



CALLEBAUT Gilles

GILLEES
CALLEBAUT

**DIGITALE
SIGNAALVERWERKING**

2015 - 2016

Inhoudsopgave

1 Discrete-Tijd Signalen en Systemen	8
1.1 Discrete-tijd signalen	8
1.2 Discrete-tijd systemen	9
1.3 Antwoord op een willekeurige ingang	9
1.4 Stabiliteit	12
2 Discrete-tijd Fouriertransformatie	15
2.1 Discrete-tijd Fourier-transformatie (DTFT) e inverse transformatie	15
2.1.1 Toepassing: Amplitude Modulatie	18
2.2 De Hilbert Transformatie	20
2.2.1 Bandbreedte van het verzonden signaal $x_u(k)$	21
2.2.2 Recuperatie oorspronkelijk signaal $m(k)$	23
2.3 Quadrature Amplitude Modulation (QAM)	23
2.3.1 Bandbreedte van het QAM-signaal	24
2.3.2 Reconstructie beide signalen $m_1(k)$ en $m_2(k)$	24
2.4 De rechthoekige vensterfunctie	24
2.4.1 Spectrum van een rechthoekige vensterfunctie	25
2.5 Toepassing op cosinus en som van cosinussen	28
3 Discrete Fouriertransformatie	29
3.1 De Discrete Fourier Reeks	29
3.2 Eigenschappen van de DFT	29
3.2.1 Lineariteit	30
3.2.2 Cyclische tijdsverschuiving	30
3.2.3 Modulatie	30
3.2.4 Cyclische convolutie in het tijdsdomein	31
3.3 Fast Fourier Tansforamtie (FFT)	31
3.3.1 De Radix-2 FFT	32

4 Systeem Analyse	34
4.1 De Z-transformatie	34
4.2 Polen en nullen als systeembeschrijving	34
4.2.1 Conclusies	37
4.2.2 Discrete oscillator	45
4.3 All-pass systemen	46
4.3.1 Weergavekarakteristiek	47
4.3.2 Fasekarakteristiek	48
4.3.3 Besluit	48
4.4 Minimum fase systemen	49
4.4.1 Tijdsdomeinbeschrijving van een MFS	50
4.5 Lineaire-fase FIR filters	51
4.5.1 Types lineair fasige FIR filters	52
4.5.2 Ligging van de polen en de nullen bij lineaire-fase filters	54
4.5.3 Oefeningen	55
5 Eindige woordlengte	62
5.1 Algemeen	62
5.2 Kwantisering en overloop	64
5.3 A/D-omzettingsruis	66
5.4 Kwantisering van de filtercoëfficiënten	69
5.5 Woordlengtebeperking van tussenresultaten	76
5.5.1 Overloop van tussenresultaten	79
5.5.2 Kwantisering van tussenresultaten	81
5.6 Opmerkingen	89
6 Systemen met meerdere bemonsteringsfrequenties	91
6.1 Inleiding	91
6.1.1 Voorbeeld: Bemonsteringsfrequentie verlaging	91
6.1.2 Voorbeeld: Bemonsteringsfrequentie verhoging	91
6.2 Verlagen van de bemonsteringsfrequentie met een gehele factor	92
6.2.1 De decimator	92
6.2.2 Een decimerend filter	93
6.2.3 Realisatie van een decimerend filter	94
6.3 Verhogen van de bemonsteringsfrequentie met een gehele factor	96
6.3.1 Realisatie van een interpolerend filter	96
6.4 Veranderen van de bemonsteringsfrequentie met een niet-gehele factor	99

6.4.1	Frequentieverhoging	100
6.4.2	LDF	100
6.4.3	Frequentieverlaging	100
6.5	Polyfase decompositie	100
6.5.1	Variatie	101
6.6	Toepassing: Polyfase decompositie - Decimatie	102
6.6.1	Definite van de M^{de} wortel uit 1	102
6.6.2	DTFT van gedecimeerd signaal	104
6.6.3	Structuren voor decimatie	105
6.7	Toepassing: Polyfase decompositie - Interpolatie	107
6.7.1	Structuren voor interpolatie	107
6.8	Analyse en synthesefilters	108
6.9	QMF filters	109
7	Herhaling basis digitale filters	112

Lijst van figuren

2.1	Digitaal laagdoorlaat filter (herhaalt zich vanaf $\frac{f_s}{2}$, als de karakteristiek wordt uitgedrukt in θ zal deze herhalen van π)	17
2.2	Het Gibb's fenomeen	18
2.3	Amplitude Modulatie	19
2.4	DTFT van een re�el basisbandsignaal $m(k)$	21
2.5	Spectrum van het complexe signaal $u(k)$	22
2.6	Spectrum van het verzonden signaal $x_u(k)$. Opmerking de verticale-as moet $X_u(e^{j\theta})$ zijn.	23
2.7	Recuperatie van de oorspronkelijke signalen $m_1(k)$ en $m_2(k)$	24
2.8	Rechthoekige vensterfunctie $w(k)$	25
2.9	Het spectrum van een rechthoekige vensterfunctie (in dB).	26
2.10	Vereenvoudigde weergave van het spectrum van een cosinus die in de tijd is beperkt door een rechthoekige vensterfunctie. Deze spectrale lijnen moeten een sinc functie voorstellen.	28
3.1	$\frac{N}{2}$ butterflies – 8-punt radix-2 FFT	33
4.1	Grafische bepaling van $A(e^{j\phi})$ en $\varphi(e^{j\theta_0})$	39
4.2	(a) Minimum-fase-systeem met drie polen in de oorsprong en drie nulpunten. (b) Niet-minimum-fase-systeem uit (a) verkregen door spiegelen van de nulpunten t.o.v. de eenheidscirkel. Dit leidt tot een grotere groeplooptijd τ_e	40
4.3	All pass netwerk	41
4.4	All pole netwerk	41
4.5	Lineair-fase systeem	42
4.6	Belangrijke polen/nullen diagrammen	44
4.7	Een discrete oscillator. (a) Blokschema; (b) polen- en nulpuntenbeeld.	45
4.8	Impulsresponsie van de discrete oscillator	46

4.9	Visuele voorstelling van het feit dat een transferfunctie kan geschreven worden als een product van een minimumfase en allpass systeem.	50
4.10	De vier types van FIR filter met een een lineaire fase karakteristiek	51
4.11	Grafische voorstelling van alle mogelijke liggingen van de nullen in een lineair fase systeem	55
4.12	Vier filters in het frequentiedomein en LDF, BDF, HDF en BSF. Als we willen werken met θ moeten we gebruik, komt $2f_c$ overeen met π	56
4.13	Laagdoorlaat filter	59
4.14	Frequentiekarakteristiek voor $f_s = 8000 \text{ Hz}$	59
5.1	Verschillende binaire getalrepresentaties	63
5.2	Voorbeelden van getalrepresentaties	63
5.3	Drie kwantiseringskarakteristieken: (a) afronden; (b) afbreken; (c) modulus afbreken	64
5.4	Drie overloopkarakteristieken (a) verzadiging; (b) nulstelling; (c) zaagtandvormige overloop	65
5.5	De combinatie van kwantisering (afronden) met verschillende vormen van overloop.	66
5.6	Kwantisering bij A/D-omzetting.	67
5.7	Kwantiseringsruis	68
5.8	De impulsresponsie van een lineair systeem met invloed van kwantiseringsruis.	68
5.9	Verandering van de weergavekarakteristiek van een digitaal filter als de coëfficiënten gekwantiseerd worden	70
5.10	Zuiver recursief tweede-orde filter (directe vorm).	72
5.11	Mogelijke poolposities als a_1 en a_2 worden gekwantiseerd tot 4 bits.	73
5.12	Tweede-orde-sectie met een zogenaamde 'gekoppelde structuur'	73
5.13	Mogelijke poolposities als a en b worden gekwantiseerd tot 4 bits.	75
5.14	Toeneming van de woordlengte in een eenvoudig FIR-filter	76
5.15	Filter met 1 kwantisor maar een minder eenvoudige optelling om tot een uitgang van B-bits te komen	77
5.16	Filter met twee kwantisatoren maar een eenvoudigere optelling om tot een uitgang van B-bits te komen	77
5.17	Eerste-orde recursief filter.	77
5.18	Recursief tweede-orde-filter met woordlengte-begrenzing	78
5.19	Schaling in een transversaal filter waardoor geen overloop meer optreedt.	79
5.20	Schaling ondergebracht in de coëfficiënten	80
5.21	Schaling bij een tweede-orde recursief filter (IIR filter).	80
5.22	Schaling bij een serieschakeling.	80
5.23	Kwantisators gemodelleerd als kwantiseringsruisbronnen. De termen e_0 en e_1 stellen de ruisbronnen voor.	82

5.24 Transversaal filter waarbij de kwantisering door één ruisbron $e_{tot}(n) = e_1(n) + e_2(n)$ wordt voorgesteld.	82
5.25 Recursieve sectie met de directe-vorm-1-structuur. (a) met kwantisatoren en (b) met kwantiseringsruisbronnen.	83
5.26 Recursieve sectie volgens 5.25 op pagina 83 waarbij de kwantisatoren door een ruisbron worden voorgesteld.	83
5.27 Recursieve sectie met de directe-vorm-2-structuur. (a) met kwantisatoren en (b) met kwantiseringsruisbronnen.	84
5.28 Recursieve sectie volgens figuur 5.27 op pagina 84, waarbij de kwantisatoren zijn vervangen door twee ruisbronnen $e_{tot,1} = e_0(n) + e_1(n) + e_2(n)$ en $e_{tot,2} = e_3(n) + e_4(n)$	85
5.29 Eerste-orde recursief filter met $a = -0,91$ waarin grenscyclussen kunnen optreden.	86
5.30 Tweede-orde recursief filter met coëfficiënten 1,9 en -0,98 waarin grenscyclussen kunnen optreden.	86
5.31 Verloop van $y(n)$ dat geldt als Q overeenkomt met afronden	87
5.32 Verloop van $y(n)$ dat geldt als Q overeenkomt met modulus afronden	87
5.33 Zuiver recursieve tweede-orde-sectie en de waarden van a_1 en a_2 waarvoor deze sectie (zonder kwantisering) stabiel is.	88
5.34 Zuiver recursieve tweede-orde-sectie die het gunstigst is in verband met grenscyclussen (Q : kwantisator volgens afbreken van modulus; gearceerd: condities waaronder grenscyclussen optreden).	88
5.35 Combineren van polen en nulpunten van: (a) zesde-orde banddoorlaat filter; (b) zesde-orde bandsperrend filter.	90
6.1 Een discreet laagdoorlatend filter met ingangssignaal $x(n)$ en uitgangssignaal $y(n)$	92
6.2 Een discrete modulator	92
6.3 Bemonsteringsfrequentie verlager in het tijdsdomein	92
6.4 Praktisch bemonsteringsverlagend filter, i.e. een decimerend filter.	93
6.5 3e orde decimerend FIR filter	94
6.6 3e orde decimerend FIR filter (verbeterd)	94
6.7 Decimerend IIR filter	95
6.8 Decimerend IIR filter (verbeterd)	95
6.9 Symbool voor een bemonsteringsfrequentieverhoger (BFH) met een factor $R = \frac{T_1}{T_2}$	96
6.10 Het verband tussen $Y(e^{j\omega T_2})$ en $X(e^{j\omega T_1})$ bij verhoging van de bemonsteringsfrequentie met een factor $R = 3$	97
6.11 Interpolerend FIR filter	97
6.12 Interpolerend FIR filter (verbeterd)	98
6.13 Interpolerend IIR filter	99

6.14 Interpolerend IIR filter (verbeterd)	99
6.15 Verandering van de bemonsteringsfrequentie met een rationele factor $\frac{T_2}{T_1} = \frac{R_2}{R_1}$	99
6.16 Twee systemen voor een decimerend en interpolerend filter op te bouwen.	105
6.17 Decimerend filter m.b.v. polyfase decompositie. (In de nota's zullen we echter werken met P_i als component tegenover de H_i hier gebruikt.)	106
6.18 Filter bank structuur, m.b.v analyse en synthese filters.	108
6.19 Analyse en synthese filter gebaseerd op uniforme DFT filters.	109
6.20 Twee kanaals QMF	110

GILLEES
CALLEBAUF

Hoofdstuk 1

Discrete-Tijd Signalen en Systemen

1.1 Discrete-tijd signalen

Een discrete-tijd signaal is een opeenvolging van getallen, dit zijn waarden die enkel gekend zijn op bepaalde tijdstippen. Om een signaal te onderscheiden van zijn individuele waarden wordt het soms voorgesteld als $\{x(k)\}$. Als het signaal bekomen wordt door een bemonstering van een continu signaal $x(t)$ dan wordt de bemonsteringsperiode (T) er soms expliciet bij vermeld: $\{x(kT)\}$. Dit is dus de verzameling van al de functiewaarden, i.e. het discreet signaal. Het k -de element van een signaal wordt voorgesteld door $x(k)$.

Een signaal dat vaak gebruikt wordt is de eenheidspuls, $\{\delta(k)\}$. Deze bevat maar één functiewaarde en wordt beschreven door:

$$\delta(k) = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases} \quad (1.1)$$

Een ander belangrijk signaal is de eenheidsstap, $\{u(k)\}$. Dit signaal wordt gebruikt om te koppelen aan signalen, zodat er geen niet-nul waarden zijn voor negatieve tijdstippen. Het wordt beschreven door:

$$u(k) = \begin{cases} 0, & k < 0 \\ 1, & k \geq 0 \end{cases} \quad (1.2)$$

Een bruikbare voorstelling om het gedrag van een discrete-tijd systeem te bestuderen is:

$$x(k) = \sum_{n=-\infty}^{\infty} x(n)\delta(k-n) \quad (1.3)$$

1.2 Discrete-tijd systemen

Een discrete-tijd systeem zorgt voor relaties tussen discrete-tijd signalen:

$$\{y(k)\} = S\{x(k)\} \quad (1.4)$$

Waarbij $S\{x(k)\}$ staat voor een systeem S waar $\{x(k)\}$ aan de ingang wordt aangelegd.

De belangrijkste eigenschappen waaraan de systemen kunnen voldoen:

- **Lineariteit** – Een systeem is lineair als de uitgang op een lineaire combinatie van de ingangen gelijk is aan de lineaire combinatie van de uitgangeren.

$$S\{a\{x_1(k)\} + b\{x_2(k)\}\} = aS\{x_1(k)\} + bS\{x_2(k)\}$$

Meer algemeen

$$S\left\{\sum_i a_i\{x_i(k)\}\right\} = \sum_i a_i S\{x_i(k)\}$$

- **Tijd-invariantie** – Een systeem is tijdinvariant als een tijdsverschuiving van de ingang zorgt voor een even grote tijdsverschuiving aan de uitgang.

$$S\{x(k-n)\} = \{y(k-n)\}$$

- **Causaliteit** – Een systeem is causaal als het niet mogelijk is op de toekomst te anticiperen.

<i>als</i>	$x(k) = y(k)$	<i>voor</i>	$k \leq k_0$
<i>dan</i>	$S\{x(k)\} = S\{y(k)\}$	<i>voor</i>	$k \leq k_0$

De klasse van systemen die voor ons van het grootste belang is, is die met de eigenschappen: lineariteit, tijdsafhankelijkheid en causaliteit (LTO-systemen).

1.3 Antwoord op een willekeurige ingang

Hoe reageert nu een LTO-systeem op een ingangssignaal?

We weten

$$x(k) = \sum_{n=-\infty}^{\infty} x(n)\delta(k-n) \quad (1.3 \text{ revisited})$$

Als gevolg van de lineariteit kan de uitgang van een systeem, $y(k)$, berekend worden als:

$$y(k) = (S\{x\})(k) \quad (1.5)$$

$$= S\left\{\sum_{n=-\infty}^{+\infty} x(n)\delta(k-n)\right\} \quad (1.6)$$

$$= \sum_{n=-\infty}^{+\infty} x(n)S\{\delta(k-n)\} \quad (1.7)$$

Het antwoord op een eenheidsimpuls is de impulsresponsie, $h(k)$. Rekening houdend met de tijdsinvariantie van het systeem kan de uitgang $y(k)$ berekend worden als:

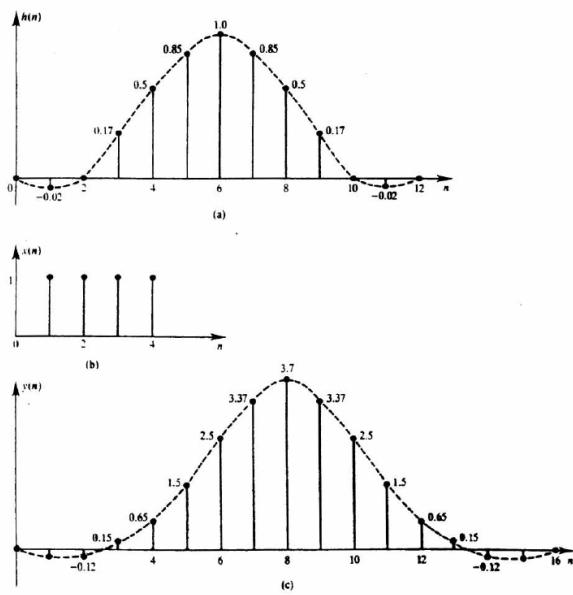
$$y(k) = \sum_{n=-\infty}^{+\infty} x(n)h(k-n) \quad (1.8)$$

$$= \sum_{n=-\infty}^{+\infty} h(n)x(k-n) \quad (1.9)$$

Deze formules zijn bekend als de lineaire convolutie; ze stellen het antwoord voor van een LTO-systeem op een willkeurige ingang $x(k)$. Dit antwoord is volledig bepaald als de impulsresponsie van het systeem bekend is.

De convolutie wordt voorgesteld als:

$$\{y(k)\} = \{h(k)\} * \{x(k)\} \quad (1.10)$$

Voorbeeld: Grafische voorstelling**Voorbeeld:**

We berekenen de convolutie van de volgende sequenties:

$$\begin{aligned}\{h(k)\} &= \{\dots, 0, 0, h(0), h(1), h(2), 0, 0, \dots\} \\ \{x(k)\} &= \{\dots, 0, 0, x(0), x(1), x(2), x(3), 0, 0, \dots\}\end{aligned}$$

Er zal geen uitgang verschillen van 0 zijn vóór $k = 0$. Vanaf $k = 0$ krijgen we:

$$\begin{aligned}y(0) &= h(0)x(0) \\ y(1) &= h(0)x(1) + h(1)x(0) \\ y(2) &= h(0)x(2) + h(1)x(1) + h(2)x(0) \\ y(3) &= h(0)x(3) + h(1)x(2) + h(2)x(1) \\ y(4) &= h(0)x(4) + h(1)x(3) + h(2)x(2) \\ y(5) &= h(0)x(5) + h(1)x(4) + h(2)x(3)\end{aligned}$$

Vanaf $k = 6$ is $y(k) = 0$.

Dit volgt uit vergelijking 1.9 op de vorige pagina.

Voorbeeld: Veronderstel dat de waarden van de impulsresponsie van een systeem gegeven worden door:

$$h(k) = a^k u(k)$$

Veronderstel verder dat de waarden van het ingangssignaal gegeven worden door:

$$x(k) = b^k u(k)$$

De waarden van de uitgang worden dan berekend m.b.v.:

$$y(k) = \sum_{n=-\infty}^{+\infty} a^n u(n) b^{k-n} u(k-n)$$

Rekening houden met de waarden van de stapfunctie ($u(n) = 1$ voor $n \geq 0$ en $u(k-n) = 1$ voor $k-n \geq 0$) bekomen we:

$$y(k) = \sum_{n=0}^k a^n b^{k-n} = b^k \sum_{n=0}^k \left(\frac{a}{b}\right)^n$$

M.b.v. de formule uit de reeksontwikkeling:

$$\sum_{n=0}^N x^n = \frac{1-x^{N+1}}{1-x} \quad (1.11)$$

Bekomen we:

$$y(k) = b^k \frac{1 - \left(\frac{a}{b}\right)^{k+1}}{1 - \left(\frac{a}{b}\right)} u(k) = \frac{b^{k+1} - a^{k+1}}{b-a} u(k)$$

In het speciale geval dat $a = b$ is de formule voor de reeksontwikkeling niet bruikbaar. In dit geval moet de uitgang $y(k)$ direct berekend worden met als resultaat:

$$y(k) = (k+1)b^k u(k)$$

1.4 Stabiliteit

Een systeem is stabiel als een eindige ingang aanleiding geeft tot een eindige uitgang.

Een ingangssignaal is eindig als er een M bestaat waarvoor geldt:

$$|x(k)| \leq M < \infty \quad \forall k$$

Een uitgangssignaal is eindig als er een N bestaat waarvoor geldt:

$$|y(k)| \leq N < \infty \quad \forall k$$

M.a.w. alle functiewaarden moeten kleiner zijn dan oneindig, dit noemt men eindige waarden.
M.b.t. verdere besprekingen hebben we het begrip norm nodig:

$$l^p \text{norm} \triangleq \|\{x(k)\}\|_p = \left(\sum_{n=-\infty}^{+\infty} |x(n)|^p \right)^{\frac{1}{p}} \quad (1.12)$$

Deze vergelijking kan worden vergeleken met een vector. De lengte van een vector kan worden beschreven als de l^2 norm, nl. $\sqrt{(x^2 + y^2)}$. Deze definitie voldoet aan de nodige eigenschappen voor een norm:

1.

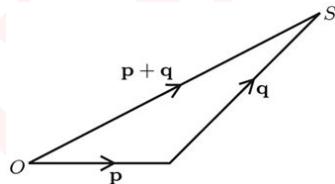
$$\|\{x(k)\}\|_p \geq 0 \quad = 0 \Leftrightarrow \{x\} = 0$$

M.a.w. de norm is altijd positief.

2.

$$\|\{x\} + \{y\}\|_p \leq \|\{x\}\|_p + \|\{y\}\|_p$$

We kunnen dit vergelijken met het optellen van twee vectoren. Waarbij de som van twee vectoren groter is dan de lengte van de som van de vectoren.



3.

$$\|a\{x\}\|_p = |a| \|\{x\}\|_p \quad \forall a$$

Voor een signaal met eindige l^p norm wordt gerefereerd naar l^p -ruimtes. Deze ruimtes zijn verzamelingen van eindig durende signalen met eindige waarden voor de l^p norm. Er zijn 3 l^p -ruimtes van belang:

- **De ruimte l^1** – Deze bevat reeksen $\{x\}$ die absoluut optelbaar zijn:

$$\|\{x\}\|_1 = \sum_{n=-\infty}^{+\infty} |x(n)| < \infty \quad (1.13)$$

- **De ruimte l^2** – Deze bevat reeksen $\{x\}$ die kwadratisch optelbaar zijn:

$$\|\{x\}\|_2 = \left(\sum_{n=-\infty}^{+\infty} |x(n)|^2 \right)^{\frac{1}{2}} < \infty \quad (1.14)$$

De l^2 norm houdt verband met de energie E_x van $\{x\}$. De definitie van energie van een signaal wordt gegeven door vergelijking:

$$E_x = \sum_{k=-\infty}^{+\infty} |x(k)|^2 \quad (1.15)$$

Men ziet in dat als $\{x\} \in l^2$ dan is $\{x\}$ een signaal met een eindige energie.

Veronderstel dat $\{x\}$ een signaal is uit de l^1 ruimte dan geldt:

$$\|\{x\}\|_1^2 = \left(\sum_{k=-\infty}^{+\infty} |x(k)| \right)^2$$

$$\|\{x\}\|_2^2 = \sum_{k=-\infty}^{+\infty} |x(k)|^2$$

We zien dus in dat $\|\{x\}\|_1^2 \geq \|\{x\}\|_2^2$

Dus:

$$\|\{x\}\|_2 \leq \|\{x\}\|_1$$

Dit betekent dat als $\{x\} \in l^1$ dan $\{x\} \in l^2$, of m.a.w. $l^2 \subset l^1$.

- **De ruimte l^∞** – Deze bevat reeksen $\{x\}$ waarvoor geldt:

$$\|\{x\}\|_\infty = \max_k |x(k)| < \infty \quad (1.16)$$

Deze nieuwe definitie is nodig vermits we geen oneindige wortel kunnen trekken.

De l^∞ norm van $\{x\}$ wordt gedefinieerd als:

$$\|\{x\}\|_\infty = \max_k |x(k)| \quad (1.17)$$

M.b.t. de stabiliteit kunnen we stellen dat een systeem stabiel is als een l^∞ ingang aanleiding geeft tot een l^∞ uitgang.

Globaal genomen zien we in dat een l^p -norm behoort tot een l^p -ruimte als de norm eindig is.

In het geval van een LTO-systeem wordt de stabiliteit eenvoudig gekarakteriseerd d.m.v. de volgende stelling:

Stelling 1 Een LTO-systeem dat gekenmerkt wordt door de impulsresponsie $\{h(k)\}$ is stabiel als en slechts als $\{h\}$ behoort tot de l^1 -ruimte.

$$\sum_{n=-\infty}^{+\infty} |h(n)| < \infty \text{ m.a.w. als } \{h\} \in l^1$$

Hoofdstuk 2

Discrete-tijd Fouriertransformatie

2.1 Discrete-tijd Fourier-transformatie (DTFT) e inverse transformatie

De DTFT van een sequentie $x(k)$ wordt gedefinieerd door:

$$X(e^{j\theta}) = \sum_{k=-\infty}^{+\infty} x(k)e^{-jk\theta} \quad (2.1)$$

uit de formule blijkt dat de DTFT een speciaal geval is van de Z-transformatie als z zich op de eenheidscirkel bevindt, nl. $z = e^{j\theta}$.

Een voldoende voorwaarde voor het bestaan van de DTFT van $x(k)$ is dat de eenheidscirkel binnen het convergentiegebied van $X(z)$ ligt. Dit komt overeen met de absolute sommeerbaarheid van $x(k)$:

$$\sum_{k=-\infty}^{+\infty} |x(k)| < \infty$$

Doordat alle stabiele sequenties per definitie sommeebaar zijn zullen alle stabiele sequenties een DTFT hebben:

$$X(e^{j\theta}) = X(z)|_{z=e^{j\theta}} = \sum_{k=-\infty}^{+\infty} x(k)e^{-jk\theta}$$

Hierdoor kunnen we een verband leggen tussen de frequentieresponsie en de DTFT. We weten dat een systeem $H(z)$ stabiel is a.s.a. zijn impulsresponsie $h(k)$ absoluut sommeerbaar is (zie stelling 1 op de pagina hiervoor). Dus de DTFT van een impulsresponsie $h(k)$ zal bestaan. Vergelijking van de formules leert dat de frequentieresponsie van een systeem gelijk is aan de DTFT van zijn impulsresponsie.

Voorbeeld:

Stel $x(k)$ wordt gegeven door:

$$x(k) = a^k u(k) |a| < 1$$

De Z-transformatie is dan:

$$X(z) = \frac{1}{1 - az^{-1}} = \frac{z}{z - a}$$

Het convergentiegebied van de Z-getransformeerde $X(z)$ is het gebied in het complexe vlak beschreven door $|z| > |a|$. Doordat $|a| < 1$ zal de eenheidscirkel binnen het convergentiegebied liggen zodat de reeks $x(k)$ absoluut sommeerbaar is.

De DTFT wordt gegeven door:

$$X(e^{j\theta}) = X(z)|_{z=e^{j\theta}} = \frac{1}{1 - ae^{-j\theta}}$$

We gebruiken de DTFT als een transformatievoorstelling van discrete-tijd signalen. Daarom hebben we ook een inverse formule nodig om het signaal te kunnen afleiden uit zijn getransformeerde. De inverse DTFT is een speciaal geval van de inverse Z-transformatie gegeven door:

$$x(k) = \frac{1}{2\pi j} \oint X(z) z^{k-1} dz \quad (2.2)$$

Om de noodzaak van contourintegratie te elimineren kunnen we de eenheidscirkel als volgt parametriseren:

We starten bij $z = -1$ wat overeenkomt met $\theta = -\pi$ en we doorlopen de eenheidscirkel in tegenwijzerzin tot $z = -1$ opnieuw is bereikt, dit komt overeen met $\theta = +\pi$.

De eenheidscirkel doorlopen in tegenwijzerzin, wordt dan geparametriseerd als:

$$z = e^{j\theta} \quad -\pi \leq \theta \leq \pi$$

Rekening houdend met $dz = je^{j\theta} d\theta$ krijgen we:

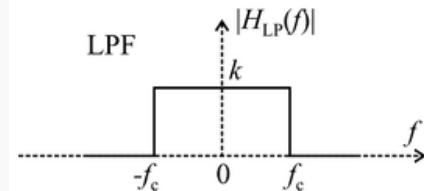
$$x(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\theta}) \cdot e^{jk\theta} d\theta \quad (2.3)$$

Dit is de formule voor de inverse DTFT.

Voorbeeld: Laagdoorlaat filter

We vertrekken van $H(e^{j\theta})$ als een ideaal laagdoorlaat filter met afsnijfrequentie θ_c :

$$H(e^{j\theta}) = \begin{cases} 1, & |\theta| \leq \theta_c \\ 0, & \theta_c < |\theta| < \pi \end{cases}$$



Figuur 2.1: Digitaal laagdoorlaat filter (herhaalt zich vanaf $\frac{f_s}{2}$, als de karakteristiek wordt uitgedrukt in θ zal deze herhalen van π)

De impulsresponsie komt overeen met de inverse DTFT (zie vergelijking 2.3 op de vorige pagina):

$$h(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H(e^{j\theta}) \cdot e^{jk\theta} d\theta$$

Hierbij kan $\int_{-\pi}^{\pi} H(e^{j\theta})$ geschreven worden als $\int_{-\theta_c}^{\theta_c} 1$, want buiten $-\theta_c$ t.e.m. θ_c is $H(e^{j\theta})$ gelijk aan 0 – per definitie.

$$h(k) = \frac{1}{2\pi} \int_{-\theta_c}^{\theta_c} e^{jk\theta} d\theta = \frac{1}{2\pi} \left[\frac{1}{jk} e^{jk\theta} \right]_{-\theta_c}^{\theta_c}$$

Dus:

$$h(k) = \frac{\sin \theta_c k}{\pi k}$$

Dit is een niet-causale impulsresponsie die niet absoluut sommeerbaar is, vermits men oneindig verschillende waarden bekomt.

De DTFT die gegeven wordt door:

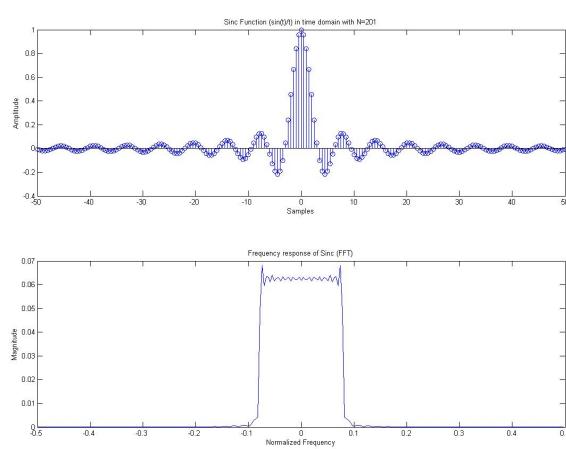
$$\sum_{k=-\infty}^{+\infty} \frac{\sin \theta_c k}{\pi k} \cdot e^{-jk\theta}$$

zal niet uniform convergeren voor alle θ .

Definiëren we echter (met beperking in de tijd):

$$H_M(e^{j\theta}) = \sum_{k=-M}^{+M} \frac{\sin \theta_c k}{\pi k}$$

dan hebben we een *mean square* convergentie; deze convergentie wordt voorgesteld in de figuur 2.2. Het oscillerend gedrag staat bekend als het Gibb's fenomeen.



Figuur 2.2: Het Gibb's fenomeen

2.1.1 Toepassing: Amplitude Modulatie

In digitale communicatie systemen kan een discrete-tijd signaal verstuurd worden m.b.v. amplitude modulatie.

Als $x(k)$ het te verzenden signaal is, dan wordt het amplitude-gemoduleerd signaal gegeven door $x_m(k) = x(k) \cos \theta_c k$ met θ_c de draaggolf frequentie.

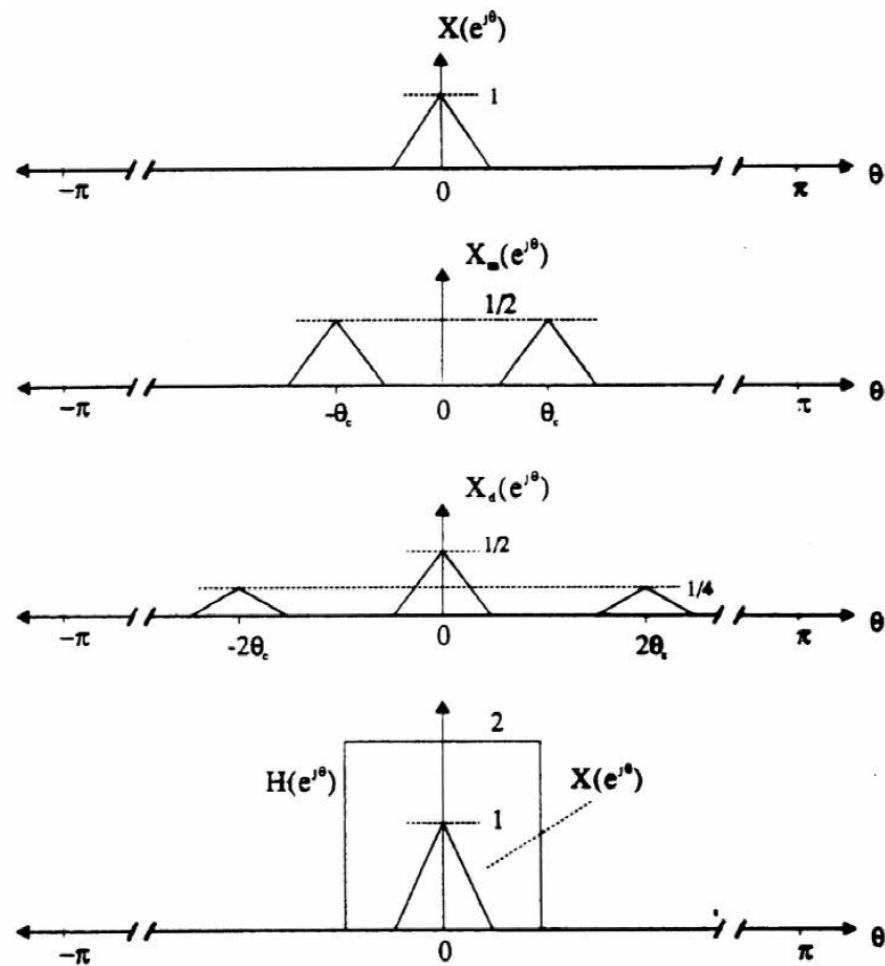
Gebruikmakend van de formule van Euler krijgen we:

$$x_m(k) = x(k) \frac{e^{j\theta_c k} + e^{-j\theta_c k}}{2}$$

Na berekening van de DTFT gebruik maken van de eigenschappen, wordt dit:

$$X_m(e^{j\theta}) = \frac{1}{2} X(e^{j(\theta-\theta_c)}) + \frac{1}{2} X(e^{j(\theta+\theta_c)})$$

Het spectrum van het gemoduleerde signaal wordt gegeven in de figuur 2.3 op de volgende pagina. Het bestaat uit 2 in frequentie verschoven kopijen van $X(e^{j\theta})$, één verschoven over θ_c en één over $-\theta_c$, met een amplitude die de helft is van die van $X(e^{j\theta})$.



Figuur 2.3: Amplitude Modulatie

Voor het versturen wordt het gemoduleerd signaal omgezet in een analoog signaal en verstuurd over een analoog kanaal.

Aan de ontvangerszijde wordt het signaal weer omgezet in een digitaal signaal zodat het spectrum overeenkomt met $X_m(e^{j\theta})$.

Om het origineel signaal te reconstrueren wordt het ontvangen signaal vermenigvuldigd met een lokale oscillator met een frequentie θ_c , dit levert het gedemoduleerd signaal op $x_d(k) = x_m(k) \cos k\theta_c$.

We beschouwen nu het spectrum van het gedemoduleerd signaal (gebruik makend van de eigenschappen van de DTFT).

$$X_d(e^{j\theta}) = \frac{1}{2}X_m(e^{j(\theta-\theta_c)}) + \frac{1}{2}X_m(e^{j(\theta+\theta_c)})$$

Uit de figuur blijkt dat deze som van verschoven kopijen van $X_m(e^{j\theta})$ resulteert in een basisband

en twee zijbanden gelokaliseerd rond $\pm 2\theta_c$. Hieruit leiden we af dat als we het gedemoduleerd signaal filteren met een laagdoorlaat filter, dat de zijbanden verdwijnen. Als we dan nog een schaal twee toepassen recupereren we $x(k)$.

Analytisch kunnen we dit als volgt aantonen:

$$x_d(k) = x_m(k) \cos \theta_c k = x(k) \cos^2 \theta_c k$$

M.b.v. :

$$\cos^2 \theta_c k = \frac{1 + \cos 2\theta_c k}{2}$$

Bekomen we:

$$x_d(k) = \frac{1}{2}x(k) + x(k) \cos 2\theta_c k$$

Dus:

$$X_d(e^{j\theta}) = \frac{1}{2}X_m(e^{j\theta}) + \frac{1}{4}X_m(e^{j(\theta-2\theta_c)}) + \frac{1}{4}X_m(e^{j(\theta+2\theta_c)})$$

Als we de zijbanden verwijderen met een laagdoorlaat filter met een versterking 2 krijgen we $x(k)$ aan de uitgang.

