Noter til signalanalyse

Jens Damgaard Andersen

12. januar 2004

Forord

Disse forelæsningsnoter til signalanalyse er overvejende baseret på tidligere kurser, hvor bogen "Signals and systems" [18] blev benyttet. Kursusdeltagerne mente dengang, at der var behov for kursusmateriale på dansk, og derfor blev noterne udarbejdet. En del af stoffet stammer fra [18], mens afsnittet om den kontinuerte deltafunktions egenskaber er inspireret af Lighthills bog om Fourieranalyse og generaliserede funktioner [12], ligesom kapitlet om den hurtige Fouriertransformation *(The Fast Fourier Transform)* stammer fra diverse kilder, bl.a. algoritmiklærebøger [7] og E. Orans Brighams bog om FFT og anvendelserne heraf [5].

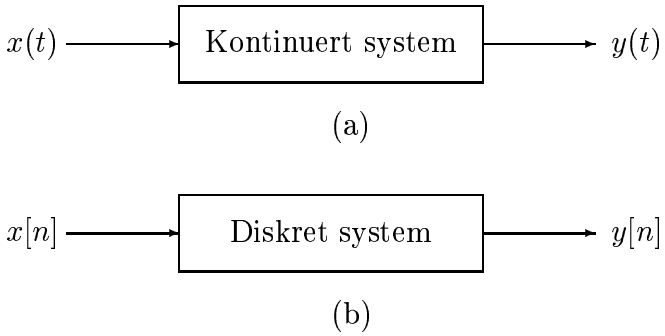
Siden noterne blev skrevet, er der sket en hel del på lærebogsfronten, især på den pædagogiske side. Inspireret af fremstillingen i to lærebøger af Steiglits [22, 23] skrev James H. McClellan, Ronald W. Schafer og Mark A. Yoder en elementær lærebog: *DSP First: A Multimedia Approach* [16], som benytter MATLAB til opgaver og øvelser og desuden indeholder en instruktiv CD-ROM. Denne bog har nu i nogle år været benyttet på det indledende kursus i multimedieteknologi på Datalogisk Instituts fuldtidsdatalogilinie. I 2003 udgav de samme forfattere en udvidet udgave af [16], som ud over diskrettidsteorien nu også omfatter kontinuerttids signalbehandling og Fouriertransformationen. Denne bog, *Signal Processing First* [17], udmærker sig ved at indeholde en hel del stof om de trigonometriske funktioner sinus og cosinus *("sinusoider")*, og en anskuelig fremstilling af sinusoiders generering som projektionen af en roterende vektor, en *fasor*. Dette stof bliver ofte forudsat bekendt på elektroingeniøruddannelsen før signalbehandlingskurset tages, og lærebøgerne udelader derfor teorien for trigonometriske funktioner m.v. I *Signal Processing First* starter man helt fra grunden, så bogen er velegnet som begynderbog. Jeg har derfor følt, at tiden er inde til at droppe noterne som undervisningsmateriale og bruge [17] i stedet. Da der dog mangler noget stof i [17], kan noterne bruges som supplerende materiale, især vedrørende dansk terminologi, teorien for kontinuerttids-deltafunktioner og endelig det sidste kapitel, der giver en anvendelse af signalbehandling, nemlig til JPEG-kompression, der efterhånden er vidt udbredt som billedkompressionsmetode i digitale kameraer og på Internettet.

J.D.A.

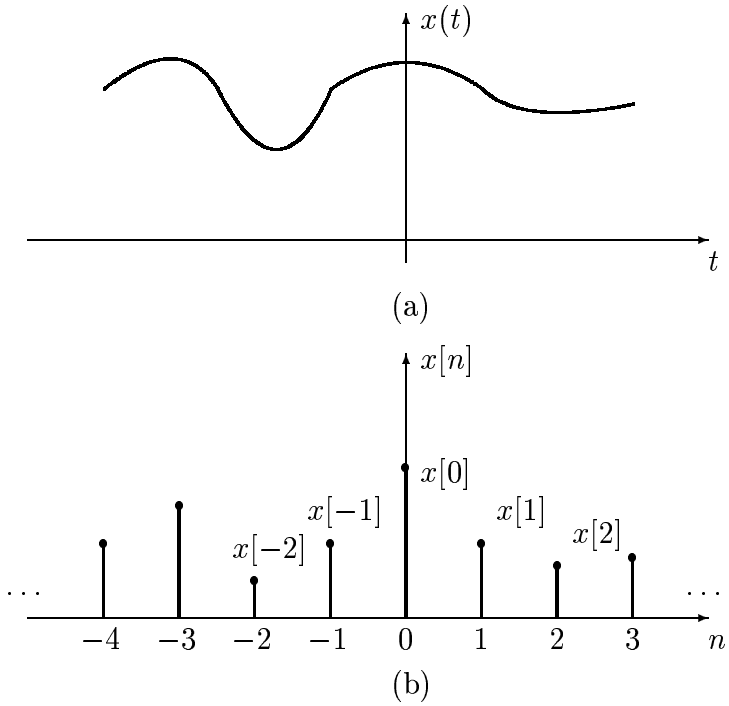
Kapitel 1

Lineære forskydningsinvariante systemer

Ved mange naturvidenskabelige og tekniske anvendelser ønsker man at kunne *simulere* et system på datamaskine, d.v.s. konstruere en *model* for systemet, således at man er i stand til at analysere systemets egenskaber og opførsel. Man vil således gerne kunne forudsige systemets *svar* på ydre påvirkninger. Dette illustreres ofte på diagramform ved at tegne systemet som en kasse ('a black box'), hvis indhold principielt er ukendt. Ikke desto mindre er det muligt at karakterisere systemets svar på ydre påvirkninger under visse antagelser om systemets opførsel. Påvirkningen på systemet betegnes systemets *input* eller indgangssignal, og systemets svar dets *output*, systemsvar eller udgangssignal. Dette er illustreret på figur 1.1(a), hvor $x(t)$ betegner input og $y(t)$ output. Denne måde at modellere systemer på finder udstrakt anvendelse indenfor en række fagområder. Man kan nævne kommunikation, datanet, musik- og taleanalyse, rumforskning, kredsløbsanalyse, akustik, seismisk analyse, medikoteknik, kemisk proceskontrol, billedanalyse og mange andre felter. Også indenfor helt andre områder end naturvidenskab som f.eks. ved finansiel analyse kan man benytte metoderne, bl.a. ved tidsrækkeanalyse af varierende størrelser som aktiekurser og andre økonomiske parametre.



Figur 1.1: (a) kontinuert system (b) diskret system

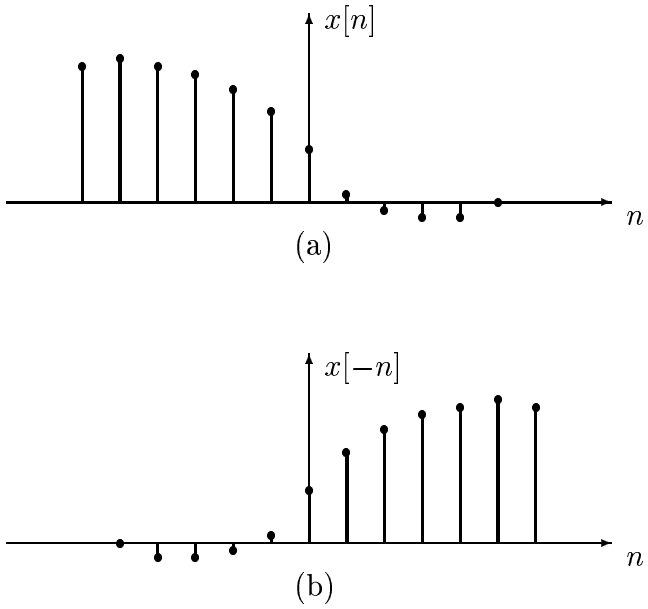


Figur 1.2: Grafisk fremstilling af (a) kontinuert og (b) diskret signal

1.1 Kontinuerte og diskrete systemer

Man skelner mellem kontinuerte og diskrete systemer. Kontinuerte systemer er karakteriseret ved at input og output er kontinuerte signaler, f.eks. en tidsvarierende spænding. En spænding som funktion af tiden kan være et resultat af en kontinuerlig måling, f.eks. af tryk, temperatur, lysintensitet pH, kraft og meget andet. Et kontinuert system kan så være et elektrisk kredsløb, der modtager det kontinuerte inputsignal og behandler det på en eller anden måde, f.eks. filtrerer det for at fjerne visse frekvenser og som output leverer det filtrerede kontinuerte signal i form af en tidsvarierende spænding.

Diskrete systemer opererer på diskrete signaler, som oftest dannet ved aftastning eller *sampling* af kontinuerte signaler. Dette gør det muligt at behandle dem ved hjælp af en datamaskine. Man tager stikprøver af kontinuerte signaler på en sådan måde at man undgår informationstab. Herved fås en række tal, en talsekvens, som kan bearbejdes med datamaskine. Afhængig af de egenskaber, man ønsker, kan man konstruere et program, der gennnemfører denne omdannelse eller transformation af talsekvensen. I teorien forestiller man sig ofte talsekvensen som uendelig lang, men i praksis kan man naturligvis kun arbejde med endelige talsekvenser. Omdannelsen fra kontinuerte signaler til diskrete signaler sker ved et elektronisk kredsløb, en *analog-til-digital omsætter* (A/D-omsætter). Man kan forstille sig en A/D-omsætter som et elektronisk voltmeter, der med passende tidsintervaller omsætter en spænding fra kontinuert (analog) form til digital form, dvs. en sekvens af tal, som kan behandles af datamaskinen. I praksis opererer man med en endelig ordlængde,



Figur 1.3: (a) Et diskret signal $x[n]$ og (b) dets spejling $x[-n]$ omkring aksen $n = 0$

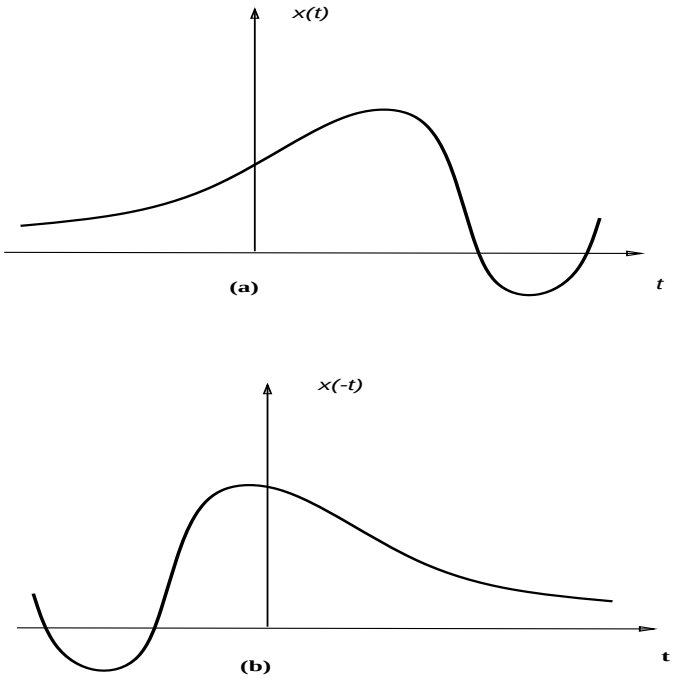
men i teorien forstiller man sig de enkelte sampleværdier som relle tal, og i teorien tager man ikke højde for den afrundingsfejl, der er resultatet af den endelige ordlængde. Man skelner derfor terminologimæssigt mellem *digitale signaler* (endelig ordlængde) og *diskrete signaler* (samplede signalværdier, der er reele tal).

En særlig klasse af systemer, de *lineære forskydningsinvariante systemer* **LFI-systemer** (engelsk: linear shift invariant, LSI-systems) er vigtig og meget ofte benyttet som model for systemer. Dette skyldes især

- Mange systemer, der optræder i virkeligheden, som f.eks. fysiske, kemiske, mekaniske og optiske systemer, kan med god nøjagtighed modelleres som lineære forskydningsinvariante systemer.
- Teorien for disse systemer er simpel i modsætning til teorien for ikke-lineære systemer.

En af de egenskaber ved lineære systemer, der gør dem lette at analysere, er gyldigheden af det såkaldte *superpositionsprincip*: kendes systemets svar på givne stimuli (påvirkninger) og benyttes som input en linearkombination af disse stimuli, vil systemets svar blive linearkombinationen af de tilhørende systemsvar.

I de næste afsnit vil teorien for lineære forskydningsinvariante systemer blive gennemgået. Men først vil vi betragte notationen for og nogle egenskaber ved *signaler*.



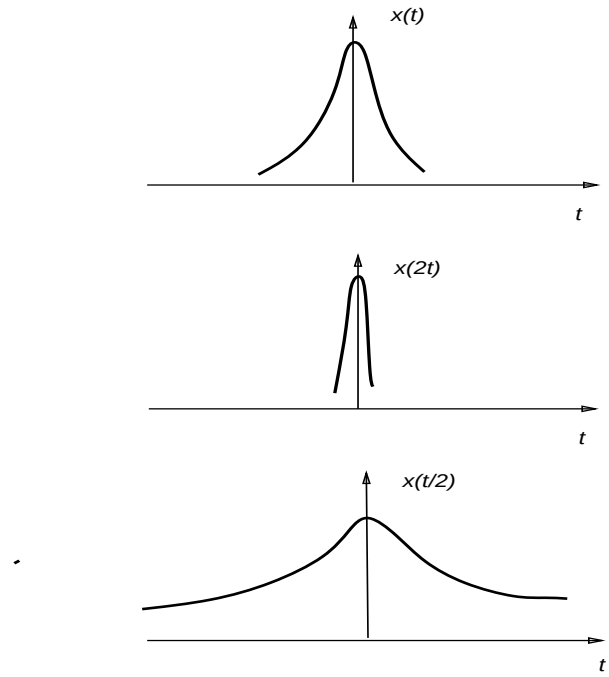
Figur 1.4: (a) Et kontinuert signal $x(t)$ og (b) dets spejling $x(-t)$ omkring aksen $t = 0$

1.2 Signaler

Betegnelsen *signal* bruges for en meget stor klasse af fysiske fænomener, der kan beskrives som en funktion med tiden eller afstand i rummet som den uafhængige variable. Eksempler på signaler er lydsignaler optaget ved hjælp af en mikrofon, signaler fra transducere, der overvåger en kemisk proces, eller den tidsmæssige fluktuation af en eller anden parameter, f.eks. vekselkursen på Euro i danske kr. Man opererer også med *flerdimensionale signaler*. Et eksempel på et to-dimensionalt signal er et monokromatisk billede, hvor forskydning i planen træder i stedet for tiden. Medens lydsignalet har en *temporal* parameter (tiden), har billedet en *spatial* (rumlig parameter). Man kan også have signaler med både temporale og spatiale parametre; et eksempel på et firedimensionalt signal med 3 spatiale og en temporal parameter er et signal bestående af en tredimensional magnetisk resonansoptagelse af et hjerte, der pumper blod.

Vi har tidligere omtalt, at der er to grundlæggende typer signaler, kontinuerte signaler, og diskrete signaler. For kontinuerte signalers vedkommende er den uafhængige variable kontinuert, og denne type signal er defineret for et kontinuum af værdier af den uafhængige variable. I modsætning hertil er diskrete signaler kun definerede for diskrete værdier af den uafhængige variable, f.eks. til diskrete tidspunkter eller i punkter i rummet. Den uafhængige variable antager kun et diskret sæt værdier. Eksempler på kontinuerttidssignaler er et talesignal og atmosfæretrykket ved DMI. Eksempler på diskrettidssignaler er Dow Jones aktieindex og det signal, der er indspillet på en CD.

For at skelne mellem kontinuerte signaler og diskrete signaler bruges bogstavet t til at betegne kontinuerttidsvariablen (de fleste kontinuerte systemer har tiden som uafhængig variabel) og n for den diskrete variable. Desuden vil vi skrive den kontinuerte variabel i



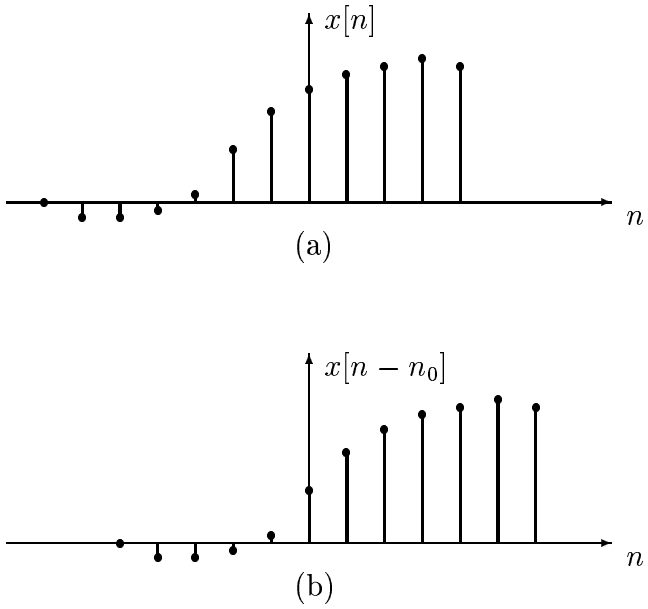
Figur 1.5: Kontinuerte signaler relaterede ved skalering

rund parentes $(\cdot)$ og den diskrete variabel i en kantet parentes $[\cdot]$ (hvilket minder om notationen for *et array*). Figur 1.2 (a) viser den grafiske fremstilling af et kontinuert signal x(t) og fig 1.2 (b) et diskret signal $x[n]$. Det er vigtigt at bemærke, at det diskrete signale $x[n]$ er *kun* defineret for heltallige værdier af den uafhængige variabel, hvilket den grafiske fremstilling for $x[n]$ også understreger. Desuden bruger man også ofte betegnelsen en diskret *følge* om $x[n]$, idet signalet kan opfattes som et række tal, en *talfølge*. Et diskret signal kan være fremkommet af et kontinuert signal ved *sampling* (stikprøvetagning), Dette er tilfældet for signalet indspillet på en Compact Disc: det er i virkeligheden et kontinuert-tidssignal (lydsignal), der er aftastet (samplet) med frekvensen 44,1 kHz, d.v.s. (stereo-) lydsignalet er aftastet 44100 gange i sekundet. Et eksempel på et todimensional samplet signal er et avisbillede: det er en rasteriseret udgave af et spatialt kontinuert signal, nemlig lysintensiteten, der rammer kameraets billedplan ved optagelsen.

I disse noter behandles kontinuerte og diskrete signaler hver for sig, men i parallel, således at man kan udnytte den indsigt, man har fået i den ene type ved forståelsen af den anden type. Der er nemlig meget store ligheder mellem kontinuerte og diskrete systemer og signaler. Desuden konvergerer de to teorier, når vi frembringer et diskret signal ved sampling af et kontinuert signal.

1.2.1 Notation og signaltransformationer

Signalanalyse er opstået som en ingeniørmæssig disciplin og notationen er beregnet på praktiske anvendelser og kan derfor forekomme at mangle en nødvendig matematisk strenghed. Et eksempel: $x[n]$ kan betyde et diskret signal (opfattet som en helhed. d.v.s. hele talfølgen) eller talfølgens værdi svarende til argumentværdien n. Desuden er det sædvane



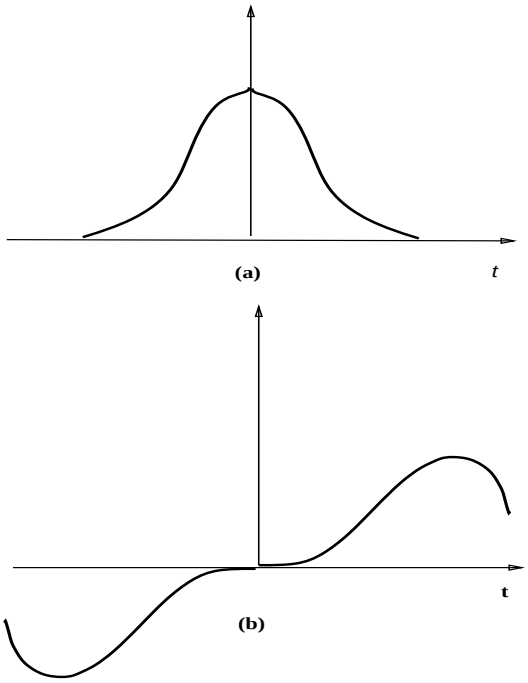
Figur 1.6: Diskretidssignaler relaterede ved tidsforskydning

at anvendelse af operationer på signaler, f.eks. tidsforskydning, spejling o.s.v., noteres i signalets argument, hvor det matematisk korrekte ville være anvendelsen af et operatorsymbol. Et eksempel: hvis signalet $h(t)$ forskydes stykket a i tid, noteres resultatet som $h(t-a)$. En mere rigoristisk notation ville være at specificere en translationsoperator τ med en værdi svarende til translationen a, d.v.s.

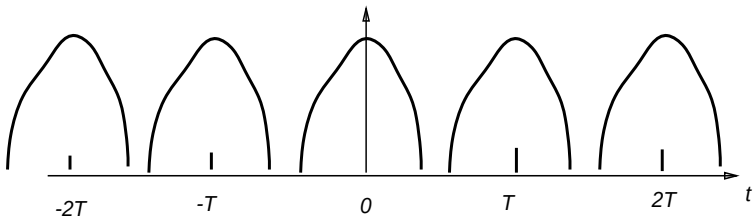
$$\tau(a)h(= h(t-a)). \tag{1.1}$$

Hvis τ betegner den generelle translationsoperator, så er $\tau(a)$ operatoren, der translaterer stykket a. At signalet x translateres stykket a kan så skrives $\tau(a)(x)$ og funktionsværdien af den a-translaterede x-funktion taget til tiden t_0 kan skrives $\tau(a)(x)(t_0)$ Denne notation er imidlertid besværlig at bruge og anvendes ikke i praksis. Af hensyn til dette kursus 'kompatibilitet' med den eksisterende litteratur vil der også her blive brugt den mest udbredte (men mindre strenge) notation. Man kan da være nødt til ud fra sammenhængen at afgøre, f.eks. om der menes signalet som helhed eller en specifik funktionsværdi.

I mange situationer er det vigtigt at betragte signaler, der er relaterede ved en modifikation af den uafhængige variable. F. eks. er signalet $x[-n]$ vist på figur 1.3, fremkommet ved at spejle signalet $x[n]$ omkring aksen $n = 0$. Man kan opfatte $x[-n]$ som givet ved den sammensatte funktion $x[-1 \cdot [n]] = x[z[n]]$. Tilsvarende som vist på figur 1.4 kan man få signalet $x(-t)$ ved spejling af $x(t)$ omkring aksen $t = 0$. Hvis $x(t)$ repræsenterer et lydsignal på et bånd, svarer $x(-t)$ til det signal, man får, hvis man afspiller båndet baglæns. Fig. 1.5 viser tre signaler, $x(t), x(2t)$ og $x(t/2)$ som svarer til hinanden ved en lineær skalaændring i den uafhængige variable. Hvis vi tænker på et båndindspillet signal,



Figur 1.7: (a) Et lige kontinuert signal (b) et ulige kontinuert signal



Figur 1.8: Kontinuert periodisk signal

svarer $x(2t)$ til det, man hører, hvis båndet afspilles med dobbelt hastighed, og $x(t/2)$ til afspilning med halv hastighed.

Endnu et eksempel på en transformation af den uafhængige variabel er vist på figur 1.6, der viser to signaler, $x[n]$ og $x[n-n_0]$ som er identiske i form, men hvor det ene signal er en tidsforskudt udgave af det andet signal. Tilsvarende repræsenterer $x(t-t_0)$ en tidsforskudt udgave af $x(t)$. Signaler, der er indbyrdes relaterede ved tidsforskydning, ses ofte i fysiske systemer, f.eks. ved sonar, radar og seismiske undersøgelser, hvor et modtaget reflekteret signal er en forsinket udgave af det udsendte signal, idet forsinkelsen fremkommer ved at signalet propagerer (udbredes) gennem et medium (jord, luft, vand). Transformationer af signaler kan også foretages med henblik på at undersøge, hvordan systemer reagerer på den transformede udgave af signalet.

Et signal kaldes *lige*, hvis det er identisk med sin spejling omkring den lodrette akse, d.v.s. (for et kontinuert signal):

$$x(-t) = x(t)$$

og for et diskret signal

$$x[-n] = x[n].$$

Et signal kaldes *ulige* hvis

$$\begin{aligned} x(-t) &= -x(t) \\ x[-n] &= -x[n], \end{aligned}$$

hvilket svarer til, at spejling omkring koordinatsystemets begyndelsespunkt ikke ændrer signalet.

Det er vigtigt at bemærke, at ethvert signal kan dekomponeres (opsplittes) i en lige og ulige komponent (del). Dette ses ved at danne

$$\mathcal{E}[x(t)] = \frac{1}{2}[x(t) + x(-t)]$$

hvilket kaldes $x(t)$'s *lige del*. Tilsvarende er $x(t)$'s *ulige del* givet ved

$$\mathcal{O}[x(t)] = \frac{1}{2}[x(t) - x(-t)]$$

Det er enkelt at se, at 'den lige del' faktisk er lige, 'den ulige del' er ulige, og at $x(t)$ er summen af de to. En nøjagtig tilsvarende definition findes for diskrete signaler og et eksempel på dekomposition af et diskret signal er vist i figur 1.7.

Endnu en klasse af signaler, man bør kende, er de *periodiske* signaler, både kontinuerte og diskrete. Et periodisk kontinuert signal $x(t)$ har den egenskab, at der findes en positiv værdi af T for hvilken

$$x(t) = x(t + T) \text{ for alle t} \tag{1.2}$$

I dette tilfælde siges $x(t)$ er periodisk med perioden T. Et eksempel på et sådant signal er vist på fig 1.8. Fra figuren eller fra ligning (1.2) kan man slutte, at dersom $x(t)$ er periodisk med perioden T, så er $x(t) = x(t + mT)$ for alle t og ethvert heltal m. $x(t)$ er derfor også periodisk med perioden $2T, 3T, 4T, \ldots$. Grundperioden (fundamentalperioden) T_0 for $x(t)$ er den mindste positive værdi af T for hvilken (1.2) gælder. Bemærk, at denne definition af grundperioden forudsætter, at $x(t)$ ikke er konstant. I dette tilfælde er grundperioden udefineret, da $x(t)$ er periodisk for ethvert valg af T, således at der ikke er en mindste værdi. Endelig benævnes et signal, der ikke er periodisk, et *aperiodisk* signal.

Periodiske diskrete signaler defineres på samme måde. Mere præcist, så er et diskret signal $x[n]$ periodisk med perioden N, hvor N er et positivt heltal, hvis

$$x[n] = x[n + N] \text{ for alle } n \tag{1.3}$$

Hvis ligning (1.3) er opfyldt, så er $x[n]$ også periodisk med perioden $2N, 3N, \ldots$, og *grundperioden* N_0 er den mindste positive værdi af N for hvilken ligning(1.3) er opfyldt.

1.3 Systemers egenskaber

Når man opstiller en model for et system skal det karakteriseres ved en række egenskaber, som kan fortolkes både matematisk og fysisk. De vigtigste er: med eller uden indre tilstande, (memory), invertibilitet, kausalitet, stabilitet, tidsinvarians, og, som den vigtigste, linearitet. Når det i det følgende tales om 'systemer' menes hermed 'modeller for systemer'.

