur lineare Abhängigkeiten. Je nach Ansatz, spricht man von einer quadratischen, kubischen... Regression.
- Treten Messwerte mit verschiedener Genauigkeit auf, so bietet es sich an diese entsprechend ihres Vertrauenswertes zu gewichten. Dies wird durch eine Gewichtungsmatrix W erreicht:

$$\underline{c} = (A^H W A)^{-1} A^H W \underline{y}$$

- Hierbei ist zu beachten, dass W positiv definit sein muss. Für indefinite Matrizen siehe Vorlesung adaptive Filter.

Least Squares Filterung

- Am Einfachsten lässt sich ein AR Prozess schätzen:

$$\underline{x}_k = a_1 \underline{x}_{k-1} + a_2 \underline{x}_{k-2} + \dots + a_p \underline{x}_{k-p} + \underline{v}_k$$

- Für den stochastischen Hintergrund, siehe hierzu auch die Vorlesung „Stochastische Signalverarbeitung“

Least Squares Filterung

- Fasst man M Beobachtungen in je einem Vektor zusammen, so kann man schreiben:

$$\begin{aligned}\underline{x}_k &= a_1 \underline{x}_{k-1} + a_2 \underline{x}_{k-2} + \dots + a_p \underline{x}_{k-p} + \underline{v}_k \\ &= [\underline{x}_{k-1}, \underline{x}_{k-2}, \dots, \underline{x}_{k-p}] \underline{a} + \underline{v}_k \\ &= X_{p,k} \underline{a} + \underline{v}_k\end{aligned}$$

- Ein Schätzwert für $\underline{a}$ kann aus der Beobachtung von $\underline{x}_k$ durch Minimierung des Schätzfehlers erfolgen

$$\min_{\underline{a}} \|\underline{x}_k - X_{p,k} \underline{a}\|^2$$

Least Squares Filterung

- Man erhält:

$$\begin{aligned}X_{p,k}^H \underline{x}_k - X_{p,k}^H X_{p,k} \underline{a} &= \underline{0} \\ \underline{a} &= \underbrace{\left[X_{p,k}^H X_{p,k} \right]^{-1}}_{\approx R_{xx}^{-1}} \underbrace{X_{p,k}^H \underline{x}_k}_{\approx \underline{r}_x}\end{aligned}$$

- Rückführung auf ein Least-Squares Problem
- Interpretation der Matrix $X_{p,k}^H X_{p,k}$ als Schätzwert der AKF Matrix und des Vektors $X_{p,k}^H \underline{x}_k$ als Schätzwert des Autokorrelationsvektors in den Yule-Walker Gleichungen.

Least Squares Filterung

- Betrachte ein unterbestimmtes Gleichungssystem, in dem mehr Parameter zu schätzen sind als Beobachtungen vorliegen.

- Beispiel 3.5

$$\begin{bmatrix} 1 & 2 & -3 \\ -5 & 4 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -4 \\ 6 \end{bmatrix}$$

$$\underline{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + t \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}; \quad t \in \mathbb{R}$$

- Dadurch, dass das Gleichungssystem unterbestimmt ist, gibt es unendlich viele Lösungen.

Least Squares Filterung

- Von allen Lösungen ist diejenige am interessantesten, welche die geringste Norm aufweist.

$$\min \|\underline{x}\| \quad ; \text{ mit Nebenbedingung : } A\underline{x} = \underline{b}$$

- Wir nehmen wieder an, dass ein Schätzer für $\underline{x}$ aus einer Linearkombination der Beobachtungen hervorgeht, also

$$\begin{aligned}\hat{\underline{x}} &= A^H \underline{c} \Rightarrow A \hat{\underline{x}} = A A^H \underline{c} = \underline{b} \Rightarrow \underline{c} = (A A^H)^{-1} \underline{b} = (A A^H)^{-1} A \hat{\underline{x}} \\ \hat{\underline{x}} &= A^H (A A^H)^{-1} \underline{b}\end{aligned}$$

Least Squares Filterung

- Beachte, dass $A^H(AA^H)^{-1}$ auch eine Pseudoinverse zu A darstellt, denn $A \times A^H(AA^H)^{-1} = I$.
- Interessanterweise liefert diese Pseudoinverse immer die Lösung mit minimaler Norm. Das liegt daran, dass alle anderen Lösungen zusätzlich Komponenten haben, die nicht durch Linearkombinationen von A^H darstellbar sind (also nicht im Spaltenraum von A^H liegen).
- Beispiel 3.6:** Betrachte das vorige Beispiel. Die Minimum Norm Lösung lautetet $\underline{x}^T = [-1, 0, 1]$ und nicht, wie vielleicht vermutet $[1, 2, 3]^T$.

Least Squares Filterung

$$\begin{bmatrix} 1 & 2 & -3 \\ -5 & 4 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -4 \\ 6 \end{bmatrix}$$

$$\underline{x} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} + t \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}; \quad t \in \mathbb{R}$$

$$2 \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} - \begin{bmatrix} -5 \\ 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 0 \\ -7 \end{bmatrix};$$

also liegt $\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ im Zeilenraum von A ; $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ nicht

Iteratives LS Problem

- Bisher haben wir uns nur mit quadratischen Formen (L_2 und l_2 -Normen) befasst. Die Frage stellt sich, ob auch andere Normen berechenbar sind, also:

$$\begin{aligned} \min_{\underline{c}} \|\underline{x} - A\underline{c}\|_p &\Rightarrow \min_{\underline{c}} \|\underline{x} - A\underline{c}\|_p^p = \min_{\underline{c}} \sum_{i=1}^m |x_i - (A\underline{c})_i|^p \\ &= \min_{\underline{c}} \sum_{i=1}^m \underbrace{|x_i - (A\underline{c})_i|^{p-2}}_{w_i} |x_i - (A\underline{c})_i|^2 \end{aligned}$$

- Das Problem wird also als quadratisches Problem mit Gewichtungsmatrix formuliert.

Iteratives LS Problem

- Algorithmus:

$$\underline{c}^{(1)} = (A^H A)^{-1} A^H \underline{x}$$

für $k = 1..$

$$\underline{e}^{(k)} = \underline{x} - A\underline{c}^{(k)}$$

$$W^{(k)} = \text{diag}\left(|e_1^{(k)}|^{p-2}, |e_2^{(k)}|^{p-2}, \dots, |e_m^{(k)}|^{p-2}\right)$$

$$\underline{c}^{(k+1)} = (A^H W^{(k)} A)^{-1} A^H W^{(k)} \underline{x}$$

Iteratives LS Problem

■ Beispiel 3.7 Filterentwurf

Ein linear-phasiges FIR Filter der Länge $2N+1$ ist so zu entwerfen, dass der Amplitudengang $|H_d(e^{j\Omega})|$ eines vorgegebenen Filters möglichst genau erreicht wird. Beachte, dass für linear-phasige FIR Filter gilt $h(k)=h(2N+1-k)$ und somit

$$H(e^{j\Omega}) = e^{-jN\Omega} H_R(e^{j\Omega}) = e^{-jN\Omega} \sum_{n=0}^N b_n \cos(n\Omega) = e^{-jN\Omega} \underline{b}^T \underline{c}(\Omega)$$