2.2 De Hilbert Transformatie

Amplitude modulatie van een basisband signaal met een draaggolf met frequentie θ_c resulteert in een banddoorlaatsignaal met herhalingen van het basisbandsignaal rond $+\theta_c$ en $-\theta_c$.

Als het oorspronkelijk signaal een bandbreedte B heeft dan heeft het banddoorlaatsignaal een bandbreedte $2B$. Verzending van dit signaal vereist tweemaal de bandbreedte van het oorspronkelijk signaal.

We kunnen een beduidende vermindering van de noodzakelijke bandbreedte bekomen als we gebruik maken van de symmetrie van het spectrum. Dit is de filosofie die schuilt achter de Hilbert transformatie.

De transferfunctie van een ideale Hilbert transformator is:

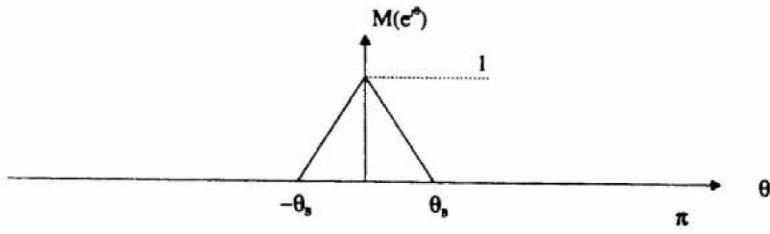
$$H(e^{j\theta}) = \begin{cases} j & -\pi \leq \theta < 0 \\ -j & 0 \leq \theta \leq \pi \end{cases} \quad (2.4)$$

De ideale Hilbert transformator is fysisch niet realiseerbaar, wel kunnen goede benaderingen gebruikt worden –in de praktijk.

Veronderstel $m(k)$ een reëel basisbandsignaal met DTFT $M(e^{j\theta})$ zoals in figuur 2.4 op de pagina hierna. De frequentie θ_B is de bandbreedte van $m(k)$. We definieren $\hat{m}(k)$ als de Hilbert getransformeerde van $m(k)$:

$$\hat{m}(k) = m(k) * h(k)$$

Het Hilbert getransformeerde signaal $\hat{m}(k)$ heeft een DTFT $\hat{M}(e^{j\theta})$. Deze heeft dezelfde amplitude als $M(e^{j\theta})$ maar heeft een fase die 90° verdraaid is voor negatieve frequenties en -90° voor positieve frequenties.

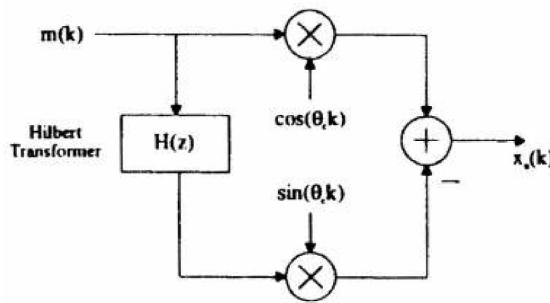


Figuur 2.4: DTFT van een reëel basisbandsignaal $m(k)$

Verzonden signaal $x_u(k)$

We realiseren nu het signaal $x_u(k)$ als:

$$x_u(k) = m(k) \cos \theta_c k - \hat{m}(k) \sin \theta_c k$$



2.2.1 Bandbreedte van het verzonden signaal $x_u(k)$

We zullen aantonen dat de bandbreedte van $x_u(k)$ slechts de helft is van de bandbreedte van een amplitude gemoduleerd signaal $m(k)$. Daarvoor definiëren we het complexe signaal $u(k)$ als:

$$u(k) = m(k) + j\hat{m}(k)$$

Het signaal $x_u(k)$ kan dan geschreven worden als:

$$x_u(k) = \Re\{u(k) \cdot e^{j\theta_c k}\}$$

Vermits we e^{jx} kunnen schrijven als $\cos x + j \sin x$, waardoor $u(k) \cdot e^{j\theta_c k} = (m(k) + j\hat{m}(k)) \cdot (\cos \theta_c k + j \sin \theta_c k)$.

Het signaal $u(k)$ heeft een DTFT $U(e^{j\theta})$ die gegeven wordt als:

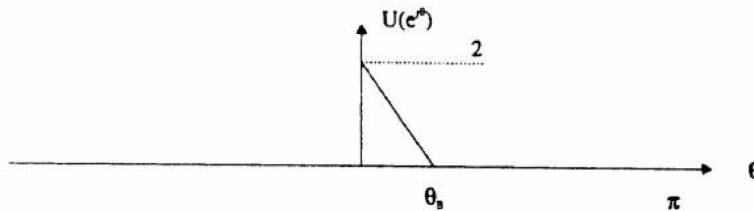
$$\begin{aligned} U(e^{j\theta}) &= M(e^{j\theta}) + j \underbrace{\hat{M}e^{j\theta}}_{= M(e^{j\theta}) \cdot H(e^{j\theta})} = M(e^{j\theta}) (1 + jH(e^{j\theta})) \\ &= M(e^{j\theta}) \cdot H(e^{j\theta}) \end{aligned}$$

Dit komt overeen met:

$$U(e^{j\theta}) = \begin{cases} 0 & -\pi \leq \theta < 0 \\ 2M(e^{j\theta}) & 0 \leq \theta \leq \pi \end{cases}$$

Dit wordt ingezien als men vergelijking 2.4 op pagina 20 bekijkt.

Het spectrum wordt weergegeven op figuur 2.5.



Figuur 2.5: Spectrum van het complexe signaal $u(k)$

Merk op dat het enkel bestaat uit de bovenste helft van $M(e^{j\theta})$.

We bepalen nu $X_u(e^{j\theta})$ via $x_u(k) = \Re\{u(k) \cdot e^{j\theta_c k}\}$. Als we het reële deel zoeken kunnen we dit via de volgende formule, $a = \frac{1}{2}((a+jb)+(a-jb))$ waarbij het complexe getal wordt voorgesteld als $a+jb$. Hierdoor vinden we voor $\Re\{u(k) \cdot e^{j\theta_c k}\}$:

$$\Re\{u(k) \cdot e^{j\theta_c k}\} = \frac{1}{2}u(k)e^{j\theta_c k} + \frac{1}{2}(u(k)e^{j\theta_c k})^* = \frac{1}{2}u(k)e^{j\theta_c k} + \frac{1}{2}u^*(k)e^{-j\theta_c k}$$

M.b.V. de eigenschap:

$$u^*(k) \xrightarrow{DTFT} U^*(e^{-j\theta})$$

Geldt:

$$X_u(e^{j\theta}) = \frac{1}{2} [U(e^{j(\theta-\theta_c)}) + U^*(e^{j(-\theta-\theta_c)})]$$

Het spectrum $U(e^{j(\theta-\theta_c)})$ komt overeen met $U(e^{j\theta})$ naar rechts verschoven over θ_c . Het spectrum $U^*(e^{j(-\theta-\theta_c)})$ komt overeen met $U(e^{j\theta})$ te spiegelen over $\theta=0$ en naar links op te schuiven over θ_c .

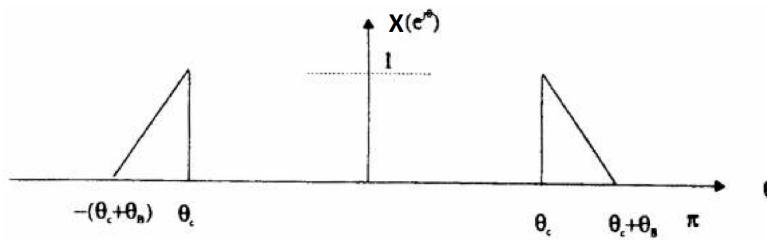
We bekomen aldus:

$$X_u(e^{j\theta}) = \begin{cases} M(e^{j(\theta-\theta_c)}) & \theta_c \leq |\theta| \leq \theta_c + \theta_B \\ 0 & elders \end{cases}$$

Het spectrum is voorgesteld in de figuur 2.6 op de volgende pagina en komt dus overeen met de bovenste zijband.

Merk op dat de bandbreedte van dit signaal de helft is van wat we zouden krijgen mocht $m(k)$ amplitude gemoduleerd zijn.

De onderste zijband kan bekomen worden door $u^*(k)$ te nemen i.p.v. $u(k)$.



Figuur 2.6: Spectrum van het verzonden signaal $x_u(k)$. Opmerking de verticale-as moet $X_u(e^{j\theta})$ zijn.

2.2.2 Recuperatie oorspronkelijk signaal $m(k)$

Om nu $m(k)$ uit de bovenste zijbandsignaal $x_u(k)$ te recupereren, moduleren we met $\cos(\theta_c k)$. We bekomen:

$$x_u(k) \cos(\theta_c k) = m(k) \cos^2(\theta_c k) - \hat{m}(k) \sin(\theta_c k) \cos(\theta_c k)$$

Dit kan omgewerkt worden tot:

$$x_u(k) \cos(\theta_c k) = \frac{1}{2}m(k) + \frac{1}{2}m(k) \cos(2\theta_c k) - \frac{1}{2}\hat{m}(k) \sin(2\theta_c k)$$

Dus als we de uitgang van de demodulator door een laagdoorlaat filter sturen met versterking twee, recupereren we $m(k)$.

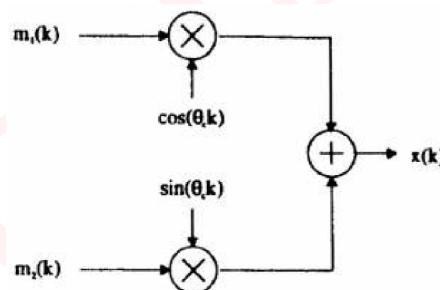
Indien nodig kan $\hat{m}(k)$ gerecupereerd worden op een gelijkaardige manier door $x_u(k)$ te moduleren met $\sin(\theta_c k)$.

2.3 Quadrature Amplitude Modulation (QAM)

De procedure voor het genereren van SSB-signalen (Single Side Band) en het herwinnen van de informatie daaruit, suggereert een ander schema voor bandbreedte compressie, nl. QAM.

Veronderstel dat $m_1(k)$ en $m_2(k)$ twee signalen zijn waarmee we een QAM-signaal $x(k)$ kunnen vormen:

$$x(k) = m_1(k) \cos(\theta_c k) + m_2(k) \sin(\theta_c k)$$



De draaggolf $\cos(\theta_c k)$ is de in-fase draaggolf en de in-fase verschoven draaggolf $\sin(\theta_c k)$ is de kwadratuur drager.

We kunnen $x(k)$ schrijven als:

$$x(k) = m_1(k) \left[\frac{e^{j\theta_c k} + e^{-j\theta_c k}}{2} \right] + m_2(k) \left[\frac{e^{j\theta_c k} - e^{-j\theta_c k}}{2j} \right]$$

2.3.1 Bandbreedte van het QAM-signaal

M.b.v. de modulatie-eigenschap vinden we voor de DTFT van het QAM signaal:

$$X(e^{j\theta}) = \frac{1}{2} \left[M_1(e^{j(\theta+\theta_c)}) + M_1(e^{j(\theta-\theta_c)}) + jM_2(e^{j(\theta+\theta_c)}) - jM_2(e^{j(\theta-\theta_c)}) \right]$$

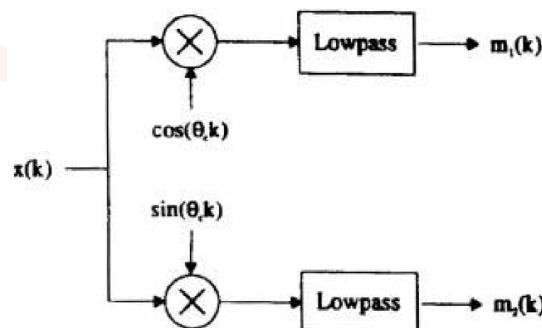
Merk op dat de bandbreedte van $x(k)$ niet groter is dan de grootste van de bandbreedtes van $m_1(k)$ en $m_2(k)$. Het spectrum bevindt zich altijd tussen $\theta_c - \theta_{max}$ en $\theta_c + \theta_{max}$.

De BB wordt bepaald door het signaal met de grootste BB, we versturen nu $2BB_{max}$ door. Dit in tegenstelling tot wanneer we de twee signalen apart zouden doorsturen via amplitude modulatie, dan wordt er $2BB_{max}$ en $2BB_{min}$ doorgestuurd.

2.3.2 Reconstructie beide signalen $m_1(k)$ en $m_2(k)$

Zelfs al overlappen de spectrale componenten van $m_1(k)$ en $m_2(k)$, toch kunnen we beiden reconstrueren.

De signalen $m_1(k)$ en $m_2(k)$ kunnen we reconstrueren uit het QAM-signaal $x(k)$ door filtering van $x(k) \cos(\theta_c k)$ of $x(k) \sin(\theta_c k)$ met een laagdoorlaatfilter met versterking twee, om zo de hoge frequentiecomponenten te verwijderen.

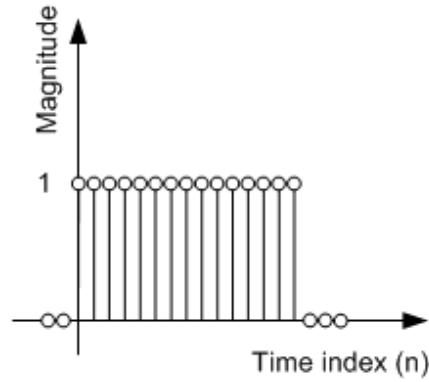


Figuur 2.7: Recuperatie van de oorspronkelijke signalen $m_1(k)$ en $m_2(k)$

2.4 De rechthoekige vensterfunctie

Om een discreet signaal te beperken in de tijd kunnen we gebruik maken van een vensterfunctie.

$$w(k) = \begin{cases} 0 & 0 \leq n \leq N-1 \\ 0 & \text{elders} \end{cases}$$



Figuur 2.8: Rechthoekige vensterfunctie $w(k)$

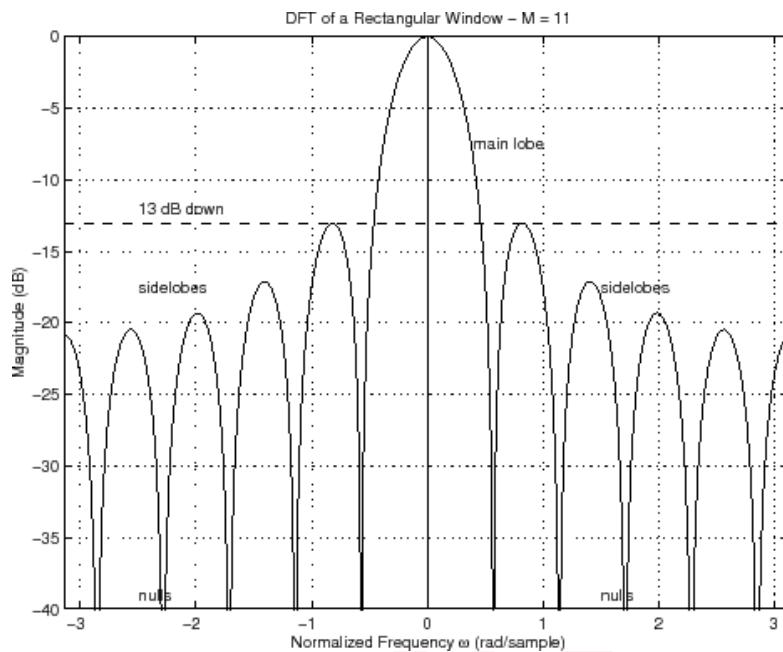
2.4.1 Spectrum van een rechthoekige vensterfunctie

Het spectrum van een rechthoekige vensterfunctie kan als volgt worden berekend:

$$\begin{aligned} W(e^{j\theta}) &= \sum_{k=-\infty}^{\infty} w(k) \cdot e^{-j\theta k} \quad (\text{zie verg.2.1}) \\ &= \sum_{k=0}^{N-1} e^{-j\theta k} \\ &= \frac{1 - e^{-jN\theta}}{1 - e^{-j\theta}} \\ &= \frac{e^{-j\frac{N\theta}{2}} \left(e^{j\frac{N\theta}{2}} - e^{-j\frac{N\theta}{2}} \right)}{e^{-j\frac{\theta}{2}} \left(e^{j\frac{\theta}{2}} - e^{-j\frac{\theta}{2}} \right)} \\ &= e^{-j\frac{\theta}{2}(N-1)} \frac{\sin \frac{N\theta}{2}}{\sin \frac{\theta}{2}} \end{aligned}$$

Algemeen kan men schrijven:

$$W_N(e^{j\theta}) = e^{-j\frac{\theta}{2}(N-1)} \frac{\sin \frac{N\theta}{2}}{\sin \frac{\theta}{2}} \quad (2.5)$$



Figuur 2.9: Het spectrum van een rechthoekige vensterfunctie (in dB).

Eigenschappen van de DTFT van een rechthoekige vensterfunctie Men kan de vensterfunctie manipuleren om bepaalde eigenschappen te verkrijgen, sommige eigenschappen liggen echter vast

- Aantal nuldoorgangen

Het aantal nuldoorgangen wordt bepaald door het aantal punten N . Dit zien we in als we vergelijking 2.5 op de pagina hiervoor gelijkstellen aan nul. Hierbij moet $\sin \frac{N\theta}{2}$ gelijk zijn aan nul. De sinus is nul bij veelvouden van π .

$$\frac{n\theta}{2} = k\pi \rightarrow \theta = \frac{2k\pi}{N}$$

- Breedte van de hoofdlobe

De breedte van de hoofdlobe is tweemaal de eerste nuldoorgang, dus $2 \cdot \frac{2\pi}{N}$. We zien hierdoor in dat de breedte van de hoofdlobe bepaald wordt door het aantal punten N .

- Volgend maximum

Het volgend maximum ligt tussen de eerste en de tweede nuldoorgang. Wegens symmetrie kunnen we de frequentie bepalen door het midden te nemen van de eerste ($\frac{2\pi}{N}$) en ($\frac{4\pi}{N}$) tweede nuldoorgang. We bekomen dus $\theta = \frac{3\pi}{N}$. De amplitude bekomen we door:

$$\left| \frac{\sin \left(\frac{3\pi}{N} \cdot \frac{N}{2} \right)}{\sin \left(\frac{3\pi}{N} \cdot \frac{1}{2} \right)} \right| = \left| \frac{-1}{\sin \left(\frac{3\pi}{N} \cdot \frac{1}{2} \right)} \right| \simeq \frac{1}{\frac{3\pi}{N}}$$

De laatste stap geldt bij kleiner waarden, $\sin x$ wordt dan nagenoeg x .

- Verhouding van zijlobe op de hoofdlobe

Deze verhouding is $\frac{2}{3\pi}$ wat aantoont dat deze verhouding onafhankelijk is van N – het verbreden van het venster zal geen invloed hebben op de rimpel.

- Verschuiven in de tijd

Er is geen gevolg voor de weergavekarakteristiek in het frequentie domein bij tijdsverschuiving. Deze verschuiving komt men door het vermenigvuldigen van een exponent in het frequentiegebied.

- Symmetrie van de rechthoekige vensterfunctie

$$v_N(k) = \begin{cases} 1 & |k| \leq \frac{N-1}{2} \\ 0 & \text{elders} \end{cases}$$

Dit komt neer op een verschuiving naar links over $\frac{N-1}{2}$.

$$v_N(k) = w_N(k + \frac{N-1}{2})$$

De DTFT van $v_N(k)$ wordt dan:

$$\begin{aligned} V_N(e^{j\theta}) &= e^{j(\frac{N-1}{2})\theta} \cdot W_N(e^{j\theta}) \\ &= \frac{\sin \frac{N\theta}{2}}{\sin \frac{\theta}{2}} \end{aligned}$$

Dus $|V_N(e^{j\theta})| = |W_N(e^{j\theta})|$.

Dit heeft echter wel een gevolg bij de fasiekarakteristiek, $\varphi V_N(e^{j\theta}) = \varphi W_N(e^{j\theta}) + \frac{N-1}{2}\theta$

2.5 Toepassing op cosinus en som van cosinussen

Toepassing: Cosinus beperking in de tijd

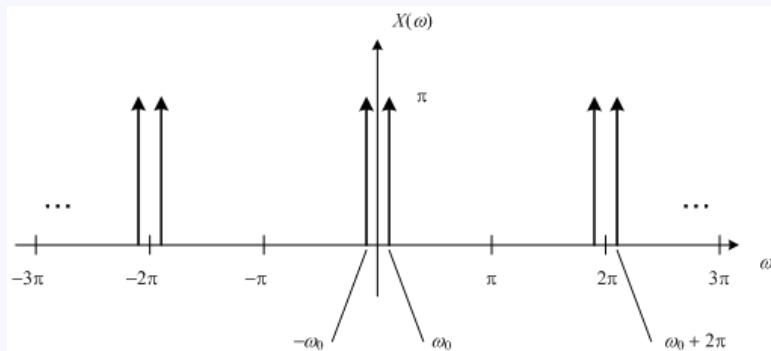
Wanneer we het spectrum van een cosinus met frequentie θ_0 bekijken zien we twee spectrale lijnen op $-\theta_0$ en θ_0 . Wat gebeurt er als we een rechthoekige vensterfunctie toepassen op de cosinus.

$$x_v(k) = \cos \theta_0 k \cdot v(k)$$

De DTFT wordt dan:

$$X_v(e^{j\theta}) = X(e^{j\theta}) * V(e^{j\theta})$$

Als we de moeilijke convolutie achterwegen laten en kijken naar het tijds-discreet signaal kunnen we deze zien als een amplitude modulatie. We weten dat het spectrum wordt opgesplitst bij AM. We verkrijgen een *sinc* functie rond $-\theta_0$ en θ_0 , i.p.v. de theoretische spectrale lijnen.



Figuur 2.10: Vereenvoudigde weergave van het spectrum van een cosinus die in de tijd is beperkt door een rechthoekige vensterfunctie. Deze spectrale lijnen moeten een *sinc* functie voorstellen.

Toepassing: Som van cosinussen met verschillende frequentie

Stel:

$$f(k) = \cos(\theta_0 k) + \cos(\theta_0 + \Delta\theta)k$$

We kunnen de cosinussen in het frequentiedomein zeker herkennen als de *sinc* functies elkaar niet overlappen. We moeten de breedte van het venster aanpassen zodat $\Delta\theta >> \frac{2\pi}{N}$.

De breedte van de hoofdlobes is, zoals gezien, $\frac{4\pi}{N}$, dus als we niet willen dat de hoofdlobes elkaar overlappen moeten de middens van de hoofdlobes zeker $\Delta\theta$ van elkaar verwijderd zijn.

Hoofdstuk 3

Discrete Fouriertransformatie

3.1 De Discrete Fourier Reeks

De discrete fouriertransformatie of DFT is een fouriertransformatie die veel wordt toegepast in de digitale signaalbewerking en verwante vakgebieden voor het analyseren van de frequenties die aanwezig zijn in een bemonsterd signaal, en voor het uitvoeren van bewerkingen zoals discrete convoluties. De DFT kan efficiënt worden berekend door gebruik te maken van het FFT-algoritme.

De discrete fouriertransformatie is een lineaire transformatie en een discrete vorm van de fouriertransformatie. Ze transformeert een periodieke (periode N) en discrete (n getallen) rij opnieuw in een periodieke discrete rij.

Een discreet signaal (met eindige duur)

De rij van N complexe getallen $x(0), \dots, x(N - 1)$ wordt door de DFT getransformeerd in de rij van N complexe getallen $X_0, \dots, X_{N - 1}$ volgens de volgende formule:

$$X(n) = \sum_{k=0}^{N-1} x(k)w_N^{nk} \quad n = 0, \dots, N - 1 \text{ met } \omega_N = \frac{-j2\pi}{N} \quad (3.1)$$

De DFT kan men ook schrijven als een speciale vorm van de DTFT:

$$X(e^{j\theta}) = \sum_{k=-\infty}^{+\infty} x(k)e^{-jk\theta} \quad (2.1 \text{ revisited})$$

$$X(n) = X(e^{j\theta})|_{\theta=\frac{2\pi n}{N}}$$

De eenheidscirkel wordt door ω_N in gelijke delen verdeeld. De DFT kan zo gezien worden als de DTFT waarbij de waarden uit het spectrum zijn gehaald.

3.2 Eigenschappen van de DFT

De eigenschappen worden kort besproken, maar zal niet tot de te-kennen leerstof behoren.

Wanneer er wordt gekeken naar $x(k)$ en $X(n)$ van eindige duur (N) dan zal de verwijzing naar k of n buiten $0 \leq k, n \leq N - 1$ anders moeten worden bekeken. Vermits we zitten met periodische herhalingen zullen we $modulo N$ in rekening moeten brengen. Een reeks met $N = 5$ kan dan worden benaderd door $x(k - 1 \bmod 5)$.

Deze cyclische verschuivingen zijn gemakkelijk in te zien in het geval van periodische signalen.

3.2.1 Lineariteit

Als beide reeksen dezelfde lengte hebben gaat de eigenschap van lineairiteit op. Zoniet zal men de kortere reeks moeten uitbreiden met nullen.

3.2.2 Cyclische tijdsverschuiving

$$\begin{aligned} x(k) &\xrightarrow{DFT} X(n) \\ x(k - m \bmod N) &\xrightarrow{DFT} \omega_N^{nm} X(n) \end{aligned}$$

Het is niet nodig om module N correctie te brengen bij een cyclische tijdsverschuiving vermits in w_N periodiek is met periode N .

$$\sum_{k=0}^{N-1} x(k - m) w_N^{nk}$$

Stel $k - m = r$:

$$= \sum_{r=-m}^{N-1-m} x(r) w_N^{n(r+m)}$$

De term $w_N^{(nm)}$ heeft geen invloed door de sommatie veranderlijken, dus:

$$= w_N^{(nm)} \sum_{r=-m}^{N-1-m} x(r) w_N^{nr}$$

Wat weer een periodieke signaal is met periode N .

$$= w_N^{(nm)} X(n)$$

3.2.3 Modulatie

Een modulatie is een cyclische frequentie-verschuiving.

$$\begin{aligned} x(k) &\xrightarrow{DFT} X(n) \\ \omega_N^{-km} x(k) &\xrightarrow{DFT} X(n - m \bmod N) \end{aligned}$$

Een verschuiving van het spectrum $X(n)$ leidt tot een complexe tijdsfunctie. We krijgen wel een reele tijdsfunctie als we een dubbele frequentie-verschuiving op $X(n)$ toepassen, dit gebeurt bijvoorbeeld bij modulatie.

3.2.4 Cyclische convolutie in het tijdsdomein

Eén van de belangrijke toepassingen van een fouriertransformatie is dat een convolutie in het tijdsdomein resulteert in een vermenigvuldiging in het frequentiedomein.

$$DFT\{x(k) * h(k)\} \Leftrightarrow X'(n) \cdot H'(n)$$

Als we met behulp van de cyclische convolutie de lineaire convolutie willen berekenen moeten we eerst de signalen uitbreiden met nullen zodat we $N + M - 1$ waarden bekomen (met N en M het aantal waarden van de twee signalen).

Het nut van de cyclische convolutie is het makkelijk berekenen van de lineaire convolutie voor periodieke functies. Als we zouden werken met de lineaire convolutie zouden we de convolutie moeten nemen van oneindig lange signalen.

De definitie voor de cyclische convolutie:

$$Z(k) = \sum_{m=0}^{N-1} x(m)y(k - m \bmod N) = x(k) \otimes y(k) \quad (3.2)$$

Als we een voorbeeld uitwerken voor de cyclische convolutie voor $f(k)$ en $g(k)$ met lengte 3:

$$\begin{aligned} y(0) &= f(0)g(-0 \bmod 3) + f(1)g(-1 \bmod 3) + f(2)g(-2 \bmod 3) \\ y(1) &= f(0)g(0) + f(1)g(2) + f(2)g(1) \end{aligned}$$

...

We zien dat er altijd een som is van drie producten. Dit kunnen we ook schrijven als een matrix, waarbij de kolommen cyclisch zijn verschoven (dit komt neer op het naar beneden 'duwen' van de kolom, zie pijlen):

$$\begin{pmatrix} y(0) \\ y(1) \\ y(2) \end{pmatrix} = \begin{pmatrix} g(0) & g(2) & g(1) \\ \downarrow & \downarrow & \\ g(1) & g(0) & g(2) \\ g(2) & g(1) & g(0) \end{pmatrix} \cdot \begin{pmatrix} f(0) \\ f(1) \\ f(2) \end{pmatrix}$$

Alternatief

Het alternatief is te werken via de Z-transformatie:

$$f(k) \otimes g(k) \rightarrow F(z) \cdot G(z) \cdot \text{mod}(Z^{-N} - 1)$$

3.3 Fast Fourier Transformatie (FFT)

We willen zo snel mogelijk de DFT kunnen berekenen van een discreet signaal. Met de huidige kennis zouden we per waarde een som moeten berekenen van N termen.

Zie formule 3.1 op pagina 29 voor de DFT van een signaal $x(k)$.

Voor een N -punt DFT zouden er N^2 complexe berekeningen vereist zijn, om deze berekeningen te verminderen zijn er FFT algoritmes ontwikkeld waarvan we er één gaan bekijken, de Radix-2 FFT.

3.3.1 De Radix-2 FFT

Bij de Radix-2 FFT moet het aantal waarden een veelvoud zijn van 2, i.e. $N = 2^K$ met K een geheel getal.

We verdelen $x(k)$ in twee subreeksen:

$$\begin{aligned} a(k) &= x(2k) \quad k = 0, 1, \dots, \frac{N}{2} - 1 \\ b(k) &= x(2k + 1) \quad k = 0, 1, \dots, \frac{N}{2} - 1 \end{aligned}$$

De DFT wordt dan:

$$X(n) = \sum_{k=0}^{\frac{N}{2}-1} a(k)W_N^{2nk} + \sum_{k=0}^{\frac{N}{2}-1} b(k)W_N^{(2k+1)n}$$

Als $W_N = e^{-j\frac{2\pi}{N}}$ → $W_N^2 = e^{-j\frac{2\pi \cdot 2}{N}} = e^{-j\frac{2\pi}{2}} = W_{\frac{N}{2}}$ dan kan dit geschreven worden als:

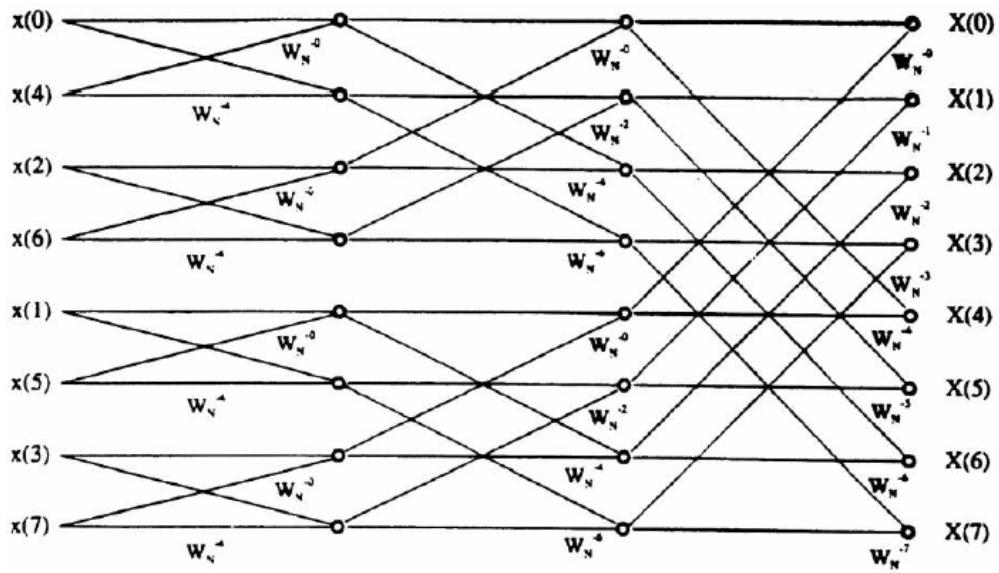
$$X(n) = \sum_{k=0}^{\frac{N}{2}-1} a(k)W_{\frac{N}{2}}^{nk} + W_N^n \sum_{k=0}^{\frac{N}{2}-1} b(k)W_{\frac{N}{2}}^{nk}$$

Deze techniek kan worden uitgevoerd tot we met een 2-punt DFT zitten:

$$Q(n) = \sum_{k=0}^1 q(k)W_2^{nk} \quad n = 0, 1$$

$$\rightarrow \begin{cases} Q(0) &= q(0) + q(1) \\ Q(1) &= q(0) - q(1) \end{cases}$$

We kunnen deze bewerkingen voorstellen door een FFT butterfly:



Figuur 3.1: $\frac{N}{2}$ butterflies – 8-punt radix-2 FFT

Via het spiegelen van de binaire woorden kunnen we de plaatsen achterhalen van de functiewaarde van het discrete signaal.

Via de Radix-2 FFT zijn we erin geslaagd tijdswinst te boeken door minder berekeningen te moeten uitvoeren. We hebben een reductie van N^2 berekeningen voor de directe DFT naar $N\log_2(N)$ voor de 2-radix FFT.

Hoofdstuk 4

Systeem Analyse

4.1 De Z-transformatie

Wordt besproken in de bijlage (cursus vorig jaar).

4.2 Polen en nullen als systeembeschrijving

Door het analyseren van de polen en nullen krijgen we een idee met welk soort systeem we te maken hebben.

De causale LTO-discrete systemen waarmee we doorgaans te maken hebben zijn te beschrijven met de volgende lineaire differentievergelijking:

$$y[n] = \sum_{i=0}^N b_i x[n-i] + \sum_{i=1}^M a_i y[n-i] \quad (4.1)$$

In dit algemeen geval zitten we met een recursief systeem, d.w.z. dat de voorgaande uitgangen bepalend zijn voor de uitgang.

De bijhorende systeemfunctie $H(z)$ is:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{i=0}^N b_i z^{-i}}{1 - \sum_{i=1}^M a_i z^{-i}} \quad (4.2)$$

Dit bekomen we door de Z-transformatie van vergelijking 4.1 te nemen.

Om te zorgen dat er rekening wordt gehouden met alle nullen en polen worden de machten

positief gemaakt:

$$H(z) = \frac{\sum_{i=0}^N b_i z^{N-i}}{z^M - \sum_{i=1}^M a_i z^{M-i}} \cdot z^{M-N}$$

Als we rekening houden met de coëfficiënt van de hoogste macht in de teller en noemer kunnen we dit ook schrijven als:

$$H(z) = b_0 \frac{(z - z_1)(z - z_2) \cdots (z - z_N)}{(z - p_1)(z - p_2) \cdots (z - p_M)} z^{M-N} \quad (4.3)$$

$$= b_0 \frac{\prod_{i=1}^N (z - z_i)}{\prod_{j=1}^M (z - p_j)} z^{M-N} \quad (4.4)$$

Hierin zijn de z_i de nullen en p_j de polen.

$H(z)$ wordt dus op een constante b_0 na, volledig bepaald door de waarden van z_i en p_j . Deze nullen en polen kunnen grafisch weergegeven worden in het complexe z-vlak.

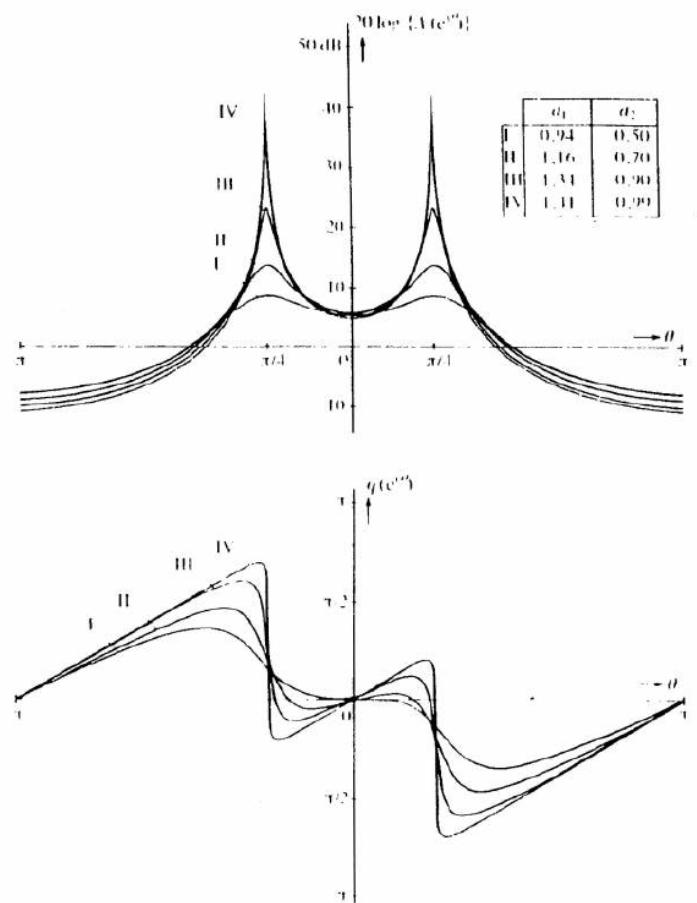
In de volgende figuren zijn respectievelijk de frequentieresponsies en de polen- en nullen diagramma gegeven van vier tweede-orde secties:

$$H(z) = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2}}$$

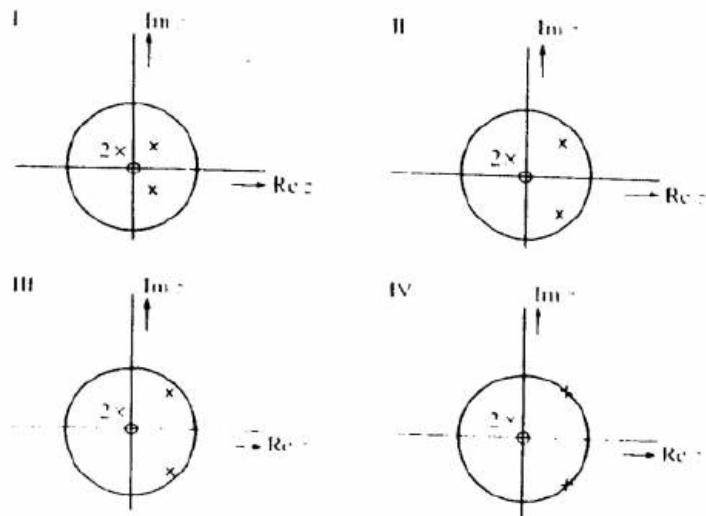
Na herschrijven voor positieve machten wordt dit:

$$H(z) = \frac{z^2}{z^2 - a_1 z - a_2}$$

Door figuren (??–?? op pagina's 36–37) ziet men dat hoe dichter de polen de eenheidscirkel naderen hoe groter de impact is van die pool bij die hoek op de weergavekarakteristiek.