Indre tilstande. Et system er uden indre tilstande, hvis dets output udelukkende afhænger af det øjeblikkelige input og er uafhængig af forhistorien. Hvis systemets output derimod er et resultat af det øjeblikkelige input i kombination med tidligere input, siges systemet at 'have indre tilstande'. Indenfor datamatarkitektur svarer den første type til et kombinatorisk netværk og den sidste til en tilstandsmaskine. Betragter man et kontinuert system, f.eks. et analogt elektrisk kredsløb, så er et system, der udelukkende indeholder modstande, af den første type. Et system, der indeholder kondensatorer, kan levere output, der er afhængig af tidligere input (fordi kondensatorer kan lades op og senere afgive denne ladning) og er af den sidste type. Et særligt simpelt tilstandsløst system er *identitetssystemet* beskrevet ved input-output relationen

$$y(t) = x(t)$$

og det tilsvarende diskrete system

$$y[n] = x[n]$$

Et eksempel på et system med indre tilstande er systemet beskrevet ved

$$y[n] = \sum_{k=-\infty}^{n} x[k] \tag{1.4}$$

Dette system indeholder for en given indeksværdi n summen af alle inputværdier fra $-\infty$ til det aktuelle indeks. Et andet eksempel er

$$y(t) = x(t-1),$$

der et system, der forsinker input én tidsenhed.

Invertibilitet og inverse systemer. Et system siges at være invertibelt, hvis man kan bestemme dets input ud fra observation af output. D.v.s. at det er muligt at konstruere et inverst system, som, hvis det sættes i serie med det oprindelige system, giver et output $z[n]$ som er lig det oprindelige input $x[n]$. Et eksempel på et invertibelt kontinuert system er

$$y(t) = 2x(t)$$

for hvilket det inverse system er

$$y(t) = \frac{1}{2}x(t)$$

Et andet eksempel er det diskrete system defineret i ligning 1.4. For dette systems vedkommende giver forskellen mellem to på hinanden følgende outputværdier netop den sidste inputværdi. Følgelig er det inverse system i dette tilfælde

$$z[n] = y[n] - y[n-1]$$

Eksempler på ikke-invertible systemer er

$$y[n] = 0$$

(det system der giver nul for enhvert inputsignal) og

$$y(t) = x^2(t)$$

Kausalitet. Et system siges at være *kausalt*, hvis output til ethvert tidspunkt kun afhænger af det nuværende og tidligere input. Systemet kaldes også 'ikke-anticiperende', fordi det ikke kan foregribe fremtidige inputværdier (skue ud i fremtiden). Selv om kausale systemer er af stor betydning, kan man også komme ud for at skulle operere med ikke-kausale systemer. Et eksempel er behandling af data, som ikke kommer direkte ind i et system, men som er lagret på et pladelager, således at man kan udnytte inputværdier efter et tidspunkt t_0 eller $t[n]$ ved beregningen af output til dette tidspunkt. Ligeledes kan man komme ud for nonkausale systemer ved billedbehandling, hvor den uafhængige variable ikke er tiden, men rumkoordinater. Et fysisk system, hvor den uafhægige variable er tiden, modelleres derimod almindeligvis kausalt. Det giver ingen mening at modellere en bils bevægelse son et nonkausalt system, idet bilen jo ikke kan forudse førerens fremtidige handlinger.

Stabilitet. Et system er *stabilt*, hvis output er begrænset, når input er begrænset. Betragt f.eks. systemet

$$y[n] = \frac{1}{2M+1} \sum_{k=-M}^{k=+M} x[n-k] \tag{1.5}$$

Antag, at input er begrænset i størrelse til f.eks. K. Det er da let at se, at $y[n]$ da aldrig kommer over K, fordi $y[n]$ er middelværdien af et endeligt antal inputværdier. Derfor er $y[n]$ også begrænset og systemet er dermed stabilt. Denne stabilitetsbetingelse kaldes også 'Bounded-Input Bounded-output' (BIBO) stabilitetsreglen. Et eksempel på et system, der ikke er stabilt, er systemet defineret ved ligning 1.4. Dette system summerer alle inputværdier og vil divergere på begrænsede inputsekvenser.

Forskydningsinvarians. Et system er *forskydningsinvariant*, hvis en forskydning af inputsignalet giver anledning til en tilsvarende forskydning af output. For et system, hvor input og output er givet som tidsfunktioner, er systemet *tidsinvariant*, hvis en tidsforskydning af inputsignalet forårsager en tilsvarende forskydning af output. For eksempel, for et diskret system, hvis $y[n]$ er systemets svar på $x[n]$, vil systemet svare med $y[n-a]$ på $x[n-a]$, En tidsforskydning af inputsignalet modsvares med andre ord af en tilsvarende forskydning af outputsignalet. Et eksempel på et system, der ikke er tidsinvariant, er $y[n] = nx[n]$. Dette systems svar afhænger af tiden (index n).

Linearitet. Et lineært system (kontinuert eller diskret) er et system, som virker som et superpositionssystem: Hvis input til systemet består af en vægtet sum af flere signaler, så er output lig superpositionen, d.v.s. den samme vægtede sum af systemets svar på hvert enkelt signal. Mere præcist, hvis $y_1(t)$ er et kontinuert systems svar på $x_1(t)$ og tilsvarende $y_2(t)$ systemets svar på $x_2(t)$, så er systemet lineært, hvis

1. Svaret på $x_1(t) + x_2(t)$ er $y_1(t) + y_2(t)$.
2. Svaret på $ax_1(t)$ er $ay_1(t)$, hvor a er en vilkårlig kompleks konstant.

Den første af de to egenskaber kaldes det lineære systems *additivitetsegenskab*. Den anden kaldes *skalerings-* eller *homogenitetsegenskaben*. Skønt definitionen her er givet for kontinuerttidssystemer, gælder en tilsvarende for diskrete systemer. Bemærk, at et system kan være lineært uden at være tidsinvariant; det er f.eks. tilfældet for systemet $y[n] = nx[n]$, og det kan være tidsinvariant uden at være lineært, som $y(t) = sin[x(t)]$ og $y(t) = x^2(t)$. De to egenskaber, der definerer et lineært system, kan kombineres til én samlet egenskab, der f.eks. for diskrete systemer lyder:

$$ax_1[n] + bx_2[n] \mapsto ay_1[n] + by_2[n]$$

hvor pilen $\mapsto$ læses "giver outputtet" og hvor a og b er vilkårlige komplekse konstanter. Tilsvarende for et vilkårligt antal superponerede signaler

$$x[n] = \sum_k a_k x_k[n] \mapsto y[n] = \sum_k a_k y_k[n]$$

Lineære systemer har en anden vigtig egenskab, nemlig at *nulinput* giver *nuloutput*. For eksempel, hvis $x[n] \mapsto y[n]$ så giver skaleringsegenskaben

$$0 = 0 \cdot x[n] \mapsto 0 \cdot y[n] = 0$$

Systemet

$$y[n] = 2x[n] + 3 \tag{1.6}$$

er ikke lineært, idet $y[n] = 3$ for $x[n] = 0$. Dette kan synes overraskende, da ligning (1.6) er en lineær ligning, men systemet overholder ikke nul-ind/nul-ud egenskaben for lineære systemer.

1.4 Signaldekomposition og foldning

Man kan benytte superpositionsprincippet til at finde et vilkårligt LFI-systems svar på et givet indgangssignal. Metoden er, at opsplitte (dekomponere) inputsignalet i en række elementarsignaler af en sådan karakter, at man let kan bestemme systemets svar på dem. Et vigtigt elementarsignal er deltafunktionen δ. Først betragtes deltafunktionen $\delta(t)$ for kontinuerte systemer.

Kontinuerte systemer: deltafunktionen. Dirac's 'delta funktion' $\delta(t)$ er egentlig ikke en funktion, men tilhører klassen af 'generaliserede funktioner', som har egenskaben

$$\int_{-\infty}^{\infty} \delta(t)x(t)dt = x(0) \tag{1.7}$$

for ethvert passende kontinuert signal $x(t)$. Ingen sædvanlig funktion har egenskaben (1.7), men man kan forestille sig en følge af funktioner, som har stadig højere og smallere toppe for $t = 0$, medens arealet under kurven forbliver 1, samtidig med at funktionens værdi går mod 0 i ethvert punkt, undtagen for $t = 0$, hvor den går mod uendelig. I grænsen vil denne funktionsfølge da have de omtalte egenskaber.

Et eksempel på en funktionsfølge med disse egenskaber er $\delta_n(t) = \sqrt{\frac{n}{\pi}}e^{-nt^2}$, der definerer en generaliseret funktion $\delta(t)$ således at

$$\int_{-\infty}^{\infty} \delta(t)x(t)dt = x(0) \tag{1.8}$$

Figur 1.9 viser eksempler på funktioner i funktionsfølgen, der bruges til at definere $\delta(t)$.

Det er også muligt at 'differentiere' $\delta(t)$, hvorved man får en funktion $\delta'(t)$ med egenskaben

$$\int_{-\infty}^{\infty} \delta'(t)x(t)dt = -\int_{-\infty}^{\infty} \delta(t)x'(t)dt = -x'(0) \tag{1.9}$$

for enhver kontinuert differentiabel funktion x(t). Opførsel som (1.9) kan som før fås som grænsen af en funktionsfølge. For eksempel de afledte funktioner fra den følge, der bruges til at repræsentere $\delta(t)$; disse er grafisk vist på figur 1.10

Man kan tilsvarende integrere $\delta(t)$. Idet det løbende integral (det bestemte integral, hvis øvre grænse er t) betegnes $u(t)$ finder man:

$$u(t) = \int_{-\infty}^{t} \delta(\tau)d\tau, \tag{1.10}$$

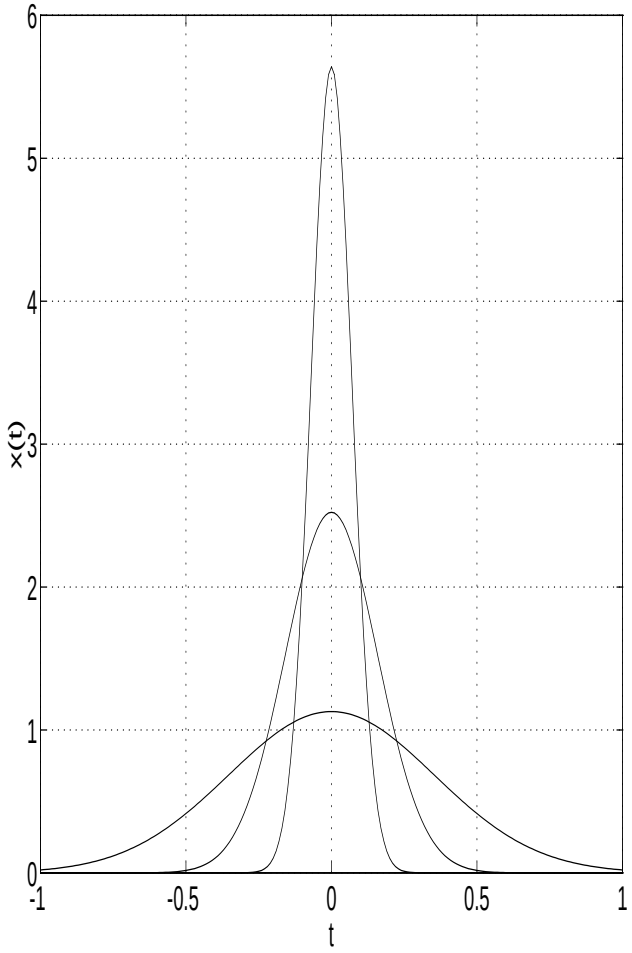
der betegnes *Heavisides enhedstrinfunktion* (the unit step function) givet ved

$$u(t) = \begin{cases} 0, & t < 0 \\ \text{udefineret}\ , & t = 0 \\ 1, & t > 0 \end{cases} \tag{1.11}$$

Bemærk, at der er en diskontinuitet for $t = 0$. Igen vil vi her betragte enhedstrinfunktionen som resultatet af en grænseovergang for $n \mapsto \infty$ ved

$$u(t) = \lim_{n\mapsto\infty} u_n(t) = \lim_{n\mapsto\infty} \int_{-\infty}^{t} \delta_n(\tau)d\tau, \tag{1.12}$$

$u_n(t)$ er vist på figur 1.11 for $n = 4, 20$ og 100. Hvis man ønsker en matematisk rigoristisk fremstilling af teorien for deltafunktioner, henvises til: M.J. Lighthill: *Fourier Analysis and Generalised Functions*, Cambridge University Press 1958.



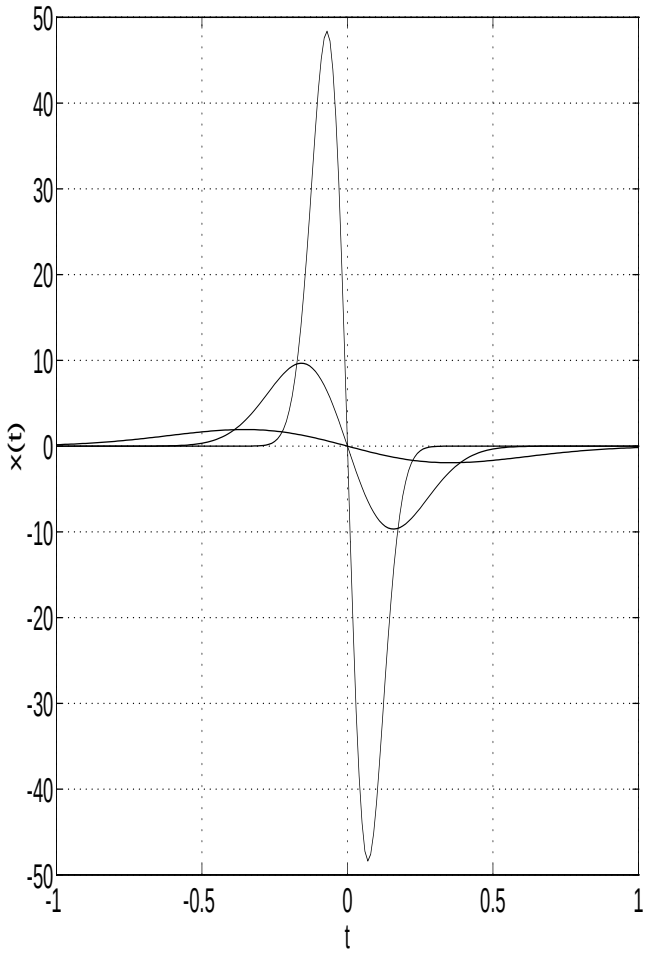
Figur 1.9: Funktioner i funktionsfølgen $\delta_n(t)$, der kan bruges til at definere deltafunktionen $\delta(t)$ med vist for $n = 100, 20$ og 4. Jo større n, desto smallere og højere impuls.

Da arealet af impulsen $\delta(\tau)$ er koncentreret ved $\tau = 0$, er det løbende integral 0 for $t < 0$ og 1 for $t > 0$. Desuden kan sammenhængen i ligning (1.10) mellem enhedsimpulsen og enhedstrinnet skrives på en anden form ved variabelsubstitutionen $\sigma = t - \tau$:

$$u(t) = \int_{-\infty}^{t} \delta(\tau)d\tau = \int_{\infty}^{0} \delta(t-\sigma)(-d\sigma) = \int_{0}^{\infty} \delta(t-\sigma)d\sigma \tag{1.13}$$

I dette tilfælde er arealet af $\delta(\tau - \sigma)$ koncentreret i punktet $\sigma = t$, således at integralet som før er 0 for $t < 0$ og 1 for $t > 0$.

Skønt denne gennemgang af enhedsimpulsen er uformel, er den tilstrækkelig for vort nuværende formål og tjener til at give intuition vedrørende dette signals egenskaber. For eksempel er det vigtigt at betragte produktet af en impulsfunktion og en 'pæn' kontinuerttidsfunktion. Fortolkningen af produktet kan baseres på definitionen af $\delta(t)$ som



Figur 1.10: Funktionsfølgen $\delta_n'(t)$ (de afledte af $\delta n(t)$) for $n = 100, 20$, og 4

$\delta(t) = \lim_{n\mapsto\infty} \delta_n(t)$, d.v.s. man betragter $x_1(t)$ givet ved

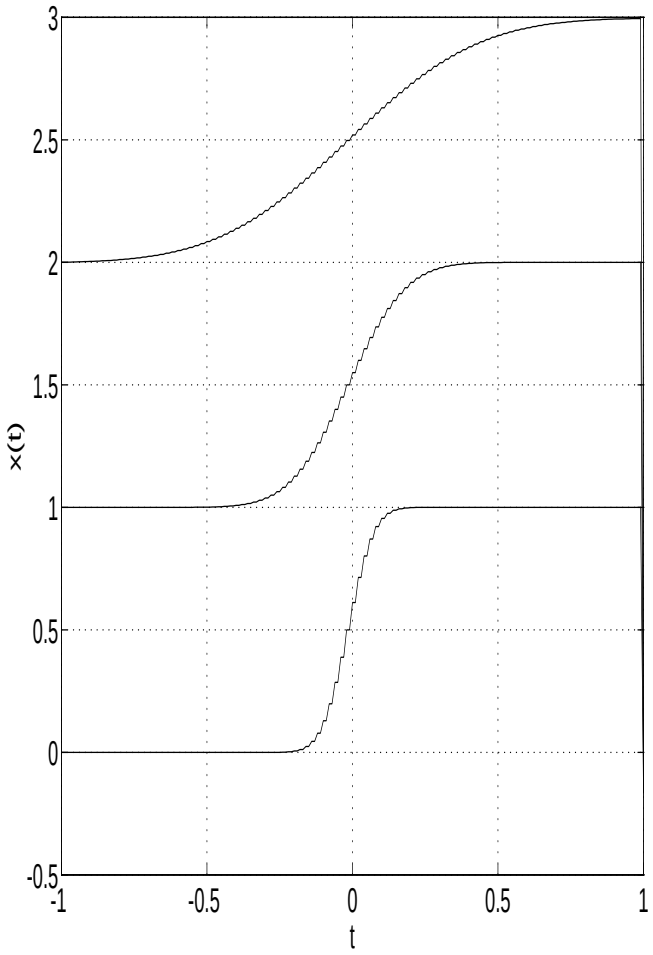
$$x_1(t) = x(t)\delta_n(t) \tag{1.14}$$

For et tilstrækkeligt stort n, således at $x(t)$ tilnærmelsesvis er konstant over det interval, hvor $\delta_n(t)$ er 'væsentlig forskellig fra nul', er

$$x(t)\delta_n(t) \sim x(0)\delta_n(t) \tag{1.15}$$

$\delta(t)$ er grænseværdien for $\delta_n(t)$ når $n \mapsto \infty$. I grænsen er impulsen 'uendelig smal' og 'uendelig høj', men bevarer arealet 1. Et integral af produktet af $\delta(t)$ og en funktion $x(t)$, vil, når integrationen inkluderer $t = 0$, give funktionsværdien af $x(t)$ for $t = 0$, d.v.s.

$$\int_{-\infty}^{\infty} x(t)\delta(t)dt = x(0). \tag{1.16}$$



Figur 1.11: Det løbende integral $u_n(t)$ af funktionsfølgen $\delta_n(t)$ for $n = 4, n = 20$ og $n = 100$ vist ovenfra og nedefter. Bemærk, at u_n nærmer sig enhedstrinfunktionen for $n \mapsto \infty$

Deltafunktionens egenskaber kan illustreres ved en udregning af integralet af $\delta_n(t)$ og af $\delta_n(t)\cos(t)$. I tabel 1.1 er vist en beregning af disse integraler udregnet fra -1 til 1 med en inddeling på 200 punkter ved hjælp af Simson's formel. Man ser, at for $n = 100$ nærmer $\int \cos(t)\delta_n(t)dt$ sig til den sande værdi 1.

Et tilsvarende udtryk fås for en impuls koncentreret i et vilkårligt andet punkt, f.eks. t_0. D.v.s.

$$\int_a^b x(t)\delta(t - t_0)dt = x(t_0), \tag{1.17}$$

såfremt integrationen strækker sig hen over $t = t_0$. Et numerisk eksempel udregnet ved hjælp af tilnærmelsen $\delta_{100}(t)$: $\cos(0.5) = 0.8776$. Ved integration findes $\int \cos(t)\delta_{100}(t - 0.5)dt = 0.8754$.

Man har altså

$$x(t) = \int_{-\infty}^{\infty} x(\tau)\delta(t - \tau)d\tau. \tag{1.18}$$

	$\int_{-1}^{1} \delta_n(t)dt$	$\int_{-1}^{1} \cos(t) \cdot \delta_n(t)dt$
$n = 4$	0.9953	0.9373
$n = 20$	1.0000	0.9876
$n = 100$	1.0000	0.9975

Tabel 1.1: Illustration af integration af produkt af en tilnærmelse til deltafunktionen og en differentiabel funktion ($\cos(t)$)

Enhedsimpulsen i positionen $\tau = t$ har derfor egenskaben, at den under integration plukker funktionsværdien af $x(\tau)$ ud for $\tau = t$. Ligning (1.18) definerer det, der kaldes kontinuerttids deltafunktionens *si-egenskab* (engl.: sifting property).

Man kan også intuitivt tænke på ligning (1.18) som definerende $x(t)$ som en "sum" af vægtede forskudte impulser, hvor impulsen $\delta(t - \tau)$ har vægten $x(\tau)d\tau$.

I almindelighed, hvis et systems svar på en deltaimpuls til tidspunktet τ er $h_\tau(t)$ som skal vægtes med inputsignalet til tiden t, $x(t)$, er systemets svar

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h_\tau(t)d\tau. \tag{1.19}$$

Med denne fortolkning repræsenterer ligning (1.19) simpelthen superpositionen af svarene på hver af disse input, og, på grund af lineariteten er vægten af svaret på den forskudte impuls $\delta(t - \tau)$ $x(\tau)d\tau$. Ligningen (1.19) repræsenterer den generelle form for et lineært kontinuerttidssystems svar. Hvis systemet ud over at være lineært også er tidsinvariant (eller i almindelighed, forskydningsinvariant) så er $h_\tau(t) = h_0(t - \tau)$, og hvis index 0 bortkastes og vi definerer systemets *enhedsimpulssvar* som

$$h(t) = h_0(t) \tag{1.20}$$

bliver ligning (1.19)

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t - \tau)d\tau. \tag{1.21}$$

Denne vigtige ligning benævnes *foldningsintegralet* eller *superpositionsintegralet* og definerer et kontinuerttidssystems svar ud fra dets svar på en enhedsimpuls. Foldningen af de to signaler $x(t)$ og $h(t)$ skrives symbolsk således

$$y(t) = x * h(t) \tag{1.22}$$

Diskrete systemer: deltafunktionen. Analogien til trinfunktionen i kontinuerte systemer er det diskrete enhedstrinnet, der betegnes $u[n]$ og defineres ved

$$u[n] = \begin{cases} 0, & n < 0 \\ 1, & n \geq 0 \end{cases} \tag{1.23}$$

Et andet vigtigt diskret signal er *enhedsimpulsen* defineret ved

$$\delta[n] = \begin{cases} 0, & n \neq 0 \\ 1, & n = 0 \end{cases} \tag{1.24}$$

Den diskrete enhedsimpulse har mange egenskaber, der ligner egenskaber ved den kontinuerte enhedsimpuls. For eksempel idet $\delta[n]$ kun er forskellig fra nul (og lig 1) for $n = 0$. Det ses umiddelbart, at

$$x[n]\delta[n] = x[0]\delta[n] \tag{1.25}$$

Ligesom den kontinuerte impuls er den første afledede af det kontinuerte enhedstrinnet, så er den diskrete impulsen *førstedifferensen* af det diskrete trin

$$\delta[n] = u[n] - u[n-1] \tag{1.26}$$

Svarende til, at det kontinuerte enhedstrin er det løbende integral af $\delta(t)$, er det diskrete enhedstrin den løbende sum af enhedsimpulsen:

$$u[n] = \sum_{m=-\infty}^{n} \delta[m] \tag{1.27}$$

Foldningssummen. Hvis man betragter et diskret lineært system og et vilkårligt input $x[n]$ til systemet, kan man udtrykke $x[n]$ som en linearkombination af forskudte diskrete enhedsimpulser:

$$x[n] = \sum_{k=-\infty}^{\infty} x[k]\delta[n-k] \tag{1.28}$$

Ved at anvende lineære systemers superpositionsegenskab kan output $y[n]$ skrives som en linearkombination af systemets svar på forskudte enhedsimpulser. Hvis h_k er systemets svar på den forskudte enhedsimpuls $\delta[n-k]$, kan systemts svar på en vilkårlig inputsekvens udtrykkes som

$$y[n] = \sum_{k=-\infty}^{\infty} x[k]h_k[n] \tag{1.29}$$

Hvis man således kender et lineært systems svar på en mængde af forskudte enhedsimpulser, kan man konstruere svaret på et vilkårligt input. I almindelighed behøver svarene $h_k[n]$ ikke at være indbyrdes relaterede for forskellige k-værdier. Hvis systemet imidlertid også er forskydningsinvariant, så er

$$h_k[n] = h_0[n-k] \tag{1.30}$$

Da $\delta[n-k]$ er en forskudt udgave af $\delta[n]$, er svaret $h_k[n]$ en forskudt udgave af $h_0[n]$. Af bekvemmelighedsgrunde bortkastes index 0, og man definerer systemets *enhedsimpulssvar* $h[n]$ som

$$h[n] = h_0[n] \tag{1.31}$$

(d.v.s. $\delta[n] \mapsto h[n]$). For et LSI system bliver 1.29 derfor

$$y[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k] \tag{1.32}$$

Dette resultat kaldes *foldningssummen* eller *superpositionssummen*, og operationen på højresiden benævnes *foldningen* af talfølgerne $x[n]$ og $h[n]$, som symbolsk skrives $x * h[n]$. Bemærk, at ligning 1.32 udtrykker et diskret LSI-systems svar på et vilkårligt inputsignal ved hjælp af dets impulssvar.

Regneregler for foldningsintegral og foldningssum. Der gælder følgende regneregler for foldningssummer:

1. Foldning er *kommutativ*:
$$x * h[n] = h * x[n] \tag{1.33}$$
2. Foldning er *associativ*:
$$x * (h_1 * h_2[n]) = (x * h_1) * h_2[n] \tag{1.34}$$
3. Foldning er *distributiv* over addition:
$$x * (h_1 + h_2) = x * h_1[n] + x * h_2[n] \tag{1.35}$$

og følgende regneregler for foldningsintegraler:

1. Foldning er *kommutativ*:
$$x * h(t) = h * x(t) \tag{1.36}$$
2. Foldning er *associativ*:
$$x * (h_1 * h_2(t)) = (x * h_1) * h_2(t) \tag{1.37}$$
3. Foldning er *distributiv* over addition:
$$x * (h_1 + h_2) = x * h_1(t) + x * h_2(t) \tag{1.38}$$

Disse regler vises let ved brug af definitionen for foldningssummen og foldningsintegralet.

1.5 Opgaver

1. Find den lige og den ulige del af signalet $x(t) = e^{-t}u(t)$. Vis, at den lige del *er* en lige funktion, og at den ulige del *er* en ulige funktion. Skitser den lige og den ulige del.

2. Et kontinuerttids LFI-system har enhedsimpulssvaret $h(t) = u(t)$. Hvad bliver outputsignalet svarende til inputsignalet $x(t) = e^{-\alpha t}u(t)$, hvor $\alpha > 0$?

3. Betragt et diskrettids LFI-system med enhedsimpulssvaret $h[n] = u[n]$. Hvad er systemets svar på inputsignalet $x[n] = \alpha^n u[n]$, hvor α er reel og $0 < \alpha < 1$? Skitser outputsignalets udseende.

4. Betragt talfølgerne
$$x[n] = \begin{cases} 1, & 0 \leq n \leq 4 \\ 0, & \text{ellers} \end{cases} \tag{1.39}$$
og
$$h[n] = \begin{cases} \alpha^n, & 0 \leq n \leq 6 \\ 0, & \text{ellers} \end{cases} \tag{1.40}$$
Hvad bliver resultatet af foldningen $x * h[n]$?

