

Einführung in die Zeitreihenanalyse Teil 1

N. Deichmann

Institut für Geophysik
ETH Zürich
Mai 2012

Inhalt

1	Zeitreihen	1
2	Diskrete Systeme	11
3	Die Fourier Transformation	19
4	Das Leistungsspektrum	44
5	Korrelation	47
6	Stochastische Zeitreihen	53
7	Anhang: MATLAB Routinen	57

Kapitel 1

Zeitreihen

Die Diskretisierung des Kontinuums

Einführung

Im Allgemeinen gehen wir davon aus, dass Veränderungen unserer makroskopischen Umwelt kontinuierlich ablaufen. Messtechnisch lässt sich das zum Beispiel für die Änderungen der Lufttemperatur an Hand der traditionellen Thermographen veranschaulichen: Die gemessene Temperatur als Funktion der Zeit wird kontinuierlich mit einer Tintenfeder auf eine sich langsam drehende Trommel aufgezeichnet. Das Resultat ist eine ohne Unterbruch durchgezogene Kurve. Obwohl eine solche Kurve eine vollständige Wiedergabe der beobachteten Temperatur darstellt, lässt sie keine weitere mathematische Bearbeitung zu. Dafür müssen die Messungen als Zahlenwerte zu diskreten Zeitpunkten vorliegen. Eine Folge solcher diskreter Messwerte nennt man eine Zeitreihe oder ein diskretes Signal. Bei der elektronischen Datenerfassung werden die Analogsignale (z. Bsp. eine kontinuierliche Spannungsänderung am Ausgang eines Messgerätes) durch Digitalisierung in digitale Signale umgewandelt. Ein Beispiel hiervon sind die digitalen Seismogramme, welche die durch ein Erdbeben oder eine Sprengung verursachte Bodenbewegung darstellen.

Zusätzlich zu den diskreten Beobachtungen kontinuierlicher Prozesse als Funktion der Zeit gibt es aber auch Zeitreihen, die einfach aus dem Zählen diskreter Ereignisse entstehen. Beispiele solcher Zeitreihen sind die zeitliche Abfolge von Erdbeben in einem bestimmten Gebiet oder die Anzahl Autos, die zu unterschiedlichen Zeiten eine gegebene Strasse befahren.

Mit der Entwicklung des Computers, der mit grosser Geschwindigkeit riesige Zahlenmengen verarbeiten kann, hat sich aus der Mathematik des Kontinuums eine Mathematik des Diskreten entwickelt. Die Zeitreihen-Analyse befasst sich mit den mathematischen Methoden zur quantitativen Erfassung und Interpretation der Eigenschaften diskreter Daten.

Einige Definitionen

Abtastperiode und Abtastrate

Bei der praktischen Verarbeitung von diskreten Messwerten stellt sich das Problem, eine kontinuierliche Zeitfunktion $x(t)$ zu diskreten Zeiten t_k zu betrachten. Die durch die *Abtastung* (englisch: *sampling*) entstandene *Folge* von Zahlen (*Stützwerte*, Engl. *samples*) nennt man *Zeitreihe* (Engl. *time series*)

Beispiel: $\{x_k\} = x_0, x_1, x_2, x_3, x_4, x_5, x_6, \dots = -23, -4, -19, -7, 5, 9, \dots$

Die zeitliche Differenz zwischen zwei aufeinanderfolgende Stützwerte $\Delta t_k = t_{k+1} - t_k$ heisst *Abtastintervall* oder *Abtastperiode* (Engl. *sampling period* oder *sampling interval*). Wenn die Abtastung zu äquidistanten Zeitpunkten vorgenommen wird, dann ist das Abtastintervall $\Delta t_k = \Delta t$ eine konstante Grösse. Oft wird das Abtastintervall auch nur mit Δ oder mit T bezeichnet. Folgende Schreibweisen sind je nach verwendeter Notation als äquivalent zu betrachten:

$$\{x_k\} = x_k = x(k) = x(kT) = x(k\Delta) = x(k\Delta t) = x(t)|_{t=k\Delta t}.$$

Der diskrete Index wird im Folgenden auch mit n oder i statt k bezeichnet.

Die *Abtastrate*, *Abtastfrequenz* oder *Digitalisierfrequenz* (Engl. *sampling rate*, *sampling frequency* oder *digitizing frequency*) ist die reziproke Abtastperiode: $f_s = 1/\Delta t$. Es ist intuitiv schon klar, dass eine kleine Abtastperiode oder grosse Abtastrate eine feinere Wiedergabe eines gegebenen kontinuierlichen Signals erlaubt, dies aber auf Kosten einer grösseren Datenmenge und entsprechendem Verarbeitungsaufwand geht. Tatsächlich zeigt es sich, dass eine zu kleine Abtastrate nicht nur eine ungenügende Auflösung zur Folge hat, sondern dass sie auch eine nicht rückgängig zu machende Verfälschung des digitalen Signals bewirkt. Somit stellt sich die grundsätzliche Frage nach einem Kriterium für die minimal notwendige Abtastrate für ein gegebenes Signal. Die Antwort ist durch das sogenannte *Abtasttheorem* gegeben, welches wir im Anschluss an eine Einführung in die Fourier Analyse behandeln werden.

Mittelwert eines Signals

$$\bar{x}_k = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x_n$$

Der Mittelwert wird auch oft als *offset* bezeichnet.

Energie eines Signals

$$E = \sum_{n=-\infty}^{\infty} |x_n|^2$$

Mittlere Leistung

$$P = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N |x_n|^2.$$

Im Englischen wird die Leistung als *power* bezeichnet.

Klassifizierung von Zeitreihen

Im Folgenden sollen diskrete Signale nach ihren Eigenschaften klassifiziert werden.¹

Einseitige und zweiseitige Signale: wir unterscheiden,

- einseitig in die Zukunft (nach rechts): $x_n = 0$ für $n < \alpha$
- einseitig in die Vergangenheit (nach links): $x_n = 0$ für $n > \beta$
- zweiseitige Signale
- Signale endlicher Länge: $x_n = 0$ für $n < \alpha$ und für $n > \beta$

Kausale und nicht-kausale Signale: wir unterscheiden,

- kausal: $x_n = 0$ für $n < 0$
- nicht-kausal: einige oder alle $x_n \neq 0$ für $n < 0$
- anti-kausal oder streng nicht-kausal: $x_n = 0$ für $n > 0$

(Man stelle sich vor, der Index $n = 0$ entspricht der Ankunftszeit des Signals.)

Periodische und nicht-periodische Signale: wir unterscheiden,

- periodisch: es gibt eine ganze Zahl p , so dass $\{x_n\} = \{x_{n+p}\}$ für alle n .
- nicht-periodisch: wenn obige Bedingung nicht erfüllt ist.

Stabile und nicht-stabile Signale: es bestehen zwei unterschiedliche Kriterien für die Definition von Stabilität:

1. absolut summierbar: $\sum_{-\infty}^{\infty} |x_n| < \infty$
2. endliche Energie: $\sum_{-\infty}^{\infty} |x_n|^2 < \infty$

Wie man aus $\sum |x|^2 \leq [\sum |x|]^2$ sieht, ist die Definition 1. restriktiver als 2. Das heisst, wenn 1. erfüllt ist, dann ist auch 2. erfüllt, aber nicht umgekehrt.