- Würde nur ein quadratisches Abstandsmaß gewählt werden,

$$\min_{H_r(e^{j\Omega})} \int_0^\pi |H_r(e^{j\Omega}) - |H_d(e^{j\Omega})||^2 d\Omega$$

Iteratives LS Problem

- so wird der Amplitudengang nur mäßig genau erreicht.
- Bei einer größeren Norm wird er um so besser erreicht, also $p \rightarrow \infty$ (engl.: equiripple design).

$$\lim_{p \rightarrow \infty} \min_{H_r(e^{j\Omega})} \int_0^\pi |H_r(e^{j\Omega}) - |H_d(e^{j\Omega})||^p d\Omega$$

Signaltransformation

- Im folgenden werden wir uns mit der Frage beschäftigen, welche Basisfunktionen besonders geeignet zur Approximation sind.
- Wir haben bisher gesehen, dass einfache Polynome zu schlecht konditionierten Problemstellungen führen.
- Auch haben wir gesehen, dass orthogonale Mengen besonders interessant sind, da dann die Inverse der Grammian Matrix besonders einfach wird.
- Wir werden uns daher auf die Suche nach geeigneten Basisfunktionen mit orthogonalen (orthonormalen) Eigenschaften machen.

Signaltransformation

- Wir approximieren im LS Sinn (L_2 - Norm) für orthonormale Funktionen $p_i(t)$:

$$\begin{aligned} \hat{x}_m(t) &= \sum_{i=1}^m c_i p_i(t) \\ \left\| x(t) - \sum_{i=1}^m c_i p_i(t) \right\|^2 &= \|x(t)\|^2 - \sum_{i=1}^m \langle x(t), p_i(t) \rangle^2 \\ &= \|x(t)\|^2 - \sum_{i=1}^m |c_i|^2 \geq 0 \end{aligned}$$

- Bessel'sche Ungleichung

Signaltransformation

- Betrachtet man die Folge mit Grenzwert:

$$\hat{x}_m(t) = \sum_{i=1}^m c_i p_i(t)$$

$$\hat{x}(t) = \hat{x}_\infty(t) = \sum_{i=1}^{\infty} c_i p_i(t)$$

- Da der Schätzwert eine Cauchy Folge darstellt und der Hilbertraum abgeschlossen ist, können wir folgern, dass der Grenzwert auch im Hilbert Raum liegt.
- Allerdings kann nicht mit jeder orthonormalen Menge eine beliebige (glatte) Funktion Punkt für Punkt nachgebildet werden.

Signaltransformation

- Wir beschränken uns auf Approximationen in der L_2 Norm.
- Auch dann kann man nicht jede Funktion nachbilden.
- Beispiel 3.8:** Die Menge $\{\sin(nt)\}$ $n=1,2,\dots,\infty$ bildet eine orthonormale Menge. Die Funktion $\cos(t)$ kann nicht approximiert werden, da alle Koeffizienten verschwinden:

$$c_n = \int_0^{2\pi} \cos(t) \sin(nt) dt = 0$$

Signaltransformation

- Wir benötigen also eine bestimmte Eigenschaft der orthonormalen Menge, um sicherzustellen, dass diese auch eine beliebige Funktion approximieren kann.
- Satz 3.3:** Eine Menge orthonormaler Funktionen kann dann eine beliebige Funktion $x(t)$ im Hilbertraum nachbilden, wenn eine der folgenden Eigenschaften erfüllt ist:

$$x(t) = \sum_{i=1}^{\infty} \langle x(t), p_i(t) \rangle p_i(t)$$

$$\left\| x(t) - \sum_{i=1}^n \langle x(t), p_i(t) \rangle p_i(t) \right\| < \varepsilon; \text{ für alle } n \geq N, N < \infty$$

$$\|x(t)\|^2 = \sum_{i=1}^{\infty} \langle x(t), p_i(t) \rangle^2; \text{ Parsevalsche Gleichung}$$

Es gibt keine von Null verschiedene Funktion $f(t) \in S$, für die die Vereinigungsmenge $\{f(t), p_1(t), p_2(t), \dots\}$ eine orthonormale Menge bildet.

Signaltransformation

- Man beachte hierbei den Unterschied zu endlich dimensional Mengen. Bei endlich dimensional Mengen genügt es zu zeigen, dass die Funktionen p_i linear unabhängig sind.
- Genügt eine unendlich dimensionale Mengen den Eigenschaften in Satz 3.3, so ist die Repräsentation von x durch die unendliche Menge der Koeffizienten c_i äquivalent.
- Man nennt diese auch verallgemeinerte Fourier Reihe.

Signaltransformation

- **Lemma 3.2:** Zwei Funktionen $x(t)$ und $y(t)$ aus S seien durch die selbe orthonormale Menge p_i durch die Koeffizienten c_i und b_i approximiert. Dann gilt:

$$\langle x, y \rangle = \sum_{i=1}^{\infty} c_i b_i$$

- **Beweis:**

$$x(t) = \sum_{i=1}^{\infty} c_i p_i(t); \quad y(t) = \sum_{i=1}^{\infty} b_i p_i(t)$$

$$\begin{aligned} \langle x(t), y(t) \rangle &= \left\langle \sum_{i=1}^{\infty} c_i p_i(t), \sum_{i=1}^{\infty} b_i p_i(t) \right\rangle \\ &= \sum_{i=1}^{\infty} c_i b_i; \end{aligned}$$

Orthogonale Funktionen

- Orthogonale Funktionen erfüllen die Bedingung, dass ihr inneres Vektorprodukt auf einem interessierenden Intervall $[a,b]$ verschwindet:

$$\langle p, q \rangle_w = 0$$

$$\langle p, q \rangle_w = \int_a^b w(t) p(t) q(t) dt$$

- Hierbei wird oftmals eine positive Gewichtsfunktion $w(t) > 0$ verwendet.

Orthogonale Funktionen

- Bekannteste Beispiele für orthonormale Mengen sind:

- **Beispiel 3.9:** Fourier Reihe

$$f(t) = \sum_{n=-\infty}^{\infty} c_n \frac{1}{\sqrt{2\pi}} e^{jnt}$$

$$c_n = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(t) e^{-jnt} dt$$

- Bei periodischen Funktionen $f(t) = f(t+mT)$ wählt man:

$$f(t) = \sum_{n=-\infty}^{\infty} c_n \frac{1}{\sqrt{2\pi}} e^{jn \frac{2\pi}{T} t}$$

$$c_n = \frac{1}{T\sqrt{2\pi}} \int_0^T f(t) e^{-jn \frac{2\pi}{T} t} dt$$

Orthogonale Funktionen

- Bekannteste Beispiele für orthonormale Mengen sind:

- **Beispiel 3.10:** Diskrete Fourier Transformation (DFT)
Eine Funktion $x(k)$ ist nur an N diskreten Punkten $k=0,1,\dots,N-1$ bekannt.

$$x(k) = \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} c_l e^{j2\pi kl/N}$$

$$c_l = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} x(k) e^{-j2\pi kl/N}$$

- Man beachte, dass bei beiden Transformationen häufig nicht orthonormale, sondern nur orthogonale Mengen verwendet werden.