Kapitel 2

Fourieranalyse og sampling

2.1 Indledning

I dette kapitel vises, hvorledes det er muligt at håndtere kontinuert varierende størrelser ved hjælp af processer på diskrete størrelser, således at man kan analysere og omforme kontinuerte signaler ved hjælp af en datamaskine. Denne teori giver baggrunden for at forstå, hvorledes man kan behandle lyd- og billedsignaler på digital form, og er derfor af stor betydning ved moderne informationsbearbejdning. Den helt grundlæggende sætning er *samplingssætningen*, som løst sagt siger, at det er muligt at gendanne et båndbegrænset signal uden informationstab, hvis det samples med en frekvens, der er større end to gange den højeste frekvens, der optræder i signalet. Et båndbegrænset signal er et signal *med et begrænset frekvensbånd*, d.v.s. det har en højeste og en laveste frekvens.

Vi begynder nu at tale om frekvenser, og derfor må vi have et redskab til analyse og karakterisering af frekvenser i signaler. Til dette formål bruges *Fourieranalyse*, og vi skal derfor først betragte Fourieranalyse, før vi ser på sampling. Først vil vi undersøge fremstilling af periodiske signaler ved hjælp af *Fourierrækker*.

2.2 Fourierrækker

Ved undersøgelse af lineære forskydningsinvariante systemer kan man med fordel benytte en opsplitning af signaler i basissignaler, der har følgende to egenskaber:

1. Mængden af basissignaler kan bruges til at konstruere en omfattende og brugbar klasse af signaler.

2. LFI systemets svar på de enkelte basissignaler er så simpel, at man får en bekvem fremstilling af systemets svar på ethvert signal, der kan konstrueres som linearkombination af basissignalerne.

For kontinuerte LFI systemer opfylder mængden af komplekse eksponentialfunktioner e^{st}, hvor s er et komplekst tal, disse betingelser. De komplekse eksponentialfunktioner har den egenskab, at et LFI systems svar på et kompleks eksponential inputsignal er det samme

komplekse eksponentialsignal, blot med en ændring i amplituden, d.v.s.

$$e^{st} \mapsto H(s)e^{st}, \tag{2.1}$$

hvor den komplekse amplitudefaktor $H(s)$ i almindelighed er en funktion af den komplekse variable s. Dette ses ved at betragte et LFI-system med impulssvaret $h(t)$. For et input $x(t)$ kan vi bestemme output ved hjælp af foldningsintegralet (ligning (1.21)).

$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau \tag{2.2}$$

Med input $x(t) = e^{st}$ fås

$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau = \int_{-\infty}^{\infty} h(\tau)e^{s(t-\tau)}d\tau = e^{st}\int_{-\infty}^{\infty} h(\tau)e^{-s\tau}d\tau = H(s)e^{st}, \tag{2.3}$$

hvor $H(s)$ er en kompleks konstant, hvis værdi afhænger af s, og som kan bestemmes fra systemets impulssvar ved

$$H(s) = \int_{-\infty}^{\infty} h(\tau)e^{-s\tau}d\tau \tag{2.4}$$

Værdien af at kunne dekomponere mere generelle signaler i komplekse eksponentialfunktioner kan ses ved et eksempel. Lad $x(t)$ være en linearkombination af tre komplekse eksponentialfunktioner, d.v.s:

$$x(t) = a_1e^{s_1t} + a_2e^{s_2t} + a_3e^{s_3t} \tag{2.5}$$

og

$$\begin{aligned} a_1e^{s_1t} &\mapsto a_1H(s_1)e^{s_1t} \\ a_1e^{s_2t} &\mapsto a_2H(s_2)e^{s_2t} \\ a_1e^{s_3t} &\mapsto a_3H(s_3)e^{s_3t} \end{aligned}$$

Ved anvendelse af superpositionsprincippet fra afsnit (1.3) ses, at systemets svar på summen af inputsignalerne bliver summen af enkeltsvarene, d.v.s.

$$y(t) = a_1H(s_1)e^{s_1t} + a_2H(s_2)e^{s_2t} + a_3H(s_3)e^{s_3t} \tag{2.6}$$

I almindelighed haves

$$\sum_k a_ke^{s_kt} \mapsto \sum_k a_kH(s_k)e^{s_kt} \tag{2.7}$$

Betragt nu to grundlæggende periodiske signaler, nemlig cosinusfunktionen $\cos\omega_0 t$ og det periodiske komplekse eksponentialsignal $x(t) = e^{i\omega_0 t}$. Begge disse signaler er periodiske med grundfrekvensen (vinkelfrekvensen) $\omega_0 = 2\pi f_0$ og grundperioden $T_0 = 2\pi/\omega_0$. Sammen med signalet $x(t) = e^{i\omega_0 t}$ kan vi betragte dets *harmoniske*

$$\phi_k(t) = e^{ik\omega_0 t}, k = 0, \pm 1, \pm 2, \ldots \tag{2.8}$$

Da hvert af de indgående signaler har en grundfrekvens, der er et multiplum af ω_0 og derfor er periodisk med perioden T_0, så er en linearkombination af harmoniske komplekse eksponentialfunktioner af formen

$$x(t) = \sum_{k=-\infty}^{+\infty} a_k e^{ik\omega_0 t} \tag{2.9}$$

også periodisk med perioden T_0. I denne ligning benævnes leddet, man får for $k = 0$, konstantleddet eller $dc - leddet$ (dc= direct current = jævnstrøm). De to led, der fås for $k = +1$ og $k = -1$, har en grundperiode lig T_0 og benævnes *grundkomponenterne* eller *den første harmoniske*. Tilsvarende kaldes de to led for $k = +2$ og $k = -2$, som er periodiske med den dobbelte frekvens (eller den halve periode), den *anden harmoniske komponent* o.s.v.

Fremstillingen af et periodisk signal ved ligning (2.9) kaldes dets *Fourierrækkefremstilling*.

Hvis man ved, at et signal kan fremstilles ved en Fourierrækkefremstilling (2.9), er problemet, hvorledes man bestemmer koefficienterne a_k. Ved at multiplicere begge sider af ligning (2.9) med $e^{-in\omega_0 t}$ fås

$$x(t)e^{-in\omega_0 t} = \sum_{k=-\infty}^{+\infty} a_k e^{ik\omega_0 t} e^{-in\omega_0 t} \tag{2.10}$$

og ved integration af begge sider fra 0 til $T_0 = 2\pi/\omega_0$ fås

$$\int_0^{T_0} x(t)e^{-in\omega_0 t}dt = \int_0^{T_0} \sum_{k=-\infty}^{+\infty} a_k e^{ik\omega_0 t} e^{-in\omega_0 t} dt. \tag{2.11}$$

T_0 er grundperioden for $x(t)$ og derfor integreres over én periode. Hvis det vides, at rækken er konvergent (dette spørgsmål diskuteres senere), så kan man ombytte integration og summation, d.v.s.

$$\int_0^{T_0} x(t)e^{-in\omega_0 t} = \sum_{k=-\infty}^{+\infty} a_k \left[\int_0^{T_0} e^{i(k-n)\omega_0 t} dt \right] \tag{2.12}$$

Udregning af integralet i skarp parentes kan ske ved anvendelse af Euler's formel:

$$\int_0^{T_0} e^{i(k-n)\omega_0 t} dt = \int_0^{T_0} [\cos(k-n)\omega_0 t] dt + i \int_0^{T_0} [\sin(k-n)\omega_0 t] dt \tag{2.13}$$

For $k \neq n$ er $\cos(k-n)\omega_0 t$ og $\sin(k-n)\omega_0 t$ periodiske trigonometriske funktioner med grundperioden $T_0/|k-n|$. I ligningen integreres over et interval (af længden T_0) som er et helt antal perioder for disse signaler. Da integralet svarer til at måle det samlede areal under disse funktioner i det givne interval, ses, at for $k \neq n$ er begge integraler på højresiden 0. For $k = n$ er integranden på venstresiden $e^0 = 1$ og derfor bliver integralet lig T_0. Derfor er

$$\int_0^{T_0} e^{i(k-n)\omega_0 t} dt = \begin{cases} T_0, & \text{for } k = n \\ 0, & \text{for } k \neq n \end{cases} \tag{2.14}$$

og følgelig reduceres højresiden af ligning (2.12) til $T_0 a_n$. Derfor er

$$a_n = \frac{1}{T_0} \int_0^{T_0} x(t) e^{-in\omega_0 t} dt, \tag{2.15}$$

hvilket er den søgte ligning til bestemmelse af koefficienterne. Vi integrerede over intervallet $[0, T_0]$ men, på grund af signalernes periodicitet, er det ligegyldigt hvilket interval af længden T_0 vi integrerer over. Lader vi $\int_{T_0}$ betegne integration over et *vilkårligt* interval af længden T_0 haves

$$\int_{T_0} e^{i(k-n)\omega_0 t} dt = \begin{cases} T_0, & \text{for } k = n \\ 0, & \text{for } k \neq n \end{cases} \tag{2.16}$$

Vi har derfor resultatet: Hvis $x(t)$ har en Fourierrækkefremstillling, d.v.s. hvis $x(t)$ kan udtrykkes som en linearkombination af harmoniske komplekse eksponentialfunktioner, så er signalets Fourierrækkefremstilling

$$x(t) = \sum_{k=-\infty}^{+\infty} a_k e^{ik\omega_0 t}, \tag{2.17}$$

hvor koefficienterne a_n er givet ved

$$a_n = \frac{1}{T_0} \int_0^{T_0} x(t) e^{-in\omega_0 t} dt \tag{2.18}$$

Ligning (2.17) kaldes ofte *synteseligningen* og ligning (2.18) for *analyseligningen*. Koefficienterne $\{a_n\}$ kaldes for $x(t)$'s *Fourierrækkekoefficienter* eller *spektralkoefficienter*. Disse komplekse koefficienter måler den del af signalet $x(t)$, der er i hver harmoniske. Koefficienten a_0 er dc- eller konstantleddet af $x(t)$ og er

$$a_0 = \frac{1}{T_0} \int_{T_0} x(t) dt \tag{2.19}$$

Eksempel. Det periodiske firkantsignal vist på figur 2.1 og defineret over en periode som

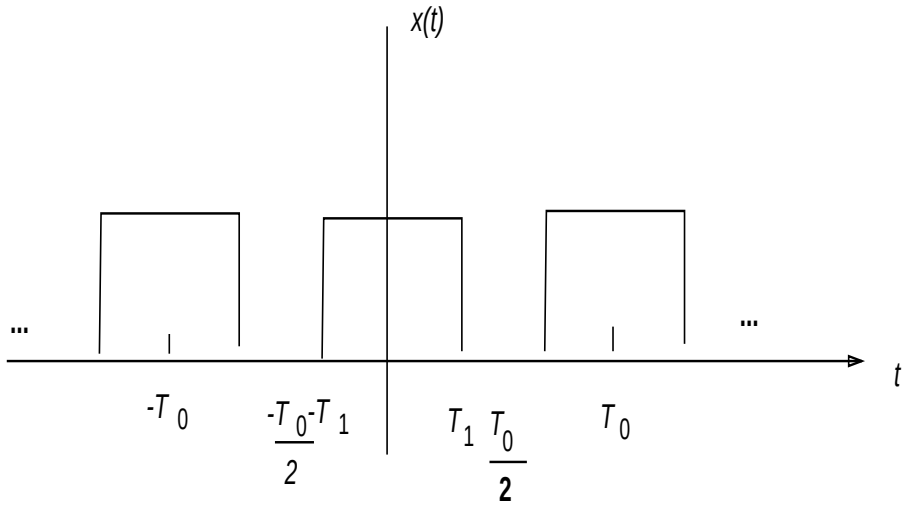
$$x(t) = \begin{cases} 1, & |t| < T_1 \\ 0, & T_1 < |t| < \frac{T_0}{2} \end{cases} \tag{2.20}$$

er periodisk med grundperioden T_0 og grundfrekvensen $\omega_0 = 2\pi/T_0$. Fourierkoefficienterne for $x(t)$ bestemmes ved hjælp af ligning (2.18). På grund af $x(t)$'s symmetri omkring $t = 0$ er det mest bekvemt at vælge at integrere over intervallet $-T_0/2 \leq t \leq T_0/2$. Man finder

$$a_0 = \frac{1}{T_0} \int_{-T_1}^{T_1} dt = \frac{2T_1}{T_0} \tag{2.21}$$

For $k \neq 0$ findes

$$a_k = \frac{1}{T_0} \int_{-T_1}^{T_1} e^{-ik\omega_0 t} dt = \left[-\frac{1}{ik\omega_0 T_0} e^{-ik\omega_0 t} \right]_{-T_1}^{T_1} \tag{2.22}$$



Figur 2.1: Periodisk firkantsignal

Dette kan omskrives således

$$a_k = \frac{2}{k\omega_0 T_0}\left[\frac{e^{ik\omega_0 T_1} - e^{-ik\omega_0 T_1}}{2i}\right] \tag{2.23}$$

Idet leddet indenfor den skarpe parantes er $\sin k\omega_0 T_1$ kan koefficienterne a_k udtrykkes som

$$a_k = \frac{2\sin(k\omega_0 T_1)}{k\omega_0 T_0} = \frac{\sin(k\omega_0 T_1)}{k\pi},\ k \neq 0, \tag{2.24}$$

idet $\omega_0 T_0 = 2\pi$. Grafisk fremstilling af Fourierrækkekoefficienterne for dette eksempel for fast værdi af T_1 og forskellige værdier af T_0 er vist på figur 2.2. For $T_0 = 4T_1$ er $x(t)$ en *symmetrisk firkantbølge* (d.v.s. som er 1 for halvdelen af perioden og 0 for resten). Da er $\omega_0 T_1 = \pi/2$ og fra ligning (2.24)

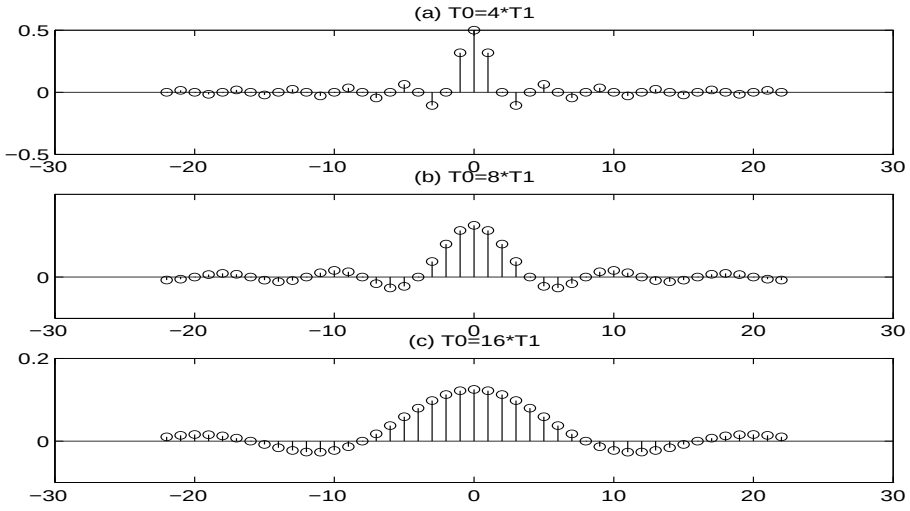
$$a_k = \frac{\sin(\pi k/2)}{k\pi},\ k \neq 0 \tag{2.25}$$

og fra ligning (2.21)

$$a_0 = \frac{1}{2} \tag{2.26}$$

Fra ligning (2.25) ses at $a_k = 0$ for k lige. Desuden svinger $\sin(\pi k/2)$ mellem ± 1 for en følge af ulige værdier af k. Derfor er

$$\begin{aligned} a_1 &= a_{-1} = \frac{1}{\pi} \\ a_3 &= a_{-3} = \frac{1}{3\pi} \\ a_5 &= a_{-5} = \frac{1}{5\pi} \\ &\vdots \end{aligned}$$



Figur 2.2: Fourierrækkekoefficienter

I dette eksempel er Fourierkoefficienterne reele og derfor kan de afbildes med en enkelt graf. I almindelighed er Fourierkoefficienterne dog komplekse og der er derfor nødvendigt med to grafer, svarende til real- og imaginærværdien eller magnituden og fasen for koefficienterne.

Fourierrækkens konvergens. $x(t)$ skal opfylde Dirichets betingelse, d.v.s. den periodiske funktion skal være af *begrænset variation* i omegnen af et punkt. Dette er tilstrækkeligt til at sikre, at funktionens Fourierrække konvergerer punktvis til middelværdien af funktionens grænser fra højre og fra venstre. Hvis funktionen er kontinuert i punktet, da til dens funktionsværdi i punktet. Begrænset variation er den egenskab ved en reel funktion $x(t)$ at dens variation er begrænset. Den kan udtrykkes som differensen mellem to monotone, ikke aftagende funktioner.

Den totale variation. En funktions totale variation er et mål for, hvor meget den svinger:

$$V_h(a,b) = sup\{\Sigma|h(t_{i+1}) - h(t_i)|\} \tag{2.27}$$

over alle inddelinger af intervallet $[a,b]$. Den totale variation er er endelig hvis og kun hvis funktionen er af begrænset variation i intervallet. Hvis funktionen dekomponeres som $f-g$, hvor f og g er monotont voksende, ved at sætte

$$\begin{aligned} 2f(t) &= V_h(a,t) + h(t) - h(a) \\ 2g(t) &= V_h(a,t) - h(t) + h(a) \end{aligned}$$

for t mellem a og b, så er den totale variation mellem a og b lig $f(a) + g(b)$.

2.3 Fouriertransformationen

Fra matematik (og fysik) vides, at den Fouriertransformerede $X(\omega) = \mathcal{F}(x(t))$ er defineret ved

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-i\omega t}dt, \tag{2.28}$$

hvor $\omega = 2\pi f = \frac{2\pi}{T}$ er vinkelfrekvensen, hvis f er frekvensen og T periodetiden og i er $\sqrt{-1}$.

Ud fra et signal $x(t)$'s Fouriertransformerede $X(\omega)$ kan vi gendanne det oprindelige signal ved den *inverse Fouriertransformation:*

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{i\omega t}d\omega, \tag{2.29}$$

2.3.1 Forskydningsreglen

Hvis

$$x(t) \overset{\mathcal{F}}{\leftrightarrow} X(\omega)$$

så er

$$X(t-t_0) \overset{\mathcal{F}}{\leftrightarrow} e^{-i\omega t_0}X(\omega) \tag{2.30}$$

Dette kaldes *forskydningsreglen*. For at vise denne betragtes

$$\mathcal{F}\{x(t-t_0)\} = \int_{-\infty}^{\infty} x(t-t_0)e^{-i\omega t}dt \tag{2.31}$$

Ved substitutionen $\sigma = t - t_0$ fås

$$\mathcal{F}\{x(t-t_0)\} = \int_{-\infty}^{\infty} x(\sigma)e^{-i\omega(\sigma+t_0)}d\sigma = e^{-i\omega t_0}X(\omega) \tag{2.32}$$

2.3.2 Foldningsreglen

En af Fouriertransformationens vigtigste egenskaber er dens sammenhæng med foldningsoperationen. For at udlede denne sammenhæng betragtes et LFI system med impulssvaret $h(t)$, output $y(t)$ og input $x(t)$, således at

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau \tag{2.33}$$

For den Fouriertransformerede $Y(\omega)$ af $y(t)$ findes

$$Y(\omega) = \mathcal{F}\{y(t)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[x(\tau)h(t-\tau)d\tau]e^{-i\omega t}dt \tag{2.34}$$

Ved ombytning af integrationsrækkefølgen, og idet $x(\tau)$ ikke afhænger af t fås

$$Y(\omega) = \int_{-\infty}^{\infty} x(\tau)\left[\int_{-\infty}^{\infty} h(t-\tau)e^{-i\omega t}dt\right]d\tau \tag{2.35}$$

Ud fra forskydningsegenskaben er leddet i skarp parentes lig $e^{-i\omega\tau}H(\omega)$. Ved indsættelse af dette i ovenstående ligning findes

$$Y(\omega) = \int_{-\infty}^{\infty} x(\tau)e^{-i\omega\tau}H(\omega)d\tau = H(\omega)\int_{-\infty}^{\infty} x(\tau)e^{-i\omega\tau}d\tau \tag{2.36}$$

Dette integral er $\mathcal{F}\{x(t)\}$, d.v.s.

$$Y(\omega) = H(\omega)X(\omega) \tag{2.37}$$

Hermed har vi udledt *foldningsreglen*

$$y(t) = h * x(t) \overset{\mathcal{F}}{\leftrightarrow} Y(\omega) = H(\omega)X(\omega) \tag{2.38}$$

2.3.3 Dualitetsreglen

Der er en vis symmetri mellem Fouriertransformationen (ligning 2.28) og den inverse Fouriertransformation (ligning 2.29). Denne symmetri er baggrunden for en egenskab ved Fouriertransformationen kendt som *dualitetsreglen*. Betragt to funktioner som er indbyrdes relaterede via integraludtrykket

$$f(u) = \int_{-\infty}^{\infty} g(v)e^{-iuv}dv \tag{2.39}$$

Ved at sammenligne ligning 2.39 med Fourier analyse og synteseligningerne (ligning 2.28 og 2.29) ses, at med $u = \omega$ og $v = t$ er

$$f(\omega) = \mathcal{F}\{g(t)\} \tag{2.40}$$

medens med $u = t$ og $v = \omega$ er

$$g(-\omega) = \frac{1}{2\pi}\mathcal{F}\{f(t)\} \tag{2.41}$$

Derfor, hvis der er givet et Fourier transformationspar for tidsfunktionen $g(t)$

$$g(t) \overset{\mathcal{F}}{\leftrightarrow} f(\omega) \tag{2.42}$$

og herefter betragter funktionen $f(t)$ som funktion af *tiden*, så er dens Fouriertransformationspar

$$f(t) \overset{\mathcal{F}}{\leftrightarrow} 2\pi g(-\omega) \tag{2.43}$$

Dette resultat kan bruges til at 'oversætte' regler fra tidsdomænet til frekvensdomænet og omvendt. F.eks. som den efterfølgende regel.

2.3.4 Modulationsreglen

Modulationsreglen kan direkte afledes fra foldningsreglen ved brug af dualitetsreglen. Foldningsreglen lyder

$$y(t) = h * x(t) \overset{\mathcal{F}}{\leftrightarrow} Y(\omega) = H(\omega)X(\omega) \tag{2.44}$$

Foldningsreglem fastslår, at foldning i tidsdomænet svarer til multiplikation i frekvensdomænet. På grund af dualiteten mellem tids- og frekvensdomænet må man forvente, at den duale egenskab også gælder, d.v.s. at multiplikation i tidsdomænet svarer til foldning i frekvensdomænet. Ved brug af dualitetsreglen fås

$$t(t) = s(t)p(t) \overset{\mathcal{F}}{\leftrightarrow} R(\omega) = \frac{1}{2\pi}[S * P(\omega)] \tag{2.45}$$

Multiplikation af et signal med et andet benævnes *amplitudemodulation* og reglen kaldes derfor modulationsreglen.

2.3.5 Fouriertransformationen af periodiske signaler

Nu ønskes Fouriertransformationen af et periodisk signal. Som man vil kunne se, kan man konstruere Fouriertransformationen af et sådant signal direkte fra dets Fourierrækkefremstilling. Den Fouriertransformerede af et periodisk signal består af et impulstog i frekvensspektret (et liniespektrum), hvor impulsernes arealer er proportionale med Fourierrækkekoefficienterne. Dette er en meget vigtig fremstilling, idet den letter anvendelsen af Fourieranalyseteknikkerne på samplingsproblemet.

For at antyde det generelle resultet betragtes et signal $x(t)$ med den Fouriertransformerede $X(\omega)$ i form af en enkelt impuls med arealet 2π ved frekvensen $\omega = \omega_0$, d.v.s

$$X(\omega) = 2\pi\delta(\omega - \omega_0) \tag{2.46}$$

For at finde det signal $x(t)$ for hvilket $X(\omega)$ er den Fouriertransformede anvendes den inverse Fouriertransformation (ligning (2.29)), hvorved fås

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} 2\pi\delta(\omega - \omega_0)e^{i\omega t}d\omega = e^{i\omega_0 t} \tag{2.47}$$

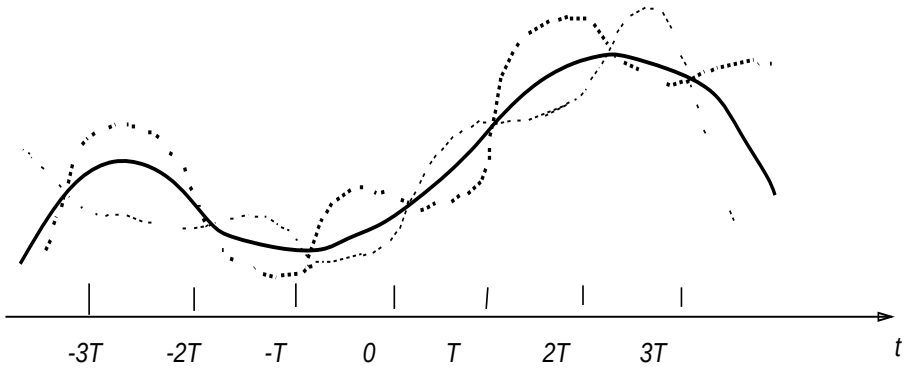
Mere generelt, hvis $X(\omega)$ er en linearkombination af impulser med ens afstand i frekvensdomænet (et liniespektrum), d.v.s.

$$X(\omega) = \sum_{k=-\infty}^{+\infty} 2\pi a_k \delta(\omega - k\omega_0) \tag{2.48}$$

giver anvendelse af ligning (2.29)

$$x(t) = \sum_{k=-\infty}^{+\infty} a_k e^{ik\omega_0 t} \tag{2.49}$$

Det ses, at dette resultat svarer til Fourierrækkefremstillingen af et periodisk signal, som givet i ligning (2.17). Det vil sige at den Fouriertransformerede af et periodisk signal med



Figur 2.3: Tre forskellige kontinuerte signaler med samme sampleværdier

Fourierkoefficienterne $\{a_k\}$ er et liniespektrum svarende til harmoniske frekvenser, hvor arealet af impulsen for den k'te harmoniske frekvens $k\omega_0$ er 2π gange den k'te Fourierrækkekoefficient a_k.

Et signal, som benyttes i analysen af samplingssætningen, er givet ved

$$x(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT) \tag{2.50}$$

Dette signal er periodisk med grundperioden T. For at finde den Fouriertransformerede af dette signal beregnes dets Fourierrækkekoefficienter

$$a_k = \frac{1}{T} \int_{-T/2}^{T/2} \delta(t) e^{ik\omega_0 t} dt = \frac{1}{T} \tag{2.51}$$

Ved indsætning i ligning 2.48 fås

$$X(\omega) = \frac{2\pi}{T} \sum_{k=-\infty}^{\infty} \delta\left(\omega - \frac{2\pi k}{T}\right) \tag{2.52}$$

d.v.s. den Fouriertransformerede af et impulstog i tidsdomænet er en række impulser i frekvensdomænet. Når afstanden mellem impulserne i tidsdomænet (d.v.s. perioden) bliver længere, bliver afstanden mellem impulserne i frekvensdomænet (d.v.s. grundfrekvensen) kortere.

2.4 Samplingssætningen

I almindelighed vil man ikke kunne forvente, at et signal på éntydig måde ville kunne rekonstrueres ud fra dets sampleværdier. For eksempel ses på figur 2.3 tre forskellige kontinuerte signaler, der alle har samme funktionsværdier for hele multipla af T, d.v.s.

$$x_1(kT) = x_2(kT) = x_3(kT)$$

Der vil normalt være en uendelighed af signaler, der kan passe til en forelagt mængde af sampleværdier. Men hvis et signal er *båndbegrænset*, og hvis sampleværdierne tages

tilstrækkelig tæt i forhold til den højeste frekvens, der findes i signalet, så vil sampleværdierne *éntydigt* definere signalet og det er muligt at rekonstruere det perfekt (d.v.s. uden informationstab).