¹Eine Uebersetzung und Kurzfassung aus *Robinson, E. A., Durrani, T. S., und Pearson, L. G.: Geophysical Signal Processing. Prentice Hall International, London, 1986.*

Beispiel: eine Folge, die nicht absolut summierbar ist, aber endliche Energie aufweist:

$$x_n = \begin{cases} 1/n & n > 0 \\ 0 & n \leq 0 \end{cases}$$

$$\sum_{n=1}^{\infty} |1/n| = \infty,$$

$$\sum_{n=1}^{\infty} |1/n|^2 = \pi^2/6 < \infty.$$

Energie- oder Leistungssignale: wir unterscheiden,

- Energiesignal (die totale Energie ist endlich):

$$E = \sum_{-\infty}^{\infty} |x_n|^2 < \infty$$

- Leistungssignal (die über alle n gemittelte Leistung ist endlich):

$$0 < \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^{n=+N} |x_n|^2 < \infty$$

Ein Energiesignal hat null Leistung und ein Leistungssignal hat unendliche Energie. Ausserdem gibt es Signale die weder das eine noch das andere sind.

Beispiel Wechselstrom

Als Beispiel eines periodischen Leistungssignals wollen wir den Wechselstrom der Frequenz f_0 betrachten. Mit der Spitzenspannung U_0 in Volt ist die Spannung als Funktion der Zeit t

$$U(t) = U_0 \sin(2\pi f_0 t)$$

und mit dem Spitzenstrom I_0 in Amp ist der Strom als Funktion der Zeit

$$I(t) = I_0 \sin(2\pi f_0 t).$$

Die momentane Leistung in Watt ist das Produkt von Spannung mal Strom:

$$P(t) = U(t)I(t).$$

Sei R ein Widerstand in Ohm dann gilt $U = IR$ oder $I = U/R$, und somit wird die momentane Leistung zu

$$P(t) = \frac{1}{R} U^2(t) = \frac{1}{R} U_0^2 \sin^2(2\pi f_0 t)$$

Die über eine gegebene Zeit T abgegebene Energie ist das Integral der Leistung

$$E = \frac{U_0^2}{R} \int_0^T \sin^2(2\pi f_0 t) dt.$$

Mit den trigonometrischen Beziehungen $\sin^2(x) + \cos^2(x) = 1$ und $\cos^2(x) - \sin^2(x) = \cos(2x)$ erhalten wir

$$\int \sin^2(x) dx = \int \left[\frac{1}{2} - \frac{1}{2} \cos(2x) \right] dx = \frac{x}{2} - \frac{1}{4} \sin(2x).$$

Somit bekommen wir für die Energie

$$E = \frac{U_0^2}{R} \left[\frac{t}{2} - \frac{1}{8\pi f_0} \sin(4\pi f_0 t) \right]_0^T$$

und

$$E = \frac{U_0^2}{2R} \left[T - \frac{1}{4\pi f_0} \sin(4\pi f_0 T) \right].$$

Mit der Periode $T_0 = 1/f_0$ und ganzzahligen Werten n verschwindet der zweite Summand für $T = nT_0$. Die entsprechende Energie ist

$$E_n = \frac{U_0^2}{2R} nT_0$$

und die über n Perioden gemittelte Leistung ist dann

$$P_n = \frac{E_n}{nT_0} = \frac{U_0^2}{2R}.$$

Wir sehen also, dass die Energie einer Sinus-Schwingung mit der Zeit unbegrenzt ist, während ihre Leistung endlich ist.

Betrachten wir noch einmal den Ausdruck für die Wechselspannung

$$U(t) = U_0 \sin(2\pi f_0 t).$$

Die Varianz von $U(t)$ ist

$$\text{Var}\{U(t)\} = \sigma_U^2 = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T [U(t) - \bar{U}]^2 dt.$$

Da der Mittelwert einer Sinus-Funktion verschwindet, ist

$$\text{Var}\{U(t)\} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T U_0^2 \sin^2(2\pi f_0 t) dt.$$

Wie wir gesehen haben, reduziert sich dieses Integral für ganzzahlige Vielfache der Periode T_0 auf

$$\text{Var}\{U(t)\} = \frac{U_0^2}{2}.$$

Die mittlere Leistung eines Wechselstromes ist somit einfach die mit $1/R$ skalierte Varianz der Spannung, und die Standardabweichung der Spannung wird als effektive Spannung oder *rms* (*root-mean-square*) bezeichnet:

$$U_{eff} = \frac{U_0}{\sqrt{2}}.$$

An diesem Beispiel sieht man, warum im verallgemeinerten Sinn die Varianz eines Signals als seine *Leistung* bezeichnet wird.

Deterministische und Stochastische Prozesse

Deterministische und stochastische Signale: wir unterscheiden,

- Deterministisch: wenn schon vor dem Auftreten des Signals keine Unsicherheit über die Form des Signals besteht, d.h. wenn zumindest im Prinzip eine mathematische Relation existiert, welche das Signal für alle Zeiten beschreibt,
- Stochastisch: wenn auch nach der Beobachtung von vergangenen Werten des Signals, eine Unsicherheit über sein zukünftiges Verhalten besteht; in diesem Fall werden zur Beschreibung des Signals statistische Kriterien verwendet.

Die Unterscheidung zwischen deterministisch und stochastisch ist nicht unbedingt naturgegeben, sondern ist abhängig von der Modellierbarkeit eines Phänomens, d.h. ist ein Merkmal unserer Kenntnisse über das Signal.

Stochastische Prozesse können als zufällige Zeitfunktionen aufgefasst werden. Die Anzahl Anrufe an einer Telefonzentrale pro Zeiteinheit kann als ein solcher Prozess betrachtet werden.

Intuitiv wird der Unterschied zwischen einer deterministischen und einer stochastischen Betrachtungsweise anhand des folgenden Beispiels verständlich.

Beispiel: Betrachten wir den freien Fall eines Körpers im Raum. Die Gesetze der Mechanik erlauben, falls die Anfangsbedingungen und der Wert von g exakt bekannt sind, die Trajektorie zu bestimmen. Das heisst, falls die *Ursachen* bekannt sind (hier : Anfangsbedingungen und der Wert von g), dann können die *Effekte* (hier: Trajektorie) bestimmt werden.

Eine sorgfältigere Analyse des Problems zeigt aber, dass im Falle, wo wir auf eine extrem genaue Bestimmung der Trajektorie angewiesen sind, unserer Kenntnis Grenzen gesetzt sind. Diese Grenzen kommen einerseits aus den Anfangsbedingungen und aus dem Wert von g , welche

fehlerbehaftet sind und aus den Einflüssen der Luftreibung und den aerodynamischen Kräften, die im allgemeinen nur begrenzt bekannt sind. Unter diesen Umständen ist die Trajektorie des Körpers nicht exakt berechenbar und das oben betrachtete *deterministische* Modell kann grundsätzlich nur Hinweise zur "exakten Trajektorie liefern.