Orthogonale Funktionen

- Beachte, dass die orthogonalen Funktionen in diesem Fall durch Trigonometrische Funktionen $e^{jnt/T}$ gebildet werden. Die Gewichtsfunktion ist hierbei $w(t)=1$.
- Es gilt:

$$\frac{1}{2\pi} \int_0^{2\pi} e^{jnt} e^{-jmt} dt = \begin{cases} 0 & ; n \neq m \\ 1 & ; n = m \end{cases}$$

$$\frac{1}{N} \sum_{n=0}^{N-1} e^{j2\pi nt/T} e^{-j2\pi mt/T} = \begin{cases} 0 & ; |n-m| \bmod N \neq 0 \\ 1 & ; \text{sonst} \end{cases}$$

Orthogonale Funktionen

- Beispiel 3.11:** Orthogonale Polynome
 - Wir hatten bereits erkannt, dass „einfache“ Polynome zu schlechter Kondition führen und zudem nicht orthogonal sind. Dennoch ist es möglich auch mit Polynomen Familien mit orthogonalen Eigenschaften aufzubauen.
 - Lemma 3.3:** Orthogonale Polynome genügen folgender Rekursionsgleichung:

$$tp_n(t) = a_n p_{n+1}(t) + b_n p_n(t) + c_n p_{n-1}(t)$$

Orthogonale Funktionen

- Beweis:** Sei $g_n(t) = tp_n(t) - a_n p_{n+1}(t)$ vom Grad n . Dann muss gelten:

$$g_n(t) = \sum_{i=0}^n d_i p_i(t); \quad d_i = \langle g_n(t), p_i(t) \rangle$$

Für $i=n$ und $i=n+1$ erhält man:

$$d_n = b_n = \langle g_n, p_n \rangle; \quad d_{n-1} = c_n = \langle g_n, p_{n-1} \rangle$$

Für $i < n-1$ sind die Koeffizienten Null, denn

$$g_n(t) = tp_n(t) - a_n p_{n+1}(t) = b_n p_n(t) + c_n p_{n-1}(t)$$

Orthogonale Funktionen

- Beispiel 3.12:** Hermite Polynome:
 - $y_n(t) = d^n/dt^n (\exp(-t^2/2)) = p_n(t) \exp(-t^2/2)$
 - $tp_n(t) = -p_{n+1}(t) - np_{n-1}(t)$
 - $p_0(t)=1, p_1(t)=-t, p_2(t)=t^2-1, p_3(t)=-t^3+3t$

$$\int_{-\infty}^{\infty} p_n(t) p_m(t) \underbrace{\frac{\exp(-t^2/2)}{\sqrt{2\pi n!}}}_{w(t)} dt = \delta_{n-m}$$

Orthogonale Funktionen

- **Beispiel 3.13:** Ebenso existieren zeitdiskrete Binomiale Hermite Folgen:
- $x_{r+1}(k) = -x_{r+1}(k-1) + x_r(k) - x_r(k-1)$; $x_r(-1) = 0$

$$x_r(-1) = 0 \quad ; 0 \leq r \leq N$$

$$x_0(k) = \binom{N}{k}$$

- Z-Transformation ergibt:
- $X_r(z) = (z-1)/(z+1) X_{r-1}(z) = [(z-1)/(z+1)]^r X_0(z)$

$$X_r(z) = (1 - z^{-1})^r (1 + z^{-1})^{N-r}$$

Orthogonale Funktionen

Diskrete Hermite - Polynome: $x_r(k) = \binom{N}{k} P_r(k)$

Beachte, dass für grosse N gilt: $\binom{N}{k} \approx \frac{2^N}{\sqrt{N\pi/2}} \exp\left(-\frac{(k - N/2)^2}{N/2}\right)$

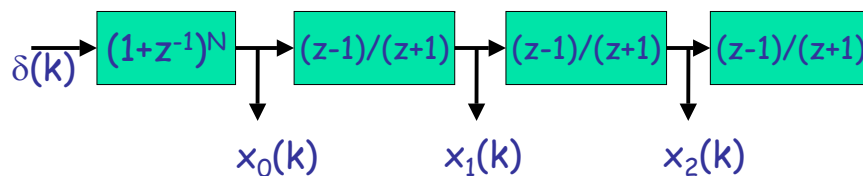
Es gilt: $P_r(k) = P_k(r)$

Orthogonalität besteht bzgl. $\binom{N}{k}$

$$\sum_{k=0}^N P_r(k) P_s(k) \binom{N}{k} = \frac{2^N}{\binom{N}{s}} \delta_{r-s}$$

Orthogonale Funktionen

- Binomiale Filterbank



Orthogonale Funktionen

- **Beispiel 3.14:** Legendre Polynome ($w(t)=1$)

$$tp_n(t) = \frac{n+1}{2n+1} p_{n+1}(t) + \frac{n}{2n+1} p_{n-1}(t)$$

$$p_0(t) = 1; \quad p_1(t) = t;$$

$$p_2(t) = 3/2 t^2 - 1/2; \quad p_3(t) = 5/2 t^3 - 3/2 t$$

Orthogonale Funktionen

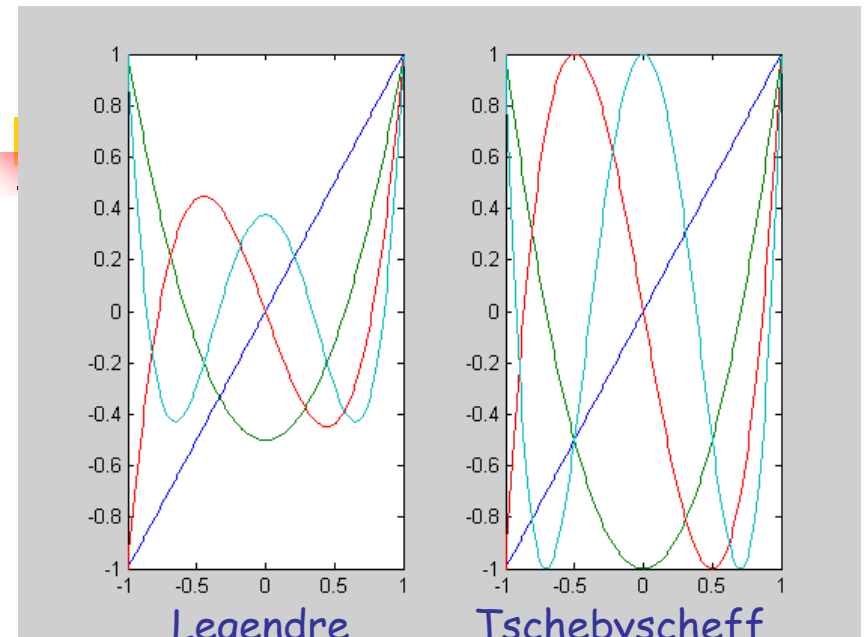
■ Beispiel 3.15: Tschebyscheff Polynome

$$\int_{-1}^1 \frac{1}{\sqrt{1-t^2}} p_n(t) p_m(t) dt = \begin{cases} 0 & n \neq m \\ \pi & n = m = 0 \\ \pi/2 & n = m \neq 0 \end{cases}$$

$$tp_n(t) = 0.5p_{n+1}(t) + 0.5p_{n-1}(t)$$

$$p_0(t) = 1; p_1(t) = t; p_2(t) = 2t^2 - 1; p_3(t) = 4t^3 - 3t$$

$$p_n(t) = \cos(n \arccos(t))$$



Orthogonale Funktionen

■ Beispiel 3.16: Betrachte

Streckung

Verschiebung

$$p_n(t) = \text{sinc}(2Bt - n)$$

$$\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$$

$$2B \int_{-\infty}^{\infty} \text{sinc}(2Bt - n) \text{sinc}(2Bt - m) dt = \begin{cases} 1 & ; n = m \\ 0 & ; n \neq m \end{cases}$$