Lad os definere et impulstog $p(t)$, som er vores samplingsfunktion. $p(t)$ består af en række deltafunktioner, således defineret, at afstanden mellem de enkelte deltaimpulser er T. Samplingsfrekvensen er da $\omega_s = 2\pi/T$. I tidsdomænet haves

$$x_p(t) = x(t)p(t) \tag{2.53}$$

hvor

$$p(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT) \tag{2.54}$$

$x_p(t)$ er et impulstog, hvor impulsernes værdi (areal under integration) er lig stikprøveværdierne af $x(t)$ med afstanden T, d.v.s.

$$x_p(t) = \sum_{n=-\infty}^{\infty} x(nT)\delta(t - nT) \tag{2.55}$$

og fra modulationsreglen

$$X_p(\omega) = \frac{1}{2\pi}[X * P(\omega)] \tag{2.56}$$

og, idet den Fouriertransformerede $P(\omega)$ af impulstoget $p(t)$ er

$$P(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \delta(\omega - k\omega_s), \tag{2.57}$$

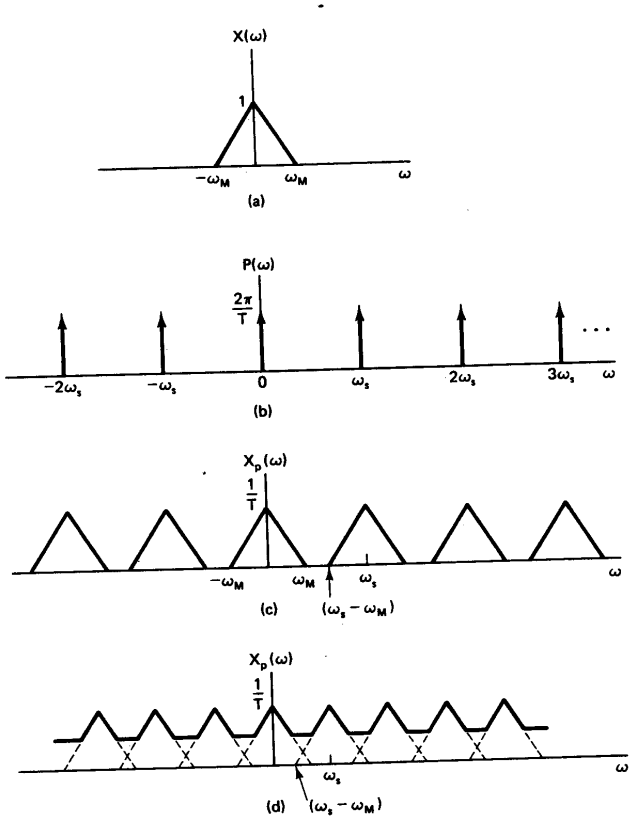
er

$$X_p(\omega) = \frac{1}{T} \sum_{k=-\infty}^{\infty} X(\omega - k\omega_s) \tag{2.58}$$

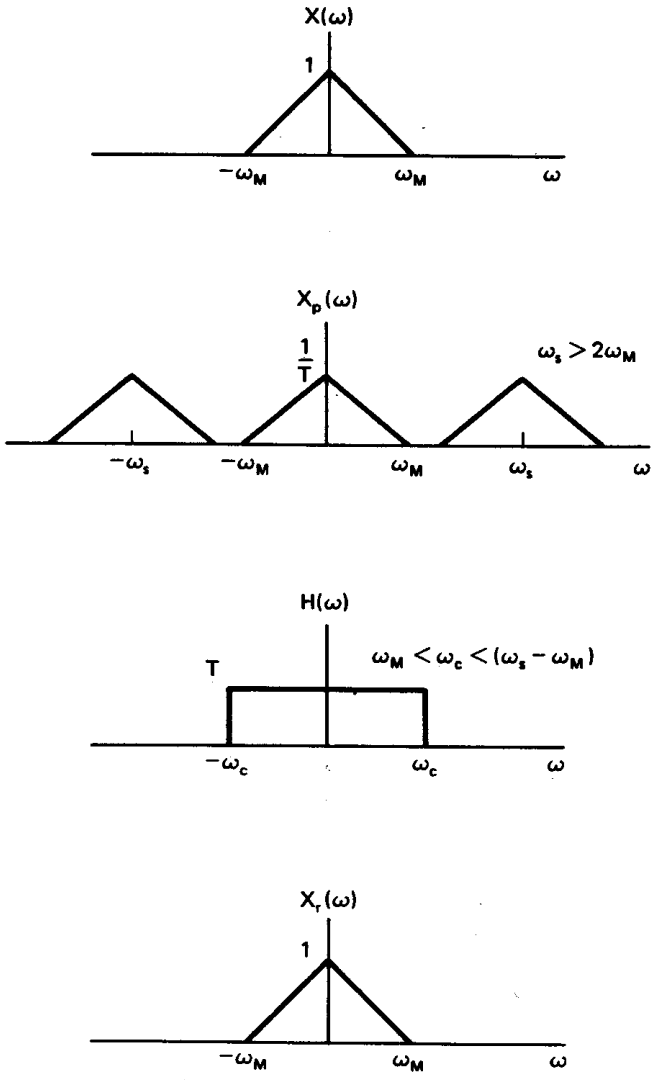
Det vil sige, at $X_p(\omega)$, frekvensspektret for den samplede funktion, består af en række forskudte kopier af $X(\omega)$ skaleret med $1/T$ som vist på figur 2.4. Lad nu ω_M betegne den højeste frekvens, der forekommer i signalet. På figur 2.4 (c) er $\omega_M < (\omega_s - \omega_M)$ d.v.s. $\omega_s > 2\omega_M$ og følgelig er der ingen overlapning af de forskudte kopier af $X(\omega)$. På figur 2.4(d), hvor $\omega_s < 2\omega_M$ er der overlapning. For tilfældet vist på figur (c) bliver $X(\omega)$ korrekt gengivet for hele multipla af ω_s. Derfor, hvis $\omega_s > 2\omega_M$, kan $x(t)$ genskabes eksakt ud fra den samplede funktion $x_p(t)$ ved at fjerne frekvenser over ω_c, hvor $\omega_M < \omega_c < \omega_s - \omega_M$, som vist på figur 2.5. Dette grundlæggende resultat, der betegnes *samplingssætningen* kan formuleres således:

> *Hvis $x(t)$ er et båndbegrænset signal med $X(\omega) = 0$ for $|\omega| > \omega_M$, så er $x(t)$ éntydigt bestemt ved dets sampleværdier $x(nt), n = 0, \pm 1, \pm 2, \ldots$ hvis samplingfrekvensen $\omega_s > 2\omega_M$, hvor $\omega_s = 2\pi/T$.*

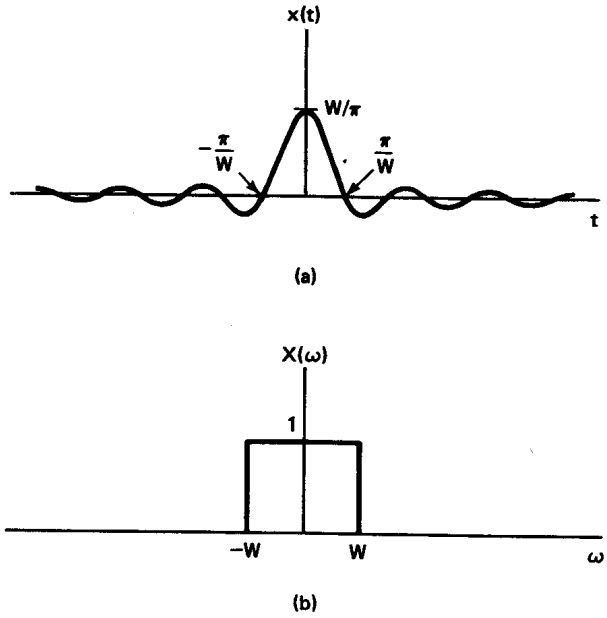
Hvis altså samplepunkterne ligger tilstrækkelig tæt, så kan det samplede signal gendannes eksakt, d.v.s. der er muligt at interpolere mellem sampleværdierne, således at vi får det oprindelige signal igen. Interpolationen skal ske med $\sin(\pi t)/\pi t$-funktioner. For at indse det betragtes det impulssvar $h(t)$, hvis Fouriertransformerede er



Figur 2.4: Frekvensspektret for den samplede funktion



Figur 2.5: Filtreret signal



Figur 2.6: Fouriertransformationspar for båndbegrænset signal

$$H(\omega) = \begin{cases} T, & |\omega| < \omega_c \\ 0, & |\omega| > \omega_c \end{cases} \tag{2.59}$$

Ved brug af den inverse Fouriertransformation (ligning (2.29) findes $x(t)$:

$$h(t) = \frac{T}{2\pi} \int_{-\omega_c}^{\omega_c} e^{i\omega t} d\omega = T \frac{\sin \omega_c t}{\pi t} = T \frac{\omega_c}{\pi} sinc\left(\frac{\omega_c}{\pi}\right), \tag{2.60}$$

som er vist på figur 2.6. For funktionene $\sin(\pi t)/\pi t$ bruges ofte betegnelsen *sinc funktionen.*

Det rekonstruerede signal $x_r(t)$ fås ved linearkombination af impulssvar på x_p (det samplede signal), d.v.s.

$$x_r(t) = \sum_{n=-\infty}^{\infty} x(nT) T \frac{\omega_c}{\pi} sinc\left(\frac{\omega_c(t - nT)}{\pi}\right). \tag{2.61}$$

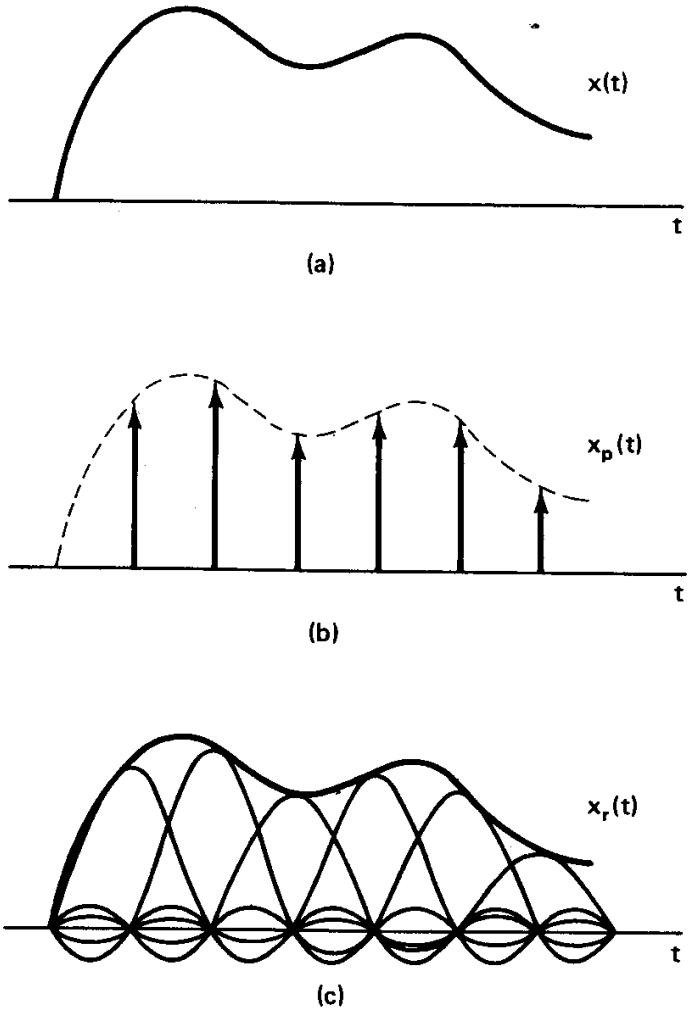
Interpolationen er vist på figur 2.7.

2.4.1 Opgaver

1. Find Fourierrækkefremstillingen for $x(t) = \sin \omega_0 t$

2. Bestem og skitser den Fouriertransformerede $X(\omega)$ af signalet $x(t) = e^{-a|t|}$, hvor $a > 0$.

3. Bestem den Fouriertransformerede af enhedsimpulsen $x(t) = \delta(t)$.



Figur 2.7: Ideel båndbegrænset interpolation ved hjælp af sinc-funktionen.

4. Find og skitsér den Fouriertransformerede $X(\omega)$ af firkantsignalet

$$x(t) = \begin{cases} 1, & |t| < T_1 \\ 0, & |t| > T_1 \end{cases} \tag{2.62}$$

5. To tidsfunktioner $x_1(t)$ og $x_2(t)$ multipliceres og deres produkt $w(t)$ samples med et periodisk impulstog. $x_1(t)$ er båndbegrænset til ω_1 og $x_2(t)$ er båndbegrænset til ω_2. Find det største sampleinterval T, således at $w(t)$ kan gendannes ud fra det samplede signal $w_p(t)$ ved hjælp af et ideelt lavpasfilter.

Kapitel 3

Den diskrete Fouriertransformation

3.1 Diskrete komplekse eksponentialfunktioner

Ligesom for kontinuerte signalers vedkommende kan man definere et *komplekst eksponentialsignal* eller *kompleks eksponentialfølge* ved

$$x[n] = C\alpha^n \tag{3.1}$$

hvor C og α i almindelighed er komplekse tal.

Hvis C og α er reelle vil signalerne se ud som vist på figur 3.1. Hvis $|\alpha| > 1$, vil signalet vokse eksponentielt, og hvis $|\alpha| < 1$ vil signalet aftage eksponentielt. Hvis α er positiv, vil alle værdier af $C\alpha^n$ have samme fortegn, men hvis α er negativ, vil $x[n]$'s fortegn skifte. Hvis $\alpha = 1$ er $x[n]$ konstant, mens hvis $\alpha = -1$ vil $x[n]$ skifte mellem $+C$ og $-C$. Reelle diskrete talfølger bruges ofte til at beskrive befolkningsvækst eller kapitalvækst.

Hvis eksponenten er rent imaginær kan $x[n]$ skrives

$$x[n] = e^{i\Omega_0 n}, \tag{3.2}$$

hvor $i = \sqrt{-1}$ og Ω_0 reel. Dette signal er nært beslægtet med signalet

$$x[n] = A\cos(\Omega_0 n + \phi) \tag{3.3}$$

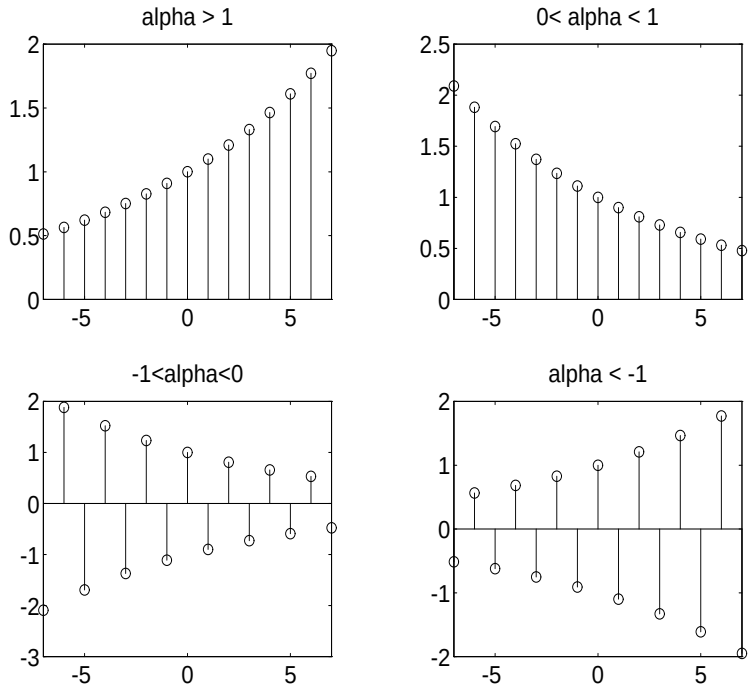
Dersom n er dimensionsløs, har både Ω_0 og ϕ enheden radianer. Eksempler på sådanne følger er vist på figur 3.2. Eulers ligning giver sammenhængen mellem komplekse eksponentialfunktioner og de trigonometriske funktioner:

$$e^{i\Omega_0 n} = \cos\Omega_0 n + i\sin\Omega_0 n \tag{3.4}$$

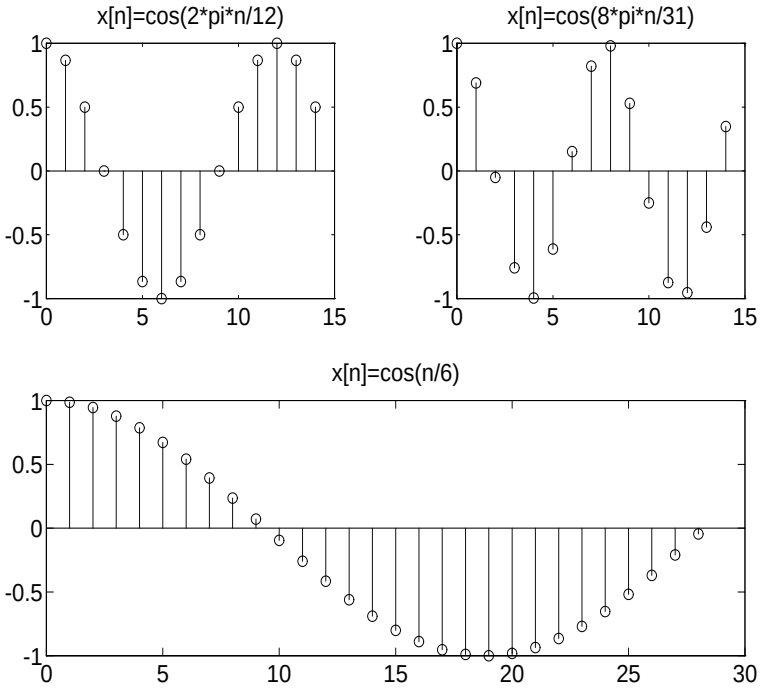
og

$$A\cos(\Omega_0 n + \phi) = \frac{A}{2}e^{i\phi}e^{i\Omega_0 n} + \frac{A}{2}e^{-i\phi}e^{-i\Omega_0 n} \tag{3.5}$$

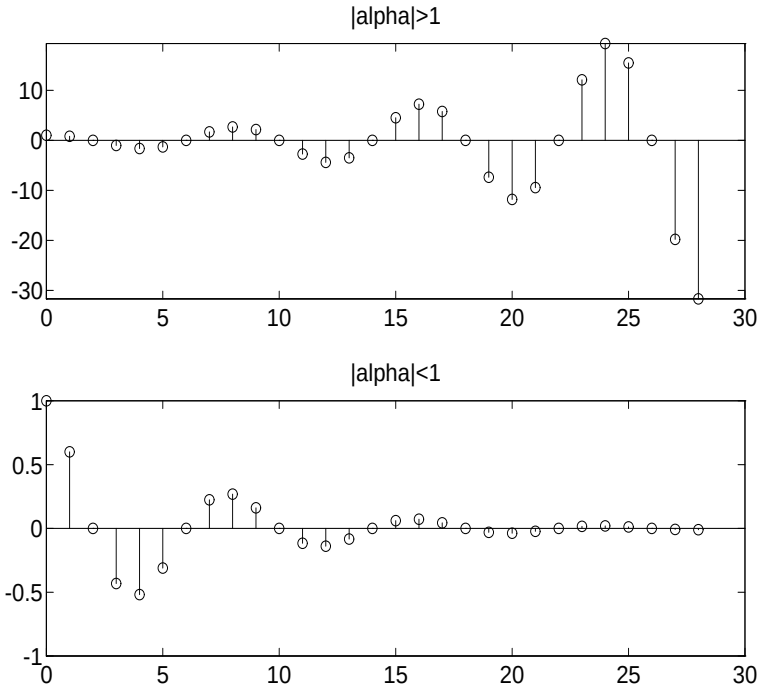
Et generelt diskret kompleks eksponentialsignal kan skrives som et produkt af reelle signaler og sinussignaler og nogle eksempler på sådanne er vist på figur 3.3



Figur 3.1: Diskrete eksponentialfunktioner $x[n] = \alpha^n$, hvor C og α er reele tal.



Figur 3.2: Diskret cosinussignal $x[n] = A\cos(\Omega_0 n + \phi)$, ($A$ er reel).



Figur 3.3: Realdelen af generelle diskrete komplekse eksponentialsignaler.

3.1.1 Diskrete komplekse eksponentialsignalers periodicitet

Det kontinuerte signal $e^{i\omega_0 t}$ har egenskaberne, at jo større værdien af ω_0 er, des hurtigere svinger signalet. Desuden er $e^{i\omega_0 t}$ periodisk for enhver værdi af ω_0. Således forholder det sig ikke med diskrete signaler. Hvis frekvensen øges med 2π fra Ω_0 til $\Omega_0 + 2\pi$ fås det samme signal, idet

$$e^{i(\Omega_0+2\pi)n} = e^{i2\pi n}e^{i\Omega_0 n} = e^{i\Omega_0 n}, \tag{3.6}$$

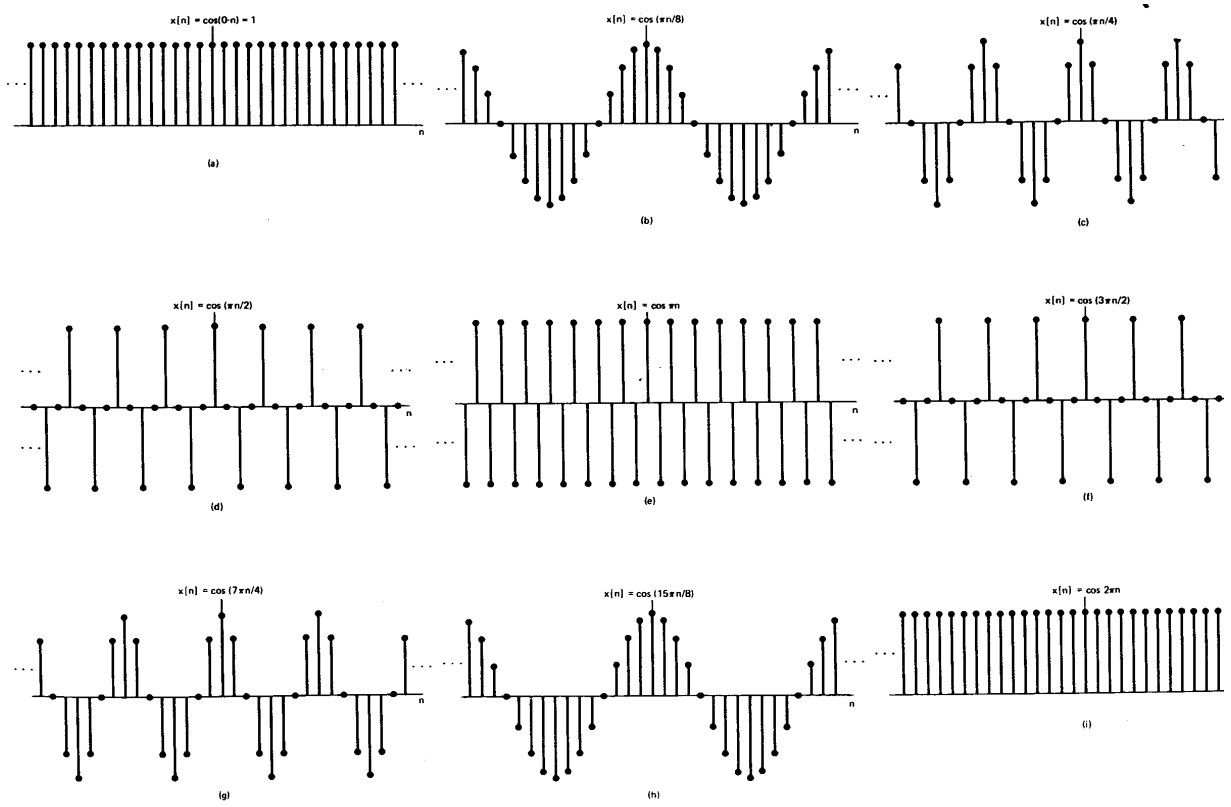
idet $e^{i2\pi n} = 1$. Det vil sige samtlige diskrete komplekse eksponentialsignaler fås ved at lade Ω_0 variere mellem 0 og 2π (eller et tilsvarende interval, f.eks. $-\pi \leq \Omega_0 < \pi$). På grund af denne periodicitet vil det diskrete eksponentialsignals svingningshastighed ikke vokse, når Ω_0 øges. Når Ω_0 øges fra 0 til π øges svingningshastigheden, men fra π til 2π aftager den igen, indtil $\Omega_0 = 2\pi$, hvor værdien er den samme som for $\Omega_0 = 0$. Dette er illustreret på figur 3.4. D.v.s. de diskrete eksponentialsignaler, der svinger langsomt, har værdier af Ω_0 nær $0, 2\pi$ eller ethvert andet multiplum af 2π og de, der svinger hurtigst, har Ω_0 nær $\Omega_0 = \pm\pi$ eller ulige multipla af π.

For at signalet $x[n] = e^{i\Omega_0 n}$ skal være periodisk med perioden N, kræves

$$e^{i\Omega_0(n+N)} = e^{i\Omega_0 n} \tag{3.7}$$

eller

$$e^{i\Omega_0 N} = 1 \tag{3.8}$$



Figur 3.4: Diskrete komplekse eksponentialfunktioner

For at denne ligning kan være opfyldt, skal $\Omega_0 N$ være et helt multiplum af 2π. D.v.s., der skal findes et heltal m, således at

$$\Omega_0 N = 2\pi m \tag{3.9}$$

eller

$$\frac{\Omega_0}{2\pi} = \frac{m}{N} \tag{3.10}$$

Det betyder, at signalet $e^{i\Omega_0 n}$ er ikke periodisk for *vilkårlige* værdier af Ω_0. Kun hvis $\Omega_0/2\pi$ er et rationelt tal, er signalet periodisk. Tilsvarende gælder diskrete sinus- og cosinussignaler. Hvis $x[n]$ er periodisk med grundperioden N, så er dets grundfrekvens $2\pi/N$. Det er let at vise, at dersom $\Omega_0 \neq 0$ og hvis N og m ingen fælles faktorer har, så er signalets grundperiode N. Derfor er grundfrekvensen for det periodiske signal $e^{i\Omega_0 n}$

$$\frac{2\pi}{N} = \frac{\Omega_0}{m} \tag{3.11}$$

Grundperioden kan også skrives

$$N = m\left(\frac{2\pi}{\Omega_0}\right) \tag{3.12}$$

Som ved kontinuerte signaler kan man betragte harmoniske periodiske eksponentialsignaler, d.v.s. periodiske eksponentialsignaler, der alle har perioden N. Fra ligning (3.10) vides, at det netop er signaler ved frekvenser, der er multipla af $2\pi/N$, d.v.s.

$$\phi_k[n] = e^{ik(2\pi/N)n}, k = 0, \pm 1, \ldots \tag{3.13}$$

For kontinuerte signaler er alle harmoniske komplekse eksponentialsignaler (formel 2.8 side 24) $e^{ik(2\pi/T)t}, k = 0, \pm 1, \pm 2, \ldots$ forskellige. Men det er ikke tilfældet for de diskrete eksponentialfunktioner, idet

$$\phi_{k+N}[n] = e^{i(k+N)(2\pi/N)n} = e^{i2\pi n} e^{ik(2\pi/N)n} = \phi_k[n] \tag{3.14}$$

Det betyder, at der kun er N forskellige periodiske eksponentialsignaler i mængden givet ved ligning (3.13) For eksempel er $\phi_0[n], \phi_1[n], , \ldots, \phi_{N-1}[n]$ alle forskellige og enhver anden $\phi_k[n]$ er identisk med en af disse.