De modulus en het argument van de frequentieresponsie $H(e^{j\theta})$ voor een tweede orde systeem met $b_0 = 1$ en $b_1 + b_2 = 0$ en a_1 en a_2 volgens de tabel.



Bij LTO-discrete systemen speelt de eenheidscirkel in het complexe z -vlak een belangrijke rol: de transferfunctie $H(z)$ op de eenheidscirkel is precies gelijk aan de frequentieresponsie $H(e^{j\theta})$. Vermits het fundamenteel interval loopt van $\theta = -\pi \rightarrow \pi$.

Omdat we altijd te maken hebben met systemen waarvoor de coëfficiënten a en b van $H(z)$ reëel zijn, kunnen de nulpunten en de polen alleen reëel zijn of voorkomen als toegevoegd complexe paar.

4.2.1 Conclusies

We kunnen uit de polen en nullen beschrijving van praktisch realiseerbare LTO-discrete systemen direct een aantal conclusies trekken.

De belangrijkste zijn:

Stabiel systeem

Bij een stabiel systeem liggen alle polen binnen de eenheidscirkel; nulpunten kunnen binnen of buiten de eenheidscirkel liggen.

Relatie tussen polen en nullen en de frequentieresponsie

De frequentieresponsie van een LTO-systeem wordt gegeven door:

$$H(e^{j\theta}) = b_0 \frac{(e^{j\theta} - z_1)(e^{j\theta} - z_2) \cdots (e^{j\theta} - z_N)}{(e^{j\theta} - p_1)(e^{j\theta} - p_2) \cdots (e^{j\theta} - p_M)} e^{j\theta(M-N)}$$

Dit bekomen we door $e^{j\theta}$ te substitueren in de Z-getransformeerde systeemfunctie (zie vergelijking 4.4 op pagina 35).

Weergavekarakteristiek (i.h. frequentie domein) De modulus van de frequentieresponsie wordt gegeven door:

$$|H(e^{j\theta})| = |b_0| \frac{|(e^{j\theta} - z_1)| \cdot |(e^{j\theta} - z_2)| \cdots |(e^{j\theta} - z_N)|}{|(e^{j\theta} - p_1)| \cdot |(e^{j\theta} - p_2)| \cdots |(e^{j\theta} - p_M)|} \cdot |e^{j\theta(M-N)}|$$

Als we bijvoorbeeld de responsie bekijken bij θ_0 dan zal $|(e^{j\theta_0} - z_1)|$ staan voor twee vectoren. Deze vectoren beginnen in de oorsprong en eindigen in respectievelijk, het punt $e^{j\theta_0}$ en z_1 . Het verschil van deze vectoren is een vector die deze twee punten verbind. De modulus van het verschil van die vectoren levert ons de afstand op tussen deze twee punten.

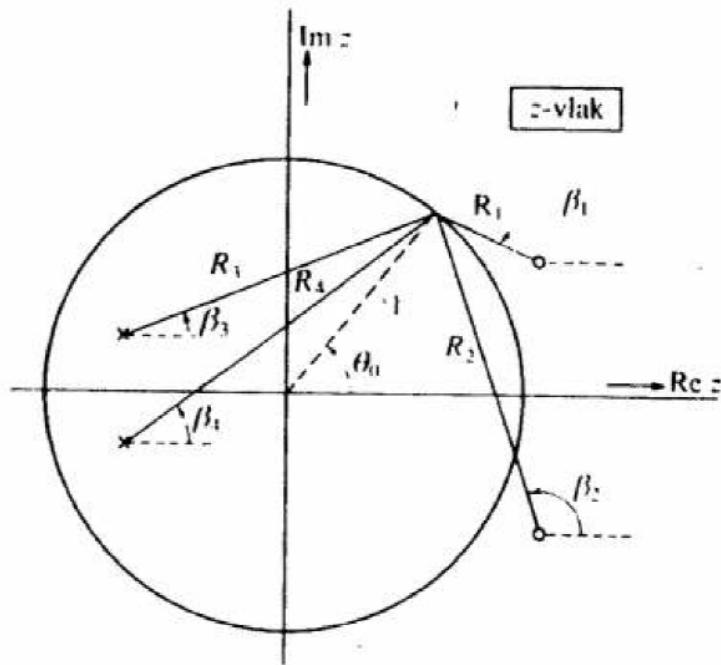
We kunnen zeggen dat de amplitude $Ae^{j\theta_0} = |H(e^{j\theta_0})|$ bepaald wordt door het product van alle afstanden tussen het punt $z = e^{j\theta_0}$ en de nullen gedeeld door het product van alle afstanden tussen het punt $z = e^{j\theta_0}$ en de polen, rekening houdend met b_0 .

Fasekarakteristiek (i.h. frequentie domein) Het argument van de frequentieresponsie wordt gegeven door:

$$\begin{aligned} \varphi(H(e^{j\theta})) &= \left\{ \begin{array}{l} 0 \\ \pi \end{array} \right\} + \left(\varphi(e^{j\theta} - z_1) + \varphi(e^{j\theta} - z_2) + \dots + \varphi(e^{j\theta} - z_N) \right) + \varphi(\theta(M-N)) \\ &\quad - \left(\varphi(e^{j\theta} - p_1) + \varphi(e^{j\theta} - p_2) + \dots + \varphi(e^{j\theta} - p_M) \right) \end{aligned}$$

De fase $\varphi(H(e^{j\theta_0}))$ vinden we dus door alle fasebijdragen van de nullen op te tellen en daar alle fasebijdragen van de polen vanaf te trekken.

Dit is grafisch weergegeven op grafiek 4.1 op de pagina hierna voor een 2^{de} orde systeem ($M = 2$) met twee nullen ($N = 2$).



Grafische bepaling van $A(e^{j\varphi})$ en $\varphi(e^{j\theta_0})$.

Figuur 4.1: Grafische bepaling van $A(e^{j\varphi})$ en $\varphi(e^{j\theta_0})$

$$A(e^{j\varphi}) = \frac{R_1 R_2}{R_3 R_4}$$

$$\varphi(H(e^{j\theta_0})) = \beta_1 + \beta_2 - \beta_3 - \beta_4$$

We zien dat iedere pool en iedere nul de meeste invloed heeft in het frequentie gebied dat overeenkomt met dat stuk van de eenheidscirkel dat het dichtst bij die pool of die nul ligt. Bovendien heeft een pool of een nul relatief gezien meer invloed naarmate hij dichter bij de eenheidscirkel ligt.

In het extreme geval dat een nul op de eenheidscirkel ligt is de amplitude van de frequentie responsie nul bij $\theta = \theta_1$ en vertoont de fase daar een sprong van π radialen. Dit duidt aan dat hier de stopband gelegen is vermits de amplitude 0 is, i.e. er wordt niets doorgelaten.

In het andere geval waarbij de polen dicht –of in het extreme geval op de eenheidscirkel – liggen, heeft men te maken met een doorlaatband. Hier zal de fasiekarakteristiek eveneens bij $\theta = \theta_2$ een fasesprong van π radialen vertonen.

minimum-fase-netwerk

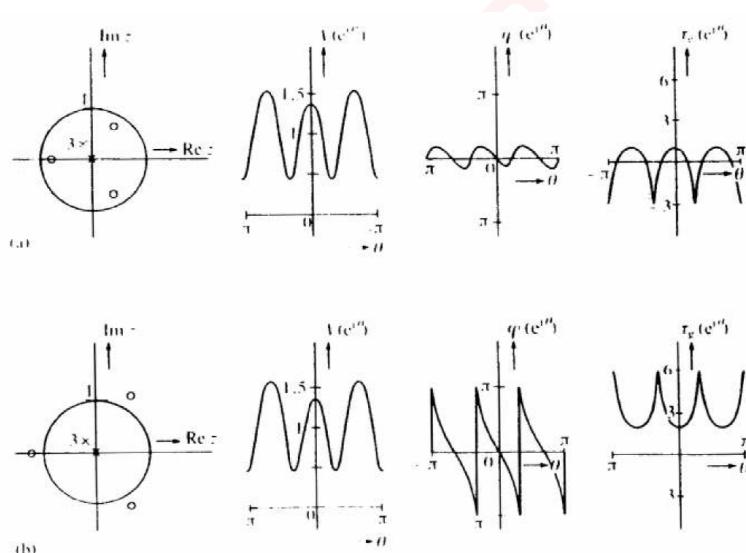
Als alle nullen binnen de eenheidscirkel liggen, hebben we te maken met een zogenaamd minimum-fase-netwerk (zie verder).

Door de nullen te spiegelen t.o.v. de eenheidscirkel veranderen we wel de fase-karakteristiek van het systeem maar niet de amplitude-karakteristiek (op een constante na).

Onder spiegelen t.o.v. de eenheidscirkel verstaan we het vervangen van een nulpunt $z_i = r_i e^{j\theta_i}$ door een nul $z'_i = \frac{1}{r_i} e^{-j\theta_i}$.

In figuur 4.2 is een voorbeeld gegeven van dit spiegelen voor een minimum-fase-systeem dat 3 polen in de oorsprong en 3 nullen bevat.

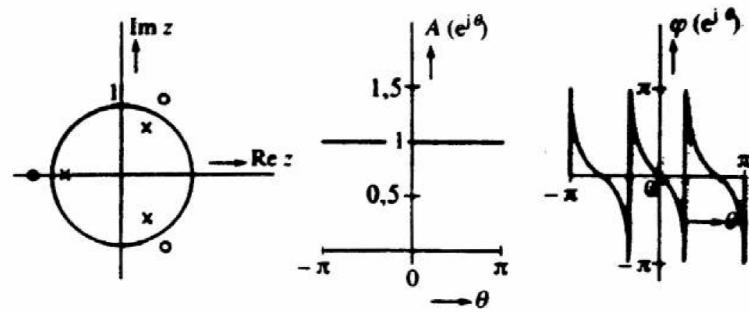
We zien duidelijk dat de fasekarakteristiek van het minimum-fase-systeem veel minder sterk varieert en de groepslooptijd dus kleiner is dan bij het niet-minimum-fase-systeem. Dit geldt algemeen.



Figuur 4.2: (a) Minimum-fase-systeem met drie polen in de oorsprong en drie nulpunten.
(b) Niet-minimum-fase-systeem uit (a) verkregen door spiegelen van de nulpunten t.o.v. de eenheidscirkel.
Dit leidt tot een grotere groepslooptijd τ_g .

All pass-netwerk

Een stabiel *all pass-netwerk* bevat alleen polen binnen de eenheidscirkel en nulpunten buiten de eenheidscirkel. Polen en nullen komen steeds in paren voor die gespiegeld zijn t.o.v. de eenheidscirkel.

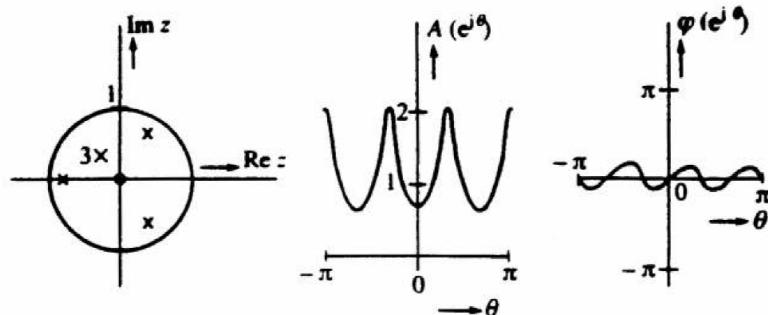


Figuur 4.3: All pass netwerk

Als we het minimum-fase (MF) systeem van hiervoor terug bekijken zien we in dat het MF-systeem in cascade met een all pass-netwerk resulteert in een niet-MF. Als men de omgekeerde redenering volgt kan men zo van een niet-MF systeem een MF systeem construeren om een minder sterk varieerde fasiekarakteristiek te bekomen.

All pole-netwerk

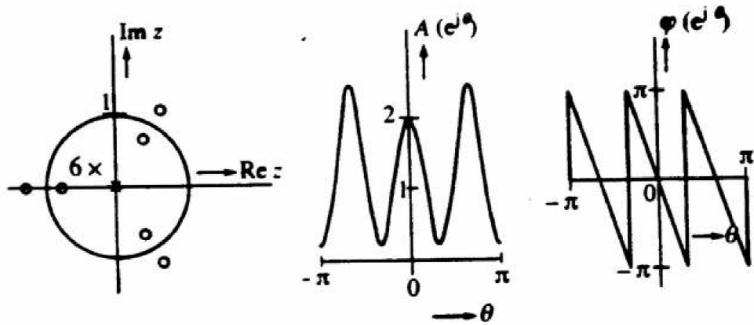
Als een systeem (buiten de oorsprong) alleen maar polen heeft spreekt men van een *all-pole netwerk*.



Figuur 4.4: All pole netwerk

Lineair-fase systeem

Als een systeem (buiten de oorsprong) alleen maar nullen heeft en die bovendien paren heeft die gespiegeld liggen t.o.v. de eenheidscirkel (plus eventueel nulpunten op de eenheidscirkel) dan heeft de fase-karakteristiek een lineair verloop en spreekt men van een lineair-fase systeem.



Figuur 4.5: Lineair-fase systeem

Polen-nullen eliminatie

Polen en nullen diagramma van systemen die in serie staan mogen bij elkaar gevoegd worden. Als een pool en een nul daarbij precies op elkaar vallen vallen ze elkaar.

Aantal polen en nullen

Van een causaal LTO-discreet systeem is het aantal polen minstens gelijk aan het aantal nullen (daarbij worden polen en nullen in $z = 0$ meegeteld). Om dat in te zien schrijven we de algemene uitdrukking voor de transferfunctie $H(z)$ op zo'n manier dat zowel in teller als in noemer alleen positieve machten van z voorkomen.

$$H(z) = \frac{c_0 z^N + c_1 z^{N-1} + \dots + c_N}{z^M + d_1 z^{M-1} + \dots + d_M}$$

Hierbij geldt dat $c_0 \neq 0$ en $c_N \neq 0$ en $d_M \neq 0$ niet beide gelijk aan 0 zijn. De andere c_i 's en d_i 's mogen een willekeurige waarde hebben (dus ook 0).

Men zegt nu dat $H(z)$ een systeem beschrijft van M e orde ; dat wordt bepaald door de hoogste macht van z in de noemer. $H(z)$ heeft M polen en N nullen.

Uitwerken van $H(z)$ d.m.v. een staartdeling levert een resultaat van de vorm op:

$$H(z) = e_0 z^{N-M} + e_1 z^{N-M-1} + \dots$$

De hierbij behorende impulsresponsie is:

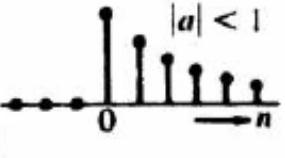
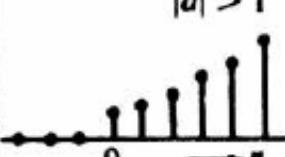
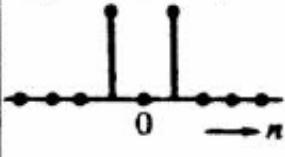
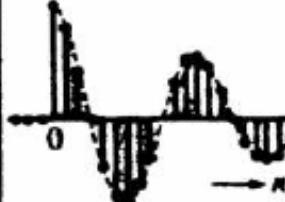
$$h(n) = e_0 \delta(n - (M - N)) + e_1 \delta(n - (M - N + 1)) + \dots$$

Dit signaal is alleen causaal als $M \geq N$. Vermits bij $M - N \leq 0$, het signaal naar links zou worden verschoven in de tijd.

Causaliteitsvoorwaarde

Als een LTO-discreet systeem wordt afgeleid van een bestaand LTO-continu systeem dan is praktisch altijd automatisch voldaan aan de causaliteitsvoorwaarde $M \geq N$.

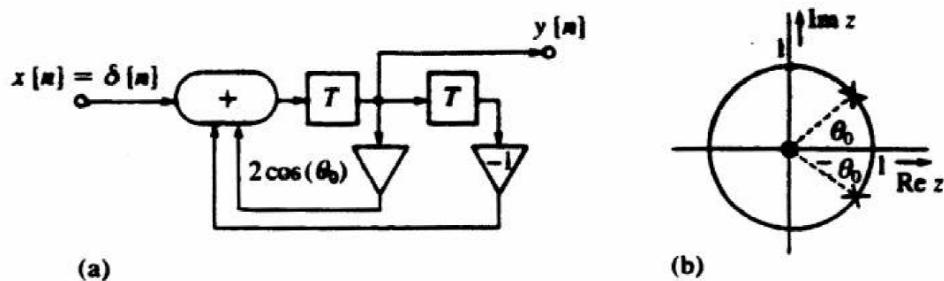
GILLES
GALLEBAU

$h[n]$	$H(z)$	
$h[n] = a^n u[n]$  $ a < 1$ $h[n] = a^n u[n]$  $ a > 1$	$H(z) = \frac{z}{z-a}$ nulpunt: $z = 0$ pool: $z = a$	
$h[n] = \delta[n+1] + \delta[n-1]$ 	$H(z) = \frac{z^2 + 1}{z}$ nulpunten: $z = \pm j$ pool: $z = 0$	
$h[n] = a^n \cos(\pi \xi) u[n]$ 	$H(z) = \frac{z(z-a \cos(\xi))}{z^2 - 2az \cos(\xi) + a^2}$ nulpunten: $z = 0$, $z = a \cos(\xi)$ polen: $z = a e^{\pm j\xi}$	
$h[n] = a^n \sin(\pi \xi) u[n]$ 	$H(z) = \frac{az \sin(\xi)}{z^2 - 2az \cos(\xi) + a^2}$ nulpunt: $z = 0$ polen: $z = a e^{\pm j\xi}$	

Figuur 4.6: Belangrijke polen/nullen diagrammen

4.2.2 Discrete oscillator

Om bruikbaar te zijn moeten systemen in het algemeen stabiel zijn. Een uitzondering hierop vormt de discrete oscillator.



Figuur 4.7: Een discrete oscillator. (a) Blokschema; (b) polen- en nulpuntenbeeld.

We kunnen de lineaire differentievergelijking opstellen a.d.h.v. het blokschema op figuur 4.7:

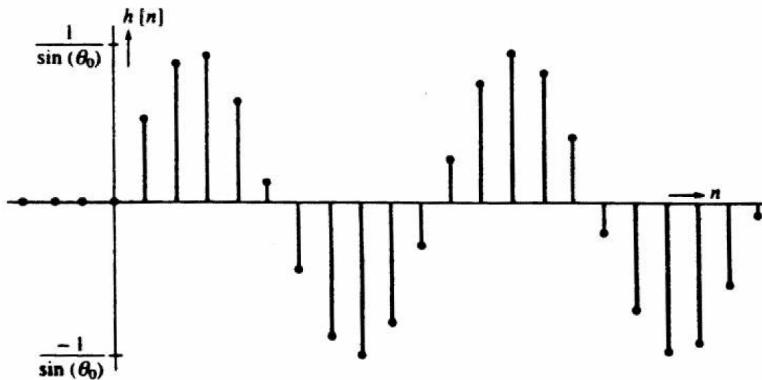
$$y(n) = x(n-1) - y(n-2) + 2 \cos \theta_0 \cdot y(n-1)$$

$$\begin{aligned} H(z) &= \frac{Y(z)}{X(z)} = \frac{z^{-1}}{1 - 2 \cos(\theta_0) z^{-1} + z^{-2}} \\ &= \frac{z}{z^2 - 2z \cos(\theta_0) + 1} \end{aligned}$$

Een tweede graadsvergelijking met als oplossing een complexe paar kan worden geschreven worden onder de vorm $z^2 + 2\Re\{p\} z + |p|^2$. Uit de noemer kan zo de (vector tot de) pool en de modulus gevonden worden. In dit geval ($z^2 - 2z \cos(\theta_0) + 1$) is de modulus 1 en de pool(vector) onder een hoek θ .

Als we deze oplossing vergelijken met de laatste rij van figuur 4.6 op de vorige pagina, dan herkennen we dat we ongeveer dezelfde vorm hebben. Als we nu kiezen voor $a = 1$ en de teller en de noemer vermenigvuldigen met $\sin \theta_0$ dan bekomen we de daarbij horende impulsresponsie:

$$h(n) = \frac{1}{\sin \theta_0} \sin(n\theta_0) u(n) \quad (4.5)$$



Figuur 4.8: Impulsresponsie van de discrete oscillator

We zien in dit voorbeeld een aantal duidelijke verschillen met tijdscontinue oscillatoren; doordat we amplitude kwantiseringsverschijnselen buiten beschouwing gelaten hebben en ervan uitgaan dat de inhoud van de vertragingselementen nul is voor $n < 0$ hebben we te maken met een lineair systeem (daarom kunnen we ook over de transferfunctie spreken).

Dit betekent enerzijds dat we geen uitgangssignaal (dus geen oscillatie) krijgen als we geen ingangssignaal aanbieden. Anderzijds hangt de amplitude van het uitgangssignaal direct af van het ingangssignaal.

4.3 All-pass systemen

In onze redenering vertrekken we steeds van een eerste orde systeem. Als er een eerste orde all-pass systeem (APS) is gevonden hoeven we enkel eerste orde all-pass systemen in cascade te plaatsen om hogere ordes te verkrijgen.

We vertrekken van een stabiele eerste-orde systeem met transferfunctie:

$$H_{ap}(z) = \frac{1 - a^* z}{z - a} \quad (4.6)$$

$H_{ap}(z)$ heeft een nul bij $z = \frac{1}{a^*}$, dit is de gespiegelde versie van de pool, en een pool bij $z = a$. De vorm van een allpass systeem wordt gekozen zodat de weergavekarakteristiek 1 zou zijn. In het andere geval zou dit omgekeerd evenredig zijn met de modulus van de nul:

$$H_{ap}(z) = \frac{z - \frac{1}{a^*}}{z - a} = \frac{1 - a^* z}{z - a} \cdot \frac{1}{-a^*}$$

Wat resulteert in een frequentieresponsie:

$$H_{ap}(e^{j\theta}) = \frac{1}{|a|}$$

4.3.1 Weergavekarakteristiek

De frequentieresponsie van $H_{ap}(z)$:

$$H_{ap}(e^{j\theta}) = \frac{1 - a^* e^{j\theta}}{e^{j\theta} - a} \quad (4.7)$$

Als we a voorstellen als $re^{j\theta_0}$ dan wordt de gespiegelde nul:

$$\frac{1}{a^*} = \frac{1}{r e^{-j\theta_0}} = \frac{1}{r} e^{j\theta_0}$$

$$\begin{aligned} |H_{ap}(e^{j\theta})| &= \left| \frac{1 - re^{-j\theta_0} e^{j\theta}}{e^{j\theta} - re^{j\theta_0}} \right| \\ &= \left| \frac{1}{e^{j\theta}} \right| \cdot \left| \frac{1 - re^{-j(\theta_0 - \theta)}}{1 - re^{j(\theta_0 - \theta)}} \right| \end{aligned}$$

De teller en de noemer zijn elkaars complex toegevoegde; d.w.z. dat

$$|H_{ap}(e^{j\theta})| = 1$$

M.a.w. de modulus van de frequentieresponsie van $H_{ap}(z)$ is constant en onafhankelijk van θ . Daarom noemen we zo'n systeem een *allpass filter*.

Wanneer een willekeurig systeem in cascade geplaatst wordt met een allpass systeem zal de weergavekarakteristiek niet beïnvloed worden. Hierdoor zijn allpass filters zeer interessant voor fase compensatie of compensatie van de groepsvertraging; als de fase of de groepsvertraging van een systeem $H(z)$ niet bevredigend is kan men het systeem in cascade plaatsen met een allpass systeem dat voor bijkomende fase of groepsvertraging zorgt en er dus voor zorgt dat voor het geheel, de fase of de groepsvertraging wel stroken met het gewenste.

Het is duidelijk dat de cascade van allpass systemen opnieuw een allpass systeem is. Meer specifiek: voor alle allpass systemen van een willekeurige orde is het zo dat de polen en nullen geplaatst zijn op complex toegevoegde reciproke¹ plaatsen. De meest algemene vorm voor een allpass systeem is:

$$H_{ap}(z) = \prod_{i=1}^{N_r} \left(\frac{1 - a_i z}{z - a_i} \right) \left(\prod_{i=1}^{N_c} \frac{(1 - b_i^* z)(1 - b_i z)}{(z - b_i)(z - b_i^*)} \right) \quad (4.8)$$

met N_r het aantal reële polen en N_c het aantal complexe polen/2, vermits je per i twee polen hebt.

Om de stabiliteit te verzekeren voor een causaal systeem is het noodzakelijk dat alle polen binnen de eenheidscirkel liggen, en omdat de nullen op de complex toegevoegde reciproke plaatsen liggen zullen alle nullen van een causaal stabiel allpass systeem buiten de eenheidscirkel liggen.

¹de reciproque (vaker: de reciproke) van een getal of grootheid is 1 gedeeld door dat getal of die grootheid

4.3.2 Fasekarakteristiek

$$H_{ap}(e^{j\theta}) = \frac{1}{e^{j\theta}} \frac{1 - re^{-j(\theta_0 - \theta)}}{1 - re^{j(\theta_0 - \theta)}} \quad (4.9)$$

Met $e^{jx} = \cos x + j \sin x$:

$$H_{ap}(e^{j\theta}) = \frac{1}{e^{j\theta}} \frac{1 - r \cos(\theta_0 - \theta) + j r \sin(\theta_0 - \theta)}{1 - r \cos(\theta_0 - \theta) - j r \sin(\theta_0 - \theta)}$$

De fase van het allpass systeem wordt dan:

$$\begin{aligned} \varphi(H_{ap}(e^{j\theta})) &= \varphi\left(\frac{1}{e^{j\theta}}\right) + \varphi(1 - r \cos(\theta_0 - \theta) + j r \sin(\theta_0 - \theta)) - \varphi(1 - r \cos(\theta_0 - \theta) - j r \sin(\theta_0 - \theta)) \\ &= -\theta + \cot\left(\frac{r \sin(\theta_0 - \theta)}{1 - r \cos(\theta_0 - \theta)}\right) - \cot\left(\frac{-r \sin(\theta_0 - \theta)}{1 - r \cos(\theta_0 - \theta)}\right) \end{aligned}$$

De fasekarakteristiek voor een eerste orde allpass systeem wordt gegeven door:

$$\varphi(H_{ap}(e^{j\theta})) = -\theta + 2 \cot\left(\frac{r \sin(\theta_0 - \theta)}{1 - r \cos(\theta_0 - \theta)}\right) \quad (4.10)$$

De groepsvertraging wordt gedefinieerd als

$$grv(H(e^{j\theta})) = -\frac{d}{d\theta} \varphi(H_{ap}(e^{j\theta})) \quad (4.11)$$

Men kan aantonen dat

$$grv(H(e^{j\theta})) = \frac{1 - r^2}{|1 - re^{j\theta_0} e^{-j\theta}|^2} \quad (4.12)$$

Voor een stabiel allpass systeem geldt dat $|r| < 1$; hieruit volgt dat de groepsvertraging voor een allpass systeem altijd niet-negatief is.

Beschouwen we verder de fase van een allpass systeem:

$$\varphi(H_{ap}(e^{j\theta})) = - \int_0^\theta grv(H_{ap}(e^{j\theta})) d\phi + \varphi(H_{ap}(e^{j0}))$$

Doordat $H_{ap}(e^{j\theta}) = 1$ zal $\varphi(H_{ap}(e^{j0})) = 0$ en bekomen we:

$$\varphi(H_{ap}(e^{j\theta})) = - \int_0^\theta grv(H_{ap}(e^{j\theta})) d\phi \quad (4.13)$$

Door de niet-negativiteit van de groepsvertraging is de fase van een allpass systeem altijd niet-positief.

4.3.3 Besluit

De weergavekarakteristiek van een allpass systeem is overal gelijk aan 1, de groepsvertraging is niet-negatief en de fase is niet-positief.

Hoewel deze conclusies afgeleid zijn uit een eerste orde allpass systeem zijn ze geldig voor een allpass systeem van willekeurige orde; immers, de weergavekarakteristiek wordt bekomen door vermenigvuldiging en de fase en de groepsvertraging zijn sommeerbaar.

4.4 Minimum fase systemen

Een causaal minimum fase systeem (MFS) wordt gedefinieerd als een causaal systeem met alle polen en nullen binnen de eenheidscirkel. Dit houdt in dat het inverse systeem ook stabiel is (en causaal als $M = N$).

We tonen eerst aan dat elke rationele transferfunctie kan geschreven worden als een cascade van een MFS en een APS.

$$H(z) = H_{min}(z) \cdot H_{ap}(z) \quad (4.14)$$

We stellen dat $H(z)$ een niet-MFS is met enkel een nul buiten de eenheidscirkel bij $z = \frac{1}{a^*}$ met $|a| < 1$. Dit resulteert in de volgende vergelijking:

$$\begin{aligned} \underbrace{H(z)}_{NMF} &= \underbrace{H_1(z)}_{MF} \cdot (1 - a^* z) \\ &= H_1(z) \cdot (1 - a^* z) \cdot \frac{(z - a)}{(z - a)} \\ \underbrace{H(z)}_{NMF} &= \underbrace{H_1(z) \cdot (z - a)}_{MF} \cdot \underbrace{\frac{(1 - a^* z)}{z - a}}_{AP} \end{aligned} \quad (4.15)$$

Doordat $H_1(z)$ een MFS is en $(z - a)$ een nul heeft binnen de eenheidscirkel, is hun product een MFS. We definiëren dus $H_{min}(z)$ als:

$$H_{min}(z) = H_1(z) \cdot (z - a)$$

De rest is duidelijk een allpass systeem:

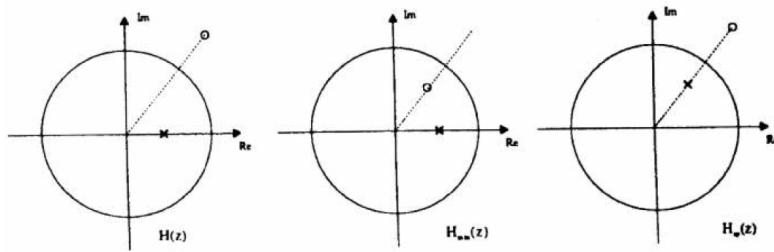
$$H_{ap}(z) = \frac{(1 - a^* z)}{z - a}$$

Hiermee is de gestelde bewering aangetoond voor het geval met één nul buiten de eenheidscirkel.

De ondernomen stappen worden op figuur 4.9 op de pagina hierna voorgesteld.

De weergavekarakteristiek van het oorspronkelijke NMFS is gelijk aan het MFS, vermits de weergavekarakteristiek van een APS 1 is.

$$\begin{aligned} |H(e^{j\theta})| &= |H_{min}(e^{j\theta})| \cdot |H_{ap}(e^{j\theta})| \\ &= |H_{min}(e^{j\theta})| \end{aligned}$$



Figuur 4.9: Visuele voorstelling van het feit dat een transferfunctie kan geschreven worden als een product van een minimumfase en allpass systeem.

Voorbeeld: Minimum Fase Systeem

$$H(z) = \frac{(z - \frac{1}{8})(z - 2)}{(z - \frac{1}{4})(z - \frac{1}{3})}$$

$H(z)$ os een NMFS want er is een nul bij $z = 2$.

Om een transferfunctie van een vorm maken zoals bij formule 4.15 op de vorige pagina, we moeten de nul buiten de eenheidscirkel omvormen tot de vorm $(1 - a^*z)$.

$$(z - 2) \rightarrow (1 - a^*z) = -2(1 - \frac{1}{2}z)$$

Een APS heeft een weergavekarakteristiek van één, dus de factor -2 moet bij het MFS gevoegd worden.

$$H(z) = \begin{cases} H_{min}(z) &= -2 \cdot \frac{(z - \frac{1}{8})(z - \frac{1}{2})}{(z - \frac{1}{4})(z - \frac{1}{3})} \\ H_{ap}(z) &= \frac{(1 - \frac{1}{2}z)}{(z - \frac{1}{2})} \end{cases}$$

4.4.1 Tijdsdomeinbeschrijving van een MFS

$$H(z) = H_{min}(z) \cdot H_{ap}(z) \quad (4.16)$$

Met de eigenschap voor de limiet-waarde ($\lim_{z \rightarrow \infty} H(z) = h(0)$) vinden we:

$$\lim_{z \rightarrow \infty} H(z) = \lim_{z \rightarrow \infty} H_{min}(z) \cdot \lim_{z \rightarrow \infty} H_{ap}(z)$$

$$h(0) = h_{min}(0) \cdot \lim_{z \rightarrow \infty} H_{ap}(z)$$

Met $\lim_{z \rightarrow \infty} H_{ap}(z) = \lim_{z \rightarrow \infty} \frac{(1-a^*z)}{z-a} = -a^*$ vinden we:

$$h(0) = h_{min}(0) \cdot -a^*$$

$$|h(0)| = |h_{min}(0)| \cdot |-a^*| = |a||h_{min}(0)|$$

Doordat $|a| < 1$ (want pool binnen eenheidscirkel) kunnen we besluiten dat:

$$|h(0)| > |h_{min}(0)|$$

Als twee systemen dezelfde amplitudekarakteristiek vertonen en het is gekend dat één van hen een MFS is, dan is het systeem met de grotere $|h(0)|$ het MFS.

4.5 Lineaire-fase FIR filters

Als eens signaal wordt aangelegd aan een filter dan kunnen zowel de amplitude als de fase veranderen. De verandering van het fasespectrum is verbonden met de groepsvertraging:

$$grv = \tau_g = -\frac{d\phi}{d\theta} \quad (4.17)$$

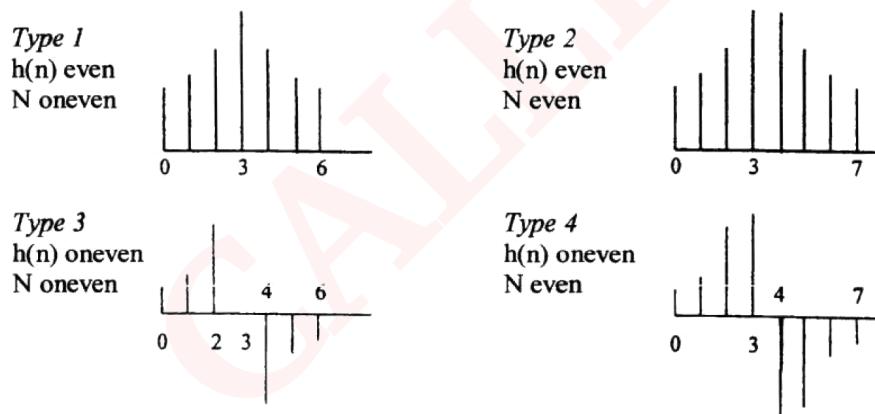
Een lineaire fase karakteristiek geeft aanleiding tot een constante groepsvertraging met een fase:

$$\phi(\theta) = \beta - \alpha \cdot \theta \quad (4.18)$$

Bepaalde symmetrieën in de impulsresponsie $h(n)$ garanderen een lineaire fase:

- $h(n)$ kan een even of oneven symmetrie hebben
- de lengte N van het filter kan even of oneven zijn

Dit geeft aanleiding tot vier types van FIR filter met een lineaire fase karakteristiek.



Figuur 4.10: De vier types van FIR filter met een een lineaire fase karakteristiek

4.5.1 Types lineair fasige FIR filters

Type 1

Een type 1 FIR filter heeft een even symmetrie voor de impulsresponsie, $h(N-n-1) = h(n)$ met een oneven lengte N , i.e. een oneven aantal impulsresponsie-waarden.

We onderzoeken de frequentieresponsie a.d.h.v een voorbeeld met een lengte van 5. (Dit voorbeeld werd gegeven in de les, voor een ander voorbeeld moet u zich wenden tot de cursus)

Voorbeeld: Type 1 FIR filter

$N = 5$ waaruit volgt dat $h(4) = h(0)$, $h(3) = h(1)$.

De frequentieresponsie wordt dan:

$$H(z) = h(0) + h(1)z^{-1} + h(2)z^{-2} + h(3)z^{-3} + h(4)z^{-4}$$

We schakelen over naar het frequentiedomein om zo de fase te kunnen onderzoeken.

$$\begin{aligned} H(e^{j\theta}) &= h(0) + h(1)e^{-j\theta} + h(2)e^{-2j\theta} + h(3)e^{-3j\theta} + h(4)e^{-4j\theta} \\ &= h(0) \left(1 + e^{-4j\theta} \right) + h(1) \left(e^{-j\theta} + e^{-3j\theta} \right) + h(2)e^{-2j\theta} \\ &= e^{-2j\theta} \left[h(0) \left(e^{2j\theta} + e^{-2j\theta} \right) + h(1) \left(e^{j\theta} + e^{-j\theta} \right) + h(2) \right] \end{aligned}$$

$$H(e^{j\theta}) = e^{-2j\theta} \cdot [h(0)2\cos(2\theta) + h(1)2\cos(\theta) + h(2)]$$

Als fase wordt dan:

$$\varphi(H(e^{j\theta})) = -2\theta + \left\{ \begin{array}{l} 0 \\ \pi \end{array} \right\}$$

$\left\{ \begin{array}{l} 0 \\ \pi \end{array} \right\}$ komt door het feit dat de laatste term van de transferfunctie enkel reëel is, dus moet de fase 0 of π (als negatief) zijn.

Algemeen wordt de fase bij een type 1 FIR filter beschreven door:

$$\varphi(H(e^{j\theta})) = -\frac{N-1}{2}\theta + \left\{ \begin{array}{l} 0 \\ \pi \end{array} \right\} \quad (4.19)$$

De belangrijkste bijdrage tot de fase komt van de complexe exponentiële term en de fase is duidelijk lineair (zie formule 4.18 op de pagina hiervoor):

$$\varphi(\theta) = \left\{ \begin{array}{l} 0 \\ \pi \end{array} \right\} - M\theta$$

Type 2

Een type 2 FIR filter heeft ook een even symmetrie voor de impulsresponsie, maar met een even lengte N .

Voorbeeld: Type 2 FIR filter

$N = 4$ waaruit volgt dat $h(3) = h(0)$, $h(2) = h(1)$.

De frequentieresponsie wordt dan:

$$\begin{aligned} H(e^{j\theta}) &= h(0) + h(1)e^{-j\theta} + h(2)e^{-2j\theta} + h(3)e^{-3j\theta} \\ &= h(0) \left(1 + e^{-3j\theta}\right) + h(1) \left(e^{-j\theta} + e^{-2j\theta}\right) \\ &= e^{-j\frac{3\theta}{2}} \left[h(0) \left(e^{j\frac{3\theta}{2}} + e^{-j\frac{3\theta}{2}}\right) + h(1) \left(e^{j\frac{\theta}{2}} + e^{-j\frac{\theta}{2}}\right) \right] \\ H(e^{j\theta}) &= e^{-2j\theta} \cdot \left[h(0)2\cos\left(\frac{3\theta}{2}\right) + h(1)2\cos\left(\frac{\theta}{2}\right) \right] \end{aligned}$$

Als fase wordt dan:

$$\varphi(H(e^{j\theta})) = -\frac{3\theta}{2} + \left\{ \begin{array}{l} 0 \\ \pi \end{array} \right\}$$

Algemeen wordt de fase bij een type 2 FIR filter beschreven door:

$$\varphi(H(e^{j\theta})) = -\frac{N-1}{2}\theta + \left\{ \begin{array}{l} 0 \\ \pi \end{array} \right\} \quad (4.20)$$

We hebben opnieuw een lineaire fase.

Bovendien geldt voor dit type dat $H(e^{j\theta}) = 0$ voor $\theta = \pi$ onafhankelijk van $h(n)$. Daardoor is een type 2 lineaire fase filter niet geschikt voor het ontwerp van een hoogdoorlaatfilter of een bandstop filter.

Type 3

Een type 3 FIR filter heeft een oneven symmetrie voor de impulsresponsie, met een oneven lengte N .

Voorbeeld: Type 3 FIR filter

$N = 5$ waaruit volgt dat $h(4) = -h(0)$, $h(3) = -h(1)$ en gegeven $h(2) = 0$.

De frequentieresponsie wordt dan:

$$\begin{aligned} H(e^{j\theta}) &= h(0) \left(1 - e^{-4j\theta}\right) + h(1) \left(e^{-j\theta} - e^{-3j\theta}\right) \\ &= e^{-j2\theta} \cdot \left[h(0) \left(e^{j2\theta} - e^{-j2\theta}\right) + h(1) \left(e^{j\theta} - e^{-j\theta}\right)\right] \\ H(e^{j\theta}) &= e^{-2j\theta} \cdot 2j \cdot [h(0) \sin(2\theta) + h(1) \sin(\theta)] \end{aligned}$$

Als fase wordt dan:

$$\varphi(H(e^{j\theta})) = -2\theta + \begin{cases} 0 \\ \pi \end{cases}$$

Algemeen wordt de fase bij een type 3 FIR filter beschreven door:

$$\varphi(H(e^{j\theta})) = \frac{\pi}{2} - \frac{N-1}{2}\theta + \begin{cases} 0 \\ \pi \end{cases} \quad (4.21)$$

De aanwezigheid van een sinus zorgt ervoor dat $H(e^{j\theta}) = 0$ voor $\theta = 0$ en π . Daardoor kan dit type alleen gebruikt worden voor het ontwerp van een banddoorlaatfilter.

Type 4

Werd niet besproken in de les, zal zien of ik dit ook nog "neerpen".

4.5.2 Ligging van de polen en de nullen bij lineaire-fase filters

Door het feit dat we te maken hebben met een FIR filter liggen alle polen van de transferfunctie in de oorsprong (in totaal N).

De symmetrie in $h(n)$ geeft aanleiding tot een specifiek nullenpatroon. We nemen als voorbeeld een type 2 systeem met $N = 6$.

De transferfunctie is:

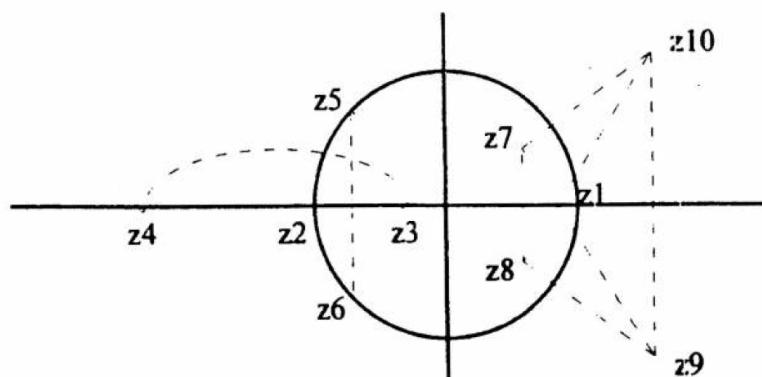
$$\begin{aligned} H(z) &= h(0) + h(1)z^{-1} + h(2)z^{-2} + h(3)z^{-3} + h(4)z^{-4} + h(5)z^{-5} \\ &= h(0) + h(1)z^{-1} + h(2)z^{-2} + h(2)z^{-3} + h(1)z^{-4} + h(0)z^{-5} \\ H\left(\frac{1}{z}\right) &= h(0) + h(1)z^1 + h(2)z^2 + h(2)z^3 + h(1)z^4 + h(0)z^5 \\ &= z^5 [h(0)z^{-5} + h(1)z^{-4} + h(2)z^{-3} + h(2)z^{-2} + h(1)z^{-1} + h(0)] \\ H\left(\frac{1}{z}\right) &= z^5 H(z) \end{aligned}$$

Dit is een zeer belangrijk resultaat: als z_0 een nul is van $H(z)$, dan is $\frac{1}{z_0}$ ook een nul.

Hetzelfde geldt voor de andere types lineaire fase filters.

Twee punten in het complexe vlak met dezelfde fase maar met een inversie modulus noemt men een gereflecteerd koppel.

Grafische voorstelling van alle mogelijke liggingen van de nullen



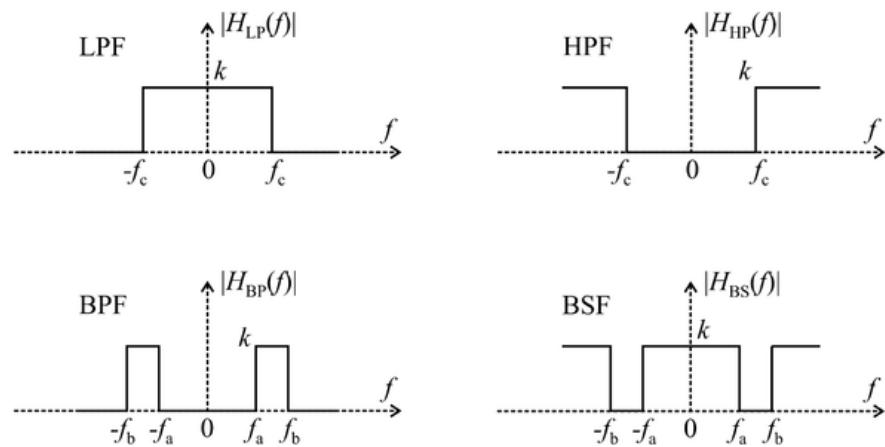
Figuur 4.11: Grafische voorstelling van alle mogelijke liggingen van de nullen in een lineair fase systeem

- enkelvoudige reële nul in $z_1 = 1$
- enkelvoudige reële nul in $z_2 = -1$
- reëel gereflecteerd koppel $z_3 = A$ en $z_4 = \frac{1}{A}$
- complex toegevoegd koppel op de eenheidscirkel (z_5 en z_6)
- complex kwartet z_7, z_8, z_9 en z_{10}

4.5.3 Oefeningen

Filter realistie

Welke filters kunnen worden gerealiseerd door welke types?



Figuur 4.12: Vier filters in het frequentiedomein en LDF, BDF, HDF en BSF. Als we willen werken met θ moeten we gebruik, komt $2f_c$ overeen met π .

Type 1 Alles, vermits men voor de hoeken 0 en π alle waarden kunnen uitkomen, eveneens voor de waarden tussen 0 en π .

Type 2 Kunnen de HDF en BSF niet worden gerealiseerd vermits de weergavekarakteristiek bij de hoek π nul is.

Type 3 Kan enkel de BDF realiseren vermits men bij de hoeken 0 en π nul uitkomt bij de weergavekarakteristiek.

FIR-filter ontwerp

Opgave Er werd een FIR-filter ontworpen met een rimpel in de doorlaatband van 3 dB en een verzwakking in de stopband van 50 dB.

De coefficienten b_0, b_1, \dots, b_{10} worden gegeven door:
 $-0.0061; 0.0620; -0.0305; -0.1024; 0.2865; 0.6360; 0.2865; -0.1024; -0.0305; 0.0620; -0.0061$

De positie van de nullen wordt gegeven door:

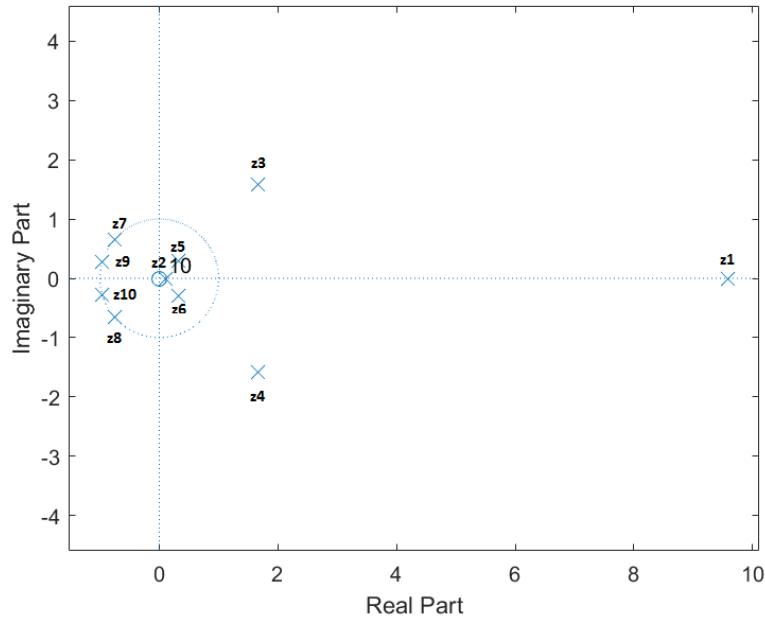
- 9.5834 en 0.1043
- $1.6711 \pm 1.5738j$
- $-0.9626 \pm 0.2709j$
- $-0.7568 \pm 0.6536j$
- $0.3171 \pm 0.2987j$

Gevraagd

- a
 1. Wat is de orde van het filter en wat is de lengte?
 2. Van welk type is dit lineaire fase filter?
 3. Teken het polen en nullen diagram en verklaar de ligging van de polen.
 4. Leidt uit dit diagram af welk soort filter het is
 5. Leidt uit datzelfde diagram de frequentiekarakteristiek af voor een $f_s = 8000 \text{ Hz}$
- b
 1. Is dit een minimumfase systeem?
 2. Verander het systeem in een MFS zonder dat de weergavekarakteristiek verandert
- c Geef een schematische voorstelling (numeriek) voor de volgende structuren:
 - transversaal (normaal en gebruik makend van symmetrie)
 - cascade structuur

Oplossing

- a
 1. **Wat is de orde van het filter en wat is de lengte?**
Er zijn 11 coëfficiënten en dus 11 impulsresponsie-waarden, i.e. de lengte is 11 en de orde is $N - 1$ en dus 10.
 2. **Van welk type is dit lineaire fase filter?**
Vermits we zitten met een oneven lengte en een even FIR filter, gaat het hier om een type 1 FIR filter.
 3. **Teken het polen en nullen diagram en verklaar de ligging van de polen.**

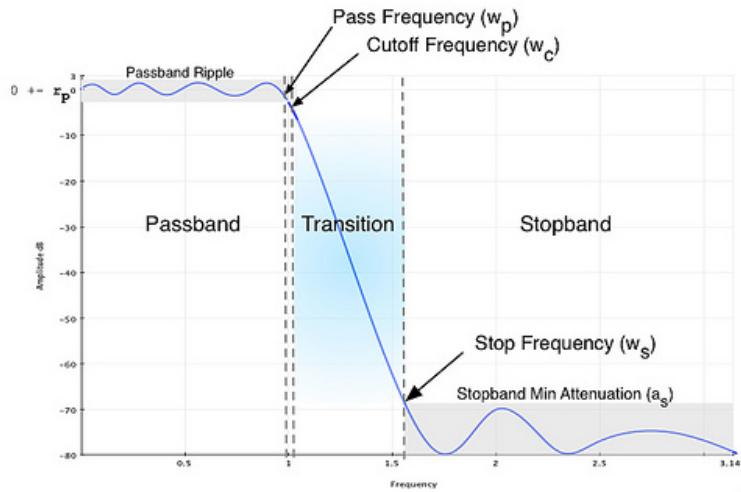


Bij een FIR filter liggen de polen in de oorsprong. De weergavekarakteristiek kan worden opgesteld door het product te nemen van de afstanden van een punt op de eenheidscirkel tot alle nullen.

$$|H(e^{j\theta})| = b_0 \cdot \frac{\prod_{\text{nullen}} |z_i|}{\underbrace{\prod_{\text{polen}} |z_i|}_{=1}}$$

4. Leidt uit dit diagram af welk soort filter het is

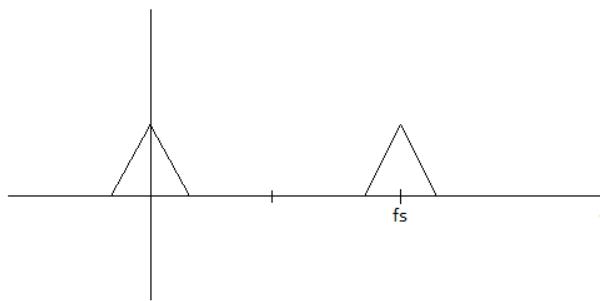
We bekijken de nullen waarbij de weergavekarakteristiek nul of één is en zetten de hoeken (θ) uit. Nullen op de eenheidscirkel zullen dus resulteren in nullen op de weergavekarakteristiek (z_7 en z_9). We bekomen dan een benaderende figuur:



Figuur 4.13: Laagdoorlaat filter

Door de symmetrie hoeven we de eenheidscirkel enkel te doorlopen van 0 tot π . De nullen op de eenheidscirkel zullen dus de eerste nuldoorgangen bepalen. We hebben dus te maken met een laagdoorlaatfilter.

5. **Leidt uit datzelfde diagram de frequentiekarakteristiek af voor een $f_s = 8000 \text{ Hz}$**
Het periodisch interval $0 \rightarrow \pi$ beslaat de helft van de sampling frequentie, dit volgt uit $\theta = \omega \cdot T \Leftrightarrow f = \frac{\theta}{2\pi} f_s$.
We verkrijgen de volgende frequentiekarakteristiek:

Figuur 4.14: Frequentiekarakteristiek voor $f_s = 8000 \text{ Hz}$

- b 1. **Is dit een minimumfase systeem?**

Het is geen MFS vermits er nullen buiten de eenheidscirkel bevinden.

2. **Verander het systeem in een MFS zonder dat de weergavekarakteristiek verandert**

We kunnen dit NMFS omzetten naar een MFS systeem door het systeem te splitsen als een MFS en een AP systeem. HET AP systeem bestaat uit een pool binnen de

eenheidscirkel die een gereflecteerde nul bevat. Het MFS zal enkel nullen en polen hebben binnen de eenheidscirkel.

We zullen dus een pool leggen op ge gespiegelde van z_3, z_4 en z_1 (dit zijn respectievelijk z_5, z_6 en z_2), deze vormen samen het APS. De overige nullen en polen zullen dan gelegen zijn binnen de eenheidscirkel en vormen dan het MFS.

$$H(z)_{ap} = \frac{(1-z_5^*)(1-z_2^*)(1-z_6^*)}{(z-z_5)(z-z_2)(z-z_6)} = \frac{(1-z_3)(1-z_1)(1-z_4)}{(z-z_5)(z-z_2)(z-z_6)}$$

De transferfunctie van de FIR filter ziet er als volgt uit:

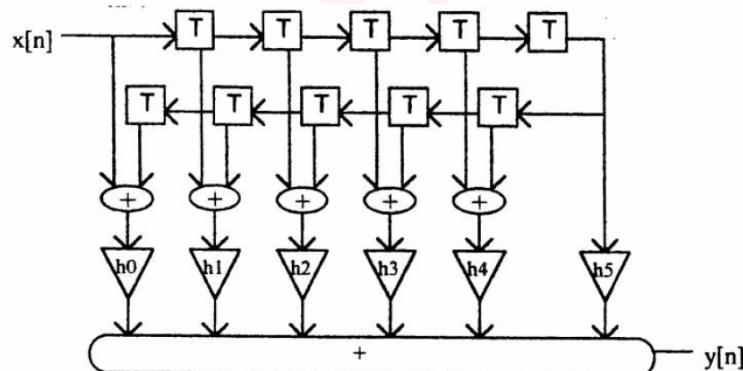
$$H(z) = b_0 \frac{(z-z_1)(z-z_2) \cdots (z-z_{10})}{z^{10}}$$

Het MFS wordt dan:

$$H(z)_{min} = b_0 \frac{(-z_3)(-z_1)(-z_4) \cdot (z-z_7)(z-z_8)(z-z_9)(z-z_{10}) \cdot (z-z_5)^2(z-z_2)^2(z-z_6)^2}{z^{10}}$$

c Geef een schematische voorstelling (numeriek) voor de volgende structuren:

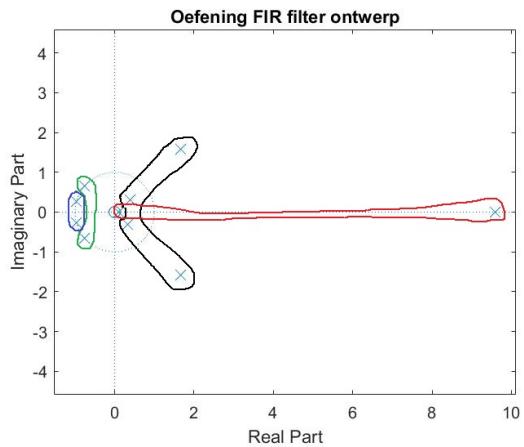
- transversaal (normaal en gebruik makend van symmetrie)



- cascade structuur

Door de symmetrie van het LDF kunnen we via cascade die fouten bij de coëfficiënten verminderen, door gespiegelde en complex toegevoegde samen te voegen.

De cascade bestaat zo uit 4 transversale structuren waarvan 3 tweede orde en 1 vierde orde.



Extra oefeningen op All-pass en MF systeem

Zoek $H(z) = H_{min}H_{ap}$

- $$H(z) = \frac{(1-z^{-2})(1+0.4z^{-1})}{1-0.85z^{-1}}$$
- $$H(z) = \frac{(3+z^{-1})(2-3z^{-1})}{1-\frac{z^{-1}}{2}}$$

Voor welke waarde van α is $Y(n) = \alpha^n X(n)$ een MFS

$$X(z) = \frac{(1-\frac{3}{2}z^{-1})(1+\frac{1}{3}z^{-1})(1+\frac{5}{3}z^{-1})}{(1-z^{-1})^2(1-\frac{z^{-1}}{4})}$$

Hoofdstuk 5

Eindige woordlengte

5.1 Algemeen

Binnen de klasse van digitale signalen en systemen wordt iedere grootheid als een combinatie van een aantal bits voorgesteld. Met B bits kunnen maximaal 2^B verschillende waarden onderscheiden worden. Als B vrij kan gekozen worden, dan kan men de digitale voorstelling zo nauwkeurig maken als gewenst is. Echter, om kosten te besparen is men gescreven in de kleinste waarde van B zonder niet-aanvaardbare fouten te introduceren. In dat geval zijn er een aantal effecten die veroorzaakt worden door de eindige woordlengte waarmee gewerkt wordt.

Situaties waarin de eindige woordlengte een belangrijke rol speelt:

- de omzetting van signalen met een continue amplitude in signalen met een discrete amplitude ; hierbij wordt ruis geïntroduceerd die meestal aangeduid wordt met de term A/D-omzettingsruis.
- het omzetten van de filtercoëfficiënten in coëfficiënten van eindige woordlengte ; dit gaat gepaard met een verandering van de frequentieresponsie (verzwakking karakteristiek en andere ligging polen).
- het uitvoeren van bewerkingen op een zodanige manier dat de woordlengte niet op een ongewenste manier toeneemt ; dit veroorzaakt ruis en soms zelfs oscillaties.

In al deze situaties hebben we te maken met kwantisering en/of overloop.

Bi een gegeven woordlengte van B bits kan nog gekozen worden welk getal overeenstemt met elk van de 2^B verschillende woorden: de getalrepresentatie. Daarvoor bestaan verschillende mogelijkheden die elk hun eigen voor- en nadelen hebben. De 3 meest gebruikte binaire getalrepresentaties zijn:

1. teken en grootte (sign and magnitude)
2. één-complement (one's complement)

3. twee-complement (two's complement)

Het verschil tussen deze 3 gevallen wordt toegelicht a.h.v. tabel 5.1.

decimale waarde	teken en grootte	één-complement	twee-complement
+ 3	011	011	011
+ 2	010	010	010
+ 1	001	001	001
+ 0	000	000	000
- 0	100	111	—
- 1	101	110	111
- 2	110	101	110
- 3	111	100	101
- 4	—	—	100

Figuur 5.1: Verschillende binaire getalrepresentaties

In bovenstaande tabel is aan de LSB de waarde 2^0 toegekend. In dat geval kunnen alleen gehele getallen voorgesteld worden. Er kan aan de LSB ook een waarde toegekend worden die overeenkomt met een gehele negatieve macht van 2, daardoor kunnen decimale breuken weergegeven worden. Voorbeeld zie tabel 5.2 ; vaste-komma-representatie (fixed point)

decimale waarde	teken en grootte	één-complement	twee-complement
+ 3,625	00011,101	00011,101	00011,101
- 3,625	10011,101	11100,010	11100,011

Figuur 5.2: Voorbeelden van getalrepresentaties

Een andere manier om een decimale waarde aan een binair woord toe te kennen wordt gebruikt in de drijvende-komma-representatie (floating point). Hierbij wordt een decimaal getal A eerst geschreven als:

$$A = \pm M \times 2^E \quad (5.1)$$

met E een geheel positief of negatief getal en $0,5 < |M| < 1$. M wordt de mantisse genoemd en E de exponent.

Een nadeel van de drijvende-komma-vorstellung is dat de bewerkingen ingewikkelder worden. Voor de vermenigvuldiging van 2 getallen moeten de beide mantissen met elkaar vermenigvuldigd worden en de exponenten opgeteld. Daarna moet ervoor gezorgd worden dat het resultaat voldoet aan de voorstellingswijze in floating-point.

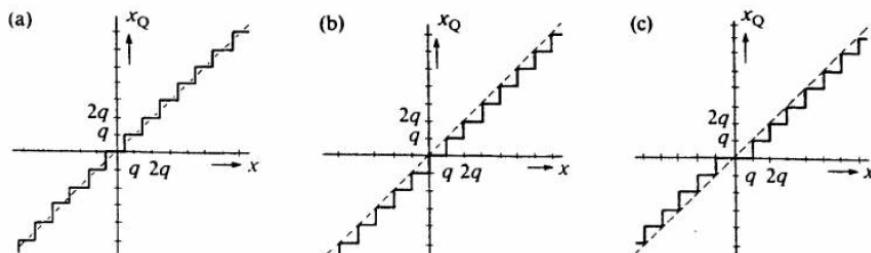
Floating-point voorstelling wordt praktisch alleen gebruikt als digitale signaalbewerking uitgevoerd wordt op een computer. In gevallen waar de digitale filters in de vorm van speciaal daarvoor ontworpen geïntegreerde schakelingen uitgevoerd worden, wordt meestal gebruik gemaakt van fixed-point voorstelling.

5.2 Kwantisering en overloop

Kwantisering is het proces waarbij een grootheid x wordt omgezet in een grootheid x_q die ongeveer gelijk is aan x , maar die minder verschillende waarden kan aannemen dan x . Kwantisering treedt op als een willekeurig reëel getal vervangen wordt door het dichtstbij gelegen gehele getal, of, bij het reduceren van de woordlengte van een binaire grootheid x door het aantal bits achter de komma te verkleinen. Het verband tussen x en x_q wordt de kwantiseringskarakteristiek genoemd.

De meest gebruikte vormen van kwantisering zijn:

- afronden (rounding)
- afbreken (value truncation)
- modulus afbreken (magnitude truncation)



Figuur 5.3: Drie kwantiseringskarakteristieken: (a) afronden; (b) afbreken; (c) modulus afbreken

De opeenvolgende waarden van x_q hebben een vaste afstand q (de kwantiseringsstap) van elkaar.

Het omzetten van x in x_q gaat praktisch altijd gepaard met het introduceren van onnauwkeurigheid. Het omzetten van een analoog signaal naar een digitaal signaal zal dus twee onnauwkeurigheden met zich meebrengen: *sampling* (\sim bemonsteringsfrequentie) en het kwantiseren van het signaal (\sim aantal bits). De grootte van deze onnauwkeurigheid van kwantisering, $x_q - x$, is echter begrensd en kan uitgedrukt worden in q :

1. Bij afronden geldt: $-\frac{q}{2} \leq x_q - x < \frac{q}{2}$
2. Bij afbreken geldt: $-q \leq x_q - x < 0$

3. Bij modulus afbreken geldt:

$$\begin{aligned} \text{als } x > 0 \quad -q \leq x_q - x < 0 \\ \text{als } x < 0 \quad 0 \leq x_q - x < q \end{aligned}$$

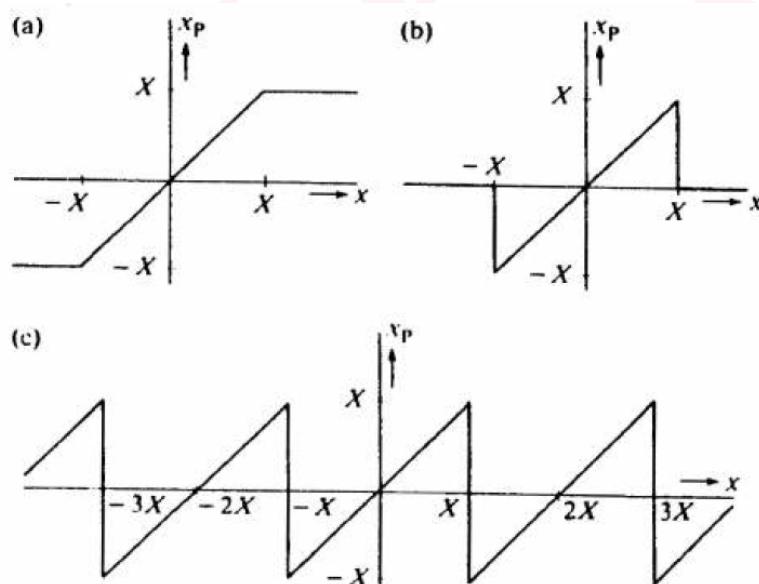
Kwantisering vertegenwoordigt in wezen een niet-lineaire bewerking omdat in het algemeen geldt: $(x+y)_q \neq x_q + y_q$

Een andere vorm van niet-lineairiteit duidt men aan met begrip overloop. Overloop is het verschijnsel dat een grootheid x een waarde zou willen aannemen buiten de grenzen ($-X$ en $+X$) die in acht dienen genomen te worden. Formeel kan dit beschouwd worden als een omzetting in x_p waarbij:

$$x_p = \begin{cases} = x & \text{als } |x| \leq X \\ \leq X & \text{als } |x| > X \end{cases}$$

Het verband tussen x en x_p wordt de overloopkarakteristiek genoemd. In figuur 5.4 worden drie voorbeelden gegeven:

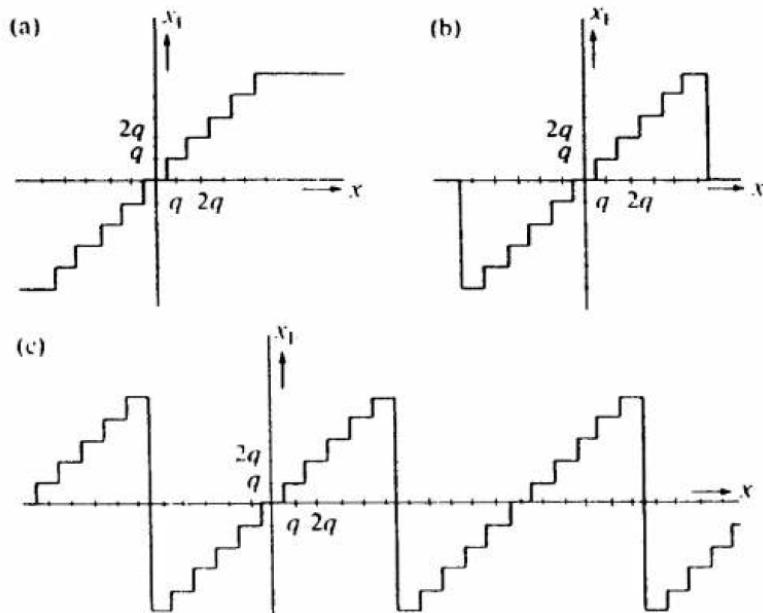
1. verzadeging (saturation)
2. nulstelling (zeroing)
3. zaagtandvormige overloop



Figuur 5.4: Drie overloopkarakteristieken (a) verzadeging; (b) nulstelling; (c) zaagtandvormige overloop

Er wordt geen fout geïntroduceerd zolang $|x| \leq X$. Voor andere waarden van x treedt er een fout $x_p - x$ op, die in principe onbegrensd is.

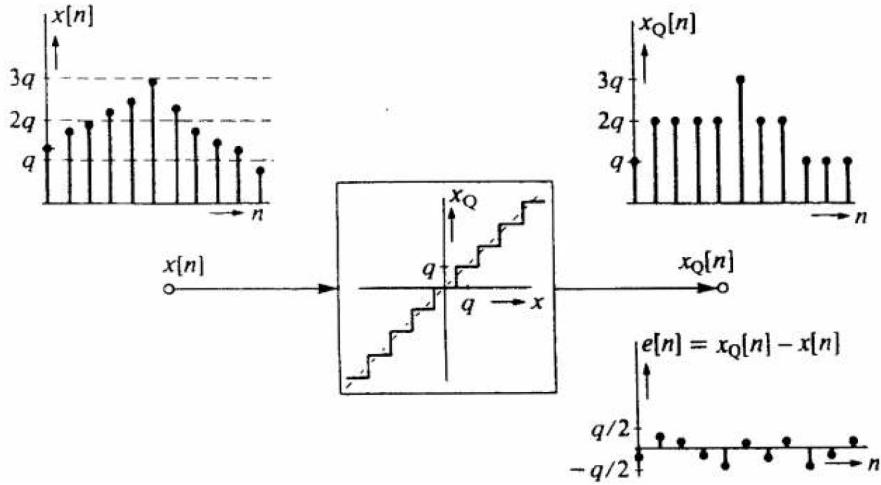
In bepaalde gevallen treedt er een combinatie van kwantisering en overloop op. Dit is bijvoorbeeld het geval wanneer men elk willekeurig reëel getal wil vervangen door het dichtst bijgelegen gehele getallen tussen $+100$ en -100 . Voor getallen met een absolute waarde kleiner dan 100 heeft men te maken met kwantisering (afronden) en voor de overige getallen met overloop (verzadeging). Deze twee effecten kunnen in één karakteristiek x_F weergegeven worden.



Figuur 5.5: De combinatie van kwantisering (afronden) met verschillende vormen van overloop.

5.3 A/D-omzettingsruis

Het eerste punt in een digitaal systeem waarop men doorgaans met kwantisering te maken heeft, vindt men bij de omzetting van een analoog ingangssignaal in een digitaal ingangssignaal. Afgezien van de overgang van continue tijd naar discrete tijd (d.m.v. bemonsteren) moet men daarbij ook de overgang van continue amplitude naar discrete amplitude realiseren (d.m.v. kwantisering).



Figuur 5.6: Kwantisering bij A/D-omzetting.

Ieder ingangsmonster $x(n)$ wordt omgezet in een uitgangsmonster $x_Q(n)$. Daarbij wordt een fout $e(n)$ geïntroduceerd die gegeven wordt door:

$$e(n) = x_Q(n) - x(n) \quad (5.2)$$

Het is direct in te zien dat in het geval van de figuur 5.6 geldt:

$$e(n) \leq \frac{q}{2}$$

Deze vergelijking gaat echter alleen op zolang er geen sprake is van overloop. Als de monsters $x_Q(n)$ een woordlengte hebben van B bits, dan geldt voor $x_Q(n)$:

$$-q \cdot 2^{B-1} \leq x_Q(n) \leq q \cdot 2^{B-1}$$

Om ernstige fouten ten gevolge van overloop te vermijden, moet men er voor zorgen dat $|x(n)|$ nooit, of praktisch nooit, de maximale waarden van $|x_Q(n)|$ overschrijdt. Men kan daarvoor zorgen door bij een gegeven signaal $x(n)$ met vermogen P_x

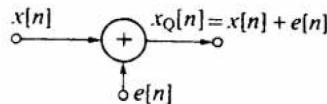
$$P_x = \lim_{N \rightarrow \infty} \left\{ \frac{1}{2N+1} \sum_{n=-N}^N x^2(n) \right\}$$

de waarde van q en B zo te kiezen dat:

$$K \cdot q \cdot 2^{B-1} = \sqrt{P_x}$$

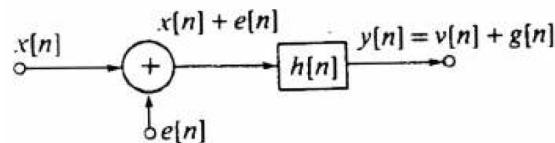
In de praktijk wordt vaak $K = \frac{1}{4}$ gekozen. De kans op overloop is dan meestal verwaarloosbaar klein. Daardoor hoeft bij de omzetting van $x(n)$ in $x_Q(n)$ alleen rekening gehouden te worden met de effecten ten gevolge van kwantisering.

De uitgang $x_Q(n)$ wordt dus: $x_Q(n) = x(n) + e(n)$. De kwantisering van $x(n)$ kan dus opgevat worden als het toevoegen van een stoorsignaal $e(n)$. Dit stoorsignaal wordt de kwantiseringsruis genoemd.



Figuur 5.7: Kwantiseringsruis

Het gekwantiseerde signaal $x_Q(n)$ wordt doorgaans toegevoerd aan een lineair systeem met impulsresponsie $h(n)$. Het uitgangssignaal $y(n)$ daarvan bestaat uit de gefilterde versie van $x(n)$ plus de gefilterde versie van $e(n)$, resp. $v(n)$ en $g(n)$.



Figuur 5.8: De impulsresponsie van een lineair systeem met invloed van kwantiseringsruis.

Uit de figuur volgt:

$$\begin{aligned} y(n) &= x(n) + e(n) * h(n) \\ &= x(n) * h(n) + e(n) * h(n) \\ &= v(n) + g(n) \end{aligned}$$

De beide signalen van $x(n)$ en $e(n)$ worden daarbij op precies dezelfde wijze gefilterd, volgends dezelfde filterkarakteristiek. Dat wil echter niet zeggen dat de signaal/ruisverhouding, P_x/P_e aan de ingang van het lineaire systeem hetzelfde is als de signaal/ruisverhouding, P_v/P_g aan de uitgang. Deze twee verhoudingen kunnen sterk verschillen als de spectra van $x(n)$ en $e(n)$ sterk verschillen.

Stochastische signalen zoals $e(n)$ en zoals een willekeurig informatiedragend signaal $x(n)$ (bijvoorbeeld spraak), hebben geen FTD en de convoluties kunnen niet direct vertaald worden naar een vermenigvuldiging in het frequentiedomein. Om correct te zijn zou hier gebruik moeten gemaakt worden van de theorie van stochastische discrete signalen. Hier worden een aantal resultaten gebruikt, uitgaande van een aantal veronderstellingen:

- de monsters van $e(n)$ nemen alle waarden tussen $+q/2$ en $-q/2$ aan met een even grote waarschijnlijkheid.
- de monsters van $e(n)$ vertonen onderling geen correlatie
- het signaal $e(n)$ is ongecorreleerd met het signaal $x(n)$

In de praktijk blijken deze aannames goed te kloppen, vooral voor enigszins grillig verlopende signalen (zoals spraak en muziek) en bij een voldoend aantal kwantiseringsstappen in de kwantiseringskarakteristiek.

Met de theorie van stochastische signalen valt te bewijzen dat een signaal waarin alle amplitudewaarden tussen $+A$ en $-A$ een even grote kans van optreden hebben, een effectieve waarde heeft van $A/\sqrt{3}$) en dus een vermogen van $1/3A^2$.

Met de eerste veronderstelling hierboven wordt het vermogen P_e gegeven door:

$$P_e = q^2/12$$

De tweede veronderstelling betekent dat het vermogen P_e gelijkmatig verdeeld is over alle frequenties in het fundamentele interval ; $e(n)$ heeft een vlak (wit) spectrum. Als een dergelijk signaal toegevoerd wordt aan een lineair netwerk met frequentieresponsie $H(e^{j\theta})$ dan vindt men aan de uitgang een signaal met een vermogen P_g dat gegeven wordt door:

$$P_g = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_e \cdot |H(e^{j\theta})|^2 d\theta = \frac{q^2}{12} \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(e^{j\theta})|^2 d\theta$$

De signaalruis-verhouding P_x/P_e wordt:

$$\frac{P_x}{P_e} = \frac{K^2 q^2 2^{2(B-1)}}{q^2/12}$$

Met $K = 1/4$ geeft dit:

$$\frac{P_x}{P_e} = \frac{3}{16} 2^{2B}$$

of uitgedrukt in decibels:

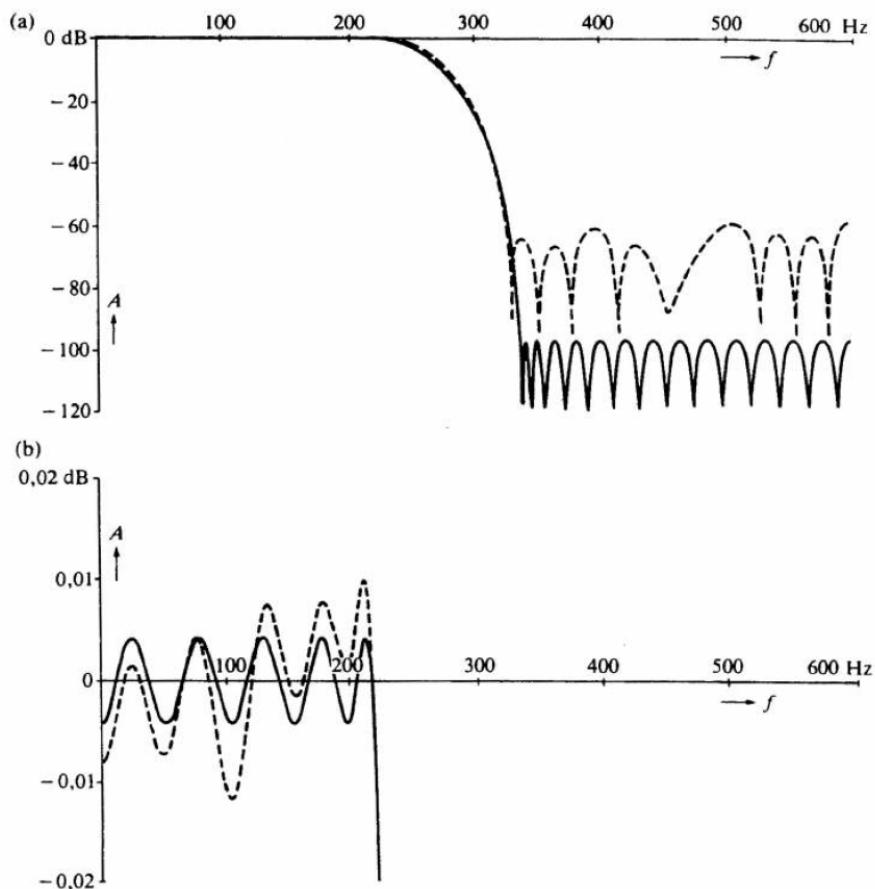
$$10 \log\left(\frac{P_x}{P_e}\right) = 10 \log\left(\frac{3}{16} 2^{2B}\right) = 20B \log(2) + 10 \log\left(\frac{3}{16}\right) = (6B - 7.3) dB$$

De constante term $-7.3 dB$ hang direct af van de keuze van K . De signaal-ruisverhouding neemt toe met $6dB$ voor elke bit die aan de woordlengte van $x_Q(n)$ wordt toegevoegd.

Men moet zich goed realiseren dat hiervoor uitgegaan werk van een ruisrijnig verondersteld signaal $x(n)$. Als $x(n)$ van een analoog signaal is afgeleid, bevat het altijd enige ruis. Het heeft geen zin om de kwantisering dan met zeer grote nauwkeurigheid uit te voeren, omdat de LSB's alleen maar een precieze voorstelling van de analoge ruis vertegenwoordigen en eigenlijk net zo goed gemist kunnen worden.

5.4 Kwantisering van de filtercoëfficiënten

Gebruikmakend van de ontwerpprocedures voor discrete filters, vindt met in het algemeen waarden van de filtercoëfficiënten met een zeer grote nauwkeurigheid. Om een praktisch realiseerbaar digitaal filter te verkrijgen, waarin de coëfficiënten slechts een beperkte woordlengte hebben



Figuur 5.9: Verandering van de weergavekarakteristiek van een digitaal filter als de coëfficiënten gekwantiseerd worden

moeten deze waarden gekwantificeerd worden. Daarmee wordt echter de bijbehorende frequentieresponsie van het filter veranderd, m.a.w. de ligging van de polen en nullen van de transferfunctie van het filter.

Deze veranderingen kunnen zeer aanzienlijk zijn. Het kan voorkomen dat na kwantisering, het filter niet meer voldoet aan de ontwerpeisen waarop de berekening van de niet-gekwantificeerde coëfficiënten gebaseerd was. In de figuur ziet men hoe de weergavekarakteristiek van een digitaal filter kan veranderen als de coëfficiënten gekwantiseerd worden. (De stippellijnen representeren het spectrum na digitalisatie)

In extreme gevallen kan zelfs een stabiel filter veranderen in een niet-stabiel filter, doordat een pool verschuift van een positie binnen de eenheidscirkel in het z -vlak naar een positie daarbuiten. Het kwantiseren van de coëfficiënten introduceren geen niet-lineariteiten en geen effecten die afhankelijk zijn van het ingangssignaal of die variëren met de tijd. Het leidt uitsluitend tot een eenmalig veranderen van de eigenschappen van het filter.

Deze verandering kan volledig geanalyseerd worden door de impulsresponsie of de frequentie-

responsie die resulteert na kwantisering, uit te rekenen. Men kan daaraan zien of nog aan de ontwerpeisen voldaan wordt. Het is gebruikelijk om de filtereisen bij de oorspronkelijke berekening van de ongekwantiseerde coëfficiënten scherper te stellen dan eigenlijk nodig is, zodat er een marge overblijft voor de afwijkingen die optreden ten gevolge van de latere kwantisering.

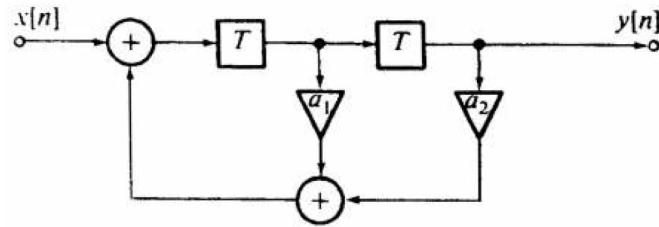
Het is gebleken dat de ene filterstructuur veel gevoeliger is voor kwantisering van de coëfficiënten dan de andere. In het algemeen kan men zeggen dat een filterstructuur minder gevoelig is naarmate de ligging van iedere pool en iedere nul afhankelijk is van minder coëfficiënten:

1. Bij de directe vorm 1 en de direct vorm 2 wordt elk van de polen beïnvloed door de waarden van alle coëfficiënten a_i en evenzo elk van de nullen door de waarden van alle coëfficiënten b_i . Daardoor kunnen kleine veranderingen in a_i en b_i grote verschuivingen in de polen en nullen veroorzaken. Dit is vooral het geval als de polen en/of de nullen dicht bij elkaar liggen zoals bijvoorbeeld bij smalbandige filters.
2. Bij de parallelenschakeling wordt elk van de polen bepaald door een klein aantal coëfficiënten in elk van de paralleltakken. De nullen daarentegen ontstaan door onderlinge compensatie van de uitgangssignalen van de verschillende takken en hangen zo van alle coëfficiënten af. Dit brengt met zich mee dat dit soort filters redelijk ongevoelig is in de doorlaatband (voornamelijk bepaald door de polen), maar erg gevoelig in de stopband (voornamelijk bepaald door de nullen).
3. In de serieschakeling worden zowel de polen en de nullen steeds door een klein aantal coëfficiënten bepaald. Filters met deze structuur hebben dus zowel in de doorlaatband als in de stopband een geringe gevoeligheid.
4. Voor het transversale filter geldt iets soortgelijks als onder punt 1. Elk van de nullen wordt bepaald door de waarde van alle coëfficiënten b_i . Dit filter is dus eveneens gevoelig voor coëfficiëntkwantisering. Het heeft echter het voordeel dat ook na kwantisering een eventueel gewenste lineaire fasekarakteristiek behouden blijft. De daarvoor vereiste symmetrie in de coëfficiëntwaarden is immers gemakkelijk te garanderen.

Invloed van filterstructuren bij kwantisering

Enig gevoel voor de invloed van kwantisering op de ligging van polen en nullen kan verkregen worden door een tamelijk eenvoudig filter eens wat nader te beschouwen.

Directe vorm Neem een zuiver recursieve tweede-orde-filtersectie met een directe-vorm-structuur.



Figuur 5.10: Zuiver recursief tweede-orde filter (directe vorm).

Als we stellen dat de knoop tussen de twee tijdsvertragingen $V(n)$ voorsteld kunnen we het volgende opstellen:

$$\begin{cases} y(n) = V(n-1) \\ V(n) = x(n-1) + a_2y(n-1) + a_1V(n-1) \end{cases}$$

Als we dit omzetten via de Z-transformatie bekomen we:

$$\begin{cases} Y(z) = V(z)z^{-1} \\ V(z) = X(z)z^{-1} + a_2Y(z)z^{-1} + a_1V(z)z^{-1} \end{cases}$$

Hieruit volgt:

$$Y(z)z = X(z)z^{-1} + a_2Y(z)z^{-1} + a_1Y(z)$$

$$Y(z)(z - a_2z^{-1} - a_1) = X(z)z^{-1}$$

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1}{z^2 - a_1z - a_2}$$

Deze functie heeft twee polen $p_{1,2}$. Als deze polen toegevoegd complex zijn, dan moet gelden:

$$p_{1,2} = \Re(e^{\pm j\varphi}) = \Re\{\cos\varphi \pm j\sin\varphi\}$$

Dus:

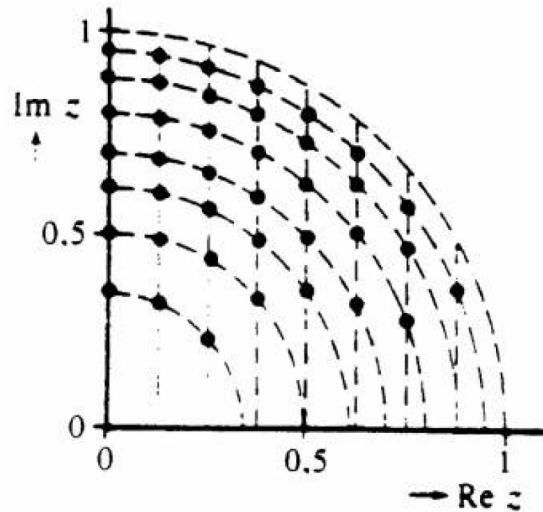
$$\begin{aligned} H(z) &= \frac{Y(z)}{X(z)} = \frac{1}{(z - \Re(e^{j\varphi}))(z - \Re(e^{-j\varphi}))} \\ &= \frac{1}{z^2 - 2\Re z \cos\varphi + R^2} \end{aligned}$$

m.a.w. $a_2 = -R^2$ en $a_1 = 2\Re \cos\varphi$

Het kwantiseren van a_1 geeft de verticale verloop weer en a_2 een cirkelvormige verloop, zoals ook te zien op figuur 5.11 op de volgende pagina.

Als a_1 en a_2 ongekwantiseerd zijn, kan men door een juiste keuze van deze twee parameters ieder gewenst paar toegevoegd complexe polen realiseren. Na kwantisering van a_1 en a_2 zijn

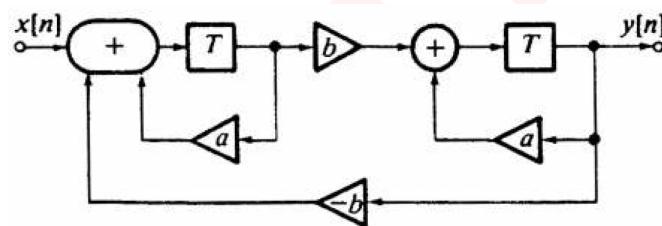
echter nog maar een beperkt aantal poolparen te realiseren, omdat a_1 en a_2 (en dus ook \Re en ϕ) slechts een eindig aantal verschillende waarden kunnen aannemen.



Figuur 5.11: Mogelijke poolposities als a_1 en a_2 worden gekwantiseerd tot 4 bits.

In figuur 5.11 is een voorbeeld gegeven van de polen die binnen de eenheidscirkel te realiseren zijn als a_1 en a_2 zo gekwantiseerd worden tot 4 bits (waarvan 1 bit voor het teken). Vanwege de symmetrie is slechts 1 kwadrant van het z -vlak getekend. Uit de figuur blijkt dat de mogelijke poolposities niet gelijkmatig over het z -vlak verdeeld zijn. De grootste afwijkingen van de frequentieresponsie kunnen vooral dan optreden als het gewenste filter polen heeft in een gebied waar de realiseerbare polen ver van elkaar liggen.

Gekoppelde structuur De mogelijke poolposities zijn sterk afhankelijk van de filterstructuur. Om dit te illustreren beschouwt men volgend filter:



Figuur 5.12: Tweede-orde-sectie met een zogenaamde 'gekoppelde structuur'

Als we stellen dat de knoop tussen de eerste tijdsvertraging en vermenigvuldiging met b voor-

gesteld wordt door $V(n)$, kunnen we het volgende opstellen:

$$\begin{cases} y(n) = ay(n-1) + bV(n-1) \\ V(n) = x(n-1) + aV(n-1) + by(n-1) \end{cases}$$

Als we dit omzetten via de Z-transformatie bekomen we:

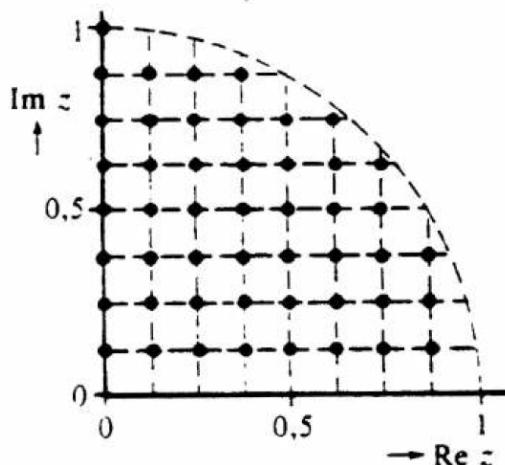
$$\begin{cases} Y(z) = aY(z)z^{-1} + bV(z)z^{-1} \\ V(z) = X(z)z^{-1} + aV(z)z^{-1} - bY(z)z^{-1} \end{cases}$$

Hieruit volgt:

$$\begin{aligned} Y(z)(1 - az^{-1}) &= bV(z)z^{-1} \\ V(z)(1 - az^{-1}) &= X(z)z^{-1} - bY(z)z^{-1} \\ Y(z)(1 - az^{-1}) &= bz^{-1} \frac{X(z)z^{-1} - bY(z)z^{-1}}{(1 - az^{-1})} \\ Y(z)(1 - az^{-1})^2 &= bX(z)z^{-2} - b^2Y(z)z^{-2} \\ Y(z)(1 - 2az^{-1} + a^2z^{-2} + b^2z^{-2}) &= bX(z)z^{-2} \end{aligned}$$

$$\begin{aligned} H(z) &= \frac{Y(z)}{X(z)} = \frac{bz^{-2}}{(1 - 2az^{-1} + (a^2 + b^2)z^{-2})} \\ &= \frac{b}{z^2 - 2az + (a^2 + b^2)} \end{aligned}$$

We krijgen twee toegevoegd complexe polen: $p = a \pm jb$. Als nu a en b worden gekwantificeerd tot 4 bit-coëfficiënten, kan men binnen de eenheidscirkel in het z-vlak polen realiseren op volgende posities:



Figuur 5.13: Mogelijke poolposities als a en b worden gekwantiseerd tot 4 bits.

De verdeling van de polen is veel gelijkmatiger. De effecten van coëfficiëntenkwantisering kan dus sterk verschillen voor beide filterstructuren.

Voorbeeld: Kwantisering van coëfficiënten

$$H(z) = \frac{1 + b_1 z^{-1} + b_2 z^{-2}}{1 - a_1 z^{-1} - a_2 z^{-2}}$$

Vragen:

1. Als we de nulpunten op de eenheidscirkel willen wat moeten b_1 en b_2 zijn?
2. Als we b_1 en b_2 willen kwantisering, wat is de invloed op de ligging?

Oplossing:

1. De teller bepaald de nulpunten, als we deze aanvullen met positieve machten bekomen we het volgende: $z^2 + b_1 z + b_2 = 0$. b_2 moet gelijk zijn aan ± 1 , want dan liggen de nulpunten op de eenheidscirkel. b_1 is gelijk aan $2\cos\phi$ en kan dus niet perfect binair voorgesteld worden, in tegenstelling tot b_2 .
2. b_1 kan perfect binair voorgesteld worden, b_2 daarentegen zal een fout bevatten door de $\cos\phi$ afronding. Deze nul zal dus verschuiven op de eenheidscirkel.

5.5 Woordlengtebeperking van tussenresultaten

De meest gecompliceerde gevolgen van het werken met eindige woordlengte vindt men bij het beperken van de woordlengte van tussenresultaten in digitale systemen. Met een tussenresultaat bedoelt men de uitkomst van een optelling of vermenigvuldiging die ergens in het digitale systeem wordt uitgevoerd. Bij wat volgt gaat men ervan uit dat een digitaal systeem gerealiseerd wordt waarin alle grootheden voorgesteld worden als woorden van B bits in een vaste-kommavoorstelling met slechts 1 bit (de tekenbit) vóór de komma. M.a.w. er kunnen alleen grootheden voorgesteld worden met een absolute waarde kleiner dan 1.

Voorbeeld optellen:

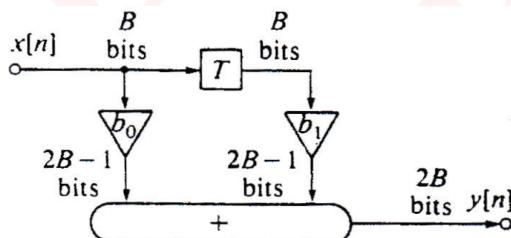
binair	decimaal
0,1101	+ 0,8125
0,1001	+ 0,5625
———— +	———— +
01,0110	+ 1,3750

Voorbeeld vermenigvuldigen:

binair	decimaal
0,1101	+ 0,8125
0,1001	+ 0,5625
———— ×	———— ×
0,01110101	+ 0,45703125

Men ziet dat de optelling van 2 B -bit-woorden kan leiden tot een som van $B + 1$ bits, en dat het product $2B - 1$ bits vereist. Het opmerkelijke verschil is hierbij dat optelling leidt tot een toename van het aantal bits vóór de komma en vermenigvuldigen tot meer bits na de komma. Teruggaand naar de oorspronkelijke getalvoorstelling kan bij optelling dus een overloop optreden en bij vermenigvuldigen een kwantisering.

Beschouw, ter illustratie, een aantal eenvoudige digitale filters.

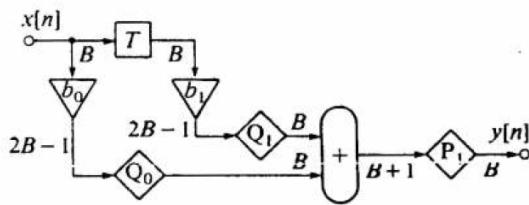


Figuur 5.14: Toename van de woordlengte in een eenvoudig FIR-filter

Dit is een eenvoudig FIR-filter waarbij geen eisen aan de woordlengte worden opgelegd. Het filter heeft een B -bit-ingangssignaal en B -bit-coëfficiënten. De woordlengte in het filter neemt geleidelijk toe en leidt tot een uitgangssignaal met een woordlengte van 28 bits.

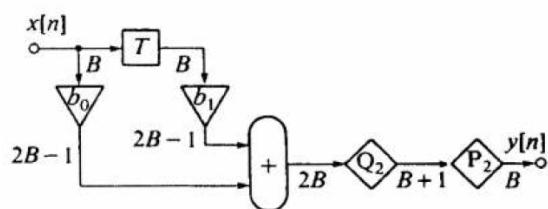
Als deze toename niet gewenst is moeten maatregelen getroffen worden. De plaats waar en de manier waarop de woordlengte beperkt wordt kan nog gekozen worden. Daarbij worden kwantificering (aangegeven met Q) en/of overloop (aangegeven met P) geïntroduceerd.

In de volgende twee figuren zijn er twee mogelijkheden aangegeven om een uitgangssignaal met woordlengte B te verkrijgen.



Figuur 5.15: Filter met 1 kwantisator maar een minder eenvoudige optelling om tot een uitgang van B-bits te komen

Na vermenigvuldigen van b_1 in figuur 5.15 hebben we $2B - 1$ bits nodig, waardoor we moeten opletten als we een eindige woordlengte hebben.

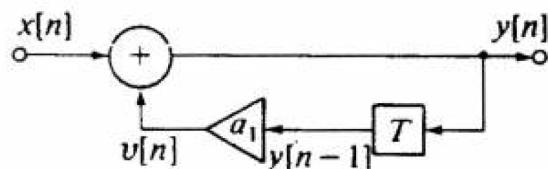


Figuur 5.16: Filter met twee kwantisatoren maar een eenvoudigere optelling om tot een uitgang van B-bits te komen

De ruiten stellen een woordlengte begrenzing voor.

De situatie ligt heel anders bij recursieve filters. Daar moet de woordlengte beperkt worden omdat deze anders onbegrensd zou toenemen. In het volgende voorbeeld beschouwt men een eerste-orde recursieve filter beschreven door de volgende differentievergelijking:

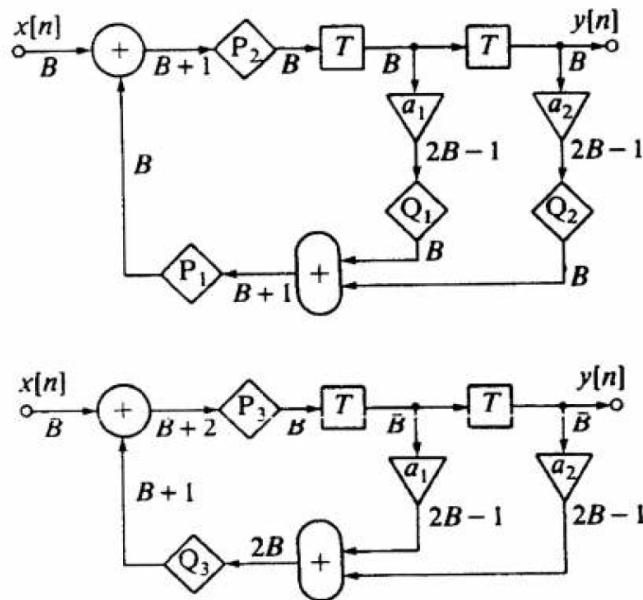
$$y(n) = v(n) + x(n) = a_1 y(n-1) + x(n)$$



Figuur 5.17: Eerste-orde recursief filter.

Als men aanneemt dat a_1 en $y(n-1)$ beide als B-bit-woorden zijn voorgesteld, dan zal zonder maatregelen $v(n)$ een $(2B-1)$ -bit-woord en $y(n)$ een $2B$ -bit-woord zijn. Bij de berekening van het volgende uitgangsmonster $y(n+1)$ moet a_1 met een $2B$ -bit-woord vermenigvuldigd worden en bekomt men een $3B$ -bit-uitgangsmonster. Bij de daaropvolgende berekening een $4B$ -bit-monster, enz.

Bij recursieve filters is begrenzing van de woordlengte van de tussenresultaten dus onvermijdelijk! In de volgende figuur is een zuiver-recursief tweede-orde-filter getekend waarin op twee verschillende manieren woordlengtebegrenzers (Q en P) zijn aangebracht.



Figuur 5.18: Recursief tweede-orde-filter met woordlengte-begrenzing

Het grote probleem bij de analyse van woordlengte-beperking van tussenresultaten is gelegen in het feit dat het filter nu strikt genomen niet-lineair is. Als de maatregelen zorgvuldig genoeg genomen worden, zal het filter in eerste benadering echter toch wel te beschrijven zijn als het lineaire filter dat ontworpen moet worden. Er kunnen echter een aantal specifieke verschijnselen optreden (zoals oscillaties bij recursieve filters) die het niet-lineaire karakter van het filter verraden. Deze verschijnselen treden op bij bepaalde soorten ingangssignalen zoals een constant ingangssignaal (speciaal $x(n) = 0$) of een periodiek (bijvoorbeeld een sinusvormig) ingangssignaal. Daarnaast veroorzaakt de niet-lineariteit ten gevolge van overloop (P) andere effecten dan de niet-lineariteit ten gevolge van kwantificering (Q).

Bovendien hangen de effecten nog af van:

- de structuur van het filter (recursief, niet-recursief, serieschakeling, parallelschakeling, direct vorm, enz. ...)
- de plaats van de woordlengtebegrenzers
- de karakteristiek van de woordlengtebegrenzers (afronden, afbreken, verzadigen, nulstelling, enz...)

Het is dus belangrijk een beetje orde te scheppen in dit oerwoud van mogelijkheden.

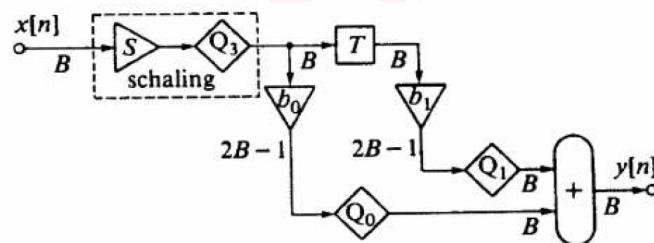
5.5.1 Overloop van tussenresultaten

Overloop in een digitaal filter kan leiden tot grote afwijkingen in het uitgangssignaal in vergelijking met het lineaire filter dat men eigenlijk zou willen hebben. Bij een filter met een niet-recursieve structuur zijn deze afwijkingen van eindige duur (nooit langer dan de duur van de impulsresponsie). Bij een recursief filter kunnen de gevolgen van het eenmalig optreden van overloop van onbeperkte duur zijn en allerlei nare effecten veroorzaken:

- Als het ingangssignaal $x(n)$ gelijk aan nul wordt vanaf een bepaalde waarde van n_{na} het optreden van overloop, kan er een blijvende oscillatie (overloop-oscillatie) optreden. Deze oscillatie heeft een grote amplitude die samenhangt met het overloopniveau.
- Als het ingangssignaal $x(n)$ periodiek is, kunnen er ten gevolge van overloop en afhankelijk van de begintoestanden van de filterregisters geheel verschillende uitgangssignalen ontstaan bij hetzelfde ingangssignaal.
- Als het ingangssignaal $x(n)$ periodiek is, kunnen ten gevolge van overloop kleine veranderingen in het ingangssignaal leiden tot grote veranderingen in het uitgangssignaal (sprongverschijnsel).
- Als het ingangssignaal $x(n)$ periodiek is, kunnen ten gevolge van overloop, signalen ontstaan met lagere frequentie (subharmonischen). Hierdoor kan een ingangssignaal in de stopband van een filter aanleiding geven tot een signaal in de doorlaatband.
- Als het ingangssignaal voldoende grillig verloop, zal overloop niet gauw aanleiding geven tot duidelijk herkenbare verschijnselen aan de uitgang. Desondanks zullen er toch extra afwijkingen ten opzichte van het ideale lineaire filter optreden.

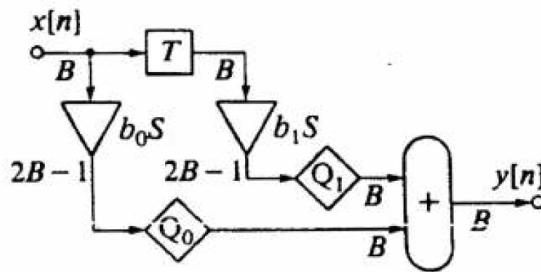
Alles bij elkaar genomen is er voldoende reden om overloop te vermijden. In principe is dit mogelijk door het toepassen van **schaling**. Dit houdt in dat het ingangssignaal van het filter met een zodanige factor $S < 1$ wordt vermenigvuldigd, dat geen overloop meer kan optreden. Bij voorkeur gebruikt men hiervoor een gehele macht van 2 omdat vermenigvuldigen dan in wezen slechts neerkomt op een verschuiving van de bits.

De overloop-niet-lineariteiten kunnen op deze manier geëlimineerd worden zoals in volgend voorbeeld:



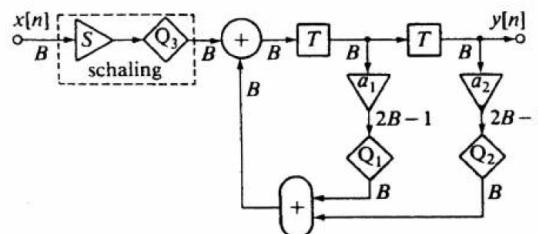
Figuur 5.19: Schaling in een transversaal filter waardoor geen overloop meer optreedt.

Het is ook mogelijk de schaalfactor S te combineren met de coëfficiënten b_0 en b_1 , in het geval van FIR filters. Want $Y(z) = H(z) \cdot X(z) \cdot S$ is gelijk aan $H_s(z) \cdot X(z)$, waar de S is verwerkt in de transferfunctie. Bij IIR filters daarentegen mogen we de coëfficiënten niet zomaar schalen vermits de S niet gemakkelijk in de transferfunctie, $H(s) = \frac{\sum_{i=0}^N b_i \cdot z^{-i}}{1 - \sum_{i=1}^N a_i \cdot z^{-i}}$, te krijgen is. In tegenstelling tot de FIR filter die wel een makkelijke transferfunctie heeft, nl. $H(s) = \sum_{i=0}^N b_i \cdot z^{-i}$.



Figuur 5.20: Schaling ondergebracht in de coëfficiënten

Men kan ook de overloop niet-lineariteiten bij recursieve filters elimineren.

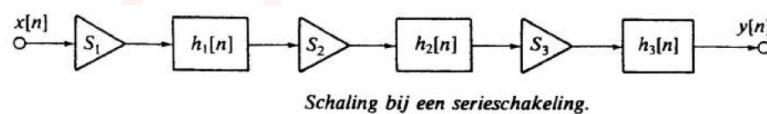


Figuur 5.21: Schaling bij een tweede-orde recursief filter (IIR filter).

Hier kan echter de schaalfactor S niet gecombineerd worden npt de filtercoëfficiënten, omdat daardoor de hele frequentieresponsie van de recursieve sectie zou veranderen.

In het algemeen geldt, dat de schaalfactoren wel gecombineerd kunnen worden met coëfficiënten die de ligging van de nulpunten bepalen en niet met coëfficiënten die de ligging van de polen bepalen.

Als we te maken hebben met een filter dat de structuur van een serieschakeling heeft, dan wordt schaling tussen elke 2 secties in gedaan. Er kunnen dan ook schaalfactoren $S > 1$ voorkomen als het uitgangssignaal van een bepaalde sectie altijd erg klein is.



Figuur 5.22: Schaling bij een serieschakeling.

De precieze grootte van iedere schaalfactor moet met enige zorg gekozen worden. Enerzijds moet hij zo gekozen worden dat overloop voorkomen word, anderzijds leidt een te kleine waarde er toe dat de meestwaardige bits in de volgende filtersectie(s) ongebruikt blijven. In dat geval spelen de onnauwkeurigheden, die geïntroduceerd worden door de kwantisators Q, relatief gezien een te grote rol. Dit geeft een slechtere signaal/ruis-verhouding aan de uitgang van het filter dan nodig is. Om de juiste waarde van S te bepalen moet men zowel het signaal kennen dat geschaald wordt als het filter (of de filtersectie) waaraan het wordt toegevoerd. In de literatuur zijn verschillende manieren bekend om op basis daarvan S te berekenen. Men kan daarbij nog verschillende criteria kiezen:

- worst-case-criterium : er mag nooit overloop optreden zolang $|x(n)|$ beneden een bepaalde waarde blijft.
- vermogencriterium : er mag nooit overloop optreden zolang het vermogen P_x beneden een bepaalde waarde blijft.
- sinus-criterium : er mag nooit overloop optreden bij sinusvormige signalen met een amplitude kleiner dan een bepaalde waarde.

Als de schaalfactor volgens het eerste criterium bepaald wordt, kan er nooit overloop optreden. Bij schaling volgens het tweede en derde criterium is dat niet zeker. Bij een goed ontwerp zal overloop een uitzondering blijven. Men moet echter wel bewust kiezen welke vorm van overloop er in deze uitzonderlijke gevallen optreedt. In het algemeen geldt dat verzadiging minder nare effecten veroorzaakt dan de andere mogelijkheden (zoals nulstelling en zaagtandvormige overloop). Een nadeel hierbij kan zijn dat verzadiging een meer ingewikkelde bewerking vraagt.

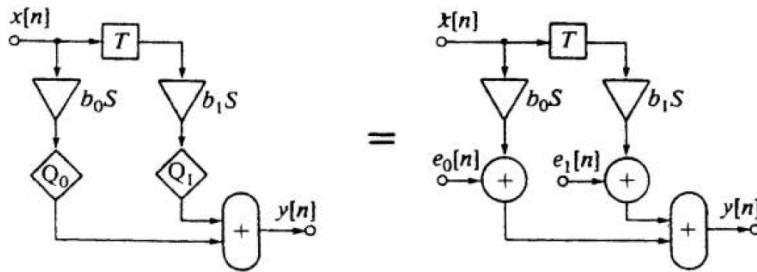
5.5.2 Kwantisering van tussenresultaten

Bij deze besprekking wordt er uitgegaan van digitale systemen waarin door de juiste manier van schaling, overloop niet kan voorkomen. Iedere kwantisering van een tussenresultaat komt overeen met de introductie van een onnauwkeurigheid binnen bepaalde grenzen. Als de opeenvolgende waarden van de tussenresultaten een voldoend grillig verloop hebben, kunnen deze onnauwkeurigheden opgevat worden als een stoorsignaal (kwantiseringsruis) dat aan het nuttige signaal wordt toegevoegd op de plaats waar de kwantisator zich bevindt. De besprekking wordt beperkt tot kwantisators die werken met afronden.

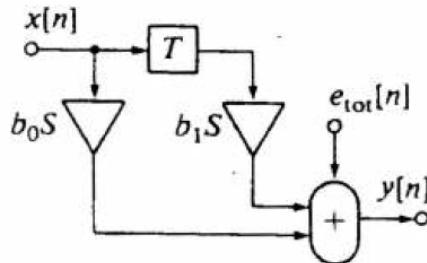
In de figuur werd ieder van de kwantisators door een opteller met een ruisbron vervangen. Ieder van de ruisbronnen levert een kwantiseringsruis met een vermogen:

$$P_e = \frac{q^2}{12}$$

Hierbij is q de kwantiseringsstap (= waarde van de LSB). Als aangenomen word dat de ruissignalen ongecorreleerd zijn, dan kunnen de 2 ruisbronnen $e_1(n)$ en $e_2(n)$ gecombineerd worden tot één ruisbron.



Figuur 5.23: Kwantisators gemodelleerd als kwantiseringsruisbronnen. De termen e_0 en e_1 stellen de ruisbronnen voor.



Figuur 5.24: Transversaal filter waarbij de kwantisering door één ruisbron $e_{tot}(n) = e_1(n) + e_2(n)$ wordt voorgesteld.

De transversale filter (5.24 op de pagina hierna) is dezelfde eerste orde FIR filter als in figuur 5.23 op de volgende pagina.

Het totale kwantiseringsruisvermogen aan de uitgang van het filter wordt dan:

$$P_{tot} = 2P_e = \frac{q^2}{6}$$

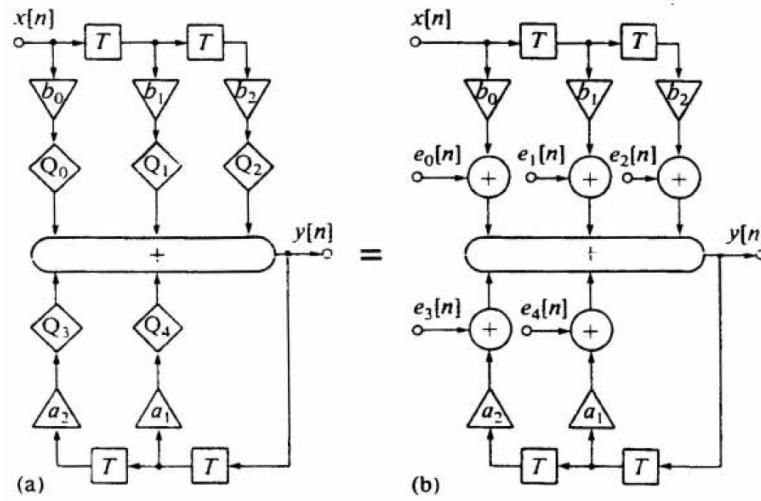
Deze benadering kan ook toegepast worden op andere soorten filters. In de volgende figuur is een tweede-orde recursief filter getekend met de directe-vorm-1-structuur. Daarin komen 5 kwantisatoren Q voor. Daarnaast is een model getekend met 5 ongecorreleerde ruisbronnen $e_0(n), \dots, e_4(n)$.

Deze 5 ruisbronnen kunnen gecombineerd worden tot één ruisbron $e_{tot}(n)$ met een ruisvermogen P_{tot} :

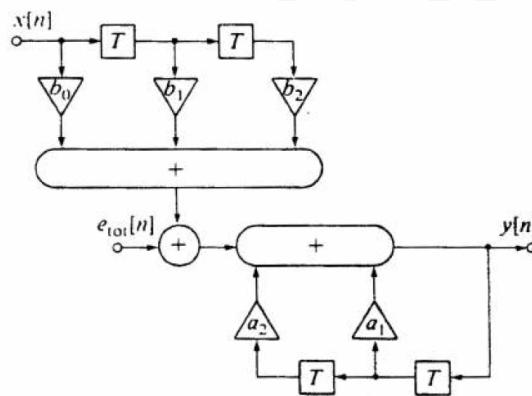
$$P_{tot} = 5P_e = 5\frac{q^2}{12}$$

Om te weten of dit ook het totale ruisvermogen is aan de uitgang van het filter wordt het schema hertekend.

We zien in figuur 5.26 op de pagina hierna dat het ingangssignaal $x(n)$ volledig wordt gefilterd, de ruis gaat enkel door het laatste deel van de filter, hierdoor moeten we bij het ruisvermogen enkel rekening houden met de noemer (waar de a -coëfficiënten staan).



Figuur 5.25: Recursieve sectie met de directe-vorm-1-structuur. (a) met kwantisatoren en (b) met kwantiseringsruisbronnen.



Figuur 5.26: Recursieve sectie volgens 5.25 waarbij de kwantisatoren door een ruisbron worden voorgesteld.

Hieruit blijkt dat de kwantiseringsruis $e_{tot}(n)$ gefilterd aan de uitgang verschijnt. De ruis wordt echter niet gefilterd volgens de totale frequentieresponsie $H(e^{j\theta})$.

$$H(e^{j\theta}) = \frac{T(e^{j\theta})}{N(e^{j\theta})} = \frac{b_0 + b_1 e^{-j\theta} + b_2 e^{-2j\theta}}{1 - a_1 e^{-j\theta} - a_2 e^{-2j\theta}}$$

Maar slechts volgens:

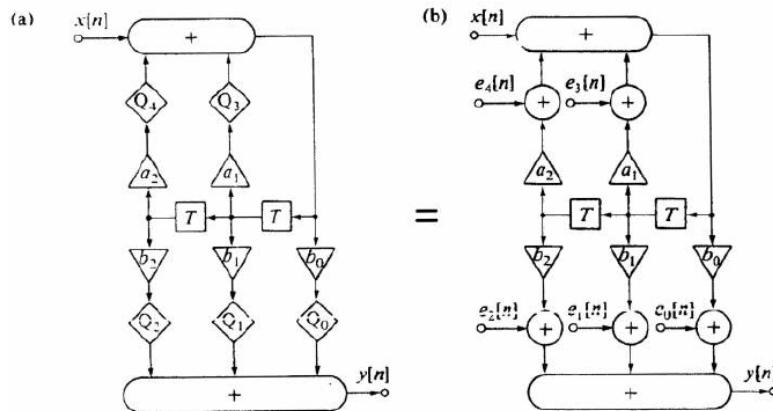
$$\frac{1}{N(e^{j\theta})} = \frac{1}{1 - a_1 e^{-j\theta} - a_2 e^{-2j\theta}}$$

Zoals eerder al besproken.

Het vermogen P_u van de kwantiseringsruis aan de uitgang van het filter wordt:

$$\begin{aligned} P_u &= \frac{1}{2\pi} \int_{-\pi}^{\pi} P_e \cdot |H(e^{j\theta})|^2 d\theta \\ &= \frac{5q^2}{12} \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{1}{N(e^{j\theta})} \right|^2 d\theta \end{aligned}$$

In de volgende figuur werd dezelfde methode toegepast op een tweede-orde recursief filter met de directe-vorm-2-structuur.



Figuur 5.27: Recursieve sectie met de directe-vorm-2-structuur. (a) met kwantisatoren en (b) met kwantiseringsruisbronnen.

De ruissignalen $e_0(n)$, $e_1(n)$ en $e_2(n)$ verschijnen rechtstreeks op de uitgang van het filter en geven daar een ruissignaal $e_{tot,1}(n)$ met een ruisvermogen:

$$P_{tot,1} = 3 \frac{q^2}{12}$$

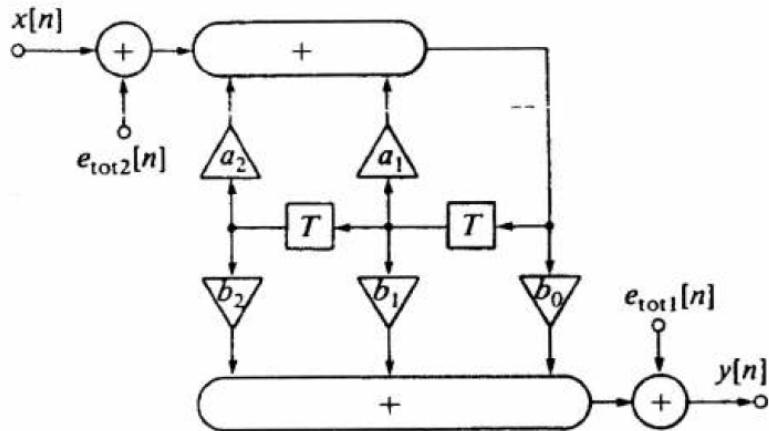
De ruissignalen $e_3(n)$ en $e_4(n)$ kunnen ook vervangen worden door een ruissignaal $e_{tot,2}(n)$ met een ruisvermogen:

$$P_{tot,2} = 2 \frac{q^2}{12}$$

Dit ruissignaal wordt echter eerst nog gefilterd voordat het aan de uitgang verschijnt. Een nauwkeurige beschouwing leert dat het met precies dezelfde frequentieresponsie $H(e^{j\theta})$ wordt gefilterd als $x(n)$.

Het totaal kwantiseringsruisvermogen P_u , aan de uitgang van het filter is gelijk aan:

$$P_u = \frac{3q^2}{12} + \underbrace{\frac{2q^2}{12} \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{T(e^{j\theta})}{N(e^{j\theta})} \right|^2 d\theta}_{\text{ruisvermogen is afhankelijk van de volledige filter}}$$



Figuur 5.28: Recursieve sectie volgens figuur 5.27 op de pagina hierboven, waarbij de kwantisatoren zijn vervangen door twee ruisbronnen $e_{tot,1} = e_0(n) + e_1(n) + e_2(n)$ en $e_{tot,2} = e_3(n) + e_4(n)$.

Bij het toepassen van deze methode op een willekeurig filter moet voor elke ruisbron apart nagegaan worden wat de frequentieresponsie is van de plaats van de ruisbron naar de uitgang van het filter.

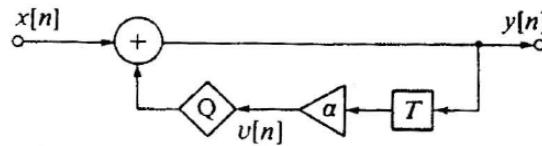
Het is interessant om de beide laatste uitdrukkingen met elkaar te vergelijken. Hieruit blijkt dat een tweede-orde-sectie in de directe vorm 1 en zo'nzelfde sectie in de directe vorm 2 met exact dezelfde coëfficiënten a_1 , a_2 , b_0 en b_2 en dus ook dezelfde $H(e^{j\theta})$, zich verschillend gedragen ten opzichte van kwantisering van de tussenresultaten.

De voorgaande analyse met betrekking tot ruisbronnen is gebaseerd op aannames die soms niet helemaal juist zijn (met name het ongecorreleerd zijn van de ruissignalen onderling en van de ruissignalen met de signalen vóór kwantisering). De resultaten zijn echter wel aanvaardbaar gezien het feit dat niet gestreefd wordt naar een 100% exacte analyse. De woordlengten kunnen toch maar in stappen van 1 bit veranderen, hetgeen overeenkomt met stappen van 6 dB in kwantiseringsruis. Een analyse met een nauwkeurigheid van 30 tot 40% is goed genoeg om de vereiste woordlengte te kunnen bepalen.

Voor ingangssignalen $x(n)$ die een voldoend grillig verloop hebben is de bovenstaande methode dan ook goed bruikbaar. De situatie verandert echter als het ingangssignaal gedurende langere tijd een constante waarde heeft of een periodiek karakter vertoont. De fouten die dan door de kwantisatoren geïntroduceerd worden zijn zeker niet meer ongecorreleerd en het uitgangssignaal kan een constante of een periodieke afwijking (grenscyclus) vertonen. Een praktische situatie waarin grenscyclussen ongewenst zijn doet zich bijvoorbeeld voor als $x(n)$ een gedigitaliseerd geluidssignaal is. In de pauzes ($x(n) = 0$) zijn dan periodieke stoorsignalen, zelfs op een zeer laag niveau, al gauw erg storend.

Om enig inzicht te krijgens in grenscyclussen wordt het volgende eerste-orde recursieve filter bekeken, waarbij Q een kwantisator met kwantiseringssstap q is en waarin afronden plaatsvindt.

Er kan een tabel worden opgesteld die $v(n)$ weergeeft op bepaalde n waarden, deze waarden worden gekwantiseerd (afgerond).



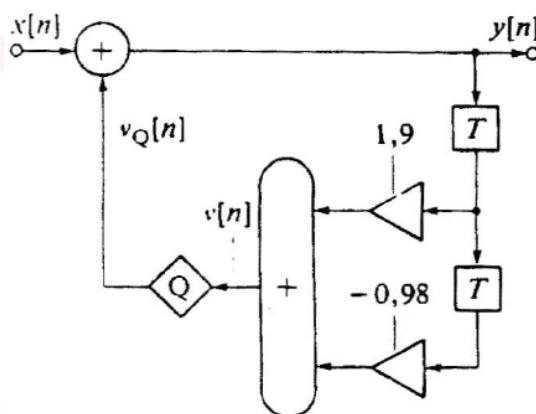
Figuur 5.29: Eerste-orde recursief filter met $a = -0,91$ waarin grenscyclussen kunnen optreden.

n	$v[n] = \alpha y[n-1]$	$y[n] = v_Q[n]$
-1	—	$7q$
0	$-0,91 \times 7q = -6,37q$	$-6q$
1	$-0,91 \times -6q = +5,46q$	$+5q$
2	$-0,91 \times 5q = -4,55q$	$-5q$
3	$-0,91 \times -5q = +4,55q$	$+5q$
4	$-0,91 \times 5q = -4,55q$	$-5q$
enz.	enz.	enz.

Uit de tabel blijkt dat $y(n)$ na enige tijd een periodiek verloop krijgt met periode 2 en amplitude $5q$, ondanks het feit dat het ingangssignaal $x(n) = 0$. De eigenschappen van de grenscyclus hangen af van de begintoestand $y(-1)$, de waarde van a en de louantiseringskarakteristiek van Q .

In dit eerste-orde recursieve filter kunnen altijd grenscyclussen optreden als de kwantiseringskarakteristiek gebaseerd is op afronden en als $|a| > 0,5$. Grenscyclussen kunnen alleen vermeden worden voor elke $|a| < 1$ als bij de kwantisering gebruik wordt gemaakt van het afbreken van de modulus. Deze laatste vorm van kwantisering geeft echter meer kwantiseringsruis dan afronden.

Ook in hogere-orde recursieve filters kunnen grenscyclussen optreden. Het volgende tweede-orde recursief filter heeft coëfficiënten 1,9 en $-0,98$.



Figuur 5.30: Tweede-orde recursief filter met coëfficiënten 1,9 en $-0,98$ waarin grenscyclussen kunnen optreden.

Het filter wordt beschreven met de volgende differentievergelijkingen:

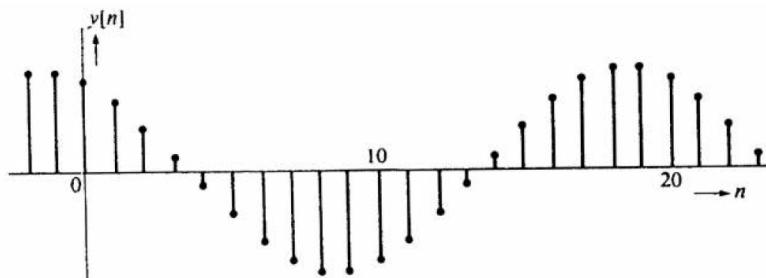
$$y(n) = x(n) + v_Q(n)$$

$$v_Q(n) = 1,9y(n-1) - 0,98y(n-2)_Q$$

Neem aan dat $x(n) = 0$ voor $n \geq 0$ en dat als beginvoorwaarden gelden:

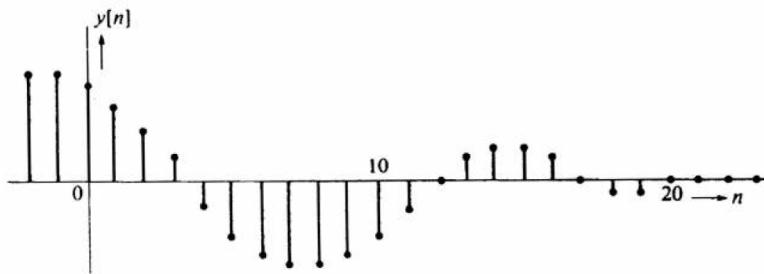
$$y(-2) = y(-1) = 11q$$

dan hangt het al of niet optreden van een grenscyclus weer af van de kwantiseringskarakteristiek van Q . De volgende figuur geeft het verloop van $y(n)$ dat geldt als Q overeenkomt met afronden.



Figuur 5.31: Verloop van $y(n)$ dat geldt als Q overeenkomt met afronden

Er treedt een grenscyclus op met een periode $N = 20$ en een amplitude van $11q$. De volgende figuur geeft $y(n)$ voor het geval dat Q overeenkomt met modulus afbreken.



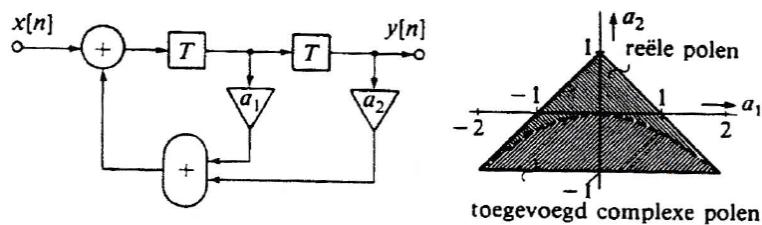
Figuur 5.32: Verloop van $y(n)$ dat geldt als Q overeenkomt met modulus afronden

Er treedt dan geen grenscyclus op.

De in de literatuur gepubliceerde resultaten hebben voor een belangrijk deel betrekking op de tweede-orde zuiver recursieve filtersectie. Het belang hiervan schuilt in het feit dat hogere-orde-filters vaak uit tweede-orde-secties worden samengesteld en dat het totale filter alleen wij van grenscyclussen kan zijn als elk van de tweede-orde-secties dat is.

Beschouw de volgende sectie en tevens de introductie van het (a_1, a_2) -vlak.

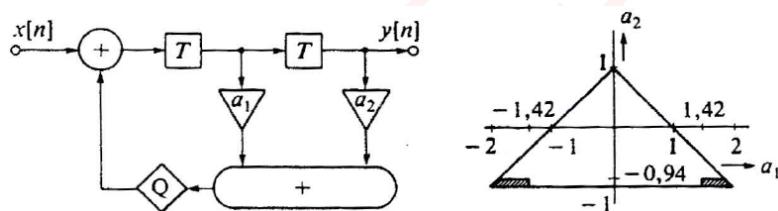
In dit (a_1, a_2) -vlak kan men bepaalde eigenschappen van het filter aanduiden die afhangen van de waarden van a_1 en a_2 .



Figuur 5.33: Zuiver recursieve tweede-orde-sectie en de waarden van a_1 en a_2 waarvoor deze sectie (zonder kwantisering) stabiel is.

In de figuur komt bijvoorbeeld het gebied binnen de driehoek overeen met stabiele filters (polen binnen de eenheidscirkel in het z-vlak), het gebied daarbuiten met instabiele filters. Verder is nog aangegeven wanneer de polen re  l zijn en wanneer complex.

Met betrekking tot grenscyclussen ten gevolge van kwantisering is gebleken dat er één type tweede-orde zuiver recursieve filtersectie is met de gunstigste eigenschappen voor wat betreft deze grenscyclussen. Die is getekend in de volgende figuur.



Figuur 5.34: Zuiver recursieve tweede-orde-sectie die het gunstigst is in verband met grenscyclussen (Q : kwantisor volgens afbreken van modulus; gearceerd: condities waaronder grenscyclussen optreden).

Hij bevat één kwantisator die werkt met het afbreken van de modulus. In het (a_1, a_2) -vlak is d.m.v. arcering aangegeven wanneer er nog grenscyclussen mogelijk zijn. Opmerking : deze structuur is niet optimaal voor wat betreft kwantiseringsruis, daarvoor zou kwantisering d.m.v. afronden moeten plaatsvinden.

In tegenstelling tot overloop-oscillaties hebben grenscyclussen over het algemeen een relatief kleine amplitude; men denkt daarbij aan de orde van grootte van enkele kwantiseringsstappen q .

Bij toenemende selectiviteit van het filter (polen dichter bij de eenheidscirkel in het z-vlak of combinaties (a_1, a_2) die dichter bij de rand van de stabiliteitsdriehoek liggen) kunnen ook grotere amplitudes voorkomen.

Als men in een gegeven filter, waarin grenscyclussen optreden de kwantiseringssstap q van alle kwantisators verkleint (door de woordlengte uit te breiden) zal ook de absolute amplitude (uitgedrukt als getal en niet als het aantal kwantiseringsstappen) afnemen. Door dan het uitgangssignaal $y(n)$ te kwantiseren kan men een filter maken dat schijnbaar vrij is van grenscyclussen.

Daarnaast bestaan er meer ingewikkelde methodes (randomized quantization, controlled quantization, error feedback) waarmee grenscyclussen gereduceerd of geëlimineerd kunnen worden.

5.6 Opmerkingen

Een overzicht van enkele belangrijke facetten van woordlengte-beperking van tussenresultaten in recursieve digitale filters is aangegeven in de volgende tabel.

	kwantisering	overloop
ingangssignaal $x[n]=0$	grenscyclussen (relatief kleine amplitude)	overlooposcillaties (relatief grote amplitude)
periodiek ingangssignaal (bijvoorbeeld sinusvormig)	grenscyclussen of kwantiseringsruis (in combinatie met gewenst uitgangssignaal)	sprongverschijnsel ('jump phenomenon')
grillig verlopend ingangssignaal	kwantiseringsruis (in combinatie met gewenst uitgangssignaal)	nog weinig over bekend
gunstige vorm van de karakteristiek	<ul style="list-style-type: none"> — voor kwantiseringsruis: afronden — voor grenscyclussen: modulus afbreken 	verzadiging

Na alles wat aan bod gekomen is over woordlengtebeperking van tussenresultaten zal het duidelijk zijn dat het ontwerpen van een digitaal filter nog lang niet klaar is, als via de ontwerp-procedures voor discrete filters de waarden van de niet-gekwantiseerde coëfficiënten gevonden zijn.

Als eerste stap worden de coëfficiënten gekwantiseerd op zo'n manier dat nog aan de ontwerp-eisen voldaan is

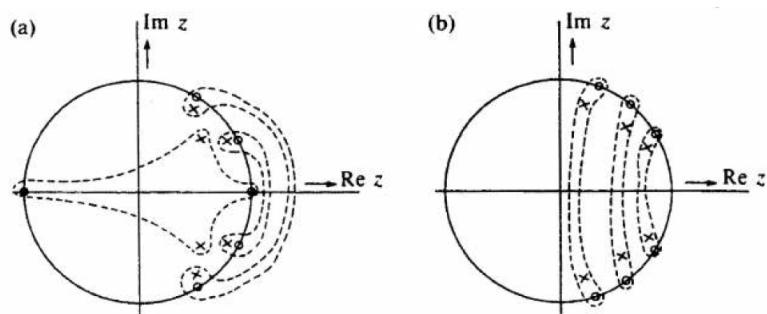
Daarna worden de effecten van de woordlengtebeperking van tussenresultaten geanalyseerd en zo veel mogelijk geminimaliseerd.

Bij de veelgebruikte cascade-schakeling valt tweede-orde-secties zijn er nog vier vrijheidsgraden:

1. combineren: welke polen worden gecombineerd met welke nullen in een bepaalde sectie
2. ordenen: in welke volgorde worden de secties in serie geschakeld
3. structureren: hoe ziet elke tweede-orde-sectie er uit (directe vorm 1, directe vorm 2, aantal en plaats van de kwantisators, de kwantiseringskarakteristiek, de over loopkarakteristiek)
4. schalen: met welke constante factoren moeten de signalen die optreden tussen de verschillende secties vermenigvuldigd worden

Over de punten schalen en structuren is hier al een en ander aan bod gekomen. De punten ordenen en combineren beperken zich tot vuistregels. Als goede vuistregel voor het combineren geldt realiseren polen en nullen die dicht bij elkaar liggen in dezelfde sectie (zie figuur 5.35 op de volgende pagina).

Met stippenlijnen is aangegeven welke polen en nullen in dezelfde sectie gerealiseerd worden.



Figuur 5.35: Combineren van polen en nulpunten van: (a) zesde-orde banddoorlaat filter; (b) zesde-orde bandsperrend filter.

Het ordenen van de verschillende secties kan vaak het best gebeuren in volgorde van oplopende selectiviteit, of juist in volgorde van aflopende selectiviteit.

Hoofdstuk 6

Systemen met meerdere bemonsteringsfrequenties

6.1 Inleiding

Bij alle tot nu toe besproken discrete systemen werd slechts één bemonsteringsfrequentie gebruikt voor alle signalen. Bij de omzetting van continue naar discrete signalen is T een parameter die vrij mag gekozen worden, zolang $\frac{1}{T}$ maar groter is dan tweemaal de hoogste frequentie die in het continue signaal voorkomt. Als aan deze voorwaarde voldaan is, dan is het discrete signaal een éénduidige voorstelling van het continue signaal. Het discrete signaal bevat alle informatie die nodig is om het continue signaal weer volledig onvervormd terug te winnen.

Het kan voordelen hebben om in een discreet systeem, waarin op verschillende punten signalen optreden waarvan de hoogste frequentie verschilt, met verschillende bemonsteringsfrequenties te werken. De bemonsteringsfrequentie wordt op iedere punt steeds zo laag mogelijk gehouden.

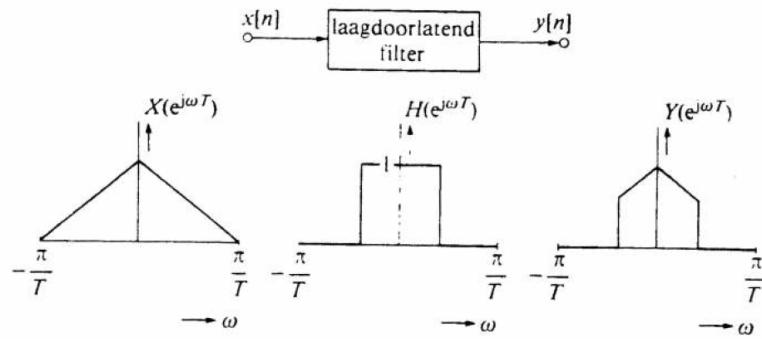
6.1.1 Voorbeeld: Bemonsteringsfrequentie verlaging

Dergelijke omzetting, zoals weergegeven op figuur 6.1 op de pagina hierna, kan gebruikt worden om een signaal voor te stellen met minder monsters per seconde waardoor bij de verwerking van het uitgangssignaal minder bewerkingen per seconde nodig zijn, waardoor op de kosten of op de vermogensconsumptie van een discreet systeem bespaard kan worden.

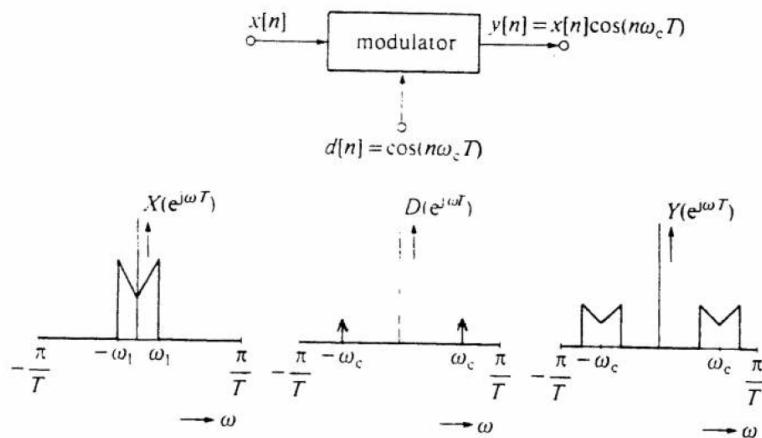
6.1.2 Voorbeeld: Bemonsteringsfrequentie verhoging

De modulator moet werken met een bemonsteringsfrequentie $\frac{1}{T}$ die hoger is dan tweemaal de hoogste frequentie in het fundamenteel interval van het uitgangssignaal.

Door het verhogen van de bemonsteringsfrequentie zal het fundamenteel interval verhogen, wat te zien is op figuur 6.2 op de volgende pagina.



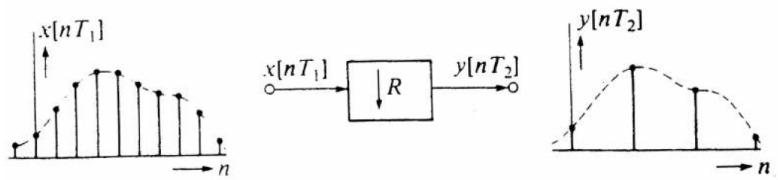
Figuur 6.1: Een discreet laagdoorlatend filter met ingangssignaal $x(n)$ en uitgangssignaal $y(n)$.



Figuur 6.2: Een discrete modulator

6.2 Verlagen van de bemonsteringsfrequentie met een gehele factor

6.2.1 De decimator



Figuur 6.3: Bemonsteringsfrequentie verlager in het tijdsdomein

De bemonsteringsfrequentie van het uitgangssignaal is dan $f_{s2} = \frac{f_{s1}}{R}$. Dit komt neer op $y(nT_2) =$

$x(nT_1R)$, we zien dus nu dat we de samplingperiode niet mogen weglaten, vermits het signaal verschillende samplingperiodes kan bevatten.

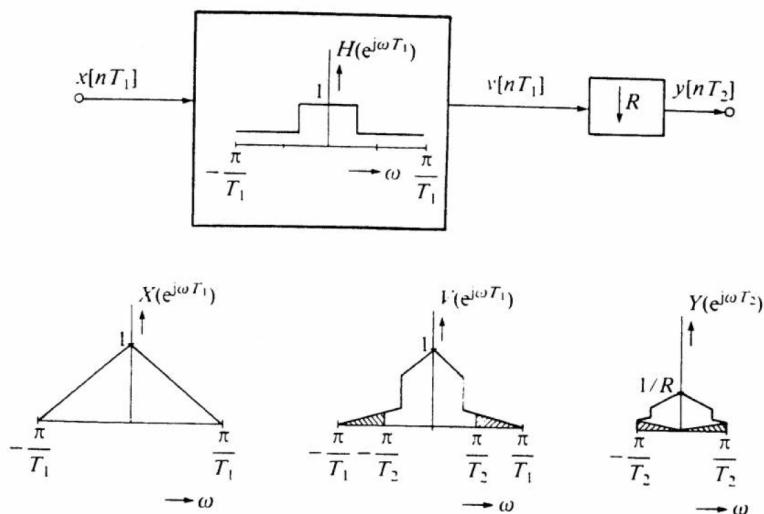
Als de eerste bemonsteringsfrequentie f_1 kleiner was dan de kritische bemonsteringsfrequentie $f_{s,krit}$ zullen we te maken krijgen met overlapping. Hierdoor moeten we eerst het signaal door een LDF sturen zodat we ruimte plaats maken in het FI. Dit zal het probleem van *folding* oplossen maar zal echter wel minder info bevatten.

We spreken van een decimator als we werken met een ideaal LDF in combinatie met bemonsteringsverlaging.

Het FI is $\pi \cdot f_s$, waardoor het FI niet gewijzigd wordt na het LDF, deze zal echter wel veranderen bij de bemonsteringsverlaging tot πf_{s2} . De amplitude zal ook wijzigen van 1 naar $\frac{1}{R}$.

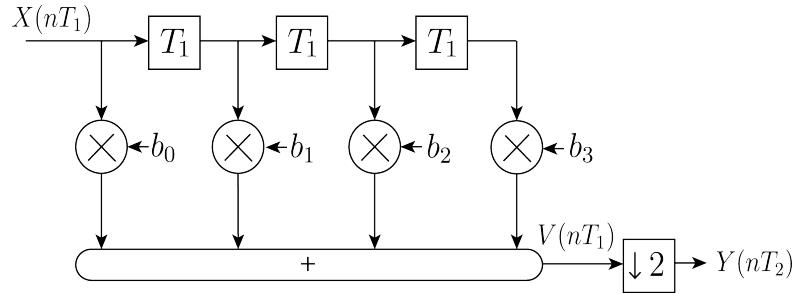
6.2.2 Een decimerend filter

In de praktijk spreekt men van een decimerend filter vermits we niet beschikken over een ideaal LDF. Hier zullen er ook nog frequenties bevinden in de stopband.



Figuur 6.4: Praktisch bemonsteringsverlagend filter, i.e. een decimerend filter.

We kunnen de bemonsteringsverlager niet schrijven als een transferfunctie omdat het niet tijds-onafhankelijk is. Een transferfunctie heeft de volgende eigenschappen: causaal, lineair en tijds-onafhankelijk.



Figuur 6.5: 3e orde decimerend FIR filter

6.2.3 Realisatie van een decimerend filter

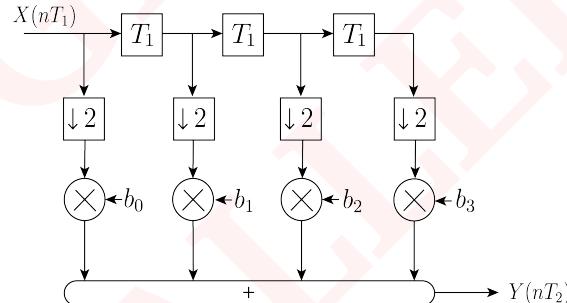
Decimerend FIR filter

In het geval van een derde orde decimerend filter, zoals weergegeven op figuur 6.5, kunnen we volgende vergelijkingen opstellen:

$$v(nT_1) = \sum_{i=0}^3 b_i x(nT_1 - iT_1)$$

$$y(nT_2) = v(2nT_1) = \sum_{i=0}^3 b_i x(2nT_1 - iT_1)$$

Het is echter beter om eerst te decimeren en dan pas een LDF toe te passen, zo besparen we kostbare bewerkingen. Dit is weergegeven in figuur 6.6. Hierbij zullen er maar vier vermenigvuldigingen nodig zijn op de helft van de waarden vermits $T_1 \rightarrow T_2$.



Figuur 6.6: 3e orde decimerend FIR filter (verbeterd)

Decimerend IIR filter

Als we figuur 6.7 op de pagina hierna bekijken, kunnen we de volgende vergelijkingen opstellen:

$$v(nT_1) = b_0 x(nT_1) + b_1 x(nT_1 - T_1) + a_1 v(nT_1 - T_1)$$

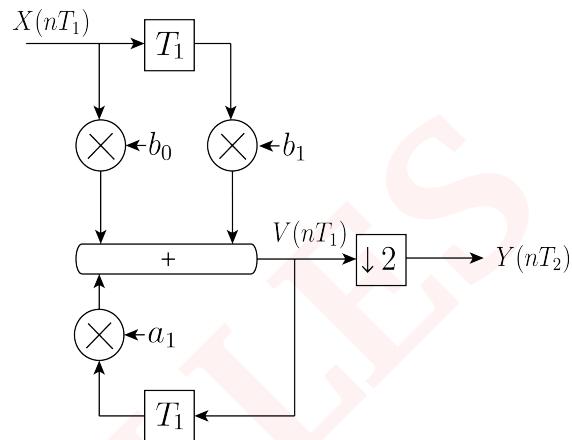
We kunnen $v(nT_1)$ omzetten naar $v(nT_1 - T_1)$ en substitueren in $v(nT_1)$, we krijgen dan:

$$v(nT_1 - T_1) = b_0x(nT_1 - T_1) + b_1x(nT_1 - 2T_1) + a_1v(nT_1 - 2T_1)$$

$$v(nT_1) = b_0x(nT_1) + (b_1 + a_1b_0)v(nT_1 - T_1) + a_1b_1x(nT_1 - 2T_1) + a_1^2v(nT_1 - 2T_1)$$

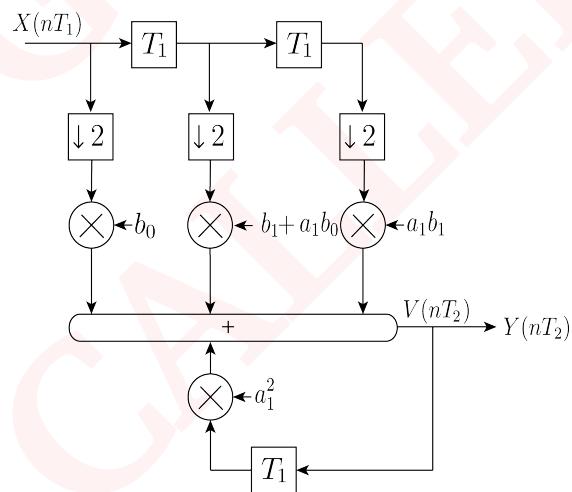
We kunnen $2T_1$ dan vervangen door T_2 en via $y(nT_2) = v(2nT_1)$ kunnen we dan het volgende schrijven:

$$y(nT_2) = v(2nT_1) = b_0x(2nT_1) + (b_1 + a_1b_0)v(2nT_1 - T_1) + a_1b_1x(2nT_1 - T_2) + a_1^2v(2nT_1 - T_2)$$



Figuur 6.7: Decimerend IIR filter

Als we echter eerst de bemonsteringsfrequentie verlagen dan moeten er minder vermenigvuldigingen gemaakt worden. Wat weergegeven is in figuur 6.8.



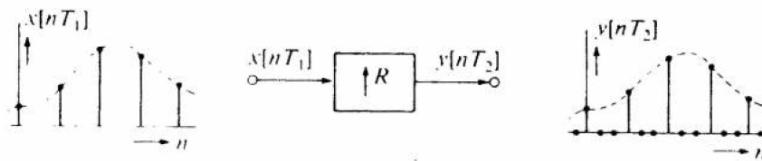
Figuur 6.8: Decimerend IIR filter (verbeterd)

Men zou denken dat er meer vermenigvuldigingen zijn 4 i.p.v. 3, maar weet dat er 4 vermenigvuldigingen wordt gedaan op een signaal met een bemonsteringsperiode van T_2 i.p.v. T_1 . Algemeen komt dit neer op het volgende:

Oude variant: 3 vermenigvuldigen in T_1 , ofwel 6 vermenigvuldigingen in T_2

Verbeterde variant: 4 vermenigvuldigingen in T_2

6.3 Verhogen van de bemonsteringsfrequentie met een gehele factor



Figuur 6.9: Symbol voor een bemonsteringsfrequentieverhoger (BFH) met een factor $R = \frac{T_1}{T_2}$.

Ook hier kunnen we geen transferfunctie koppelen aan de BFH, maar we kunnen wel het verband geven tussen de spectra van het ingangs- en uitgangssignaal.

$$\begin{aligned} Y(e^{j\omega T_2}) &= \sum_{k=0}^{\infty} y(kT_2) e^{-j\omega T_2 k} \\ &= \sum_{k=0, R, 2R, \dots}^{\infty} x(k \frac{kT_1}{R}) e^{-j\omega T_2 k} \end{aligned}$$

We schrijven $k = 0, R, 2R, \dots$ vermits we enkel niet-nul waarden moeten berekenen. Als we k vervangen door $i \cdot R$ dan krijgen we:

$$\begin{aligned} Y(e^{j\omega T_2}) &= \sum_{i=0}^{\infty} x(T_1 \cdot i) e^{-jT_1 k} \\ &= X(e^{j\omega T_1}) \end{aligned}$$

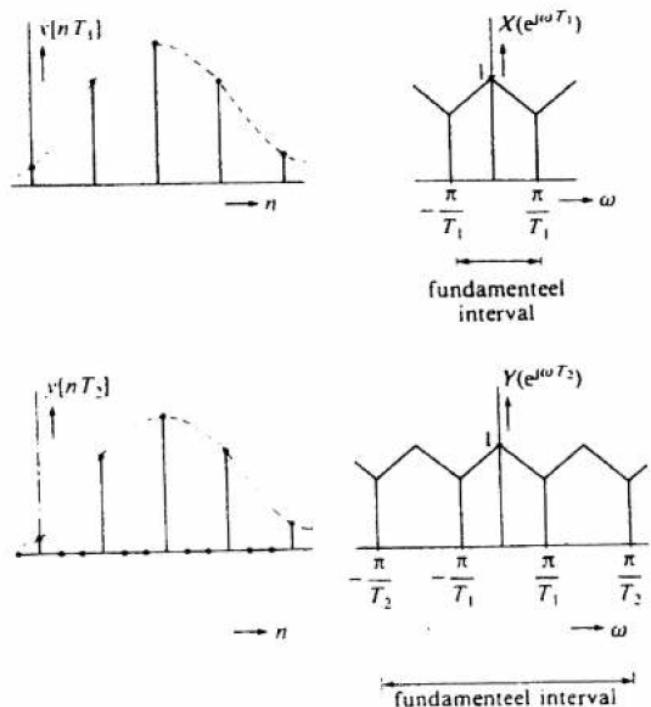
We verkrijgen dus hetzelfde spects aan de uitgang. Het FI zal echter wel veranderen, deze zal vergroot worden met een factor R (zie figuur 6.10 op de pagina hierna).

6.3.1 Realisatie van een interpolerend filter

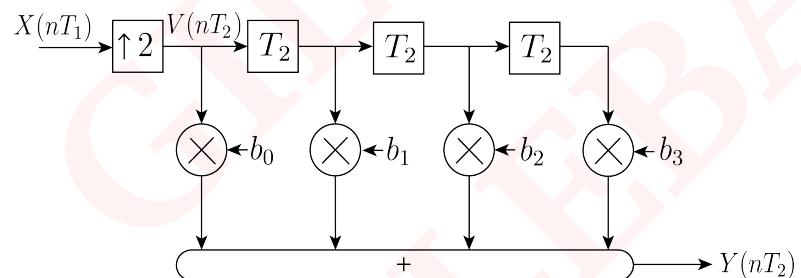
Interpolerend FIR filter

Als we eerst de bemonsteringsfrequentie verhogen zullen we dus meer bewerkingen moeten uitvoeren.

$$y(nT_2) = \sum_{i=0}^3 b_i V(nT_2 - iT_2)$$



Figuur 6.10: Het verband tussen $Y(e^{j\omega T_2})$ en $X(e^{j\omega T_1})$ bij verhoging van de bemonsteringsfrequentie met een factor $R = 3$.



Figuur 6.11: Interpolerend FIR filter

Als we een verhoging hebben van 2 weten we dat tussen de oorspronkelijke waarden telkens één nul-waarde bevindt (dit is bij $V(nT_2)$ met n -oneven).

$$\begin{aligned}
 V(0) &= x(0) \\
 V(T_2) &= 0 \\
 V(2T_2) &= x(T_1) \\
 V(3T_2) &= 0 \\
 V(4T_2) &= x(2T_1)
 \end{aligned}$$

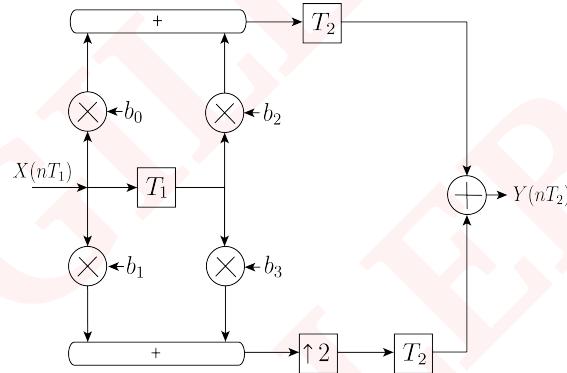
We kunnen dus de uitgang opsplitsen:

$$y(nT_2) = \begin{cases} = b_1V((n-1)T_2) + b_3V((n-3)T_2) & [n \text{ oneven}] \\ = b_0V(nT_2) + b_3V((n-2)T_2) & [n \text{ even}] \end{cases}$$

Als we dit linken met de ingang krijgen we:

$$y(nT_2) = \begin{cases} = b_1V((n-1)\frac{T_1}{2}) + b_3V((n-3)\frac{T_1}{2}) & [n \text{ oneven}] \\ = b_0V(n\frac{T_1}{2}) + b_3V((n-2)\frac{T_1}{2}) & [n \text{ even}] \end{cases}$$

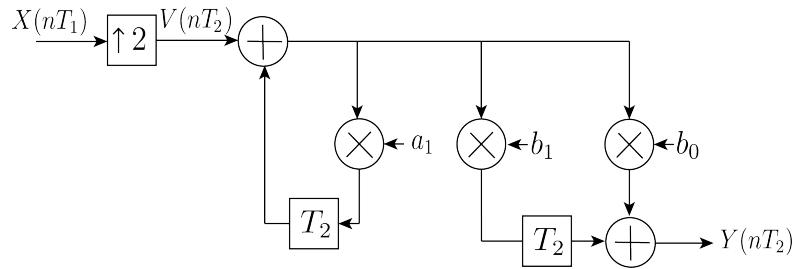
We kunnen beter eerst een LDF toepassen om nadien pas te verhogen, dan moeten er geen bewerkingen meer uitgevoerd op de toegevoegde nullen. Dit is weergegeven in figuur 6.12.



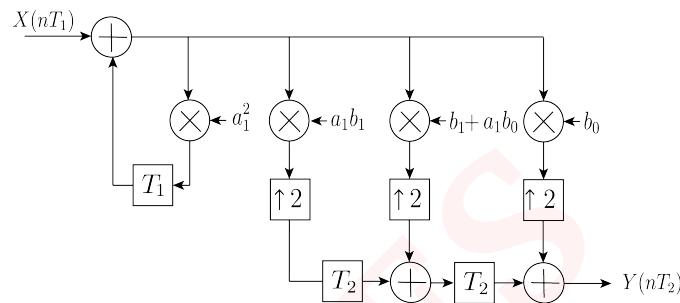
Figuur 6.12: Interpolerend FIR filter (verbeterd)

Interpolerend IIR filter

We kunnen weer een IIR filter ontwerpen die eerst interpoleert en dan de LDF toepast, maar dit zal weer leiden tot meer onnodige bewerkingen zoals weergegeven op figuur 6.13 op de pagina hierna en figuur 6.14 op de volgende pagina. Deze hebben respectievelijk drie bewerkingen in $T_2 = \frac{T_1}{2}$ en vier bewerkingen in T_1 , waarbij de laatste de minste bewerkingen zal moeten uitvoeren.



Figuur 6.13: Interpolatorend IIR filter

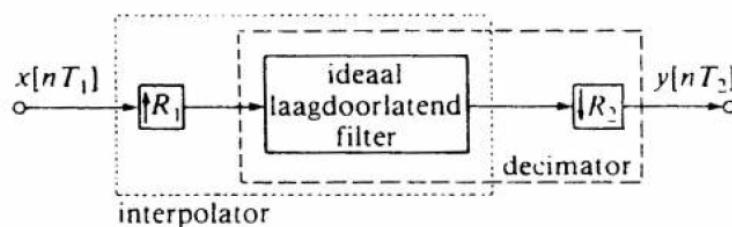


Figuur 6.14: Interpolatorend IIR filter (verbeterd)

6.4 Veranderen van de bemonsteringsfrequentie met een niet-gehele factor

We kunnen nu een decimerend en interpolerend filter combineren om de bemonsteringsfrequentie met een niet-gehele factor te veranderen.

We moeten echter weer opletten bij een globaal decimerend filter, dat we niet te maken hebben met folding, of weet dat dit kan gebeuren en dat de noodzakelijke acties worden ondernomen (i.e. instellen van f_c bij het LDF).



Figuur 6.15: Verandering van de bemonsteringsfrequentie met een rationele factor $\frac{T_2}{T_1} = \frac{R_2}{R_1}$.

6.4.1 Frequentieverhoging

We moeten eerst de frequentie verhogen, dit zal resulteren in een verbreding van het FI.

6.4.2 LDF

We passen een LDF filter toe met een $f_c = \frac{\pi}{T_1}$.

6.4.3 Frequentieverlaging

Nadien moeten de frequentie terug verlagen, wat resulteert in het versmallen van het FI. Het FI verkleint van $\frac{\pi}{T_2} \rightarrow \frac{\pi}{T_3} = \frac{R_1}{R_2} \cdot \frac{\pi}{T_1}$. Als de verhouding $\frac{R_1}{R_2}$ groter is dan 1 is er geen probleem want dan spreken we over een interpolerend filter. Is de verhouding echter kleiner dan 1 zal er folding optreden. De f_c van het LDF moet worden aangepast zodat $f_c = \frac{R_1}{R_2} \cdot \frac{\pi}{T_1}$, waardoor er geen folding meer voorkomt (wel verlies van informatie).

6.5 Polyfase decompositie

De M-component polyfase decompositie is een bruikbare voorstelling van $X(z)$ m.b.v. decimatie, interpolatie en δf_s met een niet-geheel getal.

We gaan een signaal uitdrukken in M componenten als de bemonsteringsfrequentie veranderd met een factor M .

$X(z)$ kan algemeen geschreven worden als $\sum_{k=-\infty}^{\infty} x(k)z^{-k}$.

Als we de z-transformatie nu herschrijven. Als we het signaal opdelen met indices gescheiden door M krijgen we:

$$\begin{aligned} X(z) &= \dots + x(0) + x(M)z^{-M} + x(2M)z^{-2M} + \dots \\ &\quad + x(1)z^{-1} + x(M+1)z^{-(M+1)} + x(2M+1)z^{-(2M+1)} + \dots \\ &\quad + x(2)z^{-2} + x(M+2)z^{-(M+2)} + x(2M+2)z^{-(2M+2)} + \dots \\ &\quad \dots \\ &\quad + x(M-1)z^{-(M-1)} + x(2M-1)z^{-(2M-1)} + x(3M-1)z^{-(3M-1)} + \dots \end{aligned}$$

We kunnen dit omvormen tot

$$\begin{aligned} X(z) &= \dots + x(0) + x(M)z^{-M} + x(2M)z^{-2M} + \dots \\ &\quad + z^{-1} [x(1) + x(M+1)z^{-M} + x(2M+1)z^{-2M} + \dots] \\ &\quad + z^{-2} [x(2) + x(M+2)z^{-M} + x(2M+2)z^{-2M} + \dots] \\ &\quad \dots \\ &\quad + z^{-(M+1)} [x(M-1) + x(2M-1)z^{-M} + x(3M-1)z^{-2M} + \dots] \end{aligned}$$

Dit kan geschreven worden als volgt:

$$z^{-i} \sum_{k=-\infty}^{\infty} x(kM+i)(z^M)^{-k}$$

Wat neerkomt op de Z-transformatie van de element van x gescheiden door i .

We definiëren de polyfase component $P_i(z)$.

$$P_i(z) \triangleq \sum_{k=-\infty}^{\infty} x(kM+i)(z)^{-k}$$

We kunnen dan de Z-getransformeerde schrijven als

$$X(z) = \sum_{i=0}^{M-1} z^i P_i(z^M)$$

Voorbeeld: Polyfase decompositie

Zet volgend signaal om via een 2-component polyfase decompositie.

$$X(z) = 1 + 2z^{-1} + z^{-2} + 3z^{-3}$$

$$\begin{aligned} X(z) &= (1 + z^{-2}) + (2z^{-1} + 3z^{-3}) \\ &= P_0(z^2) + z^{-1}(2 + 3z^{-2}) \\ &= P_0(z^2) + z^{-1}P_1(z^2) \end{aligned}$$

Dit kan zeer interessant zijn bij decimatie.

6.5.1 Variatie

We kunnen ook werken met de getransponeerde polyfase decompositie, die bruikbaar zal zijn bij interpolatie. Hierbij maken we gebruik van $Q_i(z) = P_{M-1-i}(z)$ met $i = 0, 1, \dots, M-1$.

We kunnen het signaal dan schrijven als:

$$X(z) = \sum_{i=0}^{M-1} Q_i(z^M)$$

6.6 Toepassing: Polyfase decompositie - Decimatie

We onderstellen dat we een gesampled signaal hebben waarbij f_s veel te hoog is. We willen deze verlagen met een gehele factor M , we moeten dus elk M^{ste} monster behouden.

Het gedecimeerd signaal ziet er dan als volgt uit:

$$x_d(k) = x(Mk)$$

De Z-getransformeerde kan dan worden beschreven via polyfase decompositie:

$$\begin{aligned} Z\{x_d(k)\} &= X_d(z) = \sum_{k=-\infty}^{\infty} x(k)z^{-k} \\ X_d(z) &= \sum_{k=-\infty}^{\infty} x(Mk)z^{-k} \triangleq P_0(z) \end{aligned} \quad (6.1)$$

6.6.1 Definite van de M^{de} wortel uit 1

We definieren de M^{de} wortel uit 1 als:

$$\omega_M = e^{-j\frac{2\pi}{M}}$$

We hadden hiervoor een uitdrukking gevonden voor $X(z)$, waarbij we het signaal verdeelden in een som van polyfase componenten. We vertrekken nu van dezelfde formule:

$$X(z) = \sum_{i=0}^{M-1} z^i P_i(z^M)$$

Maar zetten nu ω_M^k als argument, de uitdrukking die we dan vinden stelt ons in staat om deze te

linken met de eerder gevonden $P_0(z)$.

$$\begin{aligned}
 \sum_{k=0}^{M-1} X(\omega_M^k z) &= \sum_{k=0}^{M-1} \left[\sum_{i=0}^{M-1} (\omega_M^k z)^{-i} P_i((\omega_M^k z)^M) \right] \\
 &= \sum_{k=0}^{M-1} \left[\sum_{i=0}^{M-1} (\omega_M^{i k} z^{-i}) P_i(z^M) \right] \\
 &= \sum_{i=0}^{M-1} P_i(z^M) \underbrace{\sum_{k=0}^{M-1} (\omega_M^{i k} z^{-i})}_{=M\delta(i)[FFT]} \\
 &= \sum_{i=0}^{M-1} P_i(z^M) \sum_{k=0}^{M-1} M\delta(i)
 \end{aligned}$$

$$\sum_{k=0}^{M-1} X(\omega_M^k z) = M \cdot P_0 z^M$$

We kunnen hieruit $P_0(z^M)$ en $P_0(z)$ uithalen:

$$\begin{aligned}
 P_0(z^M) &= \frac{1}{M} \sum_{k=0}^{M-1} X(\omega_M^k z) \\
 P_0(z) &= \frac{1}{M} \sum_{k=0}^{M-1} X(\omega_M^k z^{\frac{1}{M}})
 \end{aligned} \tag{6.2}$$

Dit laatste stelt het gedecimeerd $X_d(z)$ signaal voor.

Voorbeeld: Stel $h(k) = a^k u(k) \Rightarrow X(z) = \frac{1}{1 - az^{-1}}$. Wat als we willen decimeren met een factor $M = 2$.

We kunnen dit op twee manieren berekenen: via directe berekening van de Z-transformatie; of via $P_0(z)$.

- via Z-transformatie

$$\begin{aligned} x(k) &= h(2k) \\ \Rightarrow X_d(z) &= \sum_{k=0}^{\infty} a^{2k} z^{-k} = \sum_{k=0}^{\infty} (a^2)^k z^{-k} = \frac{1}{1 - a^2 z^{-1}} \end{aligned}$$

- via $P_0(z)$

$$\begin{aligned} \omega_2 &= e^{-j \frac{2\pi}{2}} = -1 \\ \Rightarrow X_d(z) &= \frac{1}{2} \sum_{k=0}^{2-1} X(\omega_2^k z^{\frac{1}{2}}) \\ \Leftrightarrow X_d(z) &= \frac{1}{2} \left(X(z^{\frac{1}{2}}) + X(-z^{\frac{1}{2}}) \right) \\ &= \frac{1}{2} \left(\frac{1}{1 - az^{-\frac{1}{2}}} + \frac{1}{1 + az^{-\frac{1}{2}}} \right) \\ &= \frac{1}{1 - a^2 z^{-1}} \end{aligned}$$

6.6.2 DTFT van gedecimeerd signaal

Als we het spectrum bekijken van het gedecimeerd signaal, dan komt dit overeen met een som van kopieën van $X(e^{j\theta})$ die verschoven zijn over $2\pi k$ en herschaald zijn met een factor M .

$$X_d(e^{j\theta}) = \frac{1}{M} \sum_{k=0}^{M-1} X(e^{j(\frac{\theta}{M} - \frac{2\pi k}{M})}) \quad (6.3)$$

We kunnen dit interpreteren als volgt: we behouden alle M^{de} samples, dit geeft ons signaal $x_c(t)$ wat bemonsterd wordt aan f_s maar M keer kleiner. Er is echter vouwvervorming mogelijk als M te groot is. We moeten deze beperken door rekening te houden met de Nyquist voorwaarde: $\omega_s > 2\omega_N$ en dus moet $M < \frac{\omega_s}{2\omega_N}$.

Voorbeeld: Decimatie mogelijkheid

We onderstellen een analoog signaal met BB $\omega_N = 1000\text{rad/s}$ die is gesampled aan $\omega_s = 4000$. We zijn enkel geïnteresseerd in de spectrale inhoud van het signaal tussen 0 en 200rad/s .

We zullen dus eerst de BB beperken d.m.v. een LDF:

$$\vartheta_c = 200 \cdot \frac{2\pi}{4000} = \frac{\pi}{10}$$

Of anders: $M < \frac{4000}{2200}$, wat neerkomt op een decimatie met factor 10.

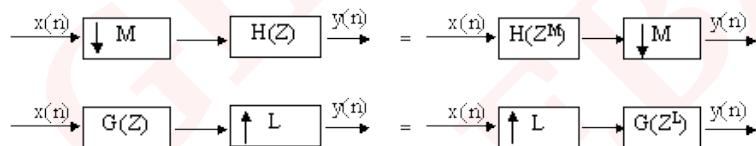
Hierbij hebben we de datastroombreedte met een factor 10 gereduceerd en kan dus interessant blijken voor real-time berekeningen. Als er na het LDF nog voldoende relevante info in het signaal aanwezig is, spreken we van datacompressie.

6.6.3 Structuren voor decimatie

In het vorig voorbeeld hadden we eerst de LDF toegepast en dan pas de decimatie, nochtans we enkel de waarden $y(Mk)$ nodig hebben.

We lossen dit op m.b.v. polyfase decompositie. We zullen dus de eerst het signaal decimeren om dan te verwerken in de LDF.

We zullen eerst bewijzen dat de volgorde geen rol speelt en dat beide systemen equivalent zijn:



Figuur 6.16: Twee systemen voor een decimerend en interpolerend filter op te bouwen.

Als we weten dat $X_d(z)$ het signaal is dat eerst door het decimerend filter is gegaan en dat $X'(z)$ het signaal is dat eerst door het LDF is gegaan dan kunnen we volgend bewijs opstellen:

Eerder werd al opgesteld dat

$$X_d(z) = \frac{1}{M} \sum_{k=0}^{M-1} X(\omega_M^k z^{\frac{1}{M}})$$

Dus dan wordt de uitgang (bij eerst decimeren):

$$y_1(z) = \frac{1}{M} \sum_{k=0}^{M-1} P_i(z) X(\omega_M^k z^{\frac{1}{M}})$$

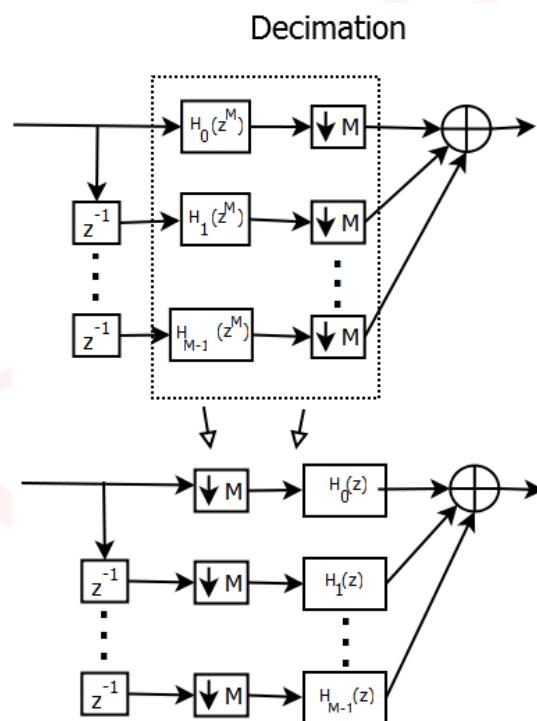
In het geval van het tweede systeem (eerst LDF):

$$X'(z) = H(z^M)X(z)$$

$$\begin{aligned} Y_2(z) &= \frac{1}{M} \sum_{k=0}^{M-1} P_i((\omega_M^k z^{\frac{1}{M}})^M) X(\omega_M^k z^{\frac{1}{M}}) \\ &= \frac{1}{M} \sum_{k=0}^{M-1} P_i(z) X(\omega_M^k z^{\frac{1}{M}}) \end{aligned}$$

Dit laatste komt omdat $\omega_M^{Mk} = 1 \forall k$.

Hieruit zien we dus dat beiden uitgangen dezelfde zijn. Echter zal het eerste efficiënter zijn. $P_i(z^M)$ heeft zelfde aantal van nul verschillende coëfficiënten als $P_i(z)$ en vereist hetzelfde aantal vermenigvuldigingen per cyclus, echter zal de datasnelheid veel hoger liggen.



Figuur 6.17: Decimerend filter m.b.v. polyphase decompositie. (In de nota's zullen we echter werken met P_i als component tegenover de H_i hier gebruikt.)

We kunnen dit bekomen resultaat gebruiken voor het afleiden van een structuur voor decimatie. Als we LDF onderstellen kunnen we schrijven:

$$H(z) = \sum_{i=0}^{M-i} z^{-i} P_i(z^M)$$

We kunnen dus onze LDF opsplitsen in een polyfase component, zoals ook weergegeven in figuur 6.17 op de pagina hiervoor.

Als $H(z)$ een FIR is met lengte N dan zal de polyfase componenten elke maximaal $\frac{N}{M}$ verschillende vermenigvuldingen uitvoeren. Met de nieuwe structuur zullen er dus $\frac{N}{M} \cdot M = N$ vermenigvuldigen per uitgangssample zijn.

Voorbeeld: Decimatie met polyfase decompositie

$$H(z) = 2 + 3z^{-1} + 3z^{-2} + 2z^{-3}$$

Onderstel filteren met $H(z)$ en decimeren met factor 2.

$$H(z) = P_0(z^2) + z^{-1}P_1(z^2)$$

Met $P_0(z) = 2 + 3z^{-1}$ en $P_1(z) = 3 + 2z^{-1}$

Als we eerst filteren zouden we 4 vermenigvuldigingen per tijdseenheid doen vermits P_0 en P_1 beide van orde 2 zijn.

Als we echter eerst decimeren zullen deze 4 vermenigvuldigingen in twee keer dezelfde tijd kunnen worden uitgevoerd.

6.7 Toepassing: Polyfase decompositie - Interpolatie

Bij interpolatie zullen we tussen twee monsters $M - 1$ nullen toevoegen.

6.7.1 Structuren voor interpolatie

We kunnen dus dezelfde redenering als hierboven volgen, waarbij we de twee systemen naast elkaar leggen en bewijzen dat ze dezelfde zijn. (zie figuur 6.16 op pagina 105)

Nu zullen we echter $H(z)$ niet vervangen door $P(z)$ maar door $Q(z)$.

We kunnen eerst het spectrum $X_e(z)$ (signaal na interpolatie) herschrijven naar $X(z^M)$ (toevoegen van de nullen):

$$X_e(e^{j\omega T_2}) = X(e^{j\omega T_1 M})$$

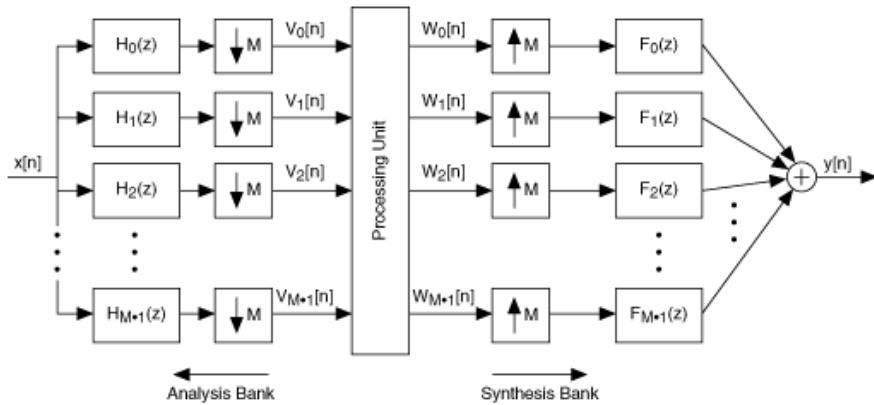
M.a.w. het spectra blijft ongewijzigd na interpolatie (wat al eerder was aangetoond).

Dus $Y_1(z) = Q(z^M) \cdot X_e(z) = Q(z^M) \cdot X(z^M)$ en is dus gelijk aan het uitgangssignaal waarbij we eerst filteren en dan interpoleren: $Y_2(z) = Q(z^M) \cdot X(z^M)$.

6.8 Analyse en synthesefilters

Bij het analyseren gaan we een spectrale analyse uitvoeren, wat neerkomt op het opsplitsen in deelbanden m.b.v. verschillende filters (FIR of IIR) $H_0(z), \dots, H_{M-1}(z)$. Deze gezamenlijke filters noemt een analyse filter bank.

In het synthese gedeelte zullen we opsplitsing terug moeten samenvoegen door filteren $F_0(z), \dots, F_{M-1}$ en deze te sommeren.



Figuur 6.18: Filter bank structuur, m.b.v analyse en synthese filters.

Een interessante analyse bank is een uniforme DFT bank. Dit is een verzameling filter $H_0(z), \dots, H_{M-1}(z)$, waarbij $H_k(z) = H_0(\omega_M^k z)$ (met $k = 1, 2, \dots, M - 1$).

In het frequentie domein kunnen we dan schrijven:

$$H_k(e^{j\theta}) = H_0(e^{j\theta - \frac{2\pi k}{M}})$$

dit is de frequentieresponsie van het k^{de} filter, dit komt neer op $H_0(e^{j\theta})$ verschoven over $\frac{\pi}{M}$.

Voorbeeld: Analyse en synthese filters

$H_0(z)$ is een LDF met afsnijfrequentie $\frac{\pi}{M}$

We weten dus dat de opvolgende componenten voldoen aan: $H_i(z)$ met $i \neq 0$ is een BDF met een BB van $\frac{2\pi}{M}$ en een centrale puls van $\frac{2\pi k}{M}$

De polyfase voorstelling zal de rekencyclomplexiteit verminderen:

$$H_0(z) = \sum_{i=0}^{M-1} z^{-i} P_{0i}(z^M)$$

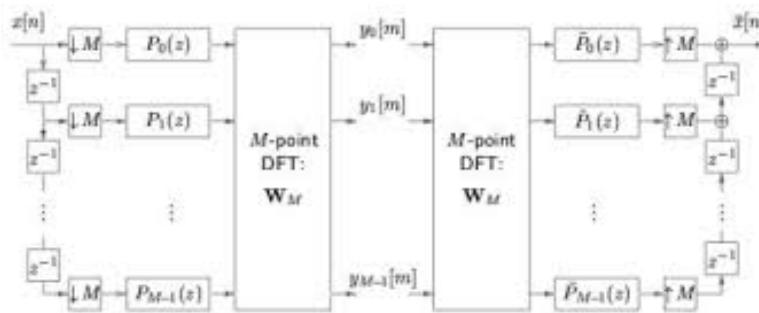
met $H_k(z) = H_0(\omega_M^k z)$

$$H_k(z) = \sum_{i=0}^{M-i} (\omega_M^k z)^{-1} P_{0i}((\omega_M^k z)^M)$$

met $\omega_{kM}^k = 1$

$$H_k(z) = \sum_{i=0}^{M-i} \omega_M^{-i k} z^{-1} P_{0i}(z^M)$$

Wat de vorm heeft van een DFT dus kunnen we gebruik maken van een FFT voor bepaalde keuzes van M .



Figuur 6.19: Analyse en synthese filter gebaseerd op uniforme DFT filters.

Voorbeeld: Rekencomplexiteit

Onderstel een 16-bands uniforme DFT bank gebaseerd op $H_0(z)$, een FIR filter met een lengte van 128.

Bij de directe implementatie zouden we dus $16 \cdot 128 = 2048$ vermenigvuldigingen moeten uitvoeren per tijdseenheid.

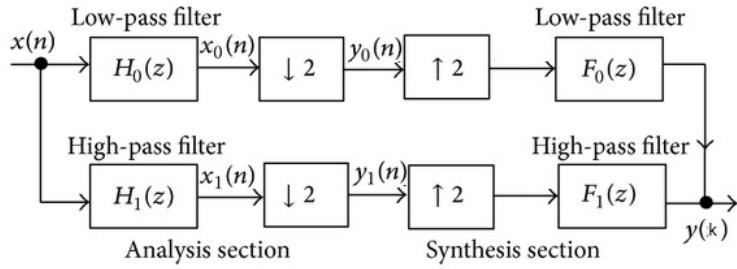
Als we echter werken met een polyfase voorstelling:

Elke polyfase filter $P_{0i}(z)$ doet 8 vermenigvuldigingen verschillende van 0 voor een totaal van $16 \cdot 8 = 128$ vermenigvuldigingen.

Als we gebruik maken van een radix-4 FFT voor de 16 punts DFT, dan zal de DFT berekenen met 16 vermenigvuldigingen. $\Rightarrow 128 + 16 = 134$ vermenigvuldigingen per tijdseenheid.

6.9 QMF filters

Quadrature mirror filters zullen het signaal in twee frequentiebanden opsplitsen m.b.v. een HD en LD filter.

**Figuur 6.20:** Twee kanaals QMF

Het analyse gedeelte bestaat uit een analyse sectie met LDF H_0 en HDF H_1 (2 decimators) en een synthese sectie met LDF F_0 en HDF F_1 (2 interpolators).

We zorgen ervoor door de keuze van onze interpolators dat we de vouwvervormingen door het decimeren verdwijnen bij de reconstructie.

De uitgangen van de decimators kunnen geschreven m.b.v. $X_d(z) = \frac{1}{M} \sum_{k=0}^{M-1} X(\omega_M^k z^{\frac{1}{M}})$:

De bovenste analyse tak:

$$X_0(z) = \frac{1}{2} \left[X(z^{\frac{1}{2}})H_0(z^{\frac{1}{2}}) + X(-z^{\frac{1}{2}})H_0(-z^{\frac{1}{2}}) \right]$$

De onderste analyse tak:

$$X_1(z) = \frac{1}{2} \left[X(z^{\frac{1}{2}})H_1(z^{\frac{1}{2}}) + X(-z^{\frac{1}{2}})H_1(-z^{\frac{1}{2}}) \right]$$

De uitgangen van de interpolaties:

$$Y_0(z) \rightarrow Y_0(z^2)$$

$$Y_1(z) \rightarrow Y_1(z^2)$$

De uitgang op de QMF:

$$Y(z) = F_0(z) Y_0(z^2) + F_1(z) Y_1(z^2)$$

We stellen dan onze uitgangen van de analyse sectie gelijk aan de ingang van de synthese sectie:

$$Y_0(z) = X_0(z) \text{ en } Y_1(z) = X_1(z)$$

$$Y(z) = F_0(z) \left[\frac{1}{2} [X(z)H_0(z) + X(-z)H_0(-z)] \right] + F_1(z) \left[\frac{1}{2} [X(z)H_1(z) + X(-z)H_1(-z)] \right]$$

We kunnen dit opdelen in de gewenste uitgang en termen die optreden ten gevolge van vouwvervorming (respectievelijk):

$$Y(z) = \frac{1}{2} [H_0(z)F_0(z) + H_1(z)F_1(z)]X(z) + \frac{1}{2} [H_0(-z)F_0(z) + H_1(-z)F_1(z)]X(-z)$$

De laatste term willen we vermijden en stellen deze dus gelijk aan nul:

$$H_0(-z)F_0(z) + H_1(-z)F_1(z) = 0$$

Dit geldt als

$$F_0(z) = H_1(-z)$$

$$F_1(z) = -H_0(-z)$$

Hierdoor wordt de TT van de QMF:

$$Y(z) = T(z) = \frac{1}{2} [H_0(z)H_1(-z) - H_1(z)H_0(-z)]X(z)$$

Hoofdstuk 7

Herhaling basis digitale filters

GILLEES
GALLEBAUF

DIGITALE SYSTEMEN

DIGITALE FILTER	1
DIGITALE SYSTEMEN.....	2
Elementaire analyse	2
Studie van LTO digitale systemen in het tijdsdomein	3
FILTERSTRUCTUREN	4
Inleiding	4
Niet-recursieve discrete filters	5
Recursieve discrete filters	6
ONTWERPMETHODES DIGITALE FILTERS.....	9
Inleiding	9
Ontwerp van IIR filters.....	9
Ontwerp van FIR filters.....	11
DE DISCRETE FOURIER TRANSFORMATIE.....	14
Discrete Fourier Reeks (DFR).....	14
Fast Fourier Transform Algorithms	15

We bespreken eerst kort hij de weergave karakteristiek eruit ziet voor een digitale filter, om vervolgens over te gaan op digitale systemen in het algemeen.

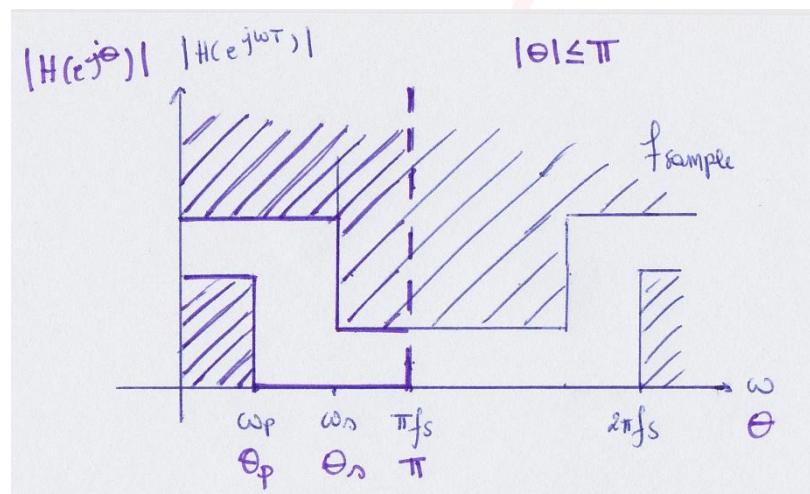
Digitale filter

Een digitale filter is zeer gelijkaardig aan een analoge filter. Hierbij gaat de frequentie niet tot oneindig maar wordt er een gedeelte –fundamenteel interval- telkens herhaalt. Het signaal waarop we werken is ook een *sample* of tijdsdiscreet signaal.

Een digitale filter wordt gekarakteriseerd door een transfer functie die is uitgedrukt als z-transformatie.

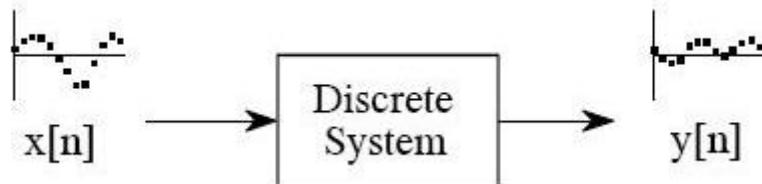
Men kan overgaan naar de relatieve –genormaliseerde- frequentie, dit door de volgende formule, $\theta = \omega T$.

Hierdoor krijgen we volgende weergavekarakteristiek:



Digitale systemen

Elementaire analyse



Eigenschappen van een discreet systeem

1. Lineair
2. Tijdsonafhankelijk (constante coëfficiënten)
Uitgangssignaal is onafhankelijk van het moment waarop de ingang wordt aangelegd.
3. Causaal systeem (geen tijdsverschuiving naar links)
Niet voorspellend.

Classificatie voor digitale systemen

We kunnen een onderscheid maken in niet-recursieve systemen en recursieve systemen. Bij een recursief systeem wordt er rekening gehouden met vorige waar(en) van het uitgangssignaal.

Niet-recursief

$$y(n) = \sum_{i=0}^N b_i x(n-i)$$

recursief

$$y(n) = \sum_{i=0}^N b_i x(n-i) + \sum_{i=1}^M a_i y(n-i)$$

Opmerking: De enige link tussen N en M is dat $N > M$ vermits er nooit meer nullen dan polen mogen voorkomen.

Bouwstenen voor een digitaal systeem

1. De eenhedsvertraging
2. De opteller
3. De vermenigvuldiger

Studie van LTO digitale systemen in het tijdsdomein

Responsie op een discrete-tijd-eenheidsimpuls $\delta(n)$

$$y(n) = \delta(n) + m y(n - 1)$$

We verkrijgen voor $n =$

$$\begin{array}{ll} 0 & y(0) = 1 + m y(-1) = 1 \\ 1 & y(1) = 0 + m y(0) = m \\ 2 & y(2) = 0 + m y(1) = m^2 \\ \dots & \dots \\ K & y(k) = 0 + m y(k - 1) = m^k \end{array}$$

$$y(n) = m^n = h(n)$$

Responsie op willekeurig ingangssignaal

$$Y(z) = H(z) \cdot X(z)$$

Via de z-transformatie kunnen we gemakkelijk de convolutie uit rekenen.

Voor recursieve systemen:

$$\begin{aligned} Y(z) &= \sum_{i=0}^N b_i Z[x(n-i)] + \sum_{i=1}^M a_i Z[y(n-i)] \\ &= \sum_{i=0}^N b_i z^{-i} X(z) + \sum_{i=1}^M a_i z^{-i} Y(z) \\ Y(z) &= \frac{\sum_{i=0}^N b_i z^{-i}}{1 - \sum_{i=1}^M a_i z^{-i}} X(z) \end{aligned}$$

Voor niet-recursieve systemen zal de noemer wegvallen.

$$Y(z) = \sum_{i=0}^N b_i z^{-i} X(z)$$

Filterstructuren

Inleiding

Bij niet-recursieve discrete filters (NRDF) komen geen terugkoppelingen voor. De impulsresponsie kan nooit meer monsters bevatten, verschillend van nul, dan het aantal vertragingselementen in het langste pad plus één.

Hierdoor zijn NRDF altijd van het FIR-type. Zoals de naam het zegt, Finite Impulse Response, zal de impulsresponsie van eindige duur zijn.

Bij een recursief discreet filter (RDF) zal de impulsresponsie, door terugkoppeling, meestal van oneindige duur zijn. Het filter is dan nu van een IIR-type (Infinite Impulse Response). Heel uitzonderlijk kan een RDF ook van het FIR-type zijn, in de praktijk is dat *helemaal* niet het geval.

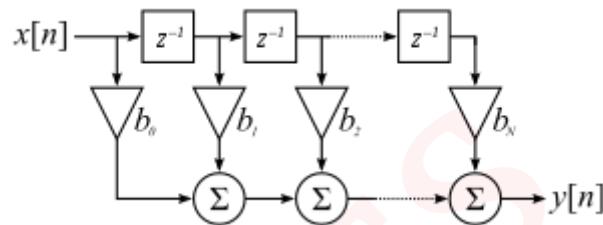
Een discreet filter wordt canoniek genoemd als het het minimum aantal vertragingselementen bevat dat nodig is om de bijhorende frequentieresponsie te realiseren.

Niet-recursieve discrete filters

$$h(n) = \sum_{i=0}^N b_i \delta(n - i)$$

Hierdoor verkrijgen we $N+1$ niet-nul waarden van $h(0)$ tot $h(N)$. De impulsresponsie op de rest ($n > N$) zal nul opleveren.

Standaardblokschema



Dit is een transversale structuur, de tijdsvertragingen worden op het originele ingangssignaal toegepast, hierna worden de signalen pas bewerkt.

Stabiliteit

$$H(z) = \sum_{i=0}^N b_i z^{-i}$$

Men zou denken dat we geen polen hebben, maar om de polen te vinden moeten we eerst zorgen dat de exponenten positief zijn.

$$H(z) = \sum_{i=0}^N b_i z^{-i} \cdot \frac{z^n}{z^n}$$

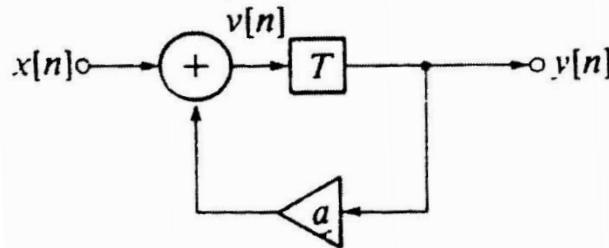
$$H(z) = \sum_{i=0}^N \frac{b_i z^{N-i}}{z^n}$$

We hebben dus N polen die allemaal liggen op de oorsprong. Hierdoor is een NRDF altijd stabiel.

Recursieve discrete filters

Vermits we een recursief discreet filter minstens één terugkoppeling heeft zullen we nu een oneindig aantal monsters hebben op de impulsresponsie. Het RDF is van het IIR-type.

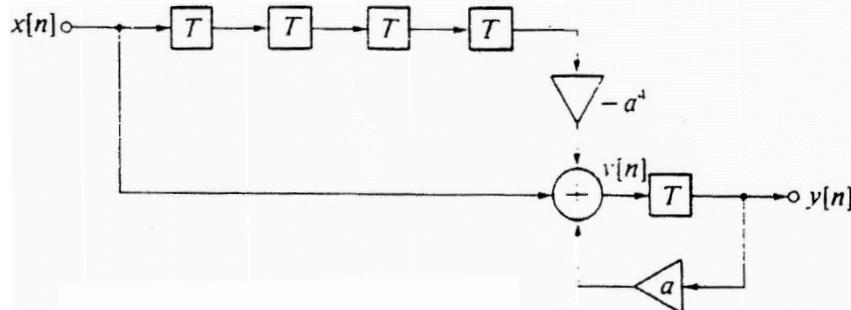
Door de impulsresponsie zal er **zeker** ooit eens op de uitgang een niet-nul monster komen te staan, deze zal door recursie terug in de schakeling terecht komen waardoor we een oneindig aantal monsters zullen bekomen.



$$y(n) = x(n-1) + a y(n-1)$$

$$\left\{ \begin{array}{l} h(0) = 0 \\ h(1) = 1 \\ h(2) = a \\ h(3) = a^2 \\ \dots \\ h(n) = a^{n-1} \end{array} \right.$$

We kunnen ook een RDF met een eindige impulsresponsie (FIR) realiseren:



$$y(n) = -a^4 x(n-5) + x(n-1) + a y(n-1)$$

$$\left\{ \begin{array}{l} h(0) = 0 \\ h(1) = 1 \\ h(2) = a \\ \dots \\ h(4) = a^3 \\ h(5) = -a^4 + a^4 = 0 \\ h(n) = 0 \text{ met } n > 5 \end{array} \right.$$

In de praktijk zal dit niet voorkomen vermits de a^4 tot stand is gekomen door een aantal vermenigvuldigingen met a, door telkens deze binaire voorstelling in de eindige woordlengte

te steken zullen er quantiseringsfouten optreden. De $-a^4$ daarentegen zal dus maar éénmalig moeten omgezet worden naar zijn binaire voorstelling, hierdoor zullen deze twee waarden niet gelijk zijn. De uitgang zal dus bij $n=5$ niet gelijk zijn aan nul waardoor we geen FIR hebben, maar terug een IIR.

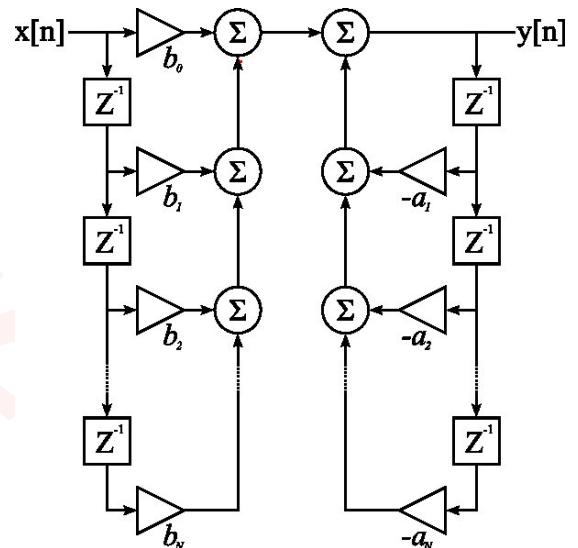
Standaardblokschema

Directe vorm 1

$$y(n) = \sum_{i=0}^N b_i x(n-i) + \sum_{i=1}^M a_i y(n-i)$$

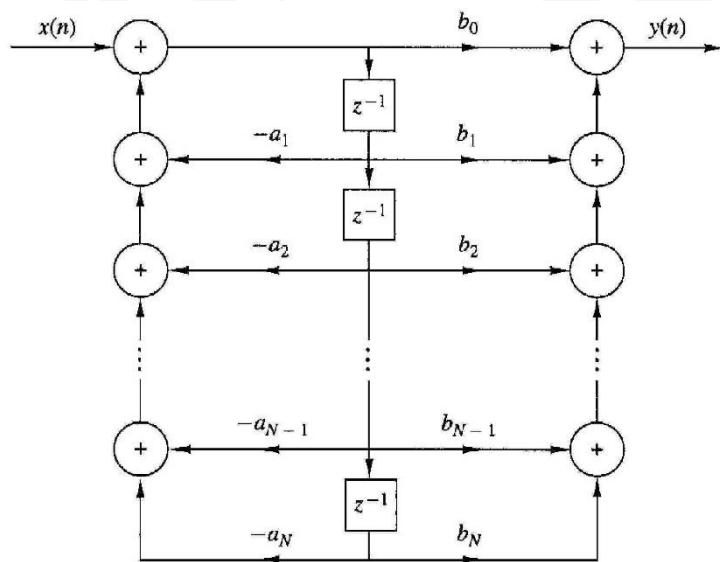
Deze structuur bevat $N+M$ vertragingselementen en $N+M+1$ vermenigvuldigers. De directe vorm 1 is dan ook niet-canoniek vermits we maar M vertragingenblokken mogen hebben.

Deze vorm is een serieschakeling van een niet-recursief en een recursief deel.



Directe vorm 2

Nu hebben we canoniek systeem vermits het aantal vertragingselementen gelijk is aan M .



Stabiliteit

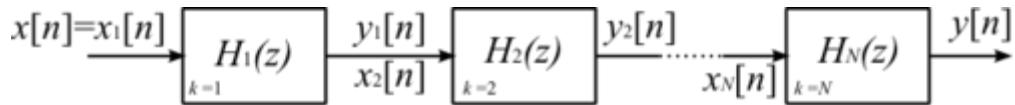
Vermits we nu te maken hebben met polen is er geen zekerheid op een stabiel systeem. Let op, maak de exponenten positief vooraleer je de polen zoekt!

Parametergevoeligheid

Een kleine variatie of afwijking in 1 van de coëfficiënten b_i heeft invloed op de ligging van alle N nulpunten en a_i heeft invloed op de ligging van de M polen.

We kunnen deze foutmarge beperken door de transferfunctie op te splitsen in kleinere deeltransferfuncties.

Serie- of cascadeschakeling



Opsplitsen van de oorspronkelijke transferfunctie in deeltransferfuncties die bestaan uit eerste orde en/of tweede orde functies.

- Eerste orde sectie (reële polen en nullen)

$$H_i(z) = \frac{b_0 + b_1 z^{-1}}{1 - a_1 z^{-1}}$$

- Tweede orde sectie (complexe polen en nullen)

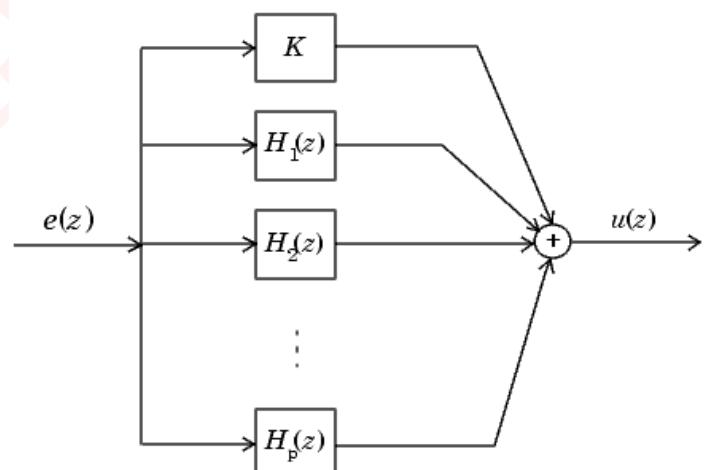
$$H_i(z) = \frac{\sum_{i=0}^2 b_i z^{-i}}{1 - \sum_{i=1}^2 a_i z^{-i}}$$

$$\begin{aligned} H(z) &= \prod_{i=1}^A H_i(z) \\ &= \frac{(z - z_1)(z - z_2)}{(z - p_1)(z - p_2)} \cdot \frac{(z - z_3)(z - z_4)}{(z - p_3)(z - p_4)} \end{aligned}$$

Parallelschakeling

Op de figuur stelt K een constante voor, de andere transferfuncties bekomen we door de methode van partieel breuken toe te passen. Hierdoor zullen de polen van de deelfuncties overeenkomen met de polen van de originele transferfunctie. Er zal geen overeenkomst zijn bij de nullen. Hierbij zijn de deelfuncties weer van eerste of tweede orde.

$$\begin{aligned} H(z) &= \sum_{i=1}^B H_i(z) \\ &= K + \frac{\alpha + \beta z^{-1}}{(z - p_1)(z - p_2)} + \frac{\gamma + \varepsilon z^{-1}}{(z - p_3)(z - p_4)} \end{aligned}$$



Ontwerpmethodes digitale filters

Inleiding

De specificaties tussen $0 < \theta < \pi$ (fundamenteel interval):

$$\begin{cases} \omega \leq \omega_p & A \leq A_{max} \\ \pi f_s \geq \omega \geq \omega_s & A \leq A_{min} \end{cases}$$

De frequentiesample zullen we iets verstrekken zodat de gevonden $H(z)$ zeker voldoet aan de originele specificaties.

Eerst kiezen we of we de gewenste frequentiekarakteristiek met een FIR of met een IIR filter willen benaderen. Een FIR filter zal een hogere orde systeem zijn. Hierna kiezen we een methode, we zien er in dit geval maar 1 per soort filter, het vensterfunctie (FIR) en bilineaire transformatie (IIR). We controleren de frequentiekarakteristiek door de Fourier te bepalen. Als de specificaties voldaan zijn zullen we over gaan naar de implementatie om nadien weer te controleren in het digitaal ontwerp.

Ontwerp van IIR filters

Inleiding

Transferfunctie van de vorm:

$$H(z) = \frac{\sum_{i=0}^2 b_i z^{-i}}{1 - \sum_{i=1}^2 a_i z^{-i}}$$

IIR filter vereist minder coëfficiënten en worden daarom vaak gekozen als men een scherpe cutoff en hoge throughput wilt.

Het nadeel echter is de aanwezigheid van polen die het systeem instabiel kunnen maken.

IIR filters worden ontworpen uitgaande van bestaande continue filters (Butterwoth, Elliptische,...).

Opmerking:

Controle achteraf is noodzakelijk, vermits men niet op voorhand weet of het systeem stabiel gaat zijn!

De bilineaire transformatie

De uitgangspunt van deze methode is de werking van een integrator.

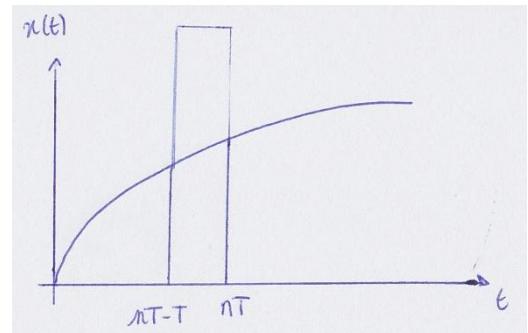
$$Y(s) = \frac{X}{s}$$

$$y(t) = y(0) + \int_0^t x(t) dt$$

$$y(nT) - y(nT - T) = \int_{nT-T}^{nT} x(t) dt$$

$$y(nT) - y(nT - T) = \frac{(nT - (nT - T))(x(nT) + x(nT - T))}{2}$$

$$y(nT) - y(nT - T) = \frac{T}{2} (x(nT) + x(nT - T))$$



Als we nu overgaan naar de z-transformatie:

$$Y(z)(1 - z^{-1}) = \frac{T}{2} X(z)(1 + z^{-1})$$

$$Y(z) = \frac{\frac{T}{2} \frac{1+z^{-1}}{1-z^{-1}}}{X(z)}$$

We kunnen dus $\frac{1}{s}$ vervangen door $\frac{\frac{T}{2} \frac{1+z^{-1}}{1-z^{-1}}}{z}$, dit is de bilineaire transformatie.

Hieruit volgt:

$$\frac{1}{s} = \frac{\frac{T}{2} \frac{1+z^{-1}}{1-z^{-1}}}{z} \rightarrow z = \frac{2+sT}{2-sT}$$

Controleren of dat we een stabiel discreet systeem hebben als we vertrekken van een stabiel continu systeem.

Stabiel in continu systeem als de polen in het LHV liggen ($Re\{s\} \leq 0$).

Stabiel in discreet systeem als de polen binnen de eenheidscirkel liggen.

$$|z| = \left| \frac{2+sT}{2-sT} \right| = \left| \frac{2+j\omega_d T}{2-j\omega_d T} \right| = 1$$

Als we negatieve reële waarden voor s invullen komen we $|z| < 1$ uit.

Omzetten digitale naar analoge specificaties

Omzetten van laplace- en z- naar frequentiedomein.

$$\begin{aligned} s &\rightarrow j\omega_a \quad \& \quad z \rightarrow e^{j\omega_d T} \\ s &= \frac{2}{T} \frac{1-z^{-1}}{1+z^{-1}} \rightarrow j\omega_a = \frac{2 e^{j\omega_d T} - 1}{T e^{j\omega_d T} + 1} \\ j\omega_a &= \frac{2}{T} \frac{e^{\frac{j(\omega_d T)}{2}} - e^{-\frac{j(\omega_d T)}{2}}}{e^{\frac{j(\omega_d T)}{2}} + e^{-\frac{j(\omega_d T)}{2}}} = \frac{2 j \sin \frac{\omega_d T}{2}}{T \cos \frac{\omega_d T}{2}} \\ \omega_a &= \frac{2}{T} \tan \frac{\omega_d T}{2} \end{aligned}$$

Oplossingsmethode

Digitale specificaties omzetten naar de digitale specificaties via de formule hierboven. Via But., Cheb., Ellip,... benadering zoeken we de transferfunctie in het Laplace domein. Door de bilineaire transformatie zetten we deze om naar een transferfunctie in het z-domein.

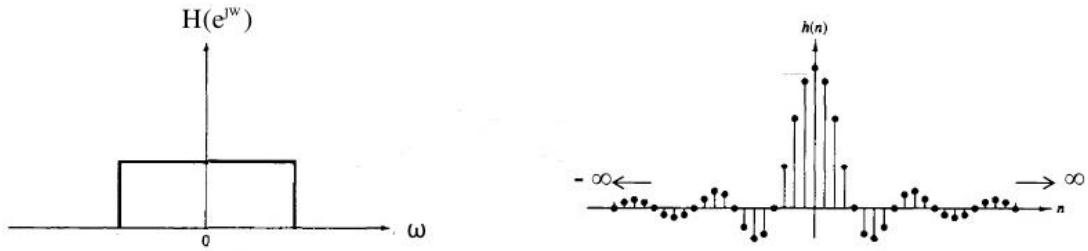
Ontwerp van FIR filters

De impulsresponsie van een FIR-filter is gelijk aan zijn filtercoëfficiënten.

$$H(z) = \sum_{i=0}^N b_i z^{-i}$$

$$Z\{h(n)\} = \sum_{i=0}^N h(i) z^{-i}$$

We vertrekken vanuit een ideaal LDF. Deze zetten we dan om naar het discreet tijdsdomein a.d.h.v. een inversie Fouriertransformatie.

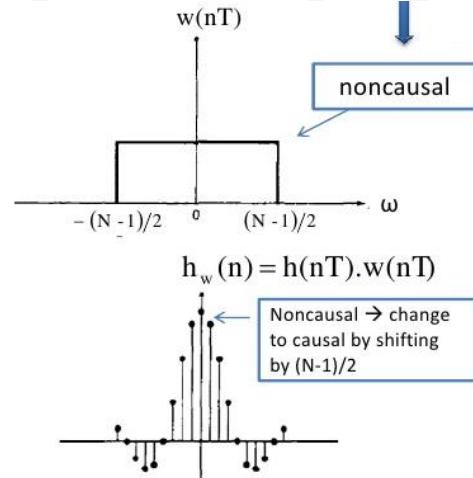


De rechthoek in het frequentiedomein wordt om gezet naar een *sinc* functie in het tijdsdiscreet domein, waarbij de waarden van n van min tot plus oneindig gaan. We hebben dus te maken met een oneindig durende impulsresponsie, een IIR. Deze waarden begrenzen we via vensterfuncties met hetzelfde bemonsteringsinterval.

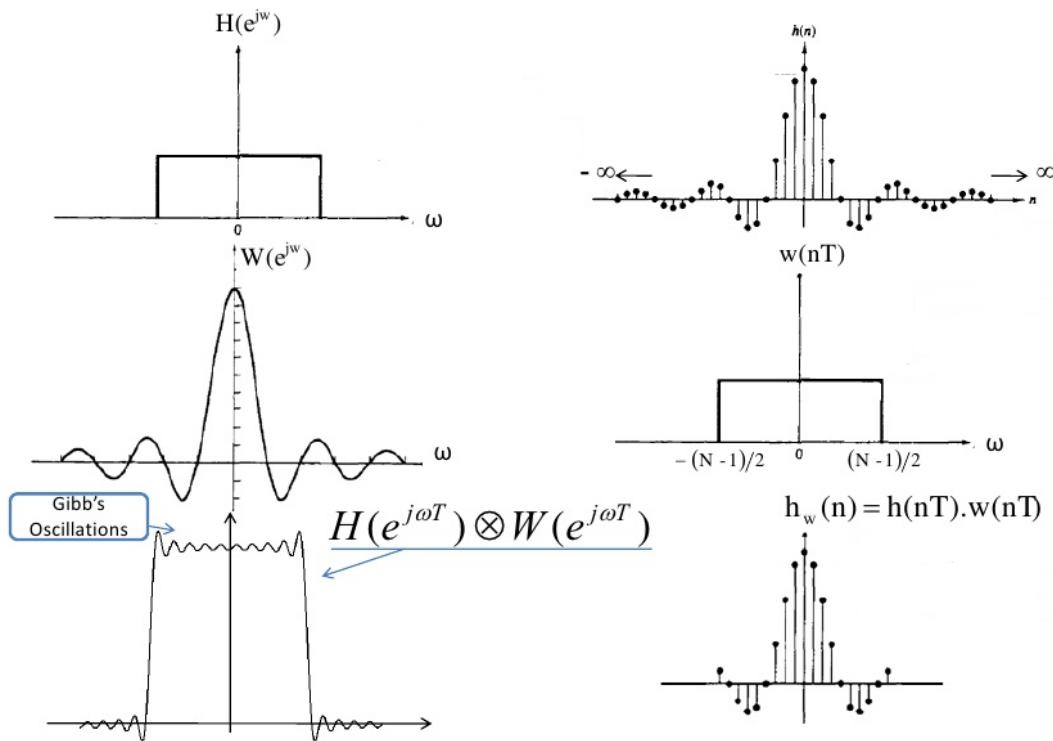
We kregen waarden op negatieve tijdsstippen waardoor we met een niet-causaal systeem zitten. Dit kunnen we voorkomen door te verschuiven naar rechts in de tijd. We verschuiven h_ω op over $\frac{N-1}{2}$ naar rechts.

Als we de beïnvloeding van deze verschuiving gaan bekijken op de weergave karakteristiek zien we dat er geen beïnvloeding is.

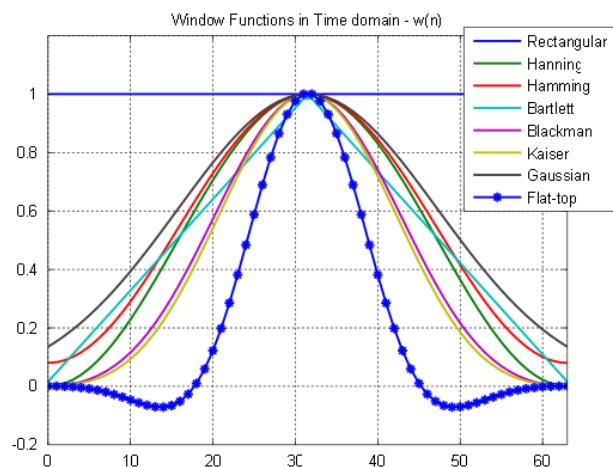
$$\mathcal{F}\left\{h_\omega\left(n - \frac{(N-1)}{2}\right)\right\} = e^{-j\frac{N-1}{2}\omega T} \mathcal{F}\{h_\omega(n)\}$$

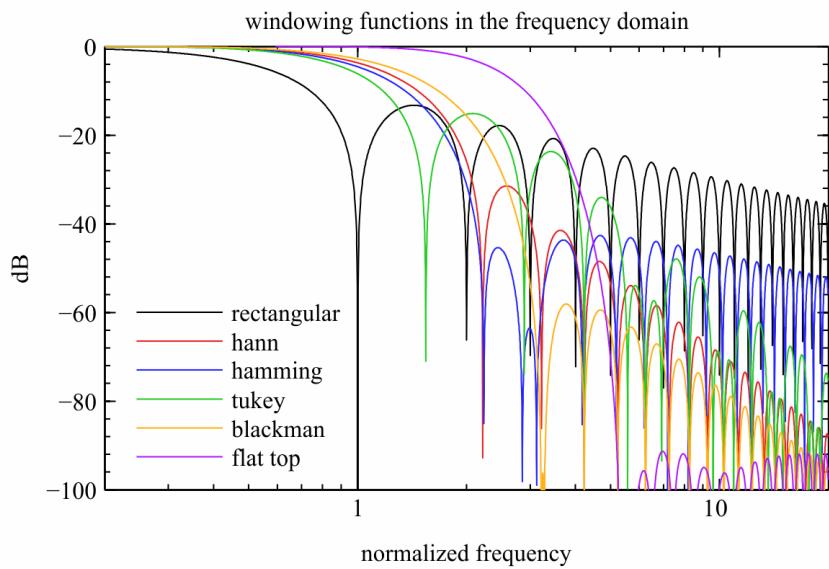


Frequentiedomein \leftrightarrow Tijdsdomein



De breedte van het venster heeft invloed op het aantal nuldoorgangen van de *sinc* functie. Hoe meer nuldoorgangen hoe smaller het transitiegebied. De verhouding van de hoofdlobe en zijn zijlobe is constant en onafhankelijk van de nuldoorgangen. De breedte heeft geen invloed op de rimpels.





Oplossingsmethode

We idealiseren onze LDF. Waarvan we de inverse fouriertransformatie berekenen. Hierbij krijgen we $h_i(n)$ met oneindig aantal waarden.

We verkiezen een gewenste vensterfunctie op basis van de vorm en breedte.

Berekenen van de FIR coëfficiënten door: $h(n) = h_i(n) \omega(n)$

We controleren of deze waarden perfect liggen binnen het niet-gearceerd gedeelte. Als dit niet het geval is berekenen we de fouriergetransformeerde van $h(n)$ om te kijken of het overgangsgebied of rimpel gebied niet OK is. Als het overgangsgebied niet OK is zullen we de breedte moeten aanpassen. Als de rimpel niet OK is zullen we de vorm moeten aanpassen, een nieuw venster kiezen. Na de aanpassingen aan het gekozen venster zullen we opnieuw de FIR coëfficiënten berekenen en opnieuw controleren, tot alles OK is.

Als alles OK is hebben we onze gewenste transferfunctie:

$$H(z) = \sum_{i=1}^N h(i) z^{-i}$$

De Discrete Fourier Transformatie

Discrete Fourier Reeks (DFR)

We beschouwen een periodische functie $x_{p(k)}$ met periode N, hierdoor:

$$x_p(k) = x_p(k + pN)$$

met p alle natuurlijke getallen.

$x_{p(k)}$ kan geschreven worden als:

$$x_p(k) = \frac{1}{N} \sum_{n=0}^{N-1} \underbrace{X_{p(n)}}_{DFR-coeff.} e^{\frac{j 2\pi kn}{N}} \equiv IDFS$$

$$X_p(n) = \sum_{k=0}^{N-1} x_p(k) \underbrace{e^{\frac{j 2\pi kn}{N}}}_{W_N^{nk}}$$

Definitie:

$$x(k) \begin{cases} x_p(k) & \text{voor } k = 0, 1, \dots, N-1 \\ 0 & \text{elders} \end{cases}$$

Hierdoor wordt de DTFT:

$$X(e^{j\theta}) = \sum_{k=0}^{N-1} x_p(k) e^{-jk\theta}$$

$$X_p(n) = X(e^{j\theta})|_{\theta=\frac{2\pi n}{N}}$$

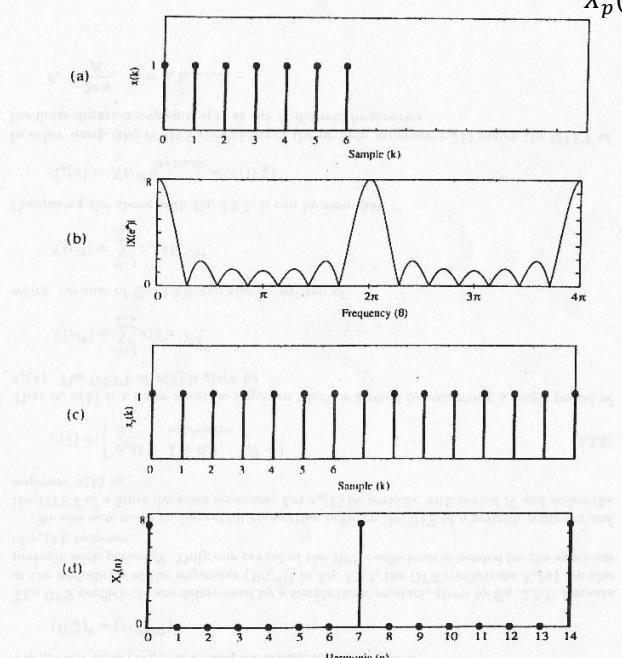


Figure 3.2.1: (a) Finite-duration sequence $x(k)$; (b) magnitude of $X(e^{j\theta})$; (c) periodically extended sequence $x_p(k)$; (d) magnitudes of DFS coefficients $X_p(n)$.

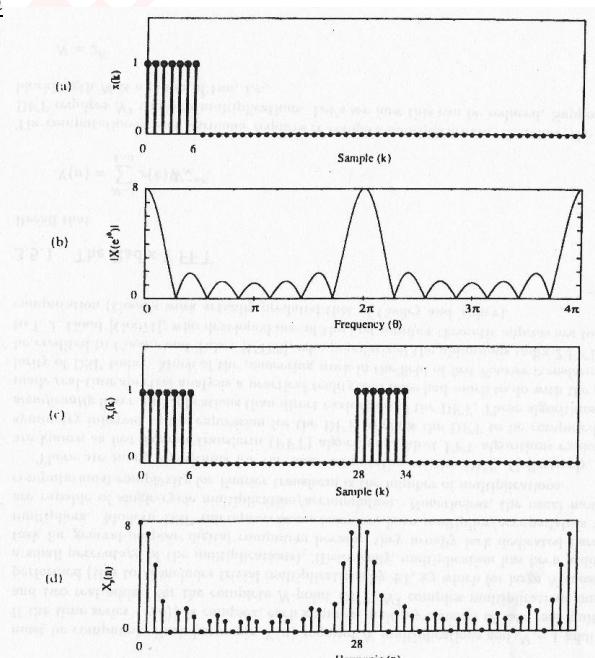


Figure 3.2.2: (a) Finite-duration sequence $x(k)$; (b) magnitude of $X(e^{j\theta})$; (c) periodically zero-padded sequence; (d) magnitude of DFS coefficients.

Fast Fourier Transform Algorithms

The Radix-2 FFT

$$X(n) = \sum_{k=0}^{N-1} x(k) W_N^{nk}$$

De berekening van elke harmonische vergt N complex multiplications, dus de hele DFT vergt N^2 complex multiplications.

We verdelen $x(k)$ in 2 subreeksen, één die bestaat uit oneven indices en &&n die bestaat uit even indices.

$$\{x(k)\} = \{x(0), x(1), \dots, x(N-1)\}$$

$$\{a\} = \{x(0), x(2), \dots\} \quad \& \quad \{b\} = \{x(1), x(3), \dots\}$$

Hieruit volgt:

$$\begin{aligned} a(k) &= x(2k) \text{ met } k = 0, 1, \dots, \frac{N}{2} - 1 \\ b(k) &= x(2k+1) \text{ met } k = 0, 1, \dots, \frac{N}{2} - 1 \end{aligned}$$

We steunen op $W_N^{-2} = e^{-j2\pi \frac{-N}{2}} = W_{\frac{N}{2}}$ en de voorgaande afleidingen om tot dit te komen:

$$X(n) = \sum_{k=0}^{N/2-1} a(k) W_{N/2}^{nk} + \underbrace{W_N^2}_{\text{komt door } (2k+1)} \sum_{k=0}^{N/2-1} b(k) W_{N/2}^{nk}$$

We verkijken dus een $N/2$ -punts discrete fouriertransformatie.

$$X(n) = A(n) + W_N^2 B(n)$$

De complexiteit van de multiplications wordt nu:

$$2 \left(\frac{N}{2}\right)^2 + N = \frac{N^2}{2} + N$$

We blijven reduceren tot men aan een 2 punt DFT komt.

$$Q(n) = \sum_{k=0}^1 q(k) W_2^{nk}$$

$$Q(n) = q(0) + q(1) W_2^n \text{ met } (n = 0, 1)$$

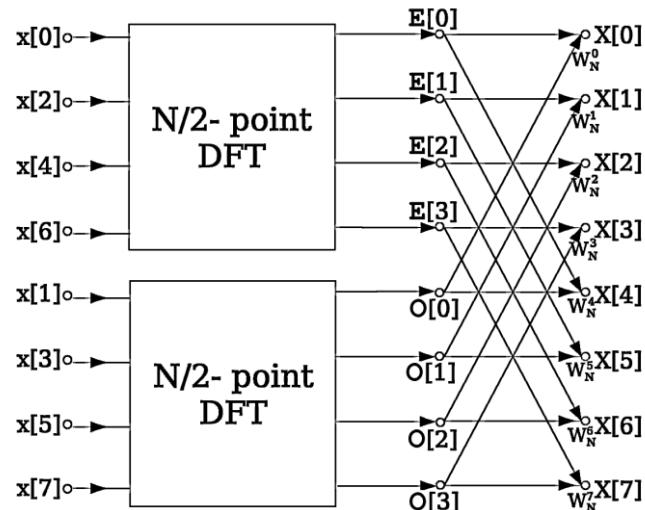
$$\begin{cases} Q(0) = q(0) + q(1) \\ Q(1) = q(0) + q(1) W_2 \end{cases}$$

Een 8-punts DFT zal dus 3 opeenvolgende stappen van elk 4 butterflies vergen.

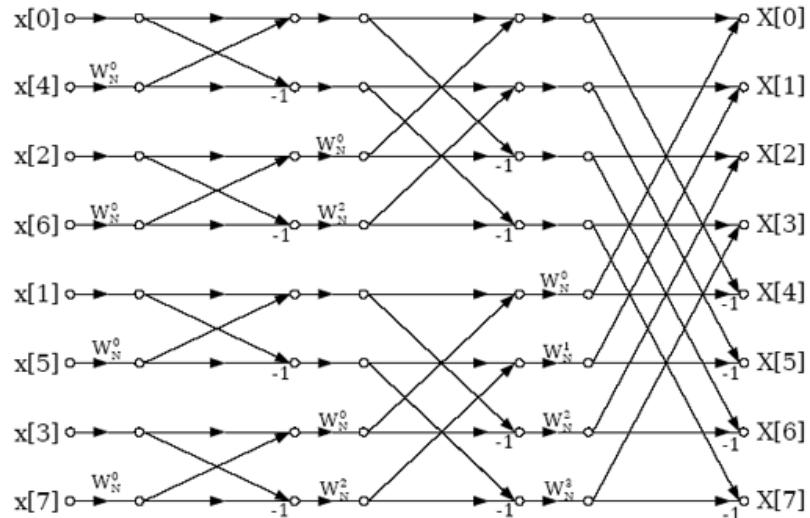
Een N-punts DFT zal $\log_2 N$ stappen met elke $\frac{N}{2}$ butterflies vergen.

Voorbeeld voor een 8-punt DFT:

Decomposing an eight-point FT into two four-point FDTS



Eight-point radix-2 decimation-in-time FFT



GILLEES
CALLEBAUT