Orthogonale Funktionen

- Für bandbegrenzte Funktionen $f(t)$, so dass $F(\omega) = 0$ für $|\omega| > 2\pi B$, finden sich die Koeffizienten zu:

$$c_n = \frac{\langle f(t), p_n(t) \rangle}{\langle p_n(t), p_n(t) \rangle} = f(n/2B)$$

$$f(t) = \sum_{n=-\infty}^{\infty} f(n/2B) \text{sinc}(2Bt - n)$$

- Abtastung kann also als Approximation im Hilbertraum gedeutet werden.

Wavelets

- Betrachte wieder Funktion

Streckung

Verschiebung

$$p_{j,k}(t) = 2^{-j/2} \phi(2^{-j}t - k)$$

- Es liegt also eine zweidimensionale Transformation vor, mit Veränderlichen bzgl. Verschiebung (position, engl.: location) und Streckung (Granularität, engl.: scale).

Wavelets

- Beachte, dass wenn $\phi(t)$ normiert ist, also $||\phi(t)||=1$, dann gilt auch $||p_{j,k}(t)||=1$.
- Man wählt nun die Funktion $\phi(t)$ derart, dass für alle Verschiebungen n ein orthonormaler Raum aufgespannt wird:

$$V_0 = \text{span}\{\phi(t - n), n \in \mathbb{Z}\}$$

Die verschobenen Funktionen bilden also eine orthonormale Basis für V_0 .

Wavelets

- Beispiel 3.17:** Betrachte den Einheitsimpuls

$$\phi(t) = u(t) - u(t-1)$$

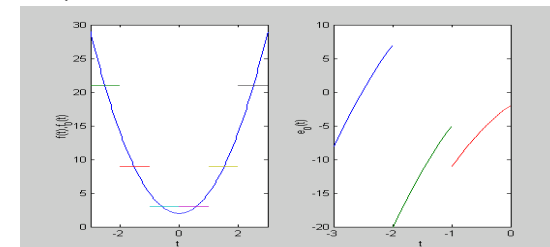
$$V_0 = \{\phi(t-n), n \in \mathbb{Z}\}$$

- Mit dieser Funktion können also alle Funktionen $f(t)$, die im Raster ganzzahliger Zahlen stückweise stetig sind exakt dargestellt werden, bzw. können glatte Funktionen mit der Granularität ganzzahliger Zahlenraster approximiert werden.
- Man schreibt:

$$f_0(t) = \sum_n \langle f(t), \phi_{0,n}(t) \rangle \phi_{0,n}(t) = f(t) - e_0(t)$$

Wavelets

- Man kann die so erhaltenen Koeffizienten
- $$c_n^{(0)} = \langle f(t), \phi_{0,n}(t) \rangle$$
- Auch als stückweise Flächen der Funktion $f(t)$ interpretieren.



Wavelets

- Interessanterweise kann man nun die Streckung dazu verwenden, Basen für andere Räume zu bilden.

$$V_{-1} = \text{span}\{\sqrt{2}\phi(2t-n), n \in \mathbb{Z}\}$$

$$V_j = \text{span}\{2^{-j/2}\phi(2^{-j}t-n), n \in \mathbb{Z}\}$$

- Weisen diese Räume eine Verschachtelung auf:

$$\dots V_2 \subset V_1 \subset V_0 \subset V_{-1} \dots$$
- So sagt man, dass $\phi(t)$ eine Skalierungsfunktion für Wavelets ist.

Wavelets

- Beispiel 3.18:** Betrachte den Einheitsimpuls

$$\phi(2t) = u(2t) - u(2t-2)$$

$$V_{-1} = \{\sqrt{2}\phi(2t-n), n \in \mathbb{Z}\}$$

- Mit dieser Funktion können also alle Funktionen $f(t)$, die im Raster $2n$ gerader, ganzzahliger Zahlen stückweise stetig sind exakt dargestellt werden, bzw. können glatte Funktionen mit der Granularität $2n$ ganzzahliger Zahlenraster approximiert werden.

$$f_{-1}(t) = \sum_n \langle f(t), \phi_{-1,n}(t) \rangle \phi_{-1,n}(t) = f(t) - e_{-1}(t)$$

Wavelets

- Die Funktion $f_{-1}(t)$ stellt also eine noch grobere Approximation von $f_0(t)$ in V_0 dar. Da aber V_0 eine Teilmenge von V_{-1} darstellt, gilt:

$$\begin{aligned} f_{-1}(t) &= \sum_n c_n^{(-1)} \phi_{-1,n}(t) \\ &= \sum_n c_n^{(0)} \phi_{0,n}(t) + e_{0,-1}(t) \\ &= \sum_n c_n^{(0)} \phi_{0,n}(t) + \sum_n d_n^{(0)} \psi_{0,n}(t) \end{aligned}$$

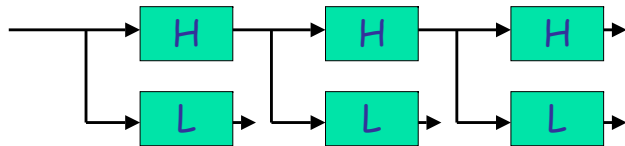
- Mit einer geeigneten Basis $\psi_{0,n}(t)$ aus W_0 mit $W_0 \cup V_0 = V_{-1}$

Wavelets

- Hierbei werden die Funktionen $\psi_{j,k}(t)$ Wavelets genannt.
- Man kann also zu einer jeden Granularitätsstufe eine Funktion in zwei Anteile ψ und ϕ zerlegen.
- Grob gesprochen entspricht der eine Anteil einem Hochpass und der anderem einem Tiefpass Anteil.
- Durch weitere Verfeinerung der Granularität kann die Funktion immer besser approximiert werden.
- Die dabei benötigte Anzahl von Koeffizienten ist stark von der Wavelet- oder Skalierungsfunktion abhängig.

Mehrraten Systeme

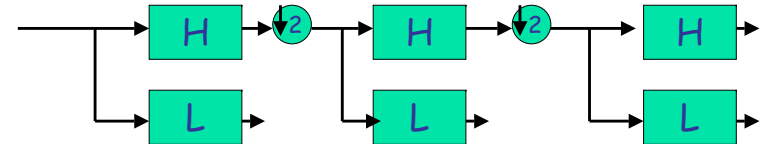
- Die Vorgehensweise bei den Wavelets könnte man auch durch folgende dyadische (oktavische) Baumstruktur beschreiben:



- Dadurch, dass aber die Tief und Hochpässe die Bandbreiten teilen, kann auch mit geringerer Datenrate weitergearbeitet werden.