Unvorhergesehene Umstände können sogar dazu führen, dass wiederholte Versuche verschiedene Trajektorien als Resultat haben. Eine sorgfältige Analyse der Resultate zeigt im allgemeinen eine gewisse statistische Regelmässigkeit in dem Sinne, dass eine nach gewissen Kriterien noch zu definierende mittlere Trajektorie die Tendenz zeigt, mit steigender Anzahl von Versuchen, sich zu stabilisieren.

Aus diesem Beispiel können wir jetzt folgende intuitive Definition eines stochastischen Prozesses betrachten:

Ein stochastischer Prozess ist ein Phänomen, dessen zeitliche Entwicklung durch die Komplexität der zugrundeliegenden Gesetze und durch die Unwägbarkeit verschiedener Faktoren nicht exakt, sondern nur innerhalb gewisser Grenzen bestimmt werden kann.

Abbildung 1.1 zeigt zwei Beispiele von stochastischen Zeitreihen: Weisses Rauschen, *white noise*, das aus einer Folge von reinen Zufallszahlen besteht, und eine Messreihe des Grundwasserspiegels bei Maienfeld. Während das weisse Rauschen sich durch maximale Regellosigkeit auszeichnet, sieht man in der Zeitreihe des Grundwasserspiegels Anzeichen einer periodischen Komponente. Die Bedeutung des Attributes *weiss* im weissen Rauschen, wird im Zusammenhang mit der Fourier Analyse klar. Weisses Rauschen hat in der Theorie der stochastischen Zeitreihen die gleiche Rolle wie der Einheitsimpuls in der Theorie der deterministischen Signale. Die Untersuchung stochastischer Prozesse erfolgt mit Methoden der Wahrscheinlichkeitstheorie, der Statistik und der mathematischen Systemtheorie.

Ueber den Begriff von Noise

Bei *Noise* (eine Störung oder das Unerklärliche) werden wir in Zukunft je nach Problem folgendes zu verstehen haben: jeder kausale oder zufällige Einfluss, der nicht modelliert worden ist, oder nicht modelliert werden kann, bei welchem keine zusätzliche Information gewonnen werden kann, der nicht reproduzierbar ist, u.s.w.

In Zusammenhang mit unserer Vorlesung werden wir das Wort Noise ver-

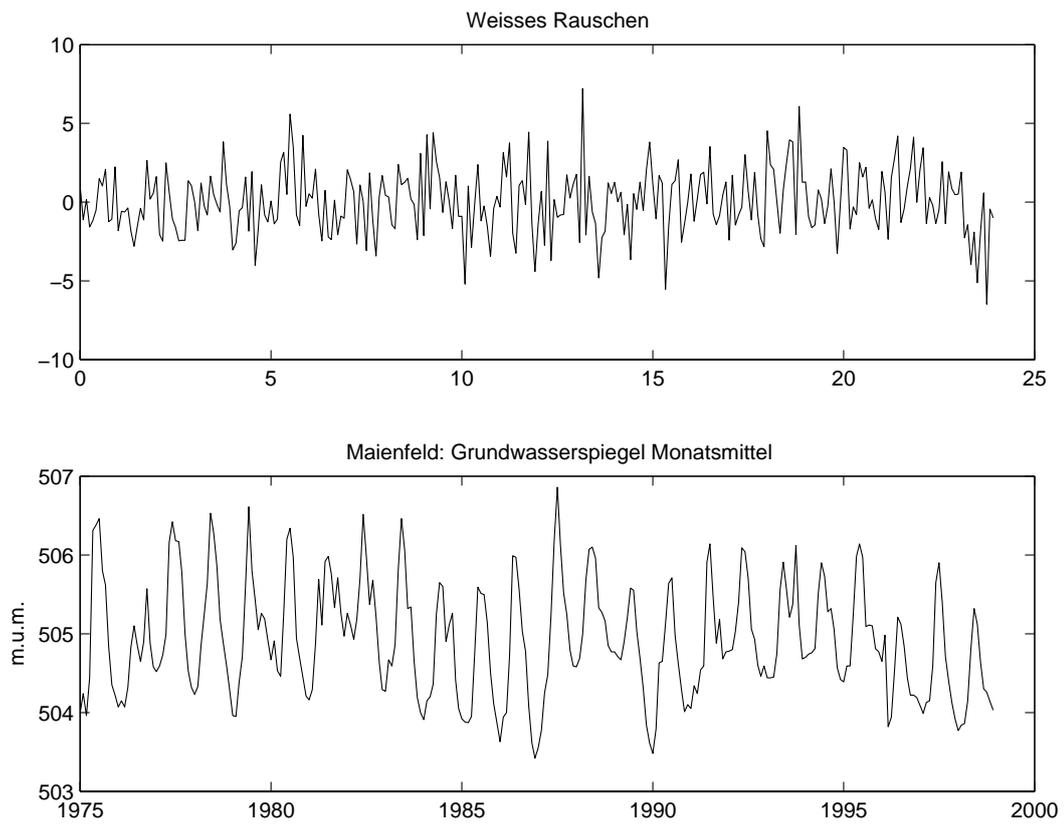


Abbildung 1.1: Zwei Beispiele stochastischer Zeitreihen: Weisses Rauschen (oben) und eine Messreihe der Monatsmittel des Grundwasserspiegels bei Maienfeld (unten).

wenden als allgemeiner Begriff für:

$$\text{NOISE} \left\{ \begin{array}{l} \text{Messfehler,} \\ \text{numerische Fehler,} \\ \text{Auswirkungen, die nicht modelliert worden sind,} \\ \text{Auswirkungen, die nicht modellierbar sind,} \\ \text{nicht reproduzierbare Störungen.} \end{array} \right.$$

Einige elementare diskrete Signale

Der diskrete δ -Impuls

Als ein diskretes Einheitssignal kann man die diskrete δ -Funktion oder *Einheitsimpuls* auffassen:

$$\delta_k = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

In der Praxis ist es zweckmässig, den Einheitsimpuls so zu definieren, dass seine Amplitude (Höhe) gleich der reziproken Abtastperiode ist:

$$\delta_k = \begin{cases} 1/\Delta t & k = 0 \\ 0 & k \neq 0 \end{cases}$$

Letztere Definition bewirkt, dass die Fläche unter dem diskreten Einheitsimpuls

$$\Delta t \sum_{k=-\infty}^{\infty} \delta_k = 1$$

ist, was äquivalent zum Integral über den δ -Impuls im Kontinuum ist. Im Kontinuum ist der δ -Impuls auch als *Dirac-Impuls* bekannt (s. später). Im Gegensatz zum kontinuierlichen Fall gibt es aber im Diskreten keine theoretischen Schwierigkeiten bei der Definition des δ -Impulses.

Die diskrete Sprungfunktion

Die diskrete Sprungfunktion, die oft entweder mit u_k oder h_k bezeichnet wird, ist wie folgt definiert:

$$u_k = \begin{cases} 1 & k \geq 0 \\ 0 & k < 0 \end{cases}$$

Die Entsprechende Funktion im Kontinuum wird auch als *Heaviside Funktion* bezeichnet.