Til slut betragtes en diskret følge, som er fremkommet ved at sample en kontinuert eksponentialfunktion $e^{i\omega_0 t}$ med ækvidistant beliggende samplingspunkter

$$x[n] = e^{i\omega_0 nt} = e^{i(\omega_0 T)n} \tag{3.15}$$

Fra denne ligning ses, at $x[n]$ selv er et diskret eksponentialsignal med $\Omega_0 = \omega_0 T$. Derfor vil, i overensstemmelse med den tidligere analyse, $x[n]$ kun være periodisk, hvis $\omega_0 T/2\pi$ er et rationelt tal. Tilsvarende gælder diskrete følger opnået ved sampling af sinus- eller cosinussignaler. F.eks. hvis

$$x(t) = \cos(2\pi t) \tag{3.16}$$

så kan de tre signaler på figur 3.4 opfattes som værende defineret ved

$$x[n] = x(nT) = \cos(2\pi nT) \tag{3.17}$$

for forskellige valg af T. Specielt er $T = \frac{1}{12}$ for figur (a), $T = \frac{4}{31}$ for figur (b) og $T = \frac{1}{12}\pi$ for figur (c). Selv om $x[n]$ ikke selv er periodisk, kan dets indhyllingskurve godt være det. Det ses på figur (c).

3.1.2 LFI systemers svar på komplekse eksponentialsignaler

Ligesom for de kontinuerte signalers vedkommende vil vi se på LFI-systemers svar på diskrete eksponentialsignaler $x[n]$. Antag, at et LFI system med impulssvaret $h[n]$ påtrykkes signalet

$$x[n] = z^n \tag{3.18}$$

på indgangen, hvor z er et komplekst tal. Systemets output kan findes ved hjælp af foldningssummen som

$$y[n] = x * h[n] = \sum_{k=-\infty}^{\infty} h[k]x[n-k] = \sum_{k=-\infty}^{\infty} h[k]z^{n-k} = z^n \sum_{k=-\infty}^{\infty} h[k]z^{-k} \tag{3.19}$$

Herved ses, at hvis input $x[n]$ er et kompleks eksponentialsignal givet ved $x[n] = z^n$, så er output det samme komplekse eksponentialsignal multipliceret med en konstant, der afhænger af z's værdi, d.v.s.

$$y[n] = H(z)z^n, \tag{3.20}$$

hvor

$$H(z) = \sum_{k=-\infty}^{\infty} h[k]z^{-k}. \tag{3.21}$$

Denne ligning, sammen med superpositionsegenskaben, kan bruges til let at kunne bestemme et LFI systems output på en linearkombination af komplekse eksponentialsignaler. D.v.s., hvis

$$x[n] = \sum_k a_k z_k^n \tag{3.22}$$

så bliver systemets output

$$y[n] = \sum_k a_k H(z_k) z_k^n \tag{3.23}$$

Med andre ord: output kan også fremstilles som en linearkombination af de samme komplekse eksponentialsignaler, og hver koefficient i denne fremstilling af output fås som produktet af den tilsvarende koefficient a_k i input og $H(z_k)$ svarende til z_k^n

3.2 Periodiske signalers Fourierrækkerefremstilling

Hvis et diskret signal er periodisk, f.eks. med perioden N, så er der kun N forskellige signaler i mængden

$$\phi_k[n] = e^{ik(2\pi/N)n}, k = 0, \pm 1, \pm 2, \ldots \tag{3.24}$$

Tilsvarende er en linearkombination af $\phi_k[n]$ af formen

$$x[n] = \sum_k a_k \phi_k[n] = \sum_k a_k e^{ik(2\pi/N)n} \tag{3.25}$$

periodisk med perioden N, og derfor er værdierne $\phi_k[n]$ kun forskellige i et interval af længden N for k, og summationen behøver derfor kun at udstrækkes over dette interval.

D.v.s. at summationen i ligning (3.25) er over k, hvor k gennemløber et interval bestående af N konsekutive heltal. Disse summationsgrænser skrives kort således: $k =< N >$. D.v.s.

$$x[n] = \sum_{k=<N>} a_k \phi_k[n] = \sum_{k=<N>} a_k e^{ik(2\pi/N)n} \tag{3.26}$$

k kan for eksempel antage værdierne $k = 0, 1, \ldots, N-1$ eller $k = 3, 4, \ldots, N+2$. På grund af periodiciteten vil nøjagtig de samme led indgå i summationen på højresiden af ligning (3.26). Ligning (3.26) betegnes *den diskrete Fourierrække* og koefficienterne a_k *Fourierrækkekoefficienter*.

Hvis der er givet en talsekvens $x[n]$, der er periodisk med perioden N, kan man undersøge, om der findes en fremstilling som ligning (3.26) og i givet fald hvad koefficienterne a_k er. Ved at evaluere (3.26) for værdier af n ses, at det svarer til at finde en løsning til det lineære ligningssystem

$$\begin{aligned} x[0] &= \sum_{k=<N>} a_k \\ x[1] &= \sum_{k=<N>} a_k e^{i2\pi k/N} \\ &\vdots \\ x[N-1] &= \sum_{k=<N>} a_k e^{i2\pi k(N-1)/N} \end{aligned}$$

Ligningen for $x[N]$ er identisk med den for $x[0]$ (fordi begge sider af (3.26) er periodiske med perioden N). D.v.s. vi har N lineære ligninger for N ukendte koefficienter a_k, hvor k løber over N på hinanden følgende heltal. Det kan vises, at ligningerne er lineært uafhængige og derfor kan løses, hvorved koefficienterne a_k kan udtrykkes ved de givne værdier af $x[n]$.

Det er muligt –ligesom for den kontinuerte transformations vedkommende – at udlede et eksplicit udtryk for koefficienterne a_k bestemt ved talfølgens værdier $x[n]$. Først vises, at

$$\sum_{n=0}^{N-1} e^{ik(2\pi/N)n} = \begin{cases} N, & k = 0, \pm N, \pm 2N, \ldots \\ 0, & \text{ellers} \end{cases} \tag{3.27}$$

Hvad denne ligning udsiger, er, at summen over en periode af værdierne af en kompleks eksponentialfunktion er nul med mindre den komplekse eksponentialfunktion er en konstant, d.v.s. ligningen er den diskrete analogi til ligningen (jfr. ligning(2.14))

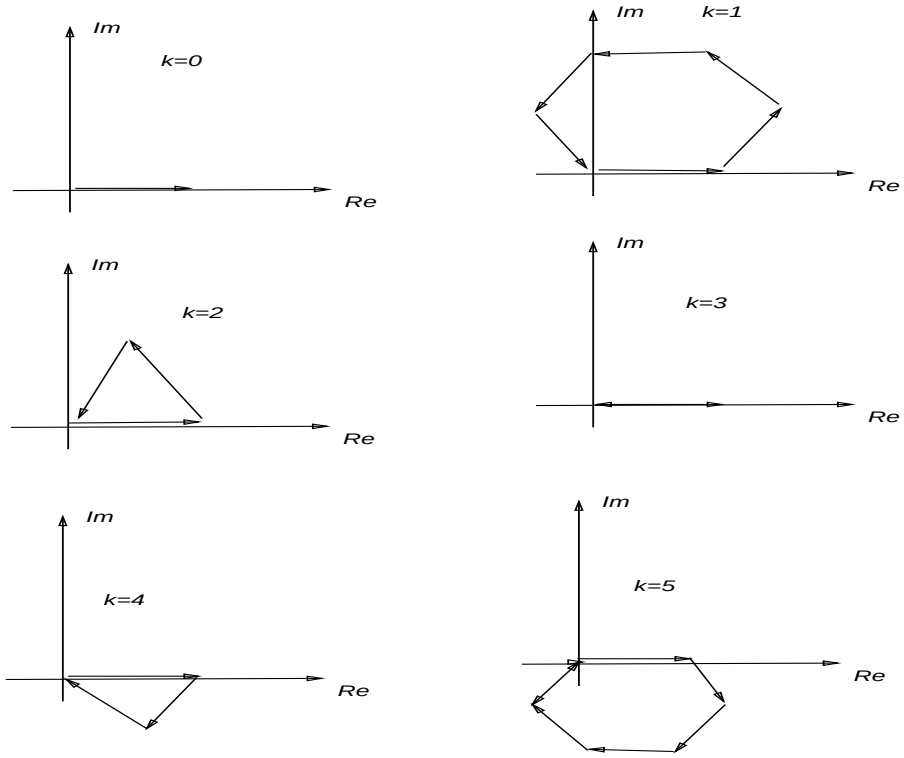
$$\int_0^T e^{ik(2\pi/T)t} dt = \begin{cases} T, & k = 0 \\ 0, & \text{ellers} \end{cases} \tag{3.28}$$

For at udlede ligning (3.27) bemærkes først, at ligningens venstreside er en sum af et endeligt antal led i en kvotientrække. Den har formen

$$\sum_{n=0}^{N-1} \alpha^n \tag{3.29}$$

hvor $\alpha = e^{ik(2\pi/N)}$. Denne sum er

$$\sum_{n=0}^{N-1} \alpha^n = \begin{cases} N, & \alpha = 1 \\ \frac{1-\alpha^N}{1-\alpha}, & \alpha \neq 1 \end{cases} \tag{3.30}$$



Figur 3.5: Illustration af sumformlen

Idet $e^{ik(2\pi/N)} = 1$ kun når k er et multiplum af N ($k = 0, \pm N, \pm 2N, \ldots$) fås

$$\sum_{n=0}^{N-1} \alpha^n = \begin{cases} N, & k = 0, \pm N, \pm 2N, \ldots \\ \frac{1-e^{ik(2\pi/N)N}}{1-e^{ik(2\pi/N)}}, & \text{ellers} \end{cases} \tag{3.31}$$

hvilket giver (3.30), idet $e^{ik(2\pi/N)N} = 1$. Idet hver af de komplekse eksponentialfunktioner i ligning (3.27) er periodiske med perioden N gælder ligning (3.27), når summationen sker over et vilkårligt interval af længden N, d.v.s.

$$\sum_{n=<N>} e^{ik(2\pi/N)n} = \begin{cases} N, & k = 0, \pm N, \pm 2N, \ldots \\ 0, & \text{ellers} \end{cases} \tag{3.32}$$

En grafisk fortolkning heraf er vist på figur 3.5 for $N = 6$. Her er tallene repræsenteret som vektorer i den komplekse plan, og hver enkelt vektor er en enhedsvektor (har længden 1). Det ses, at summen er nul, med mindre $k = 0, 6, 12, \ldots$ o.s.v. Ved at multiplicere Fourierrækkefremstillingen (3.26) på begge sider med $e^{-ir(2\pi/N)n}$ og summere over N led fås

$$\sum_{n=<N>} x[n] e^{-ir(2\pi/N)n} = \sum_{n=<N>} \sum_{k=<N>} a_k e^{i(k-r)(2\pi/N)n} \tag{3.33}$$

Ved at ombytte summationsrækkefølgen på højresiden fås

$$\sum_{n=<N>} x[n] e^{-ir(2\pi/N)n} = \sum_{k=<N>} a_k \sum_{n=<N>} e^{i(k-r)(2\pi/N)n} \tag{3.34}$$

Den inderste sum (med indeks n) er ifølge lign. (3.27) nul med mindre $k-r$ er nul eller et helt multiplum af N. Hvis man derfor bruger det samme summationsinterval for r og k, så er denne sum lig N, hvis $k=r$ og 0, hvis $k \neq r$. Højresiden af ligningen reduceres derfor til Na_r og man har

$$a_r = \frac{1}{N} \sum_{n=<N>} x[n] e^{-ir(2\pi/N)n} \tag{3.35}$$

Ved hjælp af denne formel kan Forierrækkens koefficienter udregnes. Fourierrækkeformlen for diskrete signaler ser derfor således ud:

$$x[n] = \sum_{k=<N>} a_k e^{ik(2\pi/N)n} \tag{3.36}$$

$$a_k = \frac{1}{N} \sum_{n=<N>} x[n] e^{-ik(2\pi/N)n} \tag{3.37}$$

3.3 Den diskrete Fouriertransformation

Den diskrete Fouriertransformation (DFT) for signaler af endelig længde er en vigtig metode til Fourieranalyse af diskrete sekvenser og desuden er den velegnet til implementation på en datamaskine, i form af Fast Fourier Transform (se næste kapitel). Hvis $x[n]$ er et signal af endelig længde, d.v.s. der findes et heltal N_1, således at $x[n]=0$ udenfor intervallet $0 \leq n \leq N_1 - 1$, så kan man konstruere et periodisk signal $\tilde{x}[n]$, som er lig med $x[n]$ over en periode. Mere præcist, lad $N \geq N_1$ være et givet heltal og $\tilde{x}[n]$ være periodisk med perioden N således at

$$\tilde{x}[n] = x[n],\ 0 \leq n \leq N-1 \tag{3.38}$$

$\tilde{x}[n]$'s Fourierrækkekoefficienter er givet ved

$$a_k = \frac{1}{N} \sum_{n=<N>} \tilde{x}[n] e^{-ik(2\pi/N)n} \tag{3.39}$$

Ved at vælge summationsintervallet til at være det, hvor $\tilde{x}[n] = x[n]$ fås

$$a_k = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-ik(2\pi/N)n} \tag{3.40}$$

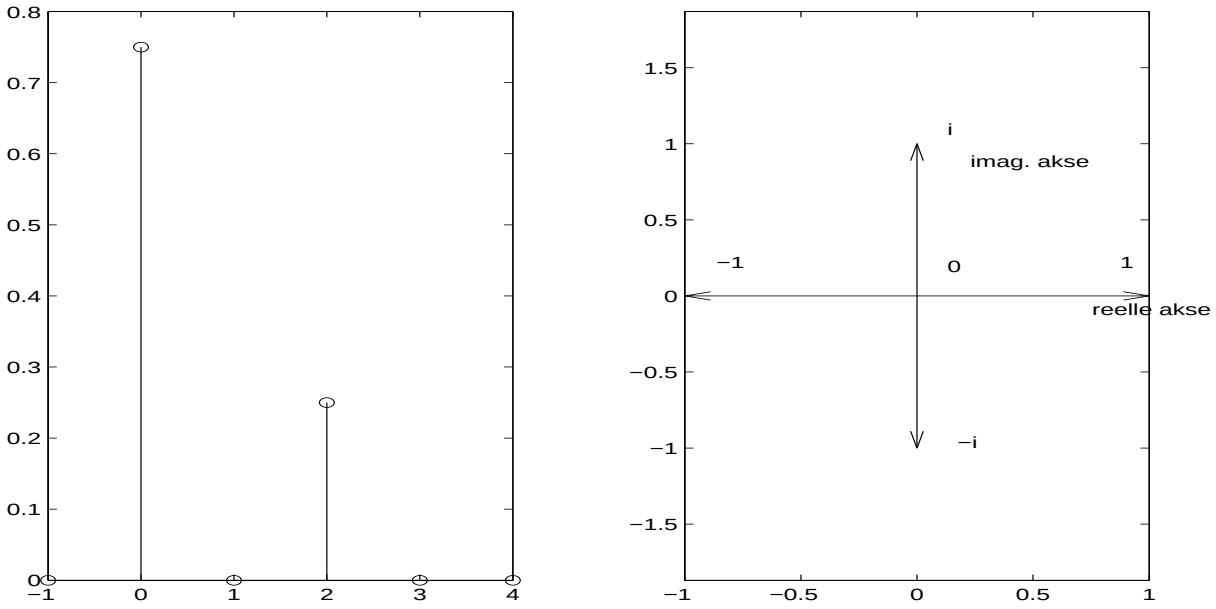
Hvis DFT af $x[n]$ betegnes $\tilde{X}(k)$, kan analyseligningen skrives

$$\tilde{X}(k) = a_k = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-ik(2\pi/N)n},\ k = 0, 1, \ldots, N-1 \tag{3.41}$$

og synteseligningen

$$x[n] = \sum_{k=0}^{N-1} \tilde{X}(k) e^{ik(2\pi/N)n},\ k = 0, 1, \ldots, N-1 \tag{3.42}$$

Dette er udtrykket for den diskrete Fouriertransformation. I det næste kapitel skal vi se på beregningsalgoritmen (Fast Fourier Transform – FFT) ved hjælp af hvilken beregningskompleksiteten af DFT-udregningen kan reduceres væsentligt.



Figur 3.6: (a) Diskret signal til eksempel 1 og (b) enhedsrødderne

Eksempel 1 Vi vil beregne den diskrete Fouriertransformation af signalet

$$x[n] = \begin{cases} x[0] = \frac{3}{4} \\ x[1] = 0 \\ x[2] = \frac{1}{4} \\ x[3] = 0 \end{cases} \tag{3.43}$$

vist på figur 3.6(a) Vi kan opskrive $x[n]$ kort på vektorform således: [3/4 0 1/4 0]. N er lig med 4 i dette eksempel. DFT beregnes ved hjælp af ligning (3.41). Faktoren $e^{-2\pi i/N}$, der indgår i udtrykket på ligningen højreside, kaldes den komplekse (primitive) enhedsrod svarende til $N = 4$. Det er anskueligt at afbilde den og dens potenser, svarende til værdierne af $n = 0, 1, 2, 3$ i et Arganddiagram (tegning af den komplekse plan), hvilket ses på figur 3.6(b). Svarende til værdierne af k kan man opstille følgende tabel for $e^{-ik(2\pi/N)n}$:

n	0	1	2	3
k=0	1	1	1	1
k=1	1	-i	-1	i
k=2	1	-1	1	-1
k=3	1	i	-1	-i

Det er nu forholdvis let at udregne $\tilde{X}(k), k = 0, 1, 2, 3$ ved produktet af vektoren [3/4 0 1/4 0] og tabellens rækker. Man finder f.eks.

$$\tilde{X}(2) = \frac{1}{4}\{1 \cdot (3/4) - 1 \cdot 0 + 1 \cdot (1/4) - 1 \cdot 0\} = \frac{1}{4} \tag{3.44}$$

Ved udregning af alle værdier findes den diskrete Fouriertransformation af [3/4 0 1/4 0] til (1/4)[1 1/2 1 1/2].

3.4 Opgaver

1. Der er givet et LFI-system med impulssvaret $h[n] = \alpha^n u[n]$, hvor $-1 < \alpha < 1$. Systemet påtrykkes signalet

$$x[n] = \cos\left(\frac{2\pi n}{N}\right) \tag{3.45}$$

   Find Fourierrækkekoefficienterne for output og bestem $y[n]$ for $N = 4$

2. Beregn den diskrete Fouriertransformerede af [3/4 0 1/4 0 0 0 0 0].

3. Beregn den inverse diskrete Fouriertransformerede af [4 12 12 12].

Kapitel 4

FFT: Den hurtige Fouriertransformation

Den diskrete Fouriertransformation (DFT) spiller en vigtig rolle i analyse, konstruktion og implementation af diskrete signal og billedbehandlingssystemer. Både på grund af anvendeligheden til frekvensanalyse, og fordi der eksisterer en effektiv algoritme til beregning af DFT, er diskret Fourieranalyse blevet vidt udbredt.

I dette kapitel skal vi se på metoder til beregning af DFT. Vi vil betragte den effektive algoritme til beregning af DFT. Der er mange måder at måle algoritmisk kompleksitet og effektivitet på, og den endelige vurdering afhænger af den teknologi, der står til rådighed og den planlagte anvendelse. Her vil vi anvende antallet af multiplikationer og additioner som et mål for algoritmisk kompleksitet. Dette mål er simpelt at anvende, og antallet af multiplikationer og additioner står i direkte forhold til beregningshastigheden, når algoritmerne implementeres på almindelige datamaskiner eller mikroprocessorer. Undertiden er andre mål mere relevante; f.eks. ved visse VLSI-implementationer er chip-arealet ofte den mest vigtige faktor, og chiparealet er ikke nødvendigvis direkte relateret til antallet af aritmetiske operationer.

Med hensyn til multiplikationer og additioner er klassen af FFT-algoritmer størrelsesordener bedre end konkurrerende algoritmer. FFT-algoritmens effektivitet er faktisk så høj, at i mange tilfælde er den mest effektive måde at beregne foldning på, at transformere de talfølger, der skal foldes, multiplicere deres transformerede og så beregne den inverse transformation af produktet af transformationerne.

4.1 Effektiv beregning af den diskrete Fouriertransformation

Som defineret i formel (3.41) er den diskrete Fouriertransformation (DFT) af en talfølge af endelig længde N

$$X[k] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-i\frac{2\pi}{N}kn}, \;\; k = 0, 1, \ldots, N-1 \tag{4.1}$$

Idet man sætter $W_N = e^{-i\frac{2\pi}{N}}$ kan formlen skrives

$$X[k] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] W_N^{kn}, \ k = 0, 1, \ldots, N-1 \tag{4.2}$$

Den inverse Fouriertransformation er givet ved

$$x[n] = \sum_{k=0}^{N-1} X[k] W_N^{-kn}, \ n = 0, 1, \ldots, N-1 \tag{4.3}$$

I ligning (4.2) og (4.3) kan både x[n] og X[k] være komplekse tal. Da de to ligninger kun adskiller sig med hensyn til fortegnet for eksponenten for W_N og skalafaktoren $\frac{1}{N}$, er de beregningsmæssige overvejelser ens for de to ligninger.

Først betragtes beregningskompleksiteten ved direkte udregning af DFT-ligningen (4.2). Da $x[n]$ kan være kompleks, kræves N komplekse multiplikationer og $(N-1)$ komplekse additioner til beregning af hver værdi af DFT, hvis ligning (4.2) anvendes direkte som beregningsformel. Til udregning af N værdier kræves derfor i alt N^2 komplekse multiplikationer og $N(N-1)$ komplekse additioner. Hver kompleks multiplikation kræver fire reele multiplikationer og to reele additioner, og hver kompleks addition kræver to reele additioner. Derfor kræver den direkte udregning af $X[k]$ for hver værdi af k $4N$ reele multiplikationer og $(4N-2)$ reele additioner [1]. Da $X[k]$ skal beregnes for N forskellige værdier af k, kræver den direkte udregning af den diskrete Fouriertransformation af en talfølge $x[n]$ $4N^2$ reele multiplikationer og $N(4N-2)$ reele additioner. Foruden de multiplikationer og additioner, der er nødvendige, kræves lagerplads til lagring af de N komplekse inputværdier samt de N komplekse koefficienter W_N^{kn}. Da antallet af beregninger og dermed beregningstiden ca. er proportional med N^2, er det klart, at antallet af aritmetiske operationer for beregning af DFT ved den direkte metode bliver meget stort for store værdier af N. Derfor er man interesseret i beregningsprocedurer, der formindsker antallet af multiplikationer og additioner.

De fleste metoder til effektivitetsforbedring ved udregning af DFT er baseret på udnyttelse af W_N^{kn}'s symmetriegenskaber, specielt

1. $W_N^{k(N-n)} = W_N^{-kn} = (W_N^{kn})^*$ (kompleks konjugeret symmetri)
2. $W_N^{kn} = W_N^{k(n+N)} = W_N^{(k+N)n}$ (periodicitet i n og k).

Den første egenskab kan benyttes til at reducere antallet af multiplikationer ca. med en faktor 2. Desuden vil, for visse værdier af produktet kn, de sinus- og cosinusfunktioner, der defineres ved den komplekse eksponentialfunktion, antage værdien 0 eller 1, hvorved multiplikation kan undgås. Den største beregningsbesparelse ligger dog i periodiciteten af følgen W_N^{kn}, som kan benyttes til den store beregningsreduktion fra $O(N^2)$ til $O(N \log_2 n)$.

FFT algoritmerne bygger på den grundlæggende egenskab at opdele beregningen af den diskrete Fouriertransformation af en talfølge af længden N i stedse mindre diskrete

[1] I denne diskussion er tallene for antallet af beregninger cirka-tal. F.eks. kræver multiplikation med W_N^0 faktisk ikke en multiplikation. Men vurderingen af de beregningsmæssige kompleksitet ved disse approksimative metoder er alligevel tilstrækkelig god til at tillade sammenligning mellem de forskellige klasser af algoritmer.

Fouriertransformationer. Den måde, hvorpå dette princip gennemføres, giver en række forskellige algoritmer med sammenlignelige forøgelser af beregningshastigheden. Vi vil her kun se på en typisk algoritme, hvor man anvender opdeling i tid (engelsk: *decimation-in-time*), hvor tidssekvensen opdeles i stadig mindre delsekvenser.

4.2 FFT algoritme

Del-og-hersk princippet går ud på at dekomponere beregningen i stadig mindre DFT-beregninger. I denne proces vil vi udnytte både symmetri-egenskaberne og periodiciteten af den komplekse eksponentialfunktion $e^{-i(2\pi/N)kn}$. Algoritmer, i hvilke dekompositionen er baseret på opdelingen af talfølgen $x[n]$ i stadig mindre delfølger, kaldes *opdeling i tid* (engelsk: decimation-in-time).