Mehrraten Systeme

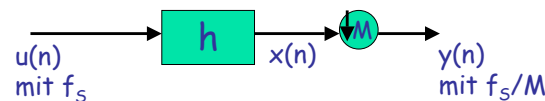
- Also:



- Durch diese Vorgehensweise kann schrittweise in jeder Stufe Komplexität eingespart werden, ohne dass dies in der Signalqualität eine Auswirkung hat.

Mehrraten Systeme

- Wir sind also zunächst einmal grundsätzlich an der Funktionsweise eines Dezimators interessiert:



$$x'(n) = \begin{cases} x(n) & ; n = 0, \pm M, \pm 2M, \dots \\ 0 & ; \text{sonst} \end{cases}$$

$$y(n) = x'(nM) = x(nM)$$

$$x'(n) = x(n) \sum_{r=-\infty}^{\infty} \delta(n - rM) = x(n) \frac{1}{M} \sum_{k=0}^{M-1} \exp\left(j \frac{2\pi}{M} nk\right)$$

Mehrraten Systeme

- also $x'(n) = x(n) \frac{1}{M} \sum_{k=0}^{M-1} \exp\left(j \frac{2\pi}{M} nk\right)$

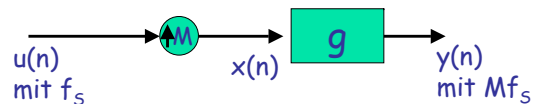
$$X'(\exp(j\Omega)) = \frac{1}{M} \sum_{k=0}^{M-1} X\left(\exp\left(j\left(\Omega - \frac{2\pi}{M}k\right)\right)\right)$$

$$Y(\exp(j\Omega)) = X'(\exp(j\Omega/M)) \\ = \frac{1}{M} \sum_{k=0}^{M-1} X\left(\exp\left(j\frac{\Omega - 2\pi k}{M}\right)\right)$$

- Das Signal $y(n)$ ist also ein Replika von $x(n)$ aber um den Faktor M in der Frequenz gestreckt.
- Ragte das Spektrum von $x(n)$ von $[-r, r]$ so ragt das Spektrum von $y(n)$ von $[-Mr, Mr]$.

Mehrraten Systeme

- Betrachten wir nun noch die umgekehrte Operation: Interpolation



$$y(n) = \begin{cases} x(n/M) & ; n = 0, \pm M, \pm 2M, \dots \\ 0 & ; \text{sonst} \end{cases}$$

$$\begin{aligned} Y(\exp(j\Omega)) &= \sum_{n=-\infty}^{\infty} x(n/M) \exp(-j\Omega n) \\ &= \sum_{k=-\infty}^{\infty} x(k) \exp(-j\Omega kM) = X(\exp(j\Omega M)) \end{aligned}$$

Re-Sampling

Beachte, dass die Interpolationsgleichung

$$f(t) = \sum_{n=-\infty}^{\infty} T f(nT) \operatorname{sinc}\left(\frac{\pi}{T}(t - nT)\right)$$

ein „Re-sampling“ erlaubt:

$$f(mT_1) = \sum_{n=-\infty}^{\infty} T f(nT) \operatorname{sinc}\left(\frac{\pi}{T}(mT_1 - nT)\right)$$

Re-Sampling

- Beispiel 3.19:** Sei $T_1 = 2T$ (Dezimator):

$$\begin{aligned} f(2mT) &= \sum_{n=-\infty}^{\infty} T f(nT) \operatorname{sinc}\left(\frac{\pi}{T}(2mT - nT)\right) = \sum_{n=-\infty}^{\infty} T f(nT) \operatorname{sinc}(\pi(2m - n)) \\ &= \begin{cases} f(2mT) & ; n = 2m \\ 0 & ; \text{sonst} \end{cases} \end{aligned}$$

- Damit liegt ein zeit-variantes System vor!

Re-Sampling

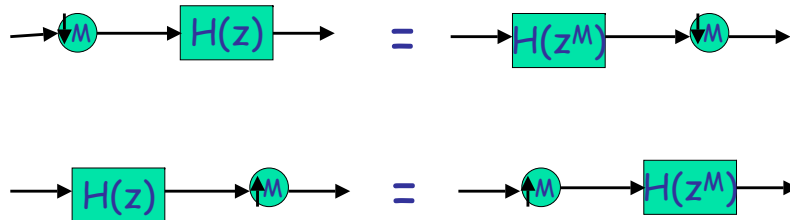
- Beispiel 3.20:** Sei $T_1 = T/2$ (Interpolator):

$$\begin{aligned} f(mT/2) &= \sum_{n=-\infty}^{\infty} T f(nT) \operatorname{sinc}\left(\frac{\pi}{T}(mT/2 - nT)\right) = \sum_{n=-\infty}^{\infty} T f(nT) \operatorname{sinc}\left(\frac{\pi}{2}(m - 2n)\right) \\ &= \begin{cases} f(mT/2) = f(nT) & ; m = 2n \\ \sum_{n=-\infty}^{\infty} T f(nT) \operatorname{sinc}\left(\frac{\pi}{2}(2k - 2n + 1)\right) & ; m = 2k + 1 \\ 0 & ; \text{sonst} \end{cases} \end{aligned}$$

- Offensichtlich liegt auch hier ein zeit-variantes System vor!

Mehrraten Systeme

- Es gelten folgende Äquivalenzen (Noble Identities):



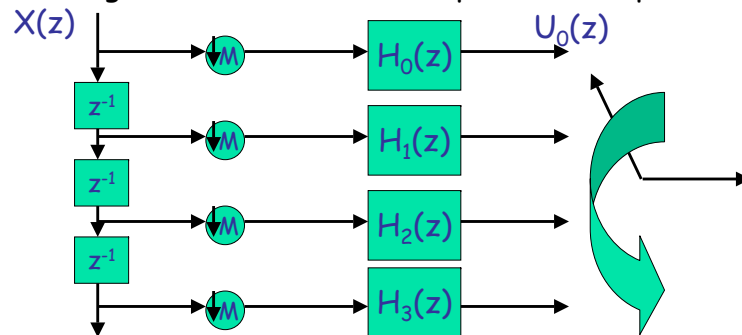
Mehrraten Systeme

- Damit lässt sich eine Polyphasen-Darstellung eines Systems beschreiben:

$$\begin{aligned}
 H(z) &= \sum_{k=0}^{MN-1} h(k)z^{-k} = \sum_{p=0}^{N-1} \sum_{k=0}^{M-1} h(k+pM)z^{-(k+pM)} \\
 &= \sum_{k=0}^{M-1} z^{-k} \left(\sum_{p=0}^{N-1} h(k+pM)z^{-pM} \right) \\
 &= \sum_{k=0}^{M-1} z^{-k} \left(\sum_{p=0}^{N-1} h_k(pM)z^{-pM} \right) \\
 &= \sum_{k=0}^{M-1} z^{-k} H_k(z^M)
 \end{aligned}$$

Mehrraten Systeme

- Allerdings lässt sich wieder Komplexität einsparen:



- Man erhält ein periodisches Filter!