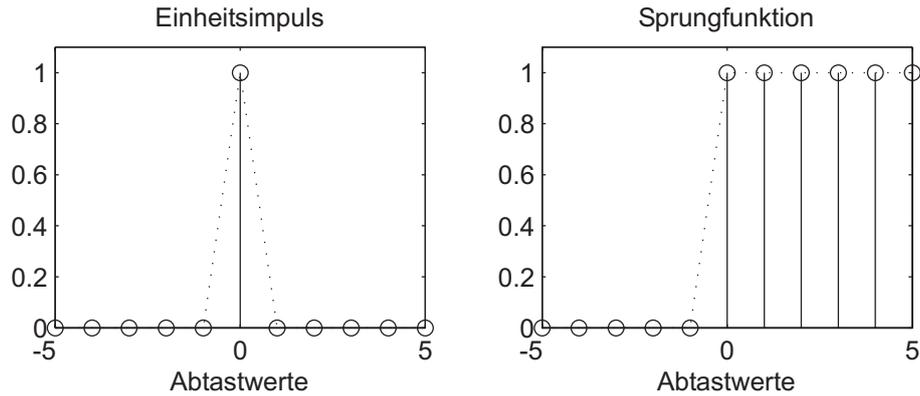


Abbildung 1.2: Der diskrete Einheitsimpuls (links) und die diskrete Sprungfunktion (rechts).

Beziehung zwischen δ -Impuls und Sprungfunktion

Aus der in der elementaren Analysis üblichen Definition der Ableitung einer Funktion ist sofort ersichtlich, dass die diskrete Ableitung als ein Differenzenquotient angenähert werden kann:

$$\frac{dx(t)}{dt} \approx \frac{x_k - x_{k-1}}{\Delta t}.$$

Angewendet auf die Sprungfunktion sehen wir, dass der δ -Impuls die diskrete Ableitung der Sprungfunktion ist:

$$\delta_k = \frac{u_k - u_{k-1}}{\Delta t}.$$

Demzufolge kann die *Sprungfunktion als Integral des δ -Impulses* betrachtet werden, was im Diskreten zu einer Summe führt:

$$u_k = \Delta t \sum_{n=-\infty}^k \delta_n.$$

Kapitel 2

Diskrete Systeme

Einleitung

In der Zeitreihenanalyse, bzw. digitalen Signalanalyse geht es häufig darum, Prozesse zu untersuchen, die eine Zeitreihe x_n in eine zweite Zeitreihe y_n verwandeln. Dabei kann es sowohl darum gehen, aus den beobachteten oder angenommenen Zeitreihen x_n und y_n auf den unbekanntem Prozess zu schließen als auch die Wirkung eines bekannten Prozesses zu simulieren, um die Eigenschaften der Zeitreihe x_n oder y_n zu untersuchen. Solche Prozesse werden in der Analyse von Zeitreihen und digitalen Signalen als *diskrete Systeme* bezeichnet.

Als Beispiel aus der Seismologie können sowohl der Ausbreitungsweg in der Erde als auch jedes Aufzeichnungsgerät ganz allgemein als Systeme betrachtet werden, welche in einer bestimmten Art die Form der beobachteten Signale bedingen. Eines der Hauptaufgaben der digitalen Signalverarbeitung ist dann die rechnerische Korrektur der dadurch bedingten Verzerrungen sowie die Unterdrückung von unerwünschten Eigenschaften im beobachteten Signal.

Einerseits geht es also darum, physikalische Systeme durch rechnerische Systeme zu simulieren, und andererseits darum, rechnerische Systeme zu entwickeln mit denen ein gegebenes Eingangssignal x_n in ein Ausgangssignal y_n mit gewünschten Eigenschaften verwandelt werden kann. Die Wirkung eines solchen Systems wird symbolisch oft wie folgt dargestellt:

$$x_n \longrightarrow y_n.$$

Ein einfaches aber auch typisches Beispiel sei die folgende Aufgabe. Gegeben ist ein langperiodisches Signal, welches durch ein zweites kurzperiodisches Signal überlagert ist. Indem wir die kurzperiodischen Signalanteile unterdrücken, wollen wir nun dieses Eingangssignal x_n in ein geglättetes Ausgangssignal y_n verwandeln. Ein mögliches Vorgehen ist die Berechnung

eines gleitenden Mittelwertes (Engl. *moving average*)¹ über drei benachbarte Punkte:

$$y_n = \frac{x_n + x_{n-1} + x_{n-2}}{3} = \frac{1}{3}x_n + \frac{1}{3}x_{n-1} + \frac{1}{3}x_{n-2}.$$

Beispiel: sei $x_n = \{1, 3, -2, 5, 3, -4, -1, 4, 1, -5, 2, \dots\}$, dann ist

$$y_n = \{0.33, 1.33, 0.67, 2, 2, 1.33, -0.67, -0.33, 1.33, 0, -0.67, \dots\},$$

wobei wir hier angenommen haben, dass $x_n = 0$ für $n < 0$.

Definieren wir nun eine Zahlenfolge

$$h_n = \{h_0, h_1, h_2\} = \{1/3, 1/3, 1/3\}$$

dann können wir obige Rechenvorschrift auch wie folgt ausdrücken:

$$y_n = h_0x_n + h_1x_{n-1} + h_2x_{n-2} = \sum_{m=0}^2 h_mx_{n-m}.$$

Man sagt dann, dass y_n das Resultat der diskreten Konvolution von h_n und x_n ist. Die Konvolution wird auch als Faltung bezeichnet und kompakt wie folgt geschrieben:

$$y_n = h_n * x_n.$$

Betrachten wir nun den Fall des Einheitsimpulses als Eingangssignal: $x_n = \delta_n$. Aus der Eigenschaft des Einheitsimpulses folgt dann

$$y_n = h_n * \delta_n = h_n$$

und h_n wird als die *Impulsantwort* des Systems bezeichnet. In obigem Beispiel ist der Glättungsoperator ein sogenanntes *lineares verschiebungsinvariantes System* und die Zahlenfolge $\{1/3, 1/3, 1/3\}$ ist die entsprechende Impulsantwort. Wie wir im Folgenden zeigen werden, lässt sich dieses Beispiel zu der Aussage verallgemeinern

*Das Ausgangssignal y_n eines linearen verschiebungsinvarianten Systems als Folge eines beliebigen Eingangssignals x_n ist das Resultat der Faltung von x_n mit der Impulsantwort h_n des Systems: $y_n = h_n * x_n$.*

Einige Definitionen

Im Folgenden wollen wir den soeben beschriebenen Sachverhalt formal begründen. Ein diskretes System ist eine Relation, welche ein diskretes Eingangssignal $\{x_n\}$ in ein diskretes Ausgangssignal $\{y_n\}$ überführt:

$$\{x_n\} \longrightarrow \{y_n\}.$$

¹Nicht zu verwechseln mit *running average*!

Einfachheitshalber schreiben wir im Folgenden jeweils nur x_n und y_n . Ob damit die gesamte Zahlenfolge oder ein einziges Element gemeint ist, ergibt sich meistens aus dem Zusammenhang. Die Zahlenfolge x_n bzw. y_n wird auch als Zeitreihe (Engl. *time series*) bezeichnet, was aber mit einer formalen Reihe nichts zu tun hat.