Princippet ved denne opdeling demonstreres mest bekvemt ved at betragte specialtilfældet, hvor N er en heltalling 2-er potens, d.v.s $N = 2^\gamma$. Da N er et lige tal kan beregningen af $X[k]$ ske ved at opdele $x[n]$ i to $(\frac{N}{2})$-punkts [2] talfølger bestående af punkterne med lige indeks i $x[n]$ og punkterne med ulige indeks i $x[n]$. Idet $X[k]$ er givet ved

$$X[k] = \sum_{n=0}^{N-1} x[n] W_N^{nk}, \;\; k = 0, 1, \ldots, N-1, \tag{4.4}$$

og ved at opdele $x[n]$ i dens lige- og ulige-indekserede punkter fås

$$X[k] = \sum_{n \;\; lige} x[n] W_N^{nk} + \sum_{n \;\; ulige} x[n] W_N^{nk}. \tag{4.5}$$

Når n er lige, kan n skrives $n = 2r$ og når n er ulige kan n skrives $n = 2r + 1$. Hvis dette indføres i den ovennævnte formel, fås

$$\begin{aligned} X[k] &= \sum_{r=0}^{(N/2)-1} x[2r] W_N^{2rk} + \sum_{r=0}^{(N/2)-1} x[2r+1] W_N^{(2r+1)k} && (4.6) \\ &= \sum_{r=0}^{(N/2)-1} x[2r] (W_N^2)^{rk} + W_N^k \sum_{r=0}^{(N/2)-1} x[2r+1] (W_N^2)^{rk}. && (4.7) \end{aligned}$$

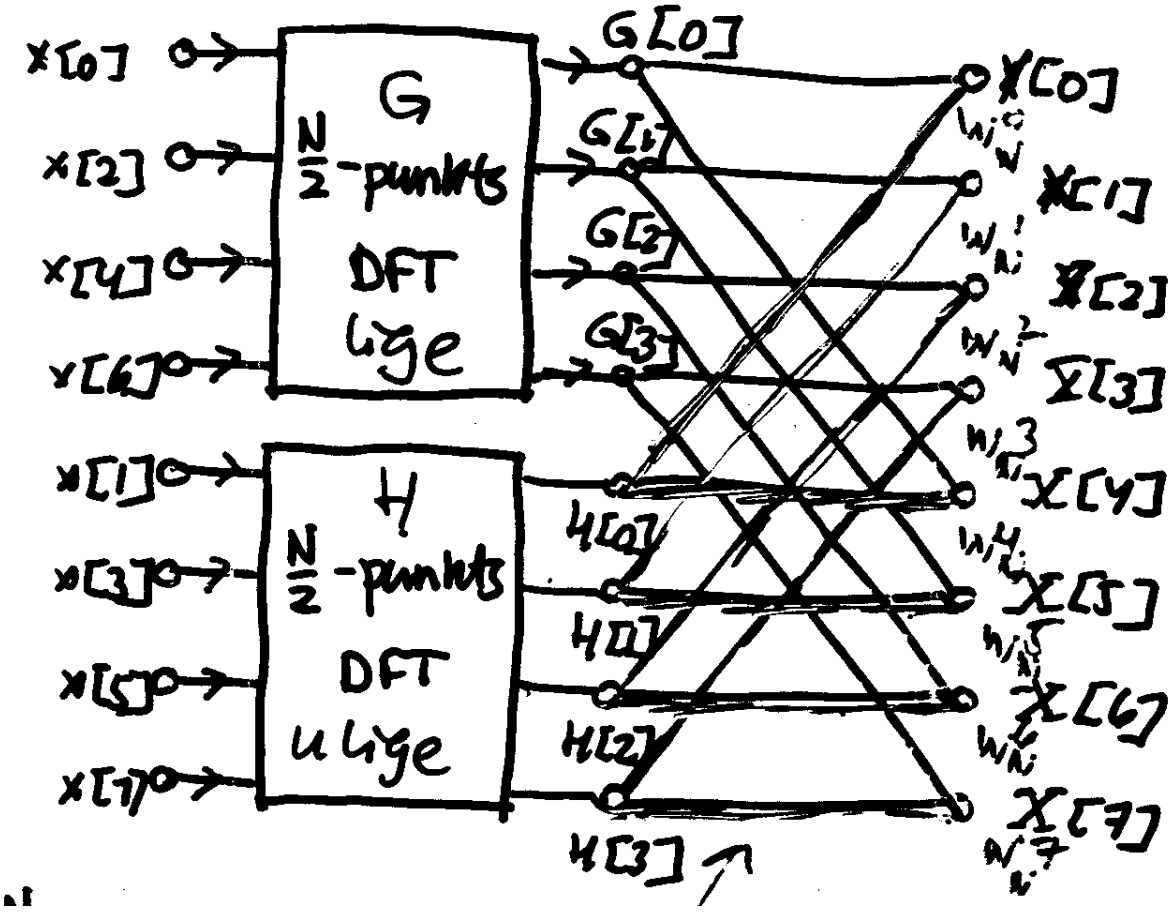
Men man kan benytte, at

$$W_N^2 = \left(e^{i\frac{2\pi}{N}}\right)^2 = e^{-2i\frac{2\pi}{N}} = e^{-i\frac{2\pi}{N/2}} = W_{\frac{N}{2}} \tag{4.8}$$

Med brug heraf kan ligning (4.7) omskrives til

$$\begin{aligned} X[k] &= \sum_{r=0}^{\frac{N}{2}-1} x[2r] W_{\frac{N}{2}}^{rk} + W_N^k \sum_{r=0}^{\frac{N}{2}-1} x[2r+1] W_{\frac{N}{2}}^{rk} && (4.9) \\ &= G[k] + W_N^k H[k] && (4.10) \end{aligned}$$

[2] Ved gennemgang af teorien for FFT-algoritmer bruges både ordet *sample* og *punkt* om værdierne i talfølgen. Desuden kaldes en talfølge af længden N en N-punkts følge, og en DFT af længden N kaldes en N-punkts DFT.



Figur 4.1: Signalgraf af dekompositionen af en N-punkts DFT i to $\frac{N}{2}$-punkts DFT beregninger ($N = 8$)

Hver af summerne i ligning(4.9) er en $\frac{N}{2}$-punkts DFT. Den første sum er en $\frac{N}{2}$-punkts DFT af punkterne med lige indeks i den oprindelige talfølge, mens den anden er en en $\frac{N}{2}$-punkts DFT af de ulige indekserede punkter i den oprindelige følge. Skønt indeks k løber over N værdier $k = 0, 1, \ldots, N-1$, skal hver af de to summer kun beregnes for k løbende mellem 0 og $\frac{N}{2} - 1$, idet $G[k]$ og $H[k]$ hver især er periodiske med perioden $\frac{N}{2}$. Når de to DFT'er er beregnet, skal de kombineres svarende til ligning (4.9) for at kunne give N-punkts DFT'en $X[k]$. Figur 4.1 viser denne beregning for $N = 8$. På denne figur bruges signalgraf vedtægterne, der tidligere er blevet indført. Kanter, der leder ind til knuder, summeres for at give knudevariablen. Når ingen koefficienter er angivet, antages kantforstærkningen at være 1. For de øvrige kanter er kantforstærkningen en heltalspotens af W_N.

På figur 4.1 er to 4-punkts DFT'er beregnet, hvor $G[k]$ angiver 4-punkts DFT'en af punkter med lige indeks, og $H[k]$ angiver 4-punkts DFT'en af punkter med ulige indeks. Herefter fås $X[0]$ ved at multiplicere $H[0]$ med W_N^0 og addere produktet til $G[0]$. $X[1]$ fås ved at multiplicere $H[1]$ med W_N^1 og addere resultatet til $G[1]$. Ligning (4.9) udsiger, at for

at beregne $X[4]$ skal $H[4]$ multipliceres med W_N^4 og resultatet skal adderes til $G[4]$. Men da både $G[k]$ og $H[k]$ er periodiske i k med perioden 4, så er $H[4] = H[0]$ og $G[4] = G[0]$. Det vil sige, at $X[4]$ fås ved at multiplicere $H[0]$ med W_N^4 og addere resultatet til $G[0]$. Som vist på figur 4.1 fås værdierne for $X[5]$, $X[6]$ og $X[7]$ på tilsvarende vis.

Når beregningerne omordnes som vist i ligning (4.9) kan man sammenligne det nødvendige antal multiplikationer og additioner med det antal, der er nødvendigt ved en direkte udregning af DFT'en. Tidligere så vi, at til direkte udregning uden udnyttelse af symmetrien krævedes N^2 komplekse multiplikationer og additioner [3]. Til sammenligning kræver ligning (4.9) beregning af to $\frac{N}{2}$-punkts DFT'er, hvilket kræver $2(\frac{N}{2})^2$ komplekse multiplikationer og ca. $2(\frac{N}{2})^2$ komplekse additioner, hvis de to $\frac{N}{2}$ DFT'er udregnes ved den direkte metode. Herefter skal de to $\frac{N}{2}$ punkts DFT'er kombineres, hvilket kræver N komplekse multiplikationer, svarende til at multiplicere den sidste sum med W_N^k, og N komplekse additioner, svarende til at addere produktet til den første sum. Følgelig kræver beregningen af ligning (4.9) for alle værdier af k højst $N + 2(\frac{N}{2})^2$ eller $N + \frac{N^2}{2}$ komplekse multiplikationer og komplekse additioner. Det er let at eftervise, at for $N > 2$ er totalen $N + \frac{N^2}{2}$ mindre end N^2.

Ligning (4.10) svarer til at opdele den originale N-punkts beregning i to $(\frac{N}{2})$-punkts DFT-beregninger. Hvis $\frac{N}{2}$ er lige, hvilket den er når N er lig en potens af 2, så kan hver af de $(\frac{N}{2})$-punkts DFT-er i ligning (4.9) beregnes ved at opsplitte hver af summerne i ligning (4.10) i to $(\frac{N}{4})$-punkts DFT-er, som herefter kan kombineres til at give $(\frac{N}{2})$-punkts DFT-erne. Derfor kan $G[k]$ i ligning (4.10) skrives som

$$G[k] = \sum_{r=0}^{\frac{N}{2}-1} g[r] W_{\frac{N}{2}}^{rk} = \sum_{l=0}^{\frac{N}{4}-1} g[2l] W_{\frac{N}{2}}^{2lk} + \sum_{l=0}^{\frac{N}{4}-1} g[2l+1] W_{\frac{N}{2}}^{(2l+1)k} \tag{4.11}$$

eller

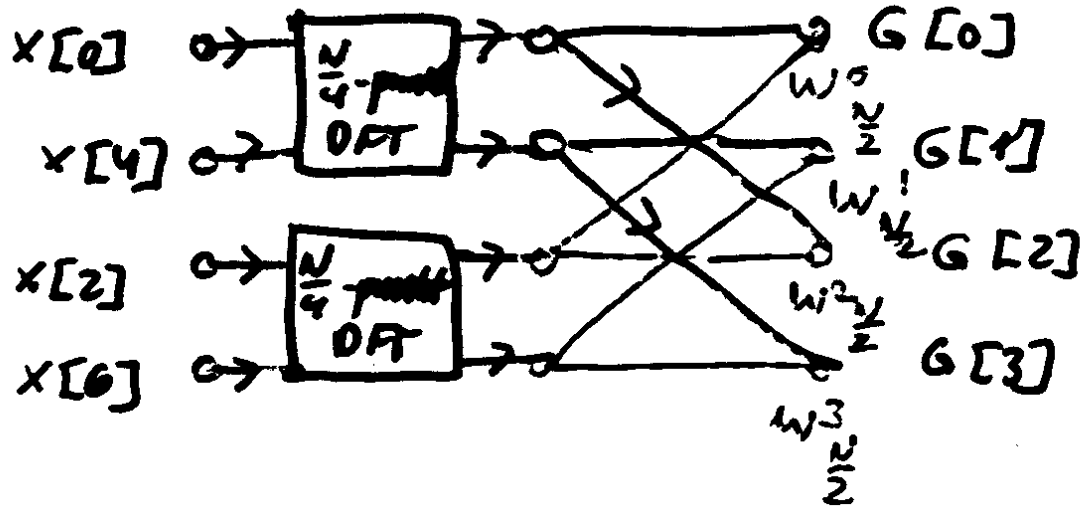
$$G[k] = \sum_{l=0}^{\frac{N}{4}-1} g[2l] W_{\frac{N}{4}}^{lk} + W_{\frac{N}{2}}^{k} \sum_{l=0}^{\frac{N}{4}-1} g[2l+1] W_{\frac{N}{4}}^{lk} \tag{4.12}$$

På samme måde kan $H[k]$ skrives

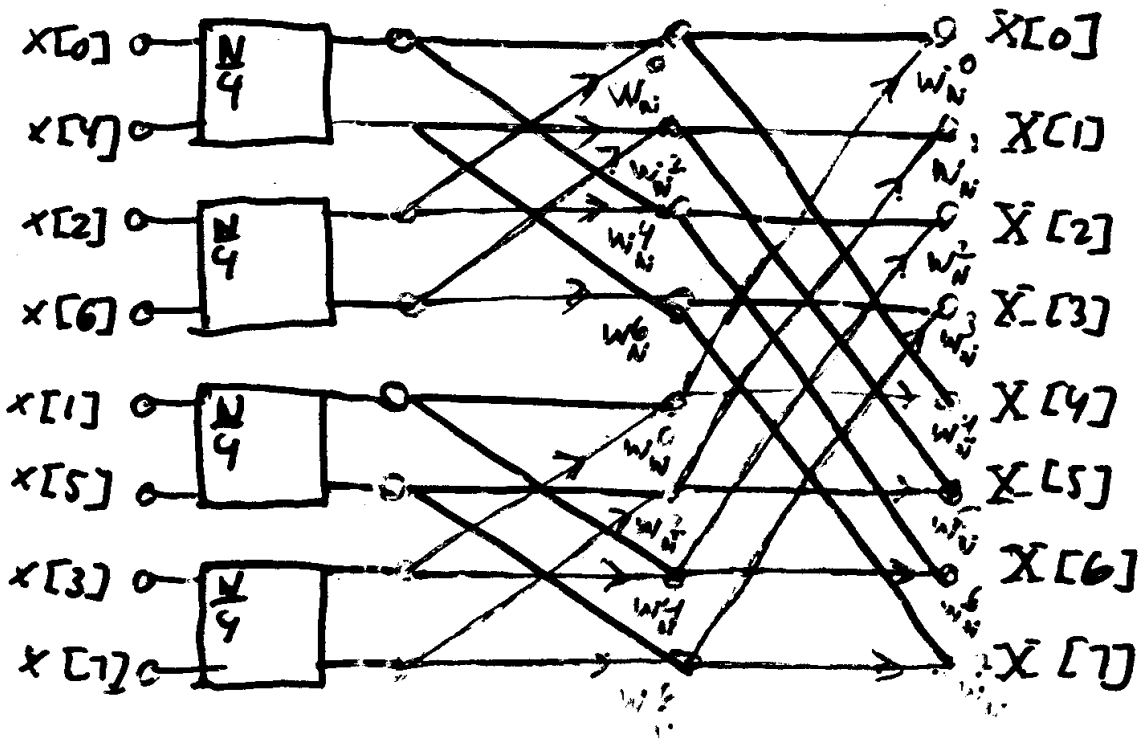
$$H[k] = \sum_{l=0}^{\frac{N}{4}-1} h[2l] W_{\frac{N}{4}}^{lk} + W_{\frac{N}{2}}^{k} \sum_{l=0}^{\frac{N}{4}-1} h[2l+1] W_{\frac{N}{4}}^{lk} \tag{4.13}$$

Det vil sige, at $\frac{N}{2}$-punkts DFT'en $G[k]$ kan fås ved at kombinere $\frac{N}{4}$-punkts DFT-erne fra følgerne $g[2l]$ og $g[2l+1]$. På samme måde kan $\frac{N}{2}$-punkts DFT'en $H[k]$ fås ved at kombinere $\frac{N}{4}$-punkts DFT-erne fra følgerne $h[2l]$ og $h[2l+1]$. Det betyder, at beregningen kan udføres som vis på figur 4.2. Hvis den beregning, der er vist på figur 4.2 indsættes i signalgrafen fra figur 4.1 fås den komplette signalgraf vist på figur 4.3, hvor koefficienterne er udtrykt i potenser af W_N i stedet for $W_{\frac{N}{2}}$, idet $W_{\frac{N}{2}} = W_N^2$. For den 8-punkts DFT, der er anvendt som illustration er beregningen reduceret til en beregning af en 2-punkts DFT, som let udregnes som vist på figur 4.4. Hvis beregningen svarende til figur 4.4 indsættes i signalgrafen figur 4.3 fås den komplette signalgraf for beregningen af 8-punkts DFT'en, som vist på figur 4.5.

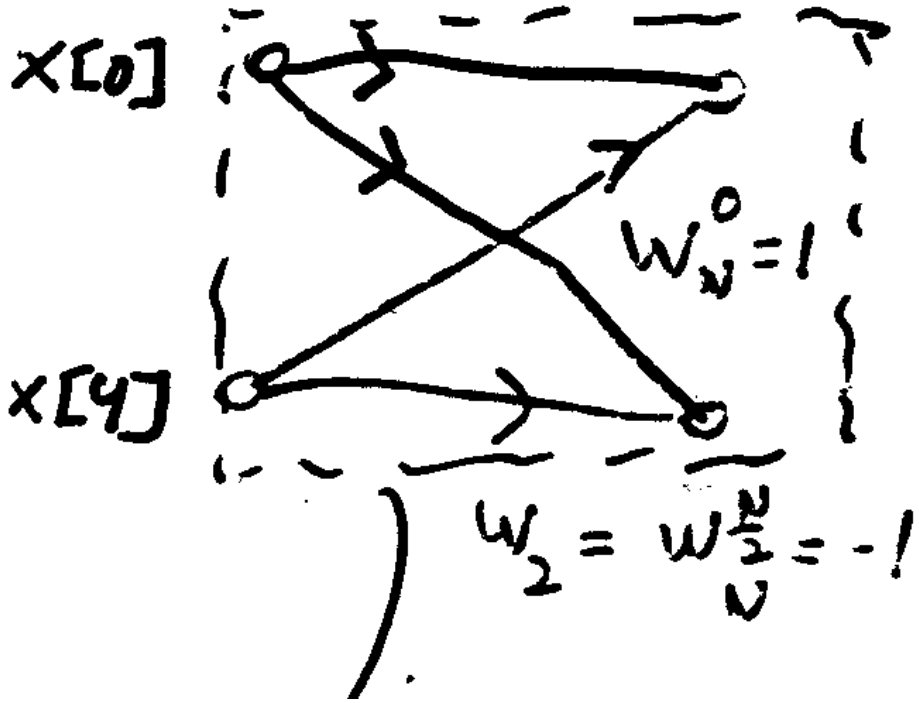
[3] For enkelhedens skyld antages at N er så stor, at $(N-1)$ kan tilnærmes med N



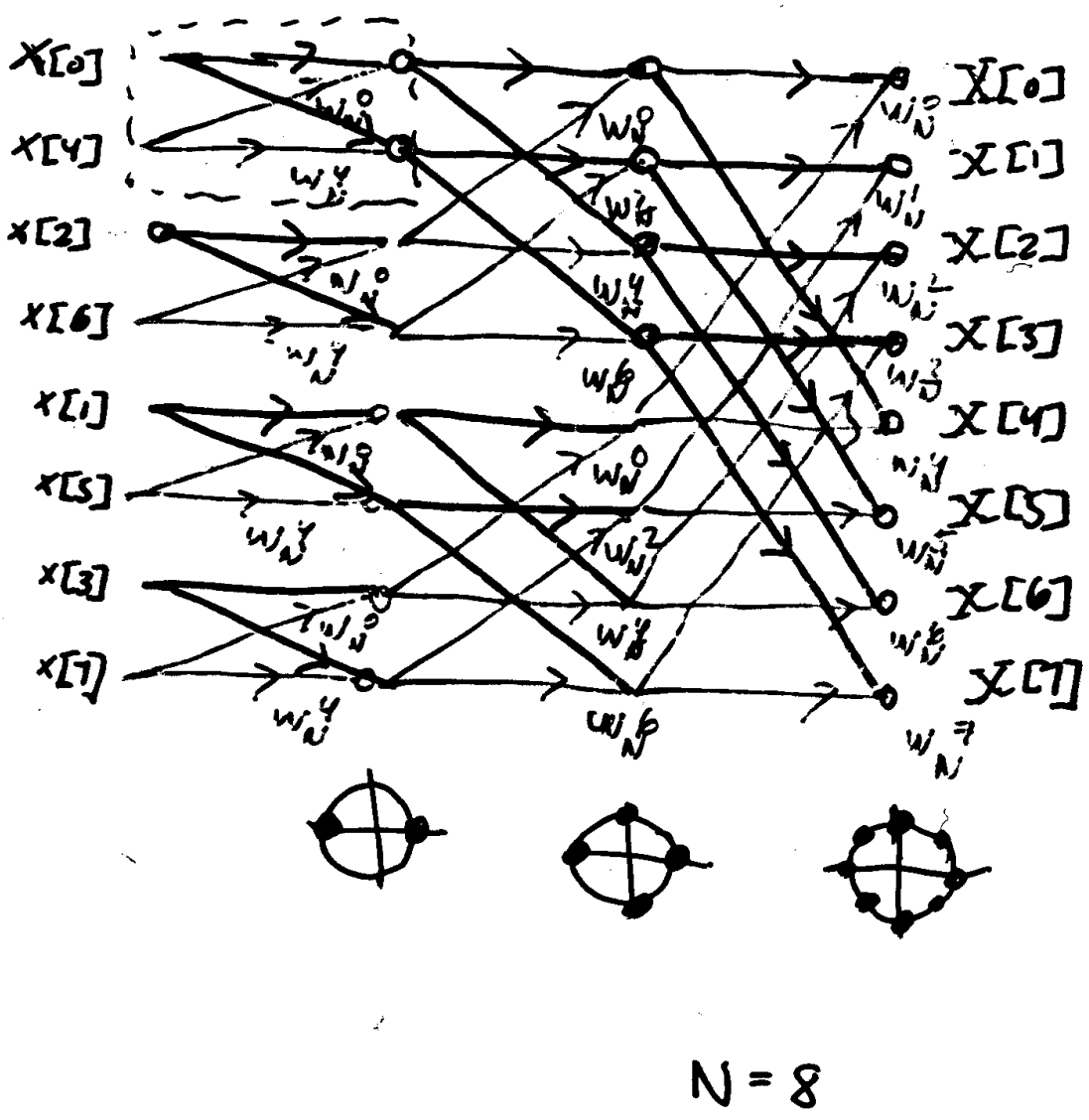
Figur 4.2: Signalgraf, der viser dekompositionene af en $\frac{N}{2}$-punkts DFT-beregning i to $\frac{N}{4}$-punkts beregninger (N=8).



Figur 4.3: Resultatet af at indsætte figur 4.2 i figur 4.1 (N=8).



Figur 4.4: Signalgraf for en 2-punkts DFT



Figur 4.5: Signalgraf for den komplette beregning af en 8-punkts DFT

For det almindelige tilfælde, men hvor N stadig er en 2-er potens, går man frem ved bestandigt at underinddele med en faktor 2 indtil der kun er en 2-punkts transformation tilbage. Signalgrafen på figur 4.5 viser eksplicit beregningsoperationerne. Ved optælling af grene med faktorer af formen W_N^r benmærkes, at der i hvert trin er N komplekse multiplikationer og N komplekse aditioner. Da der er $\log_2 N$ trin, er der i alt $n \log_2 N$ komplekse multiplikationer og additioner. Dette er en betragtelig besparelse. F.eks. hvis $N = 2^{10} = 1024$ så er $N^2 = 2^{20} = 1.048.576$, hvorimod $N \log_2 N = 10.240$.

4.3 Opgaver

1. Antag, at vi skal beregne en FFT på 16 punkter, men kun $x[0]$ og $x[1]$ er forskellige fra nul. Alle øvrige $x[n] = 0$. Optegn signalgrafen og bestem antallet af operationer.

2. Antag, at vi skal beregne en FFT på 16 punkter og vi ønsker at spare lagerplads. Vi ønsker at kunne benytte den samme array for input- og outputværdierne. Kan det lade sig gøre ? I bekræftende fald, hvorledes skal der flyttes om på outputværdierne for at de bliver lagret i en naturlig rækkefølge ?

3. At beregne en DFT kræver sædvanligvis komplekse multiplikationer. Betragt produktet $x + iy = (a + ib)(c + id) = (ac - bd) + i(bc + ad)$. På denne form kræver en kompleks multiplikation 4 reelle multiplikationer og 2 reelle additioner. Vis, at en kompleks multiplikation kan udføres ved 3 reelle multiplikationer og 5 additioner ved hjælp af algoritmen

$$\begin{aligned} x &= (a-b)d + (c-d)a \\ y &= (a-b)d + (c+d)b \end{aligned}$$

Kapitel 5

Z-transformationen

Som tidligere vist er systemsvaret $y[n]$ for et diskret forskydningsinvariant system med impulssvaret $h[n]$ på et komplekst eksponentialinput af formen z^n

$$y[n] = H(z)z^n, \tag{5.1}$$

hvor

$$H(z) = \sum_{n=-\infty}^{\infty} h[n]z^{-n} \tag{5.2}$$

Hvis $z = e^{i\Omega}$, hvor Ω er reel (d.v.s. $|z| = 1$), svarer summationen i ligning (5.2) til $h[n]$'s diskrete Fouriertransformation. I almindelighed, når $|z|$ ikke er begrænset til at være 1, kaldes summationen i ligning (5.2) for $h[n]$*'s z-transformation.*

z-transformationen af en talfølge $x[n]$ defineres således

$$X(z) \stackrel{def}{=} \sum_{n=-\infty}^{\infty} x[n]z^{-n} \tag{5.3}$$

hvor z er en kompleks variabel. z-transformationen på denne form betegnes ofte den *bilaterale* z-transformation for at skelne den fra den unilaterale z-transformation, som diskuteres senere. Vi vil henvise til $X(z)$, som defineret i ligning (5.3) som z-transformationen og kun bruge betegnelsen "bilateral" når det er nødvendigt for at undgå flertydighed. z-transformationen af $x[n]$ betegnes undertiden $\mathcal{Z}\{x[n]\}$ og forbindelsen mellem $x[n]$ og dens z-transformerede angives som

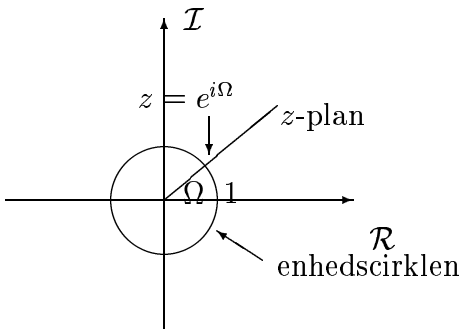
$$x[n] \stackrel{\mathcal{Z}}{\leftrightarrow} X(z) \tag{5.4}$$

Der er vigtige forbindelser mellem z-transformationen og Fouriertransformationen. Idet den komplekse variable z kan udtrykkes på polær form som

$$z = re^{i\Omega} \tag{5.5}$$

hvor r er modulus (magnitude) af z og Ω z's vinkel. Udtrykt ved r og Ω bliver ligning (5.3)

$$X(re^{i\Omega}) = \sum_{n=-\infty}^{\infty} x[n](re^{i\Omega})^{-n} \tag{5.6}$$



Figur 5.1: Den komplekse z-plan. z-transformationen reduceres til Fouriertransformationen for værdier af z på enhedscirklen.

eller

$$X(re^{i\Omega}) = \sum_{n=-\infty}^{\infty} \{x[n]r^{-n}\}e^{-i\Omega n} \tag{5.7}$$

Ud fra denne ligning ses, at $X(re^{i\Omega})$ er den Fouriertransformerede af sekvensen $x[n]$ multipliceret med en reel eksponent r^{-n}, d.v.s.

$$X(re^{i\Omega}) = \mathcal{F}\{x[n]r^{-n}\} \tag{5.8}$$

Den eksponentiale vægtning r^{-n} kan være aftagende eller voksende med voksende n, afhængig af, om r er større eller mindre end 1. Man bemærker, at for $r = 1$, d.v.s. $|z| = 1$, reduceres z-transformationen til Fouriertransformationen, d.v.s.

$$X(z)|_{z=e^{i\Omega}} = \mathcal{F}\{x[n]\} \tag{5.9}$$

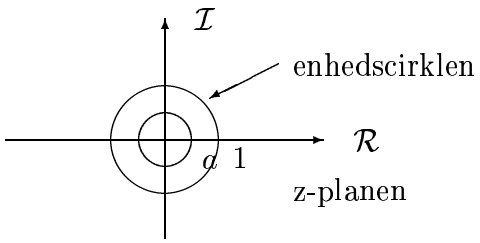
Derfor reduceres z-transformationen til Fouriertransformationen, når modulus (magnitude) af transformationsvariablen z er lig med 1, d.v.s. $r = 1$, eller ækvivalent, $|z| = 1$ Fra ligning (5.8) ses, at betingelsen for, at z-transformationen konvergerer er, at Fouriertransformationen af $x[n]r^{-n}$ konvergerer. For en given følge $x[n]$ må det forventes, at den konvergerer for nogle værdier af r og ikke for andre. Generelt er der i forbindelse med z-transformationen af en sekvens knyttet en mængde af værdier af z, for hvilke $X(z)$ konvergerer. Denne mængde kaldes konvergensområdet (region of convergence: ROC). Hvis konvergensområdet omfatter enhedscirklen, vil Fouriertransformationen også konvergere. Lad os se på et eksempel.

Eksempel 5.1 Signalet $x[n] = a^n u[n]$ har ifølge ligning (5.3)

$$X(z) = \sum_{n=-\infty}^{\infty} a^n u[n] z^{-n} = \sum_{n=0}^{\infty} (az^{-1})^n \tag{5.10}$$

Konvergens af $X(z)$ kræver $\sum_{n=0}^{\infty} |az^{-1}|^n < \infty$. Derfor er konvergensområdet de værdier af z, for hvilke $|az^{-1}| < 1$ eller $|z| > |a|$. Her findes

$$X(z) = \sum_{n=0}^{\infty} (az^{-1})^n = \frac{1}{1 - az^{-1}} = \frac{z}{z - a}, |z| > |a| \tag{5.11}$$



Figur 5.2: Pol-nulpunkts plot og konvergensområde for eksempel 5.1

Derfor konvergerer z-transformationen for enhver endelig værdi af a. Fouriertransformationen af $x[n]$ konvergerer derimod kun hvis $|a| < 1$. For $a = 1$ er $x[n]$ enhedstrinfølgen med z-transformationen

$$X(z) = \frac{1}{1 - z^{-1}}, |z| > 1 \tag{5.12}$$

På figur 5.2 er vist et pol-nulpunktsplot svarende til eksemplet. Konvergensområdet er vist skraveret.

5.1 Den unilaterale z-transformation

Den z-transformation, vi har set på indtil nu, er på den form, der kaldes den bilaterale z-transformation. Der findes en anden udgave, der betegnes den *unilaterale* z-transformation. Den unilaterale z-tranformation er specielt velegnet til at analysere kausale systemer, der er givet ved lineære differensligninger med konstante koefficienter og med begyndelsesbetingelser (d.v.s. ikke i hvile initialt).

Den unilaterale z-transformation $\mathcal{X}(z)$ af en talfølge $x[n]$ er givet ved

$$\mathcal{X}(z) = \sum_{n=0}^{\infty} x[n] z^{-n} \tag{5.13}$$

og er forskellig fra den bilaterale z-transformation derved, at summationen kun udføres over ikke-negative værdier af n, uanset om $x[n]$ er nul for $x < 0$. Derfor kan den unilaterale z-transformation opfattes som den bilaterale transformation af $x[n]u[n]$ (d.v.s. $x[n]$ multipliceret med enhedstrinfunktionen). Specielt vil for enhver følge, der er nul for $n < 0$, den unilaterale og den bilaterale z-transformation være ens. Konvergensområdet for $\mathcal{X}(z)$ vil altid være området udenfor en cirkel om origo i den komplekse plan.

Eksempel 5.2 Betragt signalet

$$x[n] = a^n u[n] \tag{5.14}$$

Da $x[n] = 0$ for $n < 0$ er den unilaterale og den bilaterale transformation ens og lig

$$\mathcal{X}(z) = \frac{1}{1 - az^{-1}}, \ |z| > |a| \tag{5.15}$$

Eksempel 5.3 Givet

$$x[n] = a^{n+1}u[n+1] \tag{5.16}$$

I dette tilfælde er den unilaterale og den bilaterale transformation ikke ens. Den unilaterale transformation er

$$\mathcal{X}(z) = \sum_{n=0}^{\infty} x[n]z^{-n} = \sum_{n=0}^{\infty} a^{n+1}z^{-n} = \frac{a}{1-az^{-1}}, \ |z| > |a| \tag{5.17}$$

Ligesom for Fouriertransformationens vedkommende kan man finde forskydningsegenskaben for den unilaterale z-transformation. For at udlede den, betragtes signalet

$$y[n] = x[n-1] \tag{5.18}$$

Heraf findes

$$\begin{aligned}
\mathcal{Y}(z) &= \sum_{n=0}^{\infty} x[n-1]z^{-n} && (5.19)\\
&= x[-1] + \sum_{n=1}^{\infty} x[n-1]z^{-n} && (5.20)\\
&= x[-1] + \sum_{n=0}^{\infty} x[n]z^{-(n+1)} && (5.21)\\
&= x[-1] + z^{-1}\sum_{n=0}^{\infty} x[n]z^{-n} && (5.22)
\end{aligned}$$

således at

$$\mathcal{Y} = x[-1] + z^{-1}\mathcal{X}(z) \tag{5.23}$$

Tilsvarende har signalet

$$w[n] = x[n-2] \tag{5.24}$$

den unilaterale z-transformerede

$$\mathcal{W}(z) = x[-2] + x[-1]z^{-1} + z^{-2}\mathcal{X}(z), \tag{5.25}$$

og tilsvarende udtryk kan findes for den unilaterale transformation af $x[n-m]$. Denne egenskab kan bruges til at løse differensligninger med begyndelsesbetingelser. For at illustrere dette, betragtes et eksempel:

Eksempel 5.4 Betragt differensligningen

$$y[n] + 3y[n-1] = x[n] \tag{5.26}$$

med $x[n] = u[n]$ og begyndelsesbetingelsen

$$y[-1] = 1 \tag{5.27}$$

Ved at tage den unilaterale z-transformation på begge sider af ligning (5.26) fås

$$\mathcal{Y}(z) + 3 + 3z^{-1}\mathcal{Y}(z) = \frac{1}{1-z^{-1}} \tag{5.28}$$

Ved at løse ligningen m.h.t. $\mathcal{Y}(z)$ findes

$$\mathcal{Y}(z) = -\frac{3}{1+3z^{-1}} + \frac{1}{(1+3z^{-1})(1-z^{-1})} \tag{5.29}$$

Ved at udvikle i partialbrøker findes

$$\mathcal{Y}(z) = -\frac{9/4}{1+3z^{-1}} + \frac{1/4}{1-z^{-1}} \tag{5.30}$$

og ved anvendelse af eksempel 5.2 på hvert led fås

$$y[n] = [\frac{1}{4} - \frac{9}{4}(-3)^n]u[n] \tag{5.31}$$

5.2 Opgaver

1. Givet differensligningen

   $$y[n] + 3y[n-1] = x[n] \tag{5.32}$$

   med begyndelsesbetingelsen $y[-1] = 1$. Bestem svaret $y[n]$ på sekvensen $x[n] = (\frac{1}{2})^n u[n]$ ved hjælp af den unilaterale z-transformation.

2. Givet differensligningen

   $$y[n] - \frac{1}{2}y[n-1] = x[n] - \frac{1}{2}x[n-1] \tag{5.33}$$

   med begyndelsesbetingelsen $y[-1] = 0$. Bestem svaret $y[n]$ på sekvensen $x[n] = u[n]$ ved hjælp af den unilaterale z-transformation.

3. Givet differensligningen

   $$y[n] - \frac{1}{2}y[n-1] = x[n] - \frac{1}{2}x[n-1] \tag{5.34}$$

   med begyndelsesbetingelsen $y[-1] = 1$. Bestem svaret $y[n]$ på sekvensen $x[n] = u[n]$ ved hjælp af den unilaterale z-transformation.

Kapitel 6

Billedkodning: JPEG

6.1 Oversigt

JPEG er ikke et dataformat men en hel familie af algoritmer til kompression af digitaliserede stillbilleder i farvekvalitet. Denne samling af forskellige metoder er vedtaget som standard i 1993 under betegnelsen ISO 10918. Fra denne værktøjskasse kan udvikleren alt efter det ønskede anvendelsesområde udvælge de nødvendige dele og implementere dem i hardware eller softwareprodukter. Udvikleren kan dermed også tilpasse kompressionsparametren efter kravene; derved falder billedkvaliteten naturligvis med stigende kompressionsforhold. Herved kan man frembringe ekstremt små billeddatamoduler, der f.eks. kan bruges som indexarkiver for billeddatabaser.

Den tabsgivende JPEG-kodning er optimeret med hensyn til fotografiske optagelser med kontinuerte farveovergange. For andre typer billeder, f.eks. billeddata der repræsenterer skarpe kontrastovergange som tegneserier, streggrafik og tekster, der indeholder store farveflader og abrupte farveskift, er den mindre egnet. For levende billeder findes en tilsvarende MPEG-standard.

6.2 Indledning

Fotokvalitet ved computerbilleder betyder mindst en opløsning på 640 x 480 billedpunkter med en farvedybde på 16 til 24 bit, hvorved der ved fotokvalitet her menes den skarphed og farveægthed, der kan opnås ved et fjernsynsbillede ved optimale betingelser. Et sådant billede svarer til en datamængde på ca. 1 Mbyte. Billeddata er således ret uhåndterlige, f.eks. når man skal oprette et større billedarkiv eller sende et farvebillede via modem og telefonledning eller via ISDN.

Det første skridt på vejen mod nedbringelse af datamængden sker ved grafikformater, der benytter interne kompressionsmetoder som runlength kodning, LZW- eller Huffmankodning, som f.eks. GIF, PCX- eller TIFF-formater. Som regel kommer man ved disse kompressionssmetoder i den mest moderne form sjældent over en kompressionsfaktor på tre. Disse metoder komprimerer dog uden informationstab, således at det er muligt at rekonstruere originalen til den sidste bit.

Man indså dog snart, at det ikke er nødvendigt at kunne genskabe billedinformationen

100 % for at det rekonstruerede billede ikke skal kunne skelnes fra originalen, for allerede ved indscanning af billedet sker en sammmenpresning af informationen i et raster (kvantisering). Dette førte til udviklingen af en effektiv metode ved navn JPEG via ISO/CCITT arbejdsgruppen af samme navn. JPEG står for Joint Photographic Expert Group.

JPEG-kompression er mindre effektiv end visse andre metoder som fraktal kompression, men har den fordel, at den som en standard er let tilgængelig og udbredt accepteret.

Ved udviklingen af JPEG-standarden var det vigtigste mål at stille en metode til rådighed, som kunne klare alle aspekter ved billedkompressionen. Herunder blev der særlig lagt vægt på følgende:

- Metoder til kompression uden datatab
- Metoder til kompression med datatab, men med justerbar kompressionsfaktor
- Algoritmerne skal have en ikke for høj kompleksitet
- Metoderne skal kunne bruges på alle typer stillbilleder uden indskrænkning i farveopløsningen

De benyttede algoritmer skal være hurtige (både implementeret i hardware og som software) og lette at implementere. Vanskelige eller dyre metoder er ikke egnede som standard.

6.3 JPEG virkemåde

JPEG er en standard, der tager sigte på flest mulige af de krav som billeddatakompressionen stiller. Da der ikke findes en enkelt algoritme som kan alt, fastlægges rammen for forskellige metoder, som hver især giver gode resultater på bestemte delområder,i JPEG-standarden. Principielt findes der to basismetoder:

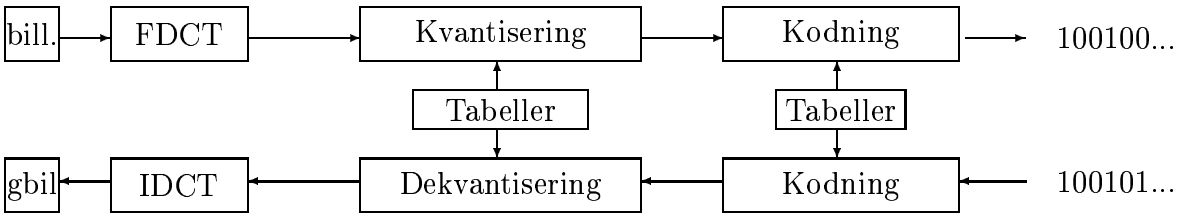
- en informationsbevarende kompressionsmetode (lossless mode)
- en kompressionsmetode med tab (lossy mode)

6.3.1 Informationsbevarende kompression

Den informationsbevarende kompressionsmåde er en metode til at komprimere billeddata, således at de efter dekompression bit for bit fuldstændig svarer til originaldata. JPEG stiller en operationsmåde til rådighed herfor, nemlig *lossless mode*.

Alle kompressionsmetoder arbejder efter samme skema: De indkomne data bliver først transformeret over i en mængde af deskriptorer, for disse deskriptorer fremstilles en statistisk model (i almindelighed en kodningstabel), og herefter kodes deskriptorerne efter modellen.

Til lossless mode bruges differentiel pulskodemodulation (DCPM) til transformationen. Til kodningen står Huffman-kodning og aritmetisk kodning til rådighed. Denne form kommer især i anvendelse, hvor der absolut ikke tolereres fejl, f.eks. ved maskinel billedanalyse. Lossless mode er imidlertid ikke den normale form for JPEG-anvendelse, idet den opnåede kompressionsfaktor ikke er særlig høj.



Figur 6.1: Princippet i baseline codec. Øverst kodning af originalbilledet (bill.) via forward DCT, kvantisering og kodning. Nederst dekodning i omvendt rækkefølge, resulterende i et gendannet billede (gbil).

6.3.2 Tabsgivende kompression

I modsætning til de informationsbevarende metoder svarer de dekodede billeddata her ikke længere nøjagtigt til de oprindelige billeddata. Begrebet tab er her lettere vildledende: det er ikke hovedsageligt billedkvaliteten, der går tabt, men information, der til en vis grad er redundant. Således er kompressionsforhold på op til 20:1 mulige uden at man kan se nævneværdig forskel i forhold til originalbilledet. Undersøgelser af JPEG-gruppen viste at 8x8 diskret cosinustransformation (DCT) gav de bedste resultater ved de tabsgivende konverteringsmetoder.. For de operationer, der bygger på DCT blev der fastlagt en minimal-algoritme (baseline codec) (codec er en forkortelse for coder-decoder) på hvilken alle JPEG-metoderne bygges. Komprimering med baseline codec består hovedsageligt af 5 trin

1. Konvertering af billedet i YC_bC_r-farverummet
2. Farvesubsampling
3. Diskret Cosinustransformation (DCT)
4. Kvantisering af DCT-koefficienterne
5. Kodning af koefficienterne

Figur 6.1 viser princippet i baseline codec. JPEG-standarden foreskriver 3 operationsmåder *(modes)*, som benyttes ved den diskrete cosinustransformation:

Sequential Mode: Billeddata kodes i et gennemløb fra øverst til venstre til nederst til højre i billedet. Hvis et billede består af flere komponenter bliver disse ikke kodet efter hinanden (altså komponent efter komponent) men komponenterne bliver behandlet overlappet. Ved denne overlappede behandling er det kun nødvendigt at have en lille buffer til rådighed, da det er muligt straks at udlæse billeleddata f.eks. til parallelt arbejdende processer, uden at det er nødvendigt at skulle vente til alle komponenter er bearbejdede. Denne måde kan bruges til de fleste anvendelser, giver den bedste kompression og er enklest at implementere.

Progressiv Mode: I den progressive mode gennemløbes billedet med flere gennemløb, hvor hver af disse kun koder en del af koefficienterne. Herunder findes igen to grundlæggende virkemåder: I den ene bliver koefficienterne samlet svarende til frekvensbånd og de laveste frekvenser bliver kodet først. I den anden bliver koefficienterne overført med stadig større nøjagtighed. JPEG kan dog også kombinere disse to grundmåder for at opnå bedre resultater. Ser man på resultatet af de enkelte gennemløb, så er billedet først "uskarpt". I løbet af transmissionen bliver det herefter mere og mere tydeligt. Denne mode kan med fordel benyttes ved billedtransmission. Man får ret hurtigt et indtryk af det overførte billede og kan afbryde overførslen, når billedkvaliteten er tilfredsstillende.

Hierarchical Mode: Denne mode er anderledes end progressiv mode. Den hierarkiske mode anvender en serie billeder med stadig større opløsning, som konstrueres ved "downsampling", d.v.s. ved filtrering med et lavpasfilter og decimering, d.v.s. udtagning af pixelværdier. Resultatet heraf bliver, at først kodes billedet med den mindste opløsning. Dette benyttes som grundlag for konstruktion af billedet med næstmindst opløsning, idet der overføres information i form af detaljering eller korrektion af billedet med lavest opløsning. Denne procedure gentages, indtil den fulde opløsning er nået. Hovedanvendelsesområdet for denne kodningsmåde kan være store billeddatabaser, hvor den lave opløsning kan bruges til indholdsfortegnelse og kun ekspanderes til høj opløsning ved behov herfor.

6.4 Grundlæggende teknikker i JPEG-standarden

6.4.1 YC_bC_r-farverummet

Fysiologiske eksperimenter og modeller har vist, at samtlige farver det menneskelige øje kan skelne, kan føres tilbage til addition af tre grundfarver, nemlig "rød", "grøn" og "blå". Dette virker også naturligt, da vort øje har farvereceptorer for netop disse tre farver. Receptorerne reagerer med nerveimpulser ved påvirkning af fotoner svarende til den pågældende farve.

Monitorer i dataterminaler udnytter disse egenskaber, idet højenergetiske elektronstråler skydes mod særligt belagte områder i billedrøret, som hver især ved beskydning udsender lys svarende til en af disse tre farver. Alt efter strålerne intensitet opstår forskellige farver. Herved tilordnes hver farve, der kan genereres, en vektor med tre elementer, RGB-værdien, der svarer til de tre farvestrålers intensitet [1]. Da disse intensiteter ikke er analoge værdier, men digitale signaler, er antallet af farver, det er muligt at vise på en billedskærm, begrænset. Dagens grafikkort har en opløsning på 255 trin (altså 8 bit) pr. grundfarve. Hermed defineres over 16 millioner genererbare farver, og da øjet kun kan skelne omkring 8 millioner forskellige farver, taler man om "true color" repræsentation.

Trefarveteorien siger også, at ikke alene rød, grøn og blå kan benyttes ved farverepræsentationen, men andre egnede farver kan også bruges som grundfarver. Der findes også andre farvemodeller, som ikke repræsenterer en farve ud fra grundfarver, men ved andre

[1] F. eks. kan den RGB-farvekodning, der benyttes i X-windows-systemet, ses i filen `/usr/local/X11R5/lib/X11/rgb.tx`

egenskaber. For eksempel Lysstyrke-farvetone-modellen (luminans-krominans modellen). Her er kriterierne farvens lysstyrke, farven med den største andel (rød, grøn eller blå) og farvens mætning, f.eks. pastel, stærk, næsten hvid o.s.v. Denne farvemodel bygger på øjets evne til bedre at kunne skelne små lysstyrkeforskelle end små farveforskelle. Således kan en grå tekst skrevet på sort bedre læses end en blå skrevet på rød baggrund ved samme farvelysstyrke. Sådanne farvemodeller kaldes lysstyrke-farvetone-modeller (luminans-krominans modellen).

YC_bC_r-modellen er en sådan lysstyrke-farvetone model. Derved bliver en RGB-farveværdi opdelt i en grundlysstyrke Y og to komponenter C_b og C_r, idet C_b er et mål for afvigelsen fra "middelfarven" grå i retning "blå". På samme måde er C_r et mål for afvigelsen mod rødt. Denne repræsentation benytter sig af den egenskab, at øjet er mere følsomt overfor grønt lys. Den meste information findes i grundlysstyrken Y, og ved anvendelserne behøver man blot at angive afvigelserne mod Rød og Blå. For at omregne en farveværdi givet i RGB-repræsentation til YC_bC_r farverummet har man brug for følgende formel:

$$\begin{pmatrix} Y \\ C_b \\ C_r \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.168 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{6.1}$$

Transformationen tilbage til RGB-værdier fra YC_bC_r-farverummet sker således

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} 1.0 & 0.0 & 1.402 \\ 1.0 & -0.344 & -0.714 \\ 1.0 & 1.772 & 0.0 \end{pmatrix} \begin{pmatrix} Y \\ C_b \\ C_r \end{pmatrix} \tag{6.2}$$

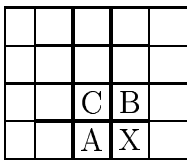
6.4.2 Farvesubsampling

Da øjet kun dårligt opfatter små positionsbestemte forskelle i C_b og C_r, kan man spare på den spatiale opløsning i C_b og C_r. Dette sker ved, at man indenfor små områder danner middelværdien af C_b og C_r og i stedet for at kode C_b og C_r i ethvert punkt benytter middelværdien i hele området. Normalt har området en størrelse på 2 x 2 pixels.

6.4.3 Differentiel pulskodemodulation

I JPEG benyttes differentiel pulskodemodulation til tabsfri kodning. Pulskodemodulation er kendt fra audioteknik. Her bliver et lydsignal ved sampling, d.v.s. digitalisering, aftastet med lige store tidsintervaller og de målte analoge værdier bliver omsat til digitale værdier. Ved billeddata er forholdet det samme. Et billede bliver her (f.eks. ved indscanning) gennemløbet med ens afstandsintervaller og de analoge farveværdier bliver digitaliserede. Audioteknikkens aftastningsfrekvens svarer til det scannede billedes opløsning. Når billeddata skal komprimeres ved hjælp af JPEG, foreligger dette allerede i et digitalt format. Der findes endog scannere og digitale fotoapparater, der direkte leverer output i JPEG-formatet.

Ved JPEG metoden lagres disse data ikke ubearbejdet. Givet data har man allerede kodet indtil et vist sted. En særlig forudsigelsesmetode giver nu en mulig værdi for de næste data. Her kodes kun forskellen mellem den forudsagte værdi og originalværdien. Jo



Nr.	Prædiktion
1	A
2	B
3	C
4	A + B - C
5	A + (B - C)/2
6	B + (A - C)/2
7	(A + B)/2

Figur 6.2: Forudsigelse ud fra nabopixels

bedre forudsigelsesmetode, desto mindre er forskellen, og desto oftere skal man kun lagre små værdier. Hvis man så tilordner korte koder til små tal, forkortes kodningslængden for data. Til forudsigelse af værdierne definerer JPEG flere beregningsmetoder, som på forskellig vis afhænger af nabopixels. Disse er illustrerede på figur 6.2.

6.5 Billedtransformationskodning

I det foregående afsnit så vi på reversible kodningsprocedurer, d.v.s. metoder med hvilke vi kan genvinde det originale billede (lossless coding). Nu vil vi se på irreversible kodningsmetoder. De er af særlig interesse til kodning af billeder, hvor et vist informationstab kan toleres, hvis billedforringelsen i forhold til originalbilledet kan accepteres af det menneskelige øje.

Billeder er kendetegnet ved en høj grad af redundans (overflødig information), fordi der normalt er en stor korrelation imellem naboelementer. Desuden, da langsomme variationer i gråtoneværdier forekommer hyppigere end hurtige variationer er det muligt at konstruere kodningsmetoder, der udnytter denne kendsgerning. Et eksempel herpå er billedtransformationskodning ved hjælp af den diskrete cosinustransformation (DCT).

Vi antager, at det digitaliserede billede er til rådighed som en matrix bestående af heltal. F.eks kan hvert matrixelement angive gråtoneværdien af det tilsvarende billedpunkt (pixel) for et gråtonebillede. Hyppigt bruges der 8 bit til at kode gråtoneværiden med (255 = hvid, 0 = sort).

6.5.1 Den diskrete cosinustransformation

Det er nemmest at illustrere metoden ved først at cosinustransformere et /'endimensionalt signal, og, efter at princippet er forstået, undersøge kodningen af todimensionale billeder. Derfor ser vi først på en et array af heltal. Dette array opdeles i enheder af fast længde (vektorer), som vi her sætter til at have længden N, hvor N er en potens af 2.

En diskret vektortransformation er blot et skift til et andet koordinatsystem for repræsentation. Som eksempel er den diskrete cosinustransformation (DCT) af en vektor

$\{x_m\}, m = 0, 1 \ldots N-1$ defineret ved

$$\underline{X} = \underline{\underline{A}}\,\underline{x}, \tag{6.3}$$

hvor $\underline{x}$ er vektoren, der indeholder signalværdierne, $\underline{X}$ er en vektor med DCT-koefficienterne, og $\underline{\underline{A}}$ er matricen indeholdende det nye systems basisvektorer. Elementerne i basisvektormatricen er givet ved

$$a_{k+1,m+1} = \sqrt{\frac{2}{N}} \cos \frac{(2m+1)k\pi}{2N}, \text{ hvor } k = 1, 2, \ldots, N-1 \text{ og } m = 1, 2, \ldots, N-1 \tag{6.4}$$

og

$$a_{1,m+1} = \frac{1}{\sqrt{N}}, \; m = 1, 2, \ldots, N-1 \tag{6.5}$$

For eksempel for $N = 8$ finder vi

$$\underline{\underline{A}} = \begin{pmatrix} 0.354 & 0.354 & 0.354 & 0.354 & 0.354 & 0.354 & 0.354 & 0.354 \\ 0.490 & 0.416 & 0.278 & 0.098 & -0.098 & -0.278 & -0.416 & -0.490 \\ 0.462 & 0.191 & -0.191 & -0.462 & -0.462 & -0.191 & 0.191 & 0.462 \\ 0.416 & -0.098 & -0.490 & -0.278 & 0.278 & 0.490 & 0.098 & -0.416 \\ 0.354 & -0.354 & -0.354 & 0.354 & 0.354 & -0.354 & -0.354 & 0.354 \\ 0.278 & -0.490 & 0.098 & 0.416 & -0.416 & -0.098 & 0.490 & -0.278 \\ 0.191 & -0.462 & 0.462 & -0.191 & -0.191 & 0.462 & -0.462 & 0.191 \\ 0.098 & -0.278 & 0.416 & -0.490 & 0.490 & -0.416 & 0.278 & -0.098 \end{pmatrix} \tag{6.6}$$

Det er let at vise, at $\underline{\underline{A}}^T \underline{\underline{A}} = \underline{\underline{I}}_8$, hvor $\underline{\underline{I}}_8$ er enhedsmatricen af orden 8.

D.v.s. for at udføre kodningen med DCT, skal signalet deles i en sekvens af blokke (vektorer) og derpå kodes ved hjælp af sættet af ortonormale basisvektorer.

6.5.2 Todimensional DCT

En todimensional (forward) diskret transformation er defineret ved

$$F(u,v) = \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} f(i,j) a(i,j,u,v) \tag{6.7}$$

Hvis en transformationskerne $a(x,y)$ kan udtrykkes som produktet af en funktion i hver af de to variable x and y, kaldes transformationen *separabel*, d.v.s. $a(x,y) = a_1(x)a_2(y)$.

Den todimensionale DCT er separabel, idet kernen $a(i,j,u,v)$ kan skrives som

$$a(i,j,u,v) = a_1(i,u)a_2(j,v). \tag{6.8}$$

Følgelig kan 2-D DCT'en udtrykkes som

$$F(u,v) = \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} f(i,j) a_1(i,u) a_2(j,v). \tag{6.9}$$

som kan skrives på matrix-formen

$$\underline{\underline{F}} = \underline{\underline{A}}_1 \underline{\underline{f}} \underline{\underline{A}}_2 \tag{6.10}$$

6.5.3 DCT-kodning i JPEG

Det menneskelige øje er ikke i stand til at opfatte svage farvenuancer nær så godt som svage forskelle i intensitet (lysstyrke). Derfor taler man ved farveforskelle om, at øjet har bedre opløsning ved lave end ved høje spatiale (rumlige) frekvenser. Analogien til frekvenser modsvarer øjets rumlige (spatiale) opløsningsevne. Ved bestemte farveforskelle kan man opfatte mere skelnelig farveinformation (d.v.s. højere spatial frekvens) end ved andre farveforskelle. DCT udnytter disse egenskaber ved det menneskelige øjes perceptionsevne ved at den bortfilterer høje stedfrekvenser og kode disse dårligt eller overhovedet ikke. Først bliver inputdata, der foreligger som fortegnsløse heltal, bragt på en for DCT egnet kurveform. Man subtraherer ganske enkelt 2^{P-1} fra hver værdi, hvor P repræsenterer den benyttede nøjagtighed i bit. I Baseline Codec er nøjagtigheden 8 bit, således at det nye "nulpunkt" lægges ved den oprindelige værdi 128. Så bliver billeddata opdelt i blokke på 8 x 8 pixel. En sådan blok bliver nu fortolket som en vektor bestående af 64 pixelværdier ud fra koefficienterne i et egnet vektorrum. DCT-en udfører herved et bassisskifte. Som basisvektorer anvendes kun 64 blokke à 8x8 pixels, som har den egenskab at de danner en ortonormalbasis med hensyn til vektorrummet. Basisvektorerne findes ved hjælp af følgende formel

$$C_{u,v} = \frac{1}{4}\gamma_u\gamma_v \sum_{m=0}^{7}\sum_{n=0}^{7} \cos\frac{(2n+1)u\pi}{16}\cos\frac{(2m+1)u\pi}{16} \tag{6.11}$$

Ved basisskiftet fås 64 entydige koefficienter, som giver antallet af tilsvarende basisblokke ud fra billeddatablokken. Koefficienterne beregnes således

$$F(u,v) = \frac{1}{4}\gamma_u\gamma_v \sum_{i=0}^{7}\sum_{j=0}^{7} f(i,j)\cos((2i+1)u\frac{\pi}{16})\cos((2j+1)v\frac{\pi}{16}) \tag{6.12}$$

For at kunne transformere disse koefficienter tilbage til deres oprindelige form har man brug for følgende relation

$$f(i,j) = \frac{1}{4}\sum_{u=0}^{7}\sum_{v=0}^{7} \gamma_u\gamma_v F(u,v)\cos((2i+1)u\frac{\pi}{16})\cos((2j+1)v\frac{\pi}{16}) \tag{6.13}$$

hvor

$$\gamma_i = \begin{cases} 1/\sqrt{2} \text{ for } i = 0 \\ 1 \text{ ellers} \end{cases}$$

og $F(u,v)$ DCT-koefficienten og $f(i,j)$ den oprindelige pixelværdi.

Man kan betragte forward transformationen (FDCT) som en harmonisk analysator og tilbagetransformationen (IDCT) som en harmonisk syntese. Ved denne ind- og afkodning (Codec) sker der tab uden ekstra behandling af koefficienterne, da de nødvendige sinus- og cosinusfunktioner kun kan repræsenteres med en begrænset nøjagtighed på en datamat. Heraf følger ligeledes, at denne metode ikke er iter/'erbar, d.v.s. hvis et DCT-kodet billede bliver dekodet og derpå atter kodet, får man et andet resultat end efter den første kodning. Fordelen ved DCT bliver især tydelig ved billeder med kontinuerte farveovergange: da nabobilledpunkter som regel næppe kan skelnes, vil kun DC-koefficienten i koefficientfremstillingen (d.v.s koefficienten for den basisvektor som har frekvensen nul i begge

retninger) og nogle få lavfrekvente AC-koefficienter antage større værdier. De øvrige er næsten nul eller for det meste lig nul. Dette betyder, at der skal kodes mindre tal og dette har ved passende kodning allerede en komprimeringseffekt. Som det ses fra formel 6.12 og 6.13 er beregningen af koefficienterne ret krævende. For en 8x8 blok kræves 63 additioner og 64 multiplikationer. FFT kan bruges som udgangspunkt for konstruktion af en hurtig DCT, men yderligere optimering kan fås ved udnyttelse af de symmetriegenskaber, der fås ved specialisering til 8 x 8 blokke. I Pennebaker og Mitchell [19] er beskrevet flere hurtige algoritmer til beregning af DCT. De udnytter bl.a. formlernes symmetriegenskaber. Med moderne hardwareteknik og effektiv flydendetalsregning er det muligt at opnå acceptable implementeringer af algoritmen.

6.5.4 Kvantisering

Ved udviklingen af JPEG-standarden var det et af målene, at kompressionsparametren skulle kunne vælges frit. Dette opnås ved hjælp af kvantiseringen. Kvantisering er en afbildning, hvor flere naboværdier afbildes på én ny værdi, hvorved koefficienterne divideres med en kvantiseringsfaktor $q(u, v)$ og afrundes til nærmeste integerværdi. Hertil bruges følgende ligning

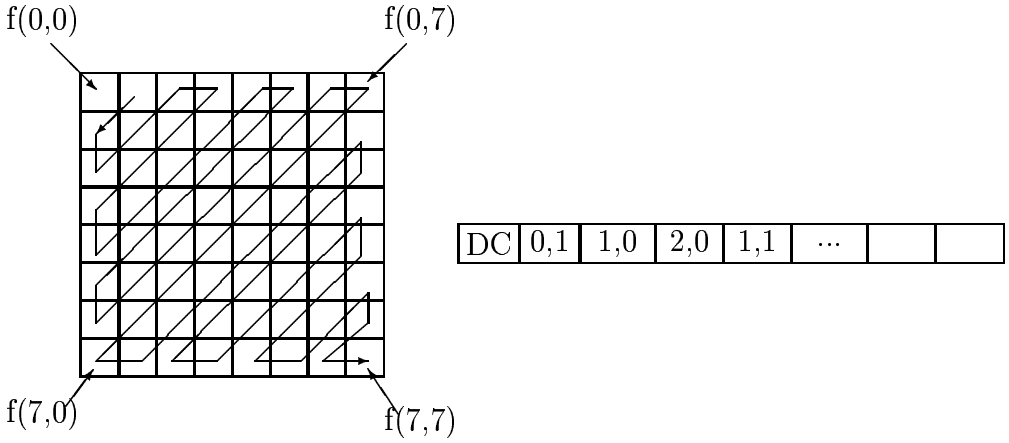
$$F^Q(u, v) = \lfloor \frac{1}{2} + \frac{F(u, v)}{q(u, v)} \rfloor \tag{6.14}$$

Ved den omvendte afbildning multipliceres den kvantiserede værdi med kvantiseringsfaktoren. Ved denne transformation frem og tilbage sker i almindelighed et informationstab, idet den kvantiserede værdi normalt ikke føres tilbage til den oprindelige værdi. Jo større kvantiseringsfaktoren er, desto er også informationstabet. Dette informationstab kan ved passende valg af kvantiseringsfaktoren holdes på så lavt et niveau, at det ikke kan observeres af øjet. Herved opnås let kompressioner under 1:10 uden at det rekonstruerede billede kan skelnes fra originalen. Om det rekonstruerede billede kan skelnes fra originalen afhænger også af gengivelsesmediet og betragtningsafstanden.

Til kvantisering uden observerbart informationstab er der blevet udviklet optimerede kvantiseringstabeller for lysstyrke og farvetoner. Disse findes bl.a. i Pennebaker og Mitchells bog [19]. I disse tabeller anvendes bedre (mindre) kvantiseringsfaktorer for DC-koefficienter og lavfrekvente AC-koefficienter end for de højere frekvenser. Herved drager man fordel af øjets frekvensfølsomhed. Ved implementering kan man som parameter angive en ønsket kompressionsfaktor (eller billedkvalitet). Ved den påfølgende kompression bliver kvantiseringsfaktoren skaleret i overensstemmelse hermed.

6.5.5 Koefficientkodning

Kodning af de kvantiserede koefficienter sker separat for DC- og AC-koefficienternes vedkommende. Ud fra 8x8 blokkene bliver der dannet en sekventiel (endimensional) bitstrøm bestående af 64 heltal (integers). DC-koefficientens første værdi er dog ikke den oprindelige værdi, men, som beskrevet under DCPM, differensen til DC-koefficienten i den foregående blok. På grund af DC-koefficienternes sammenhæng (lille indbyrdes variation) fås også her betydeligt mindre tal end ved lagring af absolutværdien. De 63 AC-koefficienter bliver



Figur 6.3: Zig-zag sekvenskodning af koefficienter

frembragt som en zig-zag kurve (se fig. 6.3), hvorved der sker en ordning mod højere spatiale frekvenser. Men da de højere frekvenskomponenter ofte er små eller nul opnås herved en for den videre kompression af billeddata gunstig rækkefølge.

6.5.6 Datakompression

De hidtil beskrevne metoder indeholder endnu ikke nogen eksplicit kompression men giver blot en, og ved kraftig kvantisering af DCT-koefficienterne, ret grov transformation af billeddata. For at kunne lagre de herved opnåede data i den mest muligt kompakt kodede form stiller JPEG-standarden flere effektive metoder til disposition. Disse er følgende:

- Repræsentation af variable-length-integers i stedet for heltal af fast længde
- Komprimering ved hjælp af Huffman-kodning
- Aritmetisk kodning

Aritmetisk kodning komprimerer bedre ned Huffman-kodning, men har den ulempe at være belagt med forskellige patenter, således at der skal betales licensgebyr for brugen heraf. Derfor benytter de fleste implementeringer Huffman-kodning.

Variable Length Integers

De hidtil beskrevne metoder har alle kun haft ét formål: man forsøger at transformere data således at man får en repræsentation med de mindst mulige tal. Men ved DCT er det muligt at tallene bliver 3 bit større. Men da mange koefficienter forsvinder i den efterfølgende kvantisering, betyder det ikke noget. Når små tal skal kodes, kan de lagres på mindre plads, således at der kan spares lagerplads. Ved fast præcision (f.eks. 8 bit) kræver små tal lige så meget plads som store tal. Her kommer metoden variable-length-integers (VLI) til hjælp, idet man kan kode heltalsværdier med variabel længde. Det er naturligvis nødvendigt for senere at kunne genskabe værdien at angive værdiens længde samtidig med

Bit	Værdi
0	0
1	-1,1
2	-3,-2,2,3
3	$-7,\ldots,-4,4,\ldots,7$
4	$-15,\ldots,-8,8,\ldots,15$
5	$-31,\ldots,-16,16,\ldots,31$
$\vdots$	$\vdots$
15	$-32767,\ldots,-16384,16384,\ldots,32767$

Tabel 6.1: Kodning af variabel-længde heltal (VLI).

Bit	Bits.Fortegn.Kode
0	000.
-1	001.1
4	011.0.00
-21	101.1.0101
124	111.0.111100

Tabel 6.2: Eksempler på VLI-er

kodningen. VLI-kodning er et specialtilfælde af *naturlig kodning*, som er gennemgået af Peter Johansen [11]. Således har et tal x følgende VLI-fremstilling

$$x \cong Bits.Fortegn.Kode \tag{6.15}$$

hvor

$$\begin{aligned} Bits &= \lfloor \log_2 |x| + 1 \text{ (Antal efterfølgende bit)} \\ Fortegn &= 1 \text{ hvis } x < 0 \text{ } 0 \text{ ellers, og} \\ Code &= |x| - 2^{Bits-1} \text{ (x's VLI-kodning)} \end{aligned}$$

Tabel 6.1 viser princippet for VLI-kodning og og 6.2 indeholder eksempler på VLI-kodning.

6.5.7 Huffman-kodning

Data, som er frembragt i lossless-mode ved hjælp af DPCM, bliver sendt uændret videre til Huffman-kodning. Fra de DCT-frembragte værdier foretages en særlig transformation for at frembringe en god Huffman-kodbar kodesekvens. For DC- og AC-koefficienter benyttes forskellige fremgangsmåder. DC-koefficienterne bliver først kodet ved hjælp af DPCM, hvor udgangspunktet for DC-koefficienten er den forudgående blok, og herefter omdannet til en VLI-værdi. AC-koefficienterne bliver kodet ved hjælp af en særlig repræsentation. Den genererede værdi har formen

(Længde, Størrelse, Værdi)

Her er *Længde* et 4-bit tal som svarer til antallet af nuller indtil næste koefficient, der er forskellig fra nul, *Størrelse* angiver antallet af bit, som værdien af ikke-nul koefficienter i VLI-repræsentation har brug for. Endelig angiver *Værdi* eksplicit koefficient-værdien. Ved begrænsningen til fire bit kan der højst angives 15 tegn på en gang.

Ud fra denne repræsentation ses det, at det er fordelagtigst, hvis længst mulige sekvenser af nuller skal kodes. Nu er det også klart, hvorfor koefficienterne kodes efter zig-zag mønstret. Efter omdannelse fra matrixform til en sekventiel bitstrøm er AC-koefficienterne sorteret, således at koefficienterne for de laveste frekvenser står først og de højeste spatiale frekvenser til sidst. Da værdierne svarende til de højeste frekvenser ofte er nul fremkommer herved lange nul-sekvenser.

Blandt disse nulfølger er der to med en særlig betydning. En nulsekvens, der går til slutningen af blokken kodes kun som blokslut (end-of-block, EOB) (undtagen når sekvensens begyndelse falder sammen med blokkens slutning). Denne kodning har formen (0,0). *Værdi* angives ikke. Kodede værdier på formen (15,0) svarer til nulsekvenser, der er længere end 15 tegn. Også her ses bort fra kodningen af værdien (man antager implicit 0) og der kodes videre i den næste sekvens. Herunder er højst 4 repræsentationer af formen *(Længde,Størrelse)* direkte bagefter mulige, idet der efter den fjerde skal angives en *Værdi*.

Efter denne omformning sker den egentlige Huffman-kodning. De data, der skal kodes, undersøges først med hensyn til deres statistiske sammensætning, idet der konstrueres en tabel, som tilordner hvert symbol den hyppighed, hvormed det optræder i datamaterialet. Ud fra denne hyppighedstabel konstruerer algoritmen en Huffman-tabel, hvori hvert symbol, der optræder, får tilknyttet en ny "digitaliseret" værdi. Den herved resulterende Huffman-kode er mere kompakt jo hyppigere et symbol i den oprindelige kode forekommer.

Denne konstruktion af Huffman-tabellen ud fra datamaterialet bruges ikke altid. Ofte anvendes prækonstruerede tabeller, som er genererede på grundlag af statistisk karakterisering af typiske datamaterialer. Hvis man skal opnå optimal Huffman-kodning er det dog nødvendigt at generere tabellen ud fra det foreliggende datamateriale.

Huffman-kodens effektivitet afhænger på den ene side af størrelsen af det benyttede kodealfabet, altså mængden af de forekommende symboler, og på den anden side af den stokastiske fordeling af data. Hvis data er ensartet fordelt er Huffman-algoritmen ikke særlig effektiv og kompressionraten tilsvarende lav.

Aritmetiske kodning

Aritmetisk kodning forsøger ligesom Huffman-kodning at kode hvert tegn med den ideale bitlængde, som svarer til tegnets informationsindhold (entropi). Huffman-kodning har imidlertid den ulempe at et bestemt tegn skal tilordnes netop en bitfølge med en given længde.

Arritmetisk kodning benytter et andet princip: den betragter fordelingen af et tegn i en sammenhæng (kontekst). F.eks. er en nulkoefficient langt mere sandsynlig for de høje frekvenser end for de lave. Nul har derfor en kontekstafhængig fordeling. Ved aritmetisk kodning får nul ikke tilknyttet en kodefølge men flere. Derved kan den gennemsnitlige tegnlængde formindskes i forhold til Huffman-kodens. Til fremstillingen af den statistiske

model tilordner den aritmetiske kode ikke et fast tal men et interval, der modsvarer sandsynlighedens størrelse. Men da man betragter en kontekst bliver intervallet delt ved hvert efterfølgende tegn. Kodealgoritmen kan kode en vilkårlig værdi i dette interval.

Den af JPEG anvendte kodning kan kun kode to symboler: nul og et. Til brug herfor bliver kodealfabetet først omdannet til et binært afkodningstræ ligesom ved fremstillingen af oversættelsestabellen ved Huffman-kodningen. For en detaljeret beskrivelse af aritmetisk kodning henvises til Peter Johansens noter [10]. ved aritmetisk kodning benyttes selvfølgelig ikke prækodning med VLI.

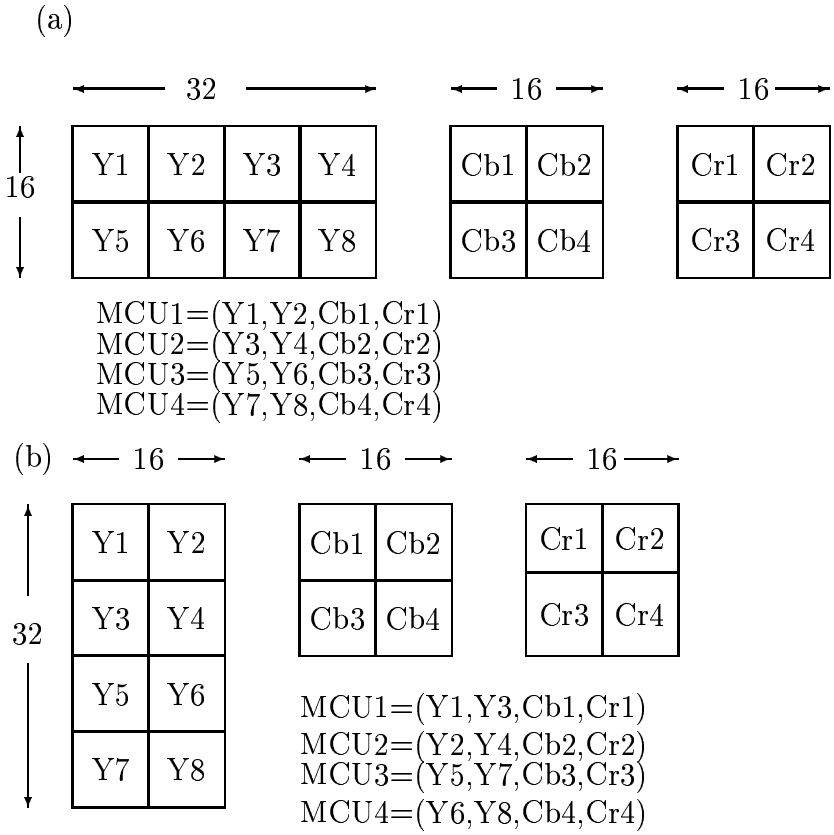
6.5.8 Kodning af billeder med flere komponenter

Den datastrøm, der genereres ved hjælp af JPEG indeholder et eneste billede, hvor billedet kan være sammensat af op til 255 billedkomponenter (kaldes også farve- eller spektralbånd eller kanaler). Den interne repræsentation af en komponent svarer til et todimensional array af billedpunkter med en vandret og lodret opløsning, der kan være forskellig fra komponent til komponent. Oftest er der ved farvebilleder de tre komponenter Y, C_b og C_r, men der er andre muligheder, f.eks. ved multispektrale billeder. Opløsningen i de tre farvekomponenter er normalt ikke ens. Således bliver forholdet Y:C_b :C_r reduceret til 2:1:1 ved 2x2 farvesubsampling og til 4:1:1 ved 4x4 farvesubsampling. Den i'te komponent har dimensionen x_i x y_i. For at håndtere formater i hvilke nogle billedkomponeneter samples med en anden oplø]sning end andre kan komponeneter have forskellige dimensioner. Dimensionerne skal dog have en indbyrdes heltallig sammenhæng (af hensyn til kodningen i M-blokke, se senere). Denne defineres ved hjælp af H_i og V_i, de relative horisontale og vertikale samplingfaktorer, der skal angives for hver komponent. Den totale billedimension X x Y defineres som maksimum af x_i og y_i for alle komponenter i billedet og kan være ethvert tal op til 2^{16}. H og V tillades kun at antage heltalsværdierne fra 1 til 4. De kodede parametre er X, Y samt H_i og V_i for hver komponent. Dekoderen rekonstruerer dimensionerne x_i og y_i for hver komponent ud fra følgende sammenhæng:

$$x_i = \lceil X \frac{H_i}{H_{max}} \rceil \tag{6.16}$$

$$y_i = \lceil Y \frac{V_i}{V_{max}} \rceil \tag{6.17}$$

hvor $\lceil\rceil$ er betegner heltalsafrunding opad (ceiling). Disse opløsningsfaktorer bliver angivet for hver komponent i frameheaderen. Medens den DCT-baserede funktionsmetode arbejder med 8x8 pixel blokke sker den tabsfri kodning pixel for pixel. JPEG definerer såkaldte dataenheder (data units, DU) til dette. I det første tilfælde er den en 8x8 blok, i det sidste tilfælde kun en pixel. Til kodningsproceduren bliver dataenhederne sammensat til minimalkodede enheder (minimal coded units, MCU). Hvis et billede kun har en komponent bliver disse kodede i rækkefølge og en MCU svarer til en dataenhed. Hvis et billede består af flere komponenter bliver disse af de ovennævnte grunde kodet overlappet. En MCU er her en pakke med H_i x V_i dataenheder for hver komponent i. Fig. 6.4 illustrerer dette ved hjælp af to små billeder. På figur 6.4(a) er $H = 2$ og $V = 1$ for Y-blokken og $H = 1, V = 1$ for de andre. På figur 6.4(b) er $H = 1, V = 2$ for Y-blokken. Entropikoderen arbejder kun



Figur 6.4: Overlappet kodning og MCU-er

på hele MCU-er. Svarer billedets opløsning ikke til et helt antal MCU-er, bliver de sidste rækker eller søjler kodet flere gange. Afkoderen tager højde for denne redundans.

6.6 JPEG Interchange Format

JPEG-komiteen har fastlagt alle enkeltheder i algoritmen men har ikke defineret noget alment anvendeligt dataformat for de komprimerede billeder. Kun det såkaldte *JPEG Interchange Format*, der benyttes til fastlæggelsen af den egentlige JPEG-datastrøm, er defineret. Et lille eksempel vil gøre den forskel mere klar: JPEG tillader farverum med 1,2,3 eller 4 komponenter. Kernealgoritmen bekymrer sig ikke om de enkelte farvers betydning, men komprimerer og dekomprimerer en datastrøm efter visse regler. Derfor siger man at JPEG er farveblind. Antallet af komponenter definerer heller ikke farverummet. F.eks. bruger RGB og YC_bC_r begge tre komponenter. Det format, der er defineret i standarden, indeholder ingen information om det anvendte farverum, så dette skal kodes ekstra ud over den egentlige datastrøm.

Komprimerede data efter JPEG Interchange Format struktureres ved hjælp af såkaldte markører (markers). Hver markør identificerer begyndelsen af et markersegment. Hvert markersegment begynder med en byte FF (hex) fulgt af en anden byte, der angiver

funktionen af den tilsvarende markør. Før den første FF-byte kan der valgfrit optræde en eller flere udfyldningsbytes med værdien FF. FF-bytes, som genereres ved Huffmankodningen udstyres med en efterfølgende nul-byte for at undgå forveksling med markører. De fleste markersegmenter har variabel længde og indeholder i starten to bytes, der angiver længden af hele segmentet (incl. de to længdebytes med excl. marker-bytes). Denne struktur tillader andre programmer at behandle datastrømmen uden at skulle tage sig af Huffman-kodningen eller den aritmetiske kodning. Ud over markersegmenter indeholder JPEG-formatet også kodesegmenter, der indeholder output fra den beskrevne entropikoder. JPEG-datastrømmen begynder med et SOI-mærke (start of image) og ender med et EOI-mærke (end of image). Imellem disse findes én (eller i hierarkisk mode) flere framesegmenter, der begynder med et SOFn-mærke (start of frame). Disse framesegmenter sætter alle de nødvendige parametre for de billedgennemløb, der forekommer i denne frame, såsom billedstørrelse, procedurernes nøjagtighed, antal komponenter og deres horisontale og vertikale opløsning. Parametren n angiver hvilken mode og kodningsmetode, der skal benyttes ved behandlingen. Frame-segmenter består af et (eller i progressiv mode flere) scansegmenter. I starten af et scansegment findes en SOS-markør (start of scan) som markerer begyndelsen af et billedgennemløb og står direkte foran kodesegmentet. Hvis et scansegment indeholder flere kodesegmenter, bliver disse adskilt af to byte lange RST-markører (restart). Dette har den fordel, at de enkelte kodesegmenter kan lokaliseres uden at data behøver at blive dekodede. Herved kan segmenterne overgives til separate processer af operativsystemet og bearbejdes i parallel. Ud over de nævnte segmenter definerer JPEG også andre segmenter, f.eks. kvantiserings- og kodetabeller samt parametre for de enkelte modi. For en nøjere gennemgang af JPEG Interchange formatet henvises til litteraturem, f.eks. [19].

6.7 Resumé

JPEG-standarden stiller de fleste af de hjælpemidler, der kræves til billedkodning til rådighed. Lossless mode er kun defineret af hensyn til standardens fuldstændighed, da de opnåelige kompressionsrater og bearbejdningshastigheder kan sammenlignes med billedkodningsmetoderne TIFF, GIF og PPM. Sin væsentligste styrke har standarden i de modi, hvor den arbejder med den diskrete cosinustransformation. Ved disse modi kan man let opnå kompressionsrater mellem 1:10 og 1:50 uden at det komprimerede billede med det blotte øje kan skelnes fra originalen. Dog har de DCT-baserede modi også nogle svagheder. Under kvantisering fjernes koefficienterne for de højfrekvente dele ofte. Ved billeder med store kontrastforskelle eller høj detailrigdom optræder synlige tab. Særligt taber billedet i skarphed. Derfor er JPEG i mindre grad egnet til tegneserier, liniegrafik og tekster, og i højere grad til billeder med glidende farveovergange. Man kan heller ikke opnå vilkårligt høje kompressionsrater med JPEG. Ved store kvantiseringsfaktorer foreligger den mulighed, at alle koefficienter bortset fra DC-koefficienten bliver nul. Derved får man et billede, der udelukkende består af 8x8 blokke, hvilket naturligvis ikke kan bruges i praksis. Trods dette er JPEG-algoritmen meget anvendelig, idet den ofte giver gode resultater, og har den fordel, at den er vedtaget som standard, og dermed kan benyttes af alle. JPEG-kompression af billeder har derfor også vundet stor udbredelse i de senere år, f.eks.

ved kodning af billeder tilgængelige via internettet (WWW) og som output fra digitale kameraer.

Litteratur

[1] N. Ahmed and K.R. Rao: *Orthogonal transforms for digital signal processing.* Berlin: Springer 1975.

[2] S. Banks: "Signal Processing, Image Processing, and Pattern Recognition". Prentice-Hall 1990 (ISBN 0-13-812579-1).

[3] D. Baumstark m.fl.: Multimedia Datenformat. Universität Karlsruhe 1995.

[4] R.N. Bracewell: " The Fourier Transform and its Applications" (2 udgave). McGraw-Hill 1978 (ISBN 0-07-007013-X).

[5] E. O. Brigham: "The Fast Fourier Transform and its Applications". Prentice-Hall 1988 (ISBN 0-13-307505-2).

[6] R.J. Clarke: *Digital Compression of Still Images and Video.* Academic Press, London 1995.

[7] T.H. Corman, C.E. Leiserson og R.L. Rivest: "An Introduction to Algorithms". MIT Press 1990 (ISBN 0-262-03141-8)

[8] A.K. Jain: *Fundamentals of Digital Image Processing.* Prentice-Hall Int. Editions 1989.

[9] J. Jájá: " An Introduction to Parallel Algorithms". Addison-Wesley 1992 (ISBN 0-201-54856-9).

[10] P. Johansen: Informationsteori og entropikodning. DIKU 1996.

[11] P. Johansen: Lempel-Ziv kodning, DIKU 1996.

[12] M.J. Lighthill: *Fourier Analysis and Generalised Functions*, Cambridge University Press 1958.

[13] J.S. Lim: *Two-dimensional signal and image processing.* Prentice-Hall 1990.

[14] G. Lindfield og J. Penny: "Numerical methods Using MATLAB". Ellis Horwood 1995 (ISBN 0-13-030966-4)

[15] A.V. Oppenheim og R.W. Shafer: "Discrete-Time Signal Processing". Prentice-Hall 1989 (ISBN 0-13-216771-9)

[16] J.H. McClellan, R.W. Schafer and M.A. Yoder: *DSP First: A Multimedian Approach*, Prentice-Hall 1998.

[17] J.H. McClellan, R.W. Schafer and M.A. Yoder: *Signal Processing First* , Prentice-Hall 2003.

[18] A.V. Oppenheim og A.S. Willsky (med I.T. Young): "Signals and Systems". Prentice-Hall 1983 (ISBN 0-13-8111758).

[19] W.B. Pennebaker og J.L. Mitchell: *JPEG Still Image Data Compression Standard*. Van Nostrand Reinhold, New York 1993.

[20] W.H. Press, B.P. Flannery, S.A. Teukolsky og W.T. Wettering: "Numerical Recipes in C". Cambridge University Press 1988 (ISBN 0-521-35465-X).

[21] K.R. Rao and P. Yip: *Discrete Cosine Transform: Algorithms, Advantages and Applications*. Academic Press 1990.

[22] K. Steiglitz: *An Introduction to Discrete Systems*, John Wiley and Sons, 1972.

[23] K. Steiglitz: *A Digital Signal Processing Primer with Applications to Computer Music*, Addison-Wesley 1996.

[24] G.K. Wallace: The JPEG still picture compression standard. *Communications of the ACM*, vol. **34**, pp. 30-44 (April 1991). Opdateret postscriptudgave: `ftp.uu.net, graphics/jpeg/wallace/ps.gz`.