Mehrraten Systeme

- Möchte man also gleich alle Bänder $H_r(z)$ der Filterbank beschreiben, so gilt:

$$H_r(z) = \sum_{k=0}^{M-1} z^{-k} H_{r,k}(z^M)$$

Die Ausgangssignale $U_r(z)$ ergeben sich dann aus dem Eingangssignal $X(z)$:

$$\begin{bmatrix} U_0(z) \\ U_1(z) \\ \vdots \\ U_{M-1}(z) \end{bmatrix} = \begin{bmatrix} H_{00}(z^M) & H_{01}(z^M) & \dots & H_{0,M-1}(z^M) \\ H_{10}(z^M) & H_{11}(z^M) & \dots & H_{1,M-1}(z^M) \\ \vdots & \vdots & \ddots & \vdots \\ H_{M-1,0}(z^M) & H_{M-1,1}(z^M) & \dots & H_{M-1,M-1}(z^M) \end{bmatrix} \begin{bmatrix} 1 \\ z^{-1} \\ \vdots \\ z^{-(M-1)} \end{bmatrix} X(z)$$

Mehrraten Systeme

- Man beachte, dass in dieser Darstellung die Spalten der Matrix eine Basis darstellen.

$$\begin{bmatrix} H_{00}(z^M) & H_{01}(z^M) & \dots & H_{0,M-1}(z^M) \\ H_{10}(z^M) & H_{11}(z^M) & \dots & H_{1,M-1}(z^M) \\ \vdots & \vdots & & \vdots \\ H_{M-1,0}(z^M) & H_{M-1,1}(z^M) & \dots & H_{M-1,M-1}(z^M) \end{bmatrix}$$

- Hierbei können die Elemente eines Bandpassbereiches orthogonal zu allen anderen Bandpassbereichen sein, die Elemente einer Phase sind orthogonal zu denen der anderen Phasen.
- Ebenso können die in Bandpasslage transformierten Signale in ihrer Polyphasendarstellung eine Basis darstellen.

Mehrraten Systeme

- Beispiel 3.21:** Gleichmäßige (engl.: uniform) Filterbank

$$H_0(z) = \sum_{p=0}^{M-1} F_p(z^M) z^{-p}; \text{ Prototyp-Tiefpass}$$

$$\begin{aligned} H_m(z) &= H_0(z \exp(-j2\pi m/M)) \\ &= \sum_{p=0}^{M-1} F_p \left(z^M \underbrace{\exp(-j2\pi m/M)^M}_1 \right) z^{-p} \underbrace{\exp(j2\pi mp/M)}_{W^{-mp}} \\ &= \sum_{p=0}^{M-1} F_p(z^M) z^{-p} W^{-mp} \end{aligned}$$

Mehrraten System

- Damit kann man alle Tiefpässe aus einem Prototyp Tiefpass formen:

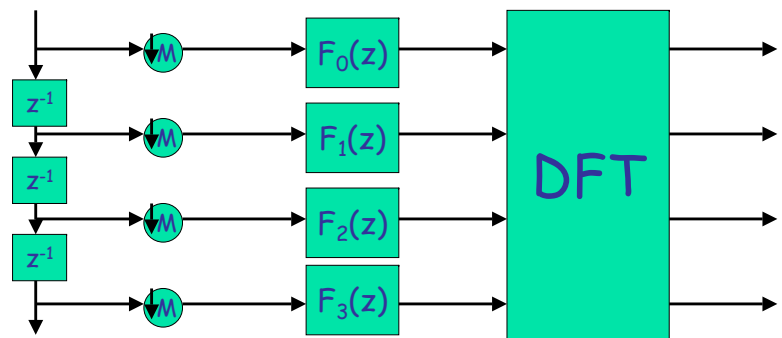
$$H_m(z) = \sum_{p=0}^{M-1} F_p(z^M) z^{-p} W^{-mp}$$

$$\begin{bmatrix} H_0(z) \\ H_1(z) \\ \vdots \\ H_{M-1}(z) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & W^{-1} & \dots & W^{-(M-1)} \\ \vdots & & \ddots & \vdots \\ 1 & W^{-(M-1)} & \dots & W^{-(M-1)^2} \end{bmatrix} \begin{bmatrix} F_0(z^M) \\ z^{-1} F_1(z^M) \\ \vdots \\ z^{-(M-1)} F_{M-1}(z^M) \end{bmatrix}$$

DFT

Mehrratensystem

- Somit erhält man eine Polyphasen Filterbank

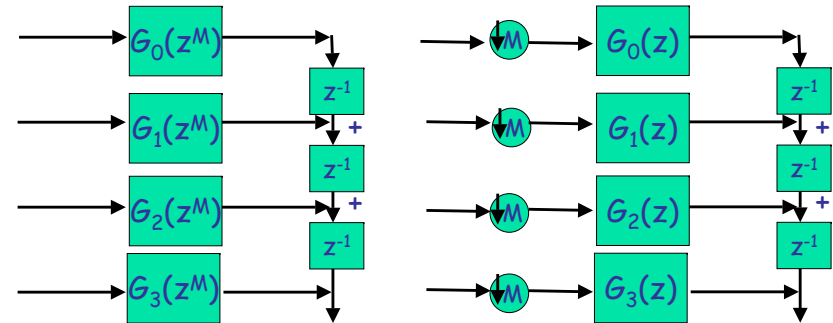


Mehrratensysteme

- Beachte, dass solche Filterbänke zwar das Aliasing im Nachbarband kompensieren, aber nicht darüber hinaus. Sie werden daher auch oft „Nahezu Perfekte Rekonstruktionsfilterbänke“ genannt (NPR).
- In verschiedenen Anwendungen genügt es nicht das Signal zu zerlegen (analysieren), sondern man möchte es auch wieder zusammenführen (synthetisieren).
- In diesem Fall ist neben der bisher besprochenen Analyse-Filterbank auch eine Synthese Filterbank erforderlich. Beide sind voneinander abhängig.

Mehrratensysteme

Beispiel 3.22: Synthesefilterbank



Mehrratensysteme

- Für die Synthesefilterbank gilt:

$$G_r(z) = \sum_{l=0}^{M-1} G_{r,M-1-l}(z^M) z^{-(M-1-l)}$$

Das Ausgangssignal $Y(z)$ ergibt sich dann aus den Eingangssignalen $U_r(z)$:

$$Y(z) = \begin{bmatrix} U_0(z) & U_1(z) & \dots & U_{M-1}(z) \end{bmatrix} \begin{bmatrix} G_{0,M-1}(z^M) & G_{0,M-2}(z^M) & \dots & G_{0,0}(z^M) \\ G_{1,M-1}(z^M) & G_{1,M-2}(z^M) & \dots & G_{1,0}(z^M) \\ \vdots & \vdots & \ddots & \vdots \\ G_{M-1,M-1}(z^M) & G_{M-1,M-2}(z^M) & \dots & G_{M-1,0}(z^M) \end{bmatrix} \begin{bmatrix} z^{-(M-1)} \\ z^{-(M-2)} \\ \vdots \\ 1 \end{bmatrix}$$