Ist das Eingangssignal der Einheitsimpuls $x_n = \delta_n$, dann bezeichnet man $y_n = h_n$ als die *Impulsantwort* des Systems:

$$\delta_n \longrightarrow h_n.$$

Diskrete Systeme können nach folgenden Kriterien klassifiziert werden:

Linearität Ein System (Operator oder Filter) wird als *linear* bezeichnet wenn es folgende Bedingung erfüllt:

Wenn $x_n \longrightarrow y_n$ und $q_n \longrightarrow r_n$, dann muss

$$\alpha x_n + \beta q_n \longrightarrow \alpha y_n + \beta r_n,$$

wobei α und β beliebige (im Allgemeinen komplexe) Konstanten sind. Alle anderen Systeme werden als *nichtlinear* bezeichnet.

Die Bedingung der Linearität besagt folgendes:

Sei

$$y = f(x) \text{ und } r = f(q)$$

dann ist die Funktion f linear wenn

$$f(\alpha x + \beta q) = \alpha f(x) + \beta f(q) = \alpha y + \beta r$$

Die Linearität ist eine besonders wichtige Eigenschaft, weil sie es erlaubt eine oft komplizierte Beziehung als Überlagerung (*Superposition*) von einfacheren Beziehungen auszudrücken.

Beispiel 1: Zeigen Sie durch anwenden obiger Definition, dass die Funktion $f(x) = mx$ linear ist.

Sei

$$y = mx \text{ und } r = mq$$

dann ist

$$f(\alpha x + \beta q) = m(\alpha x + \beta q) = \alpha mx + \beta mq$$

und

$$\alpha y + \beta r = \alpha mx + \beta mq$$

Somit ist die Funktion $f(x) = mx$ linear.

Beispiel 2: Ist die Funktion $f(x) = x^2$ linear?

Sei

$$y = x^2 \text{ und } r = q^2$$

dann ist

$$f(\alpha x + \beta q) = (\alpha x + \beta q)^2 = \alpha^2 x^2 + 2\alpha x \beta q + \beta^2 q^2$$

und

$$\alpha y + \beta r = \alpha x^2 + \beta q^2$$

Somit ist

$$f(\alpha x + \beta q) \neq \alpha y + \beta r$$

und die Funktion $f(x) = x^2$ ist, wie erwartet, nicht linear.

Beispiel 3: Ist die Funktion

$$f(x) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt$$

linear? (Diese Funktion ist das Fourier Integral)

Sei

$$y = \int_{-\infty}^{\infty} x e^{-j\omega t} dt \text{ und } r = \int_{-\infty}^{\infty} q e^{-j\omega t} dt$$

dann ist

$$f(\alpha x + \beta q) = \int (\alpha x + \beta q) e^{-j\omega t} dt = \int \alpha x e^{-j\omega t} dt + \int \beta q e^{-j\omega t} dt$$

und

$$\alpha y + \beta r = \alpha \int x e^{-j\omega t} dt + \beta \int q e^{-j\omega t} dt = \int \alpha x e^{-j\omega t} dt + \int \beta q e^{-j\omega t} dt$$

Somit ist (wie wahrscheinlich nicht erwartet) das Fourier Integral eine lineare Operation.

Frage: Ist die allgemeine Form der Geradengleichung $y = mx + n$ eine lineare Beziehung?

Invarianz gegenüber Verschiebung Ein System wird als *verschiebungsinvariant* (Englisch: *shift-invariant*) oder, falls der Index n die Zeit darstellt, als *zeit-invariant* bezeichnet, wenn die Beziehung zwischen Ein- und Ausgang unabhängig von n ist, d.h. wenn aus $x_n \rightarrow y_n$, $x_{n-k} \rightarrow y_{n-k}$ folgt.

Kausalität Ein System wird als *kausal* bezeichnet, wenn der Ausgang y_n zum Zeitpunkt n nur von gegenwärtigen und allenfalls vergangenen Eingangswerten $x_n, x_{n-1}, x_{n-2}, \dots$ und möglicherweise von den früheren

Ausgangswerten $y_{n-1}, y_{n-2}, y_{n-3}, \dots$ abhängt. Falls y_n durch eine Relation gegeben ist, welche auch spätere (zukünftige) Werte x_{n+k} miteinbezieht, dann wird das System als *nichtkausal* oder *akausal* bezeichnet. Physikalisch ist natürlich nur ersteres realisierbar. Rechnerisch hingegen kann man auch nichtkausale Systeme realisieren, wie folgendes Beispiel zeigt.

Beispiel: Der oben schon eingeführte Glättungsoperator mit drei Koeffizienten (gleitender Mittelwert) kann auch wie folgt als akausales System definiert werden:

$$y_n = \sum_{k=-1}^1 \frac{1}{3} x_{n-k} = \frac{1}{3} (x_{n+1} + x_n + x_{n-1}).$$

In diesem Fall ist der Ausgangswert eine Linearkombination vom gegenwärtigen Eingangswert mit einem vergangenen und einem zukünftigen Eingangswert. Um dieses System kausal zu machen, genügt es den Ausgang um einen Wert zu verzögern:

$$y_n = \sum_{k=0}^2 \frac{1}{3} x_{n-k} = \frac{1}{3} (x_n + x_{n-1} + x_{n-2}).$$

Stabilität Ein System wird als *stabil* bezeichnet, wenn ein beschränktes Eingangssignal auch ein beschränktes Ausgangssignal zur Folge hat. Insbesondere ist die Impulsantwort eines stabilen Systems eine beschränkte Zeitreihe. Analog zur Definition der Stabilität von diskreten Signalen, gibt es auch für diskrete Systeme zwei unterschiedliche Stabilitätskriterien, bzw. Definitionen von Beschränktheit:

1. Ein System ist stabil wenn die Impulsantwort absolut summierbar ist:

$$\sum_{n=-\infty}^{\infty} |h_n| \leq C_1 < \infty$$

2. Ein System ist stabil wenn die Impulsantwort endliche Energie aufweist:

$$\sum_{n=-\infty}^{\infty} |h_n|^2 \leq C_2 < \infty.$$

Diskrete Konvolution

Eines der Hauptprobleme der digitalen Signalverarbeitung ist die Bestimmung des Ausgangssignal eines Systems als Folge eines beliebigen Eingangssignals. Im Falle von linearen und verschiebungsinvarianten Systemen lässt

sich das Superpositionsprinzip anwenden, wobei der Ausgang als Ueberlagerung der Antworten des Systems auf einzelne elementare Eingangssignale dargestellt wird. Die entsprechende Operation ist die *Konvolution* oder *Faltung*.

Das Einheitssignal ist der diskrete Einheitsimpuls δ_n und der entsprechende Ausgang des Systems ist die Impulsantwort h_n . Wir wollen die Wirkung des Systems durch einen Pfeil symbolisieren und schreiben somit $\delta_n \longrightarrow h_n$. Da das System verschiebungsinvariant ist, gilt auch $\delta_{n-k} \longrightarrow h_{n-k}$. Wir suchen nun einen Ausdruck, der ein beliebiges Eingangssignal x_n in ein entsprechendes Ausgangssignal y_n überführt: $x_n \longrightarrow y_n$.

Wir können jede beliebige Zeitreihe als Summe von skalierten und verschobenen Einheitsimpulsen darstellen:

$$x_n = \sum_{k=-\infty}^{\infty} x_k \delta_{n-k}.$$

Hier ist nur für $k = n$ der Summand $\neq 0$, so dass sich für jeden Wert von n die Summe auf ein einziges Element reduziert und wir die Identität $x_n = x_n$ haben. Diese scheinbar komplizierte Darstellung eines banalen Resultates wird uns gleich recht nützlich sein.

Aus der Eigenschaft der Linearität unseres Systems folgt erstens,

$$x_k \delta_{n-k} \longrightarrow x_k h_{n-k}$$

und zweitens,

$$\sum_{k=-\infty}^{\infty} x_k \delta_{n-k} \longrightarrow \sum_{k=-\infty}^{\infty} x_k h_{n-k}.$$

Auf der linken Seite dieser Korrespondenz steht also ein beliebiges Eingangssignal x_n . Somit entspricht die rechte Seite dem gesuchten Ausgangssignal y_n mit der Beziehung $x_n \longrightarrow y_n$. Der Ausdruck

$$y_n = \sum_{k=-\infty}^{\infty} x_k h_{n-k} = x_n * h_n$$

wird als diskrete Konvolution oder Faltung bezeichnet.

Dies besagt, dass der Ausgang y_n eines linearen und verschiebungsinvarianten Systems als Folge eines beliebigen Eingangssignals x_n die Faltung dieses Eingangssignals mit der Impulsantwort des Systems ist.

Eigenschaften der Konvolution

Ersetzt man $m = n - k$, dann erhält man für die Konvolution:

$$y_n = \sum_{m=-\infty}^{\infty} h_m x_{n-m} = h_n * x_n.$$

Die Konvolution ist also *kommutativ*.

Im weiteren ist die Konvolution auch *distributiv*,

$$(x_n + q_n) * h_n = x_n * h_n + q_n * h_n$$

und *assoziativ*,

$$(x_n * q_n) * h_n = x_n * (q_n * h_n) = x_n * q_n * h_n.$$

Ausführung der Konvolution

Zur Veranschaulichung des praktischen Vorgehens bei der Konvolution betrachten wir die zwei Folgen:

$$x_0, x_1, x_2, x_3, x_4, x_5$$

und

$$h_0, h_1, h_2.$$

Die Konvolution ist dann,

$$y_n = \sum_{k=0}^n x_k h_{n-k},$$

mit dem Resultat,

$$\begin{aligned} y_0 &= x_0 h_0 \\ y_1 &= x_0 h_1 + x_1 h_0 \\ y_2 &= x_0 h_2 + x_1 h_1 + x_2 h_0 \\ y_3 &= x_1 h_2 + x_2 h_1 + x_3 h_0 \\ y_4 &= x_2 h_2 + x_3 h_1 + x_4 h_0 \\ y_5 &= x_3 h_2 + x_4 h_1 + x_5 h_0 \\ y_6 &= x_4 h_2 + x_5 h_1 \\ y_7 &= x_5 h_2 \end{aligned}$$

Wie wir im Folgenden sehen werden, entspricht die Konvolution von zwei Zeitreihen einer Korrelation, bei der die Reihenfolge der zweiten Folge invertiert wird. Im Allgemeinen wird das Resultat der Konvolution von zwei endlich langen Folgen mit N bzw. M Elementen die Länge $N + M - 1$ haben. Mit $N \geq M$ wird in der Praxis jedoch y_n meistens auf die Länge N beschränkt. Wegen der endlichen Länge der zwei Folgen, erstreckt sich tatsächlich für ein bestimmtes y_n die Summe über k jeweils vom grösseren der zwei Werte $n - (M - 1)$ und 0 bis zum kleineren der zwei Werte n und $N - 1$. Das heisst, dass für die ersten und letzten $M - 1$ Werte des Outputs die Summe nicht

über die volle Länge M der kürzeren Folge geht. Sind h_n die Impulsantwort eines Filters der Länge M und x_n die N Abtastwerte eines gegebenen Signals, dann entsprechen die ersten $M - 1$ Werte des Ausgangs der Einschwingzeit des Filters.

Die diskrete Konvolution kann auch in Matrix-Schreibweise als lineares Problem formuliert werden. Die Zahlenfolgen h_n , x_n und y_n können auch als Vektoren betrachtet werden und dann kann man eine Matrix H bilden, so dass die Faltung als Multiplikation einer Matrix mit einem Vektor dargestellt werden kann. Für das obige Beispiel sieht das wie folgt aus:

$$\begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix} = \begin{pmatrix} h_0 & 0 & 0 & 0 & 0 & 0 \\ h_1 & h_0 & 0 & 0 & 0 & 0 \\ h_2 & h_1 & h_0 & 0 & 0 & 0 \\ 0 & h_2 & h_1 & h_0 & 0 & 0 \\ 0 & 0 & h_2 & h_1 & h_0 & 0 \\ 0 & 0 & 0 & h_2 & h_1 & h_0 \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}$$

Hier haben wir die Länge von y_n , wie oben erwähnt, auf N beschränkt. In Kurzschreibweise kann man die Faltungsoperation also schreiben:

$$\mathbf{y} = \mathbf{H}\mathbf{x}.$$

Diese Art die Faltung auszuführen, erlaubt dann eine sehr effiziente Formulierung der später zu besprechenden Dekonvolution im Zeitbereich als lineares Inversionsproblem, zu dessen Lösung uns eine Vielzahl von Methoden aus der numerischen linearen Algebra zur Verfügung stehen².

Das Faltungsintegral

Im Kontinuum wird die Faltung sinngemäss als Integral formuliert:

$$g(t) = f(t) * h(t) = \int_{-\infty}^{\infty} f(\tau)h(t - \tau)d\tau.$$

In dieser Form ist die Operation wesentlich schwieriger nachzuvollziehen als im Diskreten. Eine anschauliche Darstellung findet sich aber im Buch von Scherbaum³.

²Gurrola, H., Baker, G.E., Minster, J. B. (1995). Simultaneous time-domain deconvolution with application to the computation of receiver functions. *Geophys. J. Int.*, 120, 537-543.

O'Dowd, R. J. (1990). Ill-conditioning and pre-whitening in seismic deconvolution. *Geophys. J. Int.*, 101, 489-491.

Oldenburg, D. W. (1981). A comprehensive solution to the linear deconvolution problem. *Geophys. J. R. astr. Soc.*, 65, 331-357.

Sipkin, A. A., Lerner-Lam, A. L. (1992). Pulse-shape distortion introduced by broadband deconvolution. *Bull. seism. Soc. Am.*, 82, 238-258.

³Scherbaum, F. (1996). *Of Poles and Zeros, Fundamentals of Digital Seismology*. Kluwer Academic Publishers, Fig. 2.4, p. 17.

Kapitel 3

Die Fourier Transformation

Einleitung

Die Fourier Transformation ist immer noch eines der wichtigsten Instrumente der Signalanalyse. Einerseits erlaubt sie es, Gesetzmässigkeiten eines Signals zu erkennen, die aus den Daten in ihrer ursprünglichen Form als Funktion z.Bsp. der Zeit oder des Raumes nicht direkt ersichtlich sind. Andererseits lassen sich gewisse mathematische Probleme (z.Bsp. Differentialgleichung) mittels Fourier Transformation einfacher und eleganter lösen und gewisse mathematische Operationen lassen sich mit deutlich geringerem Aufwand ausführen.

Die Fourier Analyse beruht auf der Tatsache, dass sich eine mathematische Funktion oder ein beobachtetes Signal als eine Summe (Ueberlagerung) von harmonischen Schwingungen darstellen lässt. Eine harmonische Schwingung $h_\omega(t)$ lässt sich mathematisch ganz allgemein durch

$$h_\omega(t) = C_\omega \cos(\omega t - \phi_\omega)$$

beschreiben. Sie ist eindeutig bestimmt durch ihre Amplitude C_ω , durch ihre Phase ϕ_ω und durch ihre Kreisfrequenz ω oder Frequenz f , wobei $\omega = 2\pi f$ ist. Unter Anwendung der Summationsformel für den Cosinus, erhält man

$$h_\omega(t) = C_\omega [\cos(\omega t) \cos \phi_\omega + \sin \phi_\omega \sin(\omega t)].$$

$C_\omega \cos \phi_\omega$ und $C_\omega \sin \phi_\omega$ können wir für eine gegebene Schwingung der Kreisfrequenz ω als zwei Konstanten a_ω und b_ω betrachten. Dann kann man die allgemeine harmonische Schwingung $h_\omega(t)$ auch als Linearkombination einer Cosinus- und einer Sinus-Schwingung darstellen:

$$h_\omega(t) = a_\omega \cos(\omega t) + b_\omega \sin(\omega t).$$

Die zugehörige Phase ergibt sich dann einfach aus

$$\tan \phi_\omega = \frac{\sin \phi_\omega}{\cos \phi_\omega} = \frac{b_\omega}{a_\omega}.$$

Da bekanntlich $\sin^2 \phi_\omega + \cos^2 \phi_\omega = 1$ ist, erhält man wieder die Amplitude von $h_\omega(t)$ als

$$C_\omega = \sqrt{a_\omega^2 + b_\omega^2}.$$

Wie man sich durch anwenden der obigen Summationsformel vergewissern kann, entspricht $\phi_\omega = 0$ einer reinen Cosinus-Schwingung und $\phi_\omega = \pi/2$ einer reinen Sinus-Schwingung. Man sagt, dass zwischen dem Sinus und dem Cosinus eine Phasenverschiebung oder Phasendrehung von $\phi = \pi/2$ besteht.

Die nebenstehende Abbildung soll nun an einem Beispiel veranschaulichen, wie die Ueberlagerung von einzelnen harmonischen Schwingungen ein Signal oder eine Funktion entstehen lässt, dessen Form weit entfernt von einer harmonischen Schwingung zu sein scheint. Links oben ist das knapp 1 Sekunde lange, aus 16 diskreten Punkten bestehende Signal abgebildet. Rechts oben ist das sogenannte Amplitudenspektrum, welches mittels der diskreten Fourier Transformation aus dem Signal berechnet wurde. Es stellt die Amplitude C als Funktion der Frequenz f dar. Für die vollständige Beschreibung der Fourier Transformation fehlt noch die Darstellung der Phase ϕ als Funktion der Frequenz.

Stattdessen sind unter dem Amplitudenspektrum die einzelnen harmonischen Schwingungen mit den Frequenzen 0 bis 8 Hz dargestellt. Da die Frequenz das Reziproke der Periode ist und die Dauer des dargestellten Zeitfensters genau 1 Sekunde beträgt, kann man durch abzählen der Maxima oder Minima die Frequenzangaben leicht verifizieren. Ausserdem entsprechen die Amplitudenverhältnisse zwischen den einzelnen Schwingungen genau dem Verlauf des abgebildeten Amplitudenspektrums. Die Phase als Funktion der Frequenz wirkt sich darin aus, dass jede einzelne Schwingung an einem unterschiedlichen Punkt relativ zum jeweiligen ersten Maximum oder Minimum anfängt. Im linken Teil der Abbildung ist das Resultat der laufenden Aufsummierung der nebenstehenden harmonischen Schwingung dargestellt. Mit jeder zusätzlichen Frequenz gleicht sich die Summe der Schwingungen mehr und mehr dem ursprünglichen Signal an, bis die Uebereinstimmung mit dem gegebenen Signal bei 8 Hz perfekt ist. Dabei ist zu bemerken, dass aus darstellerischen Gründen die harmonischen Schwingungen im Bild rechts mit 16-facher Abtastrate gegenüber den Signalen im Bild links abgebildet sind.

Mathematisch ausgedrückt haben wir nun das diskrete Signal x_n , ($n = 0, \dots, 15$) als eine Summe

$$x_n = a_0 + a_{N/2} \cos(\pi n) + 2 \sum_{k=1}^{N/2-1} a_k \cos\left(\frac{2\pi kn}{N}\right) + b_k \sin\left(\frac{2\pi kn}{N}\right)$$

erzeugt. Der Index k bezeichnet die einzelnen Frequenzen, welche das gegebene Signal x_n bilden. Der Witz der Fourier Transformation ist nun der, dass sie uns ermöglicht in eindeutiger Weise die Koeffizienten a_k und b_k aus den gegebenen Werten von x_n zu berechnen.

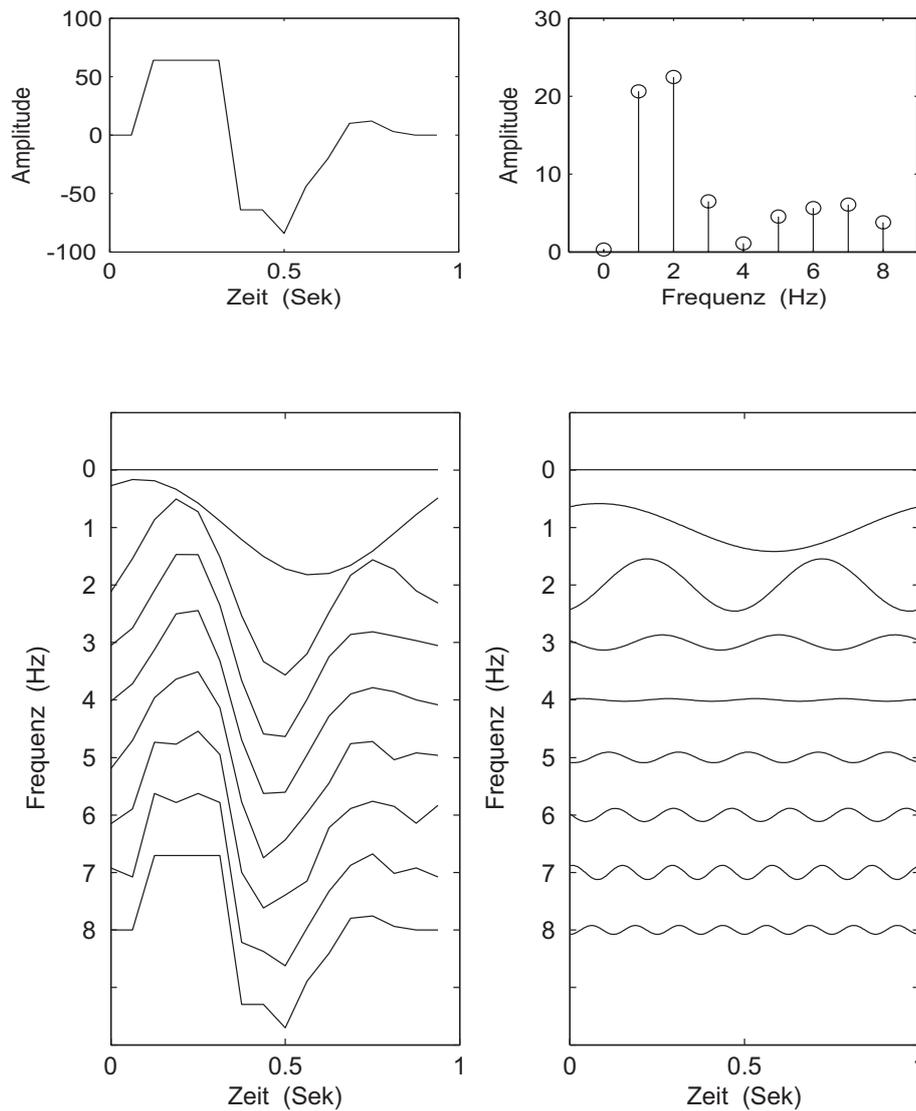


Abbildung 3.1: Beispiel einer Signalsynthese aus harmonischen Schwingungen: links oben das gegebene diskrete Signal, rechts oben das aus der Diskreten Fourier Transformation berechnete Amplitudenspektrum, rechts unten die einzelnen harmonischen Schwingungen die im gegebenen Signal enthalten sind, und links unten die laufende Summe der rechts abgebildeten Schwingungen. Die Amplituden der Schwingungen mit Frequenzen 1 bis 7 wurden vor der Aufsummierung im Bild links mit 2 multipliziert – warum wird im Abschnitt über die Diskrete Fourier Transformation erklärt. Man beachte, dass aus darstellerischen Gründen die einzelnen Schwingungen im rechten Bild mit 16-facher Abtastrate gegenüber dem Bild links gezeigt sind.

Vor der Behandlung der diskreten Fourier Transformation, die in der digitalen Signalverarbeitung Anwendung findet, ist es sinnvoll sich mit der Theorie der Fourier Reihe und des Fourier Integrals im Kontinuum vertraut zu machen.

Die Fourier Reihe

Sei $h(t)$ eine kontinuierliche Funktion von t , welche periodisch oder periodisch fortgesetzt ist mit Periode T und für $0 < t \leq T$ stückweise stetig sowie von begrenzter Schwankung ist. Dann lässt sich $h(t)$ als eine unendliche Reihe von harmonischen Schwingungen mit diskreten Frequenzen darstellen:

$$h(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos\left(\frac{2\pi kt}{T}\right) + b_k \sin\left(\frac{2\pi kt}{T}\right).$$

Wie lassen sich nun die Koeffizienten a_0 , a_k und b_k , ($k = 1, \dots, \infty$) in eindeutiger Weise aus der Funktion $h(t)$ bestimmen?

Um die Notation zu vereinfachen, schreiben wir obige Gleichung für $h(t)$ wie folgt um:

$$h(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(\omega_n t) + b_n \sin(\omega_n t),$$

wobei wir $2\pi k/T = \omega_n$ mit $k = n$ gesetzt haben. Multiplizieren wir nun $h(t)$ mit $\cos(\omega_m t)$, wobei $\omega_m = 2\pi m/T$, und integrieren wir von 0 bis T , dann erhalten wir

$$\begin{aligned} \int_0^T h(t) \cos(\omega_m t) dt &= \frac{a_0}{2} \int_0^T \cos(\omega_m t) dt + \\ &\sum_{n=1}^{\infty} a_n \int_0^T \cos(\omega_n t) \cos(\omega_m t) dt + \\ &\sum_{n=1}^{\infty} b_n \int_0^T \sin(\omega_n t) \cos(\omega_m t) dt. \end{aligned}$$

Auf Grund der sogenannten Orthogonalitätsbeziehung der Sinus und Cosinus Funktionen (durch Anwendung von elementaren trigonometrischen Beziehungen und Integrationsregeln zu beweisen) erhalten wir

$$\int_0^T h(t) \cos(\omega_m t) dt = \begin{cases} \frac{a_m T}{2} & (m = n) \\ 0 & (m \neq n) \end{cases}$$

Durch analoges Vorgehen wie oben aber mit $\sin(\omega_m t)$ an Stelle von $\cos(\omega_m t)$ erhalten wir

$$\int_0^T h(t) \sin(\omega_m t) dt = \begin{cases} \frac{b_m T}{2} & (m = n) \\ 0 & (m \neq n) \end{cases}$$

Da also nur die Terme mit $m = n$ etwas zu den Integralen beitragen, können wir wieder m und n durch k ersetzen und mit $\omega_k = 2\pi k/T$ erhalten wir die gesuchten Beziehungen für a_k und b_k :

$$a_k = \frac{2}{T} \int_0^T h(t) \cos\left(\frac{2\pi kt}{T}\right) dt \quad \text{und} \quad b_k = \frac{2}{T} \int_0^T h(t) \sin\left(\frac{2\pi kt}{T}\right) dt.$$

In komplexer Schreibweise lässt sich die Fourier Reihe auch wie folgt schreiben:

$$h(t) = \sum_{k=-\infty}^{\infty} c_k e^{j\frac{2\pi kt}{T}},$$

$$c_k = \frac{1}{T} \int_0^T h(t) e^{-j\frac{2\pi kt}{T}} dt.$$

Die komplexen Koeffizienten c_k ergeben sich aus den reellen Koeffizienten a_k und b_k durch Anwendung der Euler'schen Gleichungen für Sinus und Cosinus:

$$\frac{e^{jx} - e^{-jx}}{2j} = \frac{(\cos x + j \sin x) - (\cos x - j \sin x)}{2j} = \sin x$$

$$\frac{e^{jx} + e^{-jx}}{2} = \frac{(\cos x + j \sin x) + (\cos x - j \sin x)}{2} = \cos x$$

Das ergibt

$$c_0 = \frac{a_0}{2}; \quad c_k = \frac{a_k - jb_k}{2}; \quad c_{-k} = \frac{a_k + jb_k}{2}, \quad k > 0.$$

Kehrt man bei der Herleitung der komplexen Darstellung die Zuordnung der Koeffizienten um,

$$c'_k = \frac{a_k + jb_k}{2}; \quad c'_{-k} = \frac{a_k - jb_k}{2}, \quad k > 0,$$

dann erhält man für $h(t)$ eine Reihenentwicklung mit negativem Exponenten,

$$h(t) = \sum_{k=-\infty}^{\infty} c'_k e^{-j\frac{2\pi kt}{T}},$$

mit einem entsprechend modifizierten Ausdruck für c'_k . Die Wahl zwischen der einen oder anderen Formulierung ist absolut willkürlich und hat lediglich auf die Richtung der Phasendrehung einen Einfluss.

Betrachten wir noch einmal die erste Formulierung

$$h(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos\left(\frac{2\pi kt}{T}\right) + b_k \sin\left(\frac{2\pi kt}{T}\right)$$

und bezeichnen die einzelnen Summanden dieser Reihenentwicklung mit h_k . Aus der Beziehung

$$a_k = \frac{2}{T} \int_0^T h(t) \cos\left(\frac{2