Mehrratensysteme

- Da es sich bei den meisten (dyadischen, binären) Mehrratensystemen um Datenratenänderungen um den Faktor 2^k handelt, wollen wir diesen näher betrachten:

$$\begin{aligned} H(z) &= h_0 + h_1 z^{-1} + h_2 z^{-2} + \dots \\ &= h_0 + h_2 z^{-2} + \dots + z^{-1} [h_1 + h_3 z^{-2} + \dots] \\ &= H_0(z^2) + z^{-1} H_1(z^2) \end{aligned}$$

$$H_0(z^2) = \frac{1}{2} [H(z) + H(-z)]$$

$$H_1(z^2) = \frac{1}{2} [H(z) - H(-z)]$$

Mehrratensysteme

- Hinter $H_0(z)$ und $H_1(z)$ kann man eine Dezimierung um den Faktor 2 ausführen. Dadurch verändert sich die ursprüngliche Relation:

$$H_0(z^2) = \frac{1}{2} [H(z) + H(-z)]$$

$$H_1(z^2) = \frac{1}{2} [H(z) - H(-z)]$$

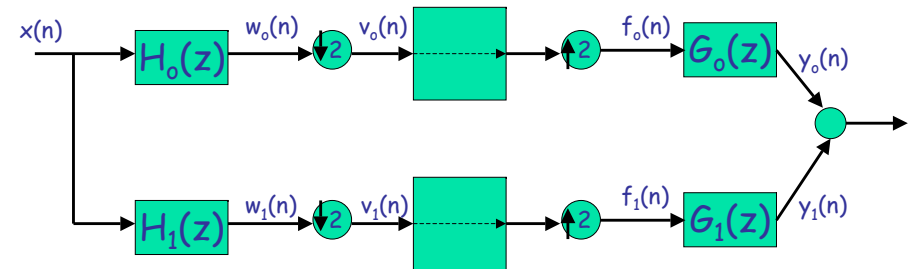
- zu:

$$H_0(z) = \frac{1}{2} [H(z^{1/2}) + H(-z^{1/2})]$$

$$H_1(z) = \frac{1}{2} [H(z^{1/2}) - H(-z^{1/2})]$$

Mehrratensysteme

- Betrachte folgende Anordnung einer zweikanaligen Filterbank:



Mehrratensysteme

- Wir finden folgende Relationen:

$$W_i(z) = H_i(z)X(z); i = 1, 2$$

$$V_i(z) = \frac{1}{2} [W_i(z^{1/2}) + W_i(-z^{1/2})]; i = 1, 2$$

$$F_i(z) = V_i(z^2); i = 1, 2$$

$$Y_i(z) = G_i(z)F_i(z); i = 1, 2$$

- Dies lässt sich kombinieren zu:

$$Y_i(z) = \frac{1}{2} G_i(z) [H_0(z)X(z) + H_0(-z)X(-z)]; i = 1, 2$$

Mehrratensysteme

- Damit finden wir für das Ausgangssignal der zweikanaligen Filterbank:

$$\hat{X}(z) = Y_0(z) + Y_1(z)$$

$$= T(z)X(z) + S(z)X(-z)$$

$$T(z) = \frac{1}{2} [H_0(z)G_0(z) + H_1(z)G_1(z)]$$

$$S(z) = \frac{1}{2} [H_0(-z)G_0(z) + H_1(-z)G_1(z)]$$

- Für perfekte Rekonstruktion (PR) muss gelten:

$$S(z) = 0; \quad T(z) = cz^{-K}$$

Mehrratensysteme

- Das gibt uns schon eine erste Bedingung:

$$G_0(z)H_0(-z) = -G_1(z)H_1(-z)$$
- Allerdings erlaubt uns diese Bedingung viele Freiheitsgrade. Wählen wir beispielsweise N stufige FIR Filter:

$$G_0(z) = -H_1(-z); \quad G_1(z) = H_0(-z)$$

$$H_1(z) = z^{-(N-1)}H_0(-z^{-1})$$
- Man erhält eine Konjugierte Quadratur-Filterbank (CQF= conjugate quadrature filter bank)

Mehrratensysteme

- Dann folgt für $T(z)$:

$$T(z) = \frac{1}{2} z^{-(N-1)} [H_0(z)H_0(z^{-1}) + H_0(-z)H_0(-z^{-1})]$$

man kann zeigen, dass: $\sum_{k=0}^{N-1} h_0(k)h_0(k+2n) = \delta(n)$

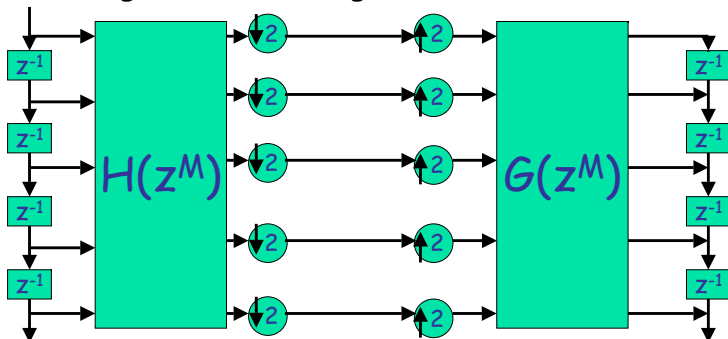
für normierte Impulsantworten: $\sum_{k=0}^{N-1} |h_0(k)|^2 = 1$
- Wählt man hingegen:

$$G_0(z) = H_1(-z); \quad G_1(z) = -H_0(-z)$$

$$H_1(z) = z^{-(N-1)}H_0(-z^{-1})$$
- So erhält man eine Quadratur Spiegel Filterbank (QMF= quadrature mirror filter bank).

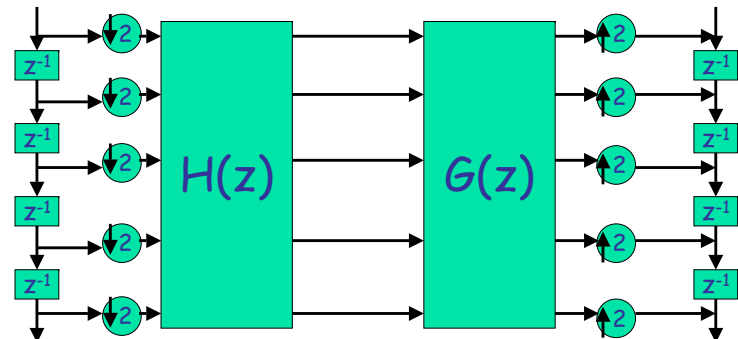
Mehrratensysteme

- Solche Filterbänke lassen sich allgemein durch folgende Anordnung beschreiben:



Mehrratensysteme

- Was ja äquivalent zu folgender Struktur ist:



Mehrratensysteme

- Wir erkennen damit die Bedingung für eine Filterbank mit perfekter Rekonstruktion (PR) in Matrixschreibweise:

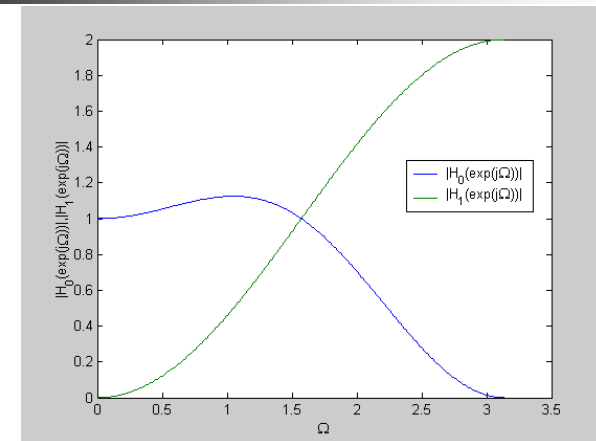
$$G(z)H(z) = z^{-D}I$$

- Beispiel 3.23:** Seien die linear-phasigen Analysefilter $H_0(z)$ und $H_1(z)$ gegeben. Finde die zugehörigen linear-phasigen Synthesefilter $G_0(z)$ und $G_1(z)$.

$$H_0(z) = -\frac{1}{8} + \frac{1}{4}z^{-1} + \frac{3}{4}z^{-2} + \frac{1}{4}z^{-3} - \frac{1}{8}z^{-4}$$

$$H_1(z) = \frac{1}{2} - z^{-1} + \frac{1}{2}z^{-2}$$

Mehrratensysteme



Mehrratensysteme

- Die Analysefilter werden in ihre Polyphasenkomponenten zerlegt:

$$H_0(z) = -\frac{1}{8} + \frac{1}{4}z^{-1} + \frac{3}{4}z^{-2} + \frac{1}{4}z^{-3} - \frac{1}{8}z^{-4}$$

$$= H_{00}(z^2) + z^{-1}H_{01}(z^2)$$

$$H_{00}(z) = -\frac{1}{8} + \frac{3}{4}z^{-1} - \frac{1}{8}z^{-2}; \quad H_{01}(z) = \frac{1}{4} + \frac{1}{4}z^{-1}$$

$$H_1(z) = \frac{1}{2} - z^{-1} + \frac{1}{2}z^{-2}$$

$$= H_{10}(z^2) + z^{-1}H_{11}(z^2)$$

$$H_{10}(z) = \frac{1}{2} + \frac{1}{2}z^{-1}; \quad H_{11}(z) = -1$$

Mehrratensysteme

- Damit ergibt sich als Forderung für eine PR Filterbank:

$$\begin{bmatrix} G_{00}(z) & G_{01}(z) \\ G_{10}(z) & G_{11}(z) \end{bmatrix} \begin{bmatrix} -\frac{1}{8} + \frac{3}{4}z^{-1} - \frac{1}{8}z^{-2} & \frac{1}{4} + \frac{1}{4}z^{-1} \\ \frac{1}{2} + \frac{1}{2}z^{-1} & -1 \end{bmatrix} = z^{-D}I$$

- Mit $D=1$ erhält man:

$$\begin{bmatrix} G_{00}(z) & G_{01}(z) \\ G_{10}(z) & G_{11}(z) \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{4} + \frac{1}{4}z^{-1} \\ \frac{1}{2} + \frac{1}{2}z^{-1} & \frac{1}{8} - \frac{3}{4}z^{-1} + \frac{1}{8}z^{-2} \end{bmatrix}$$

Mehrratensysteme

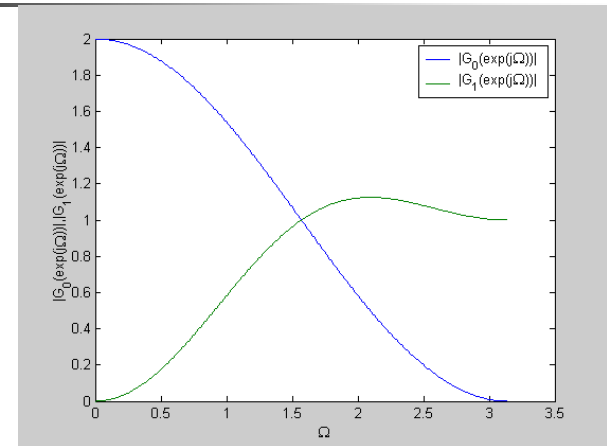
- Also lauten die Synthesefilter:

$$\begin{bmatrix} G_{00}(z) & G_{01}(z) \\ G_{10}(z) & G_{11}(z) \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{4} + \frac{1}{4}z^{-1} \\ \frac{1}{2} + \frac{1}{2}z^{-1} & \frac{1}{8} - \frac{3}{4}z^{-1} + \frac{1}{8}z^{-2} \end{bmatrix}$$

$$G_0(z) = G_{10}(z^2) + z^{-1}G_{00}(z) = \frac{1}{2} + z^{-1} + \frac{1}{2}z^{-2}$$

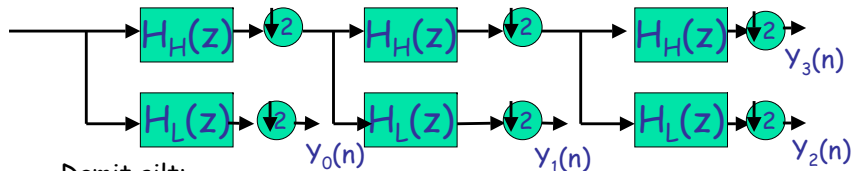
$$G_1(z) = G_{11}(z^2) + z^{-1}G_{01}(z) = \frac{1}{8} + \frac{1}{4}z^{-1} - \frac{3}{4}z^{-2} + \frac{1}{4}z^{-3} + \frac{1}{8}z^{-4}$$

Mehrratensysteme

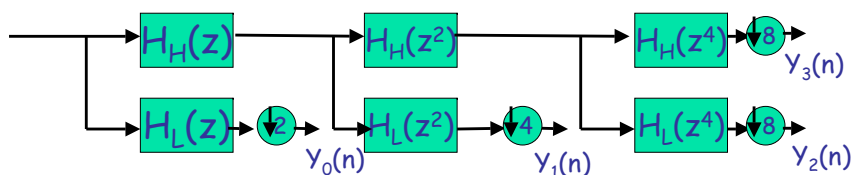


Mehrraten Systeme

- Zurück zur Wavelettransformation:

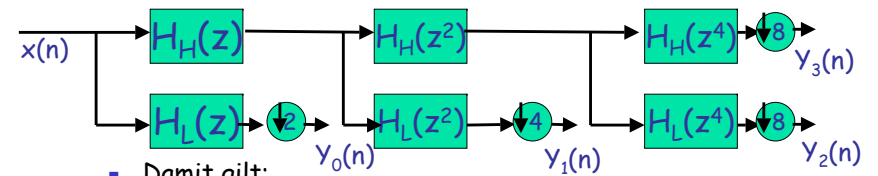


- Damit gilt:



Mehrraten Systeme

- Also:



- Damit gilt:

$$Y_0(z) = H_L(z)X(z)$$

$$Y_1(z) = H_H(z)H_L(z^2)X(z)$$

$$Y_2(z) = H_H(z)H_H(z^2)H_L(z^4)X(z)$$

$$Y_3(z) = H_H(z)H_H(z^2)H_H(z^4)X(z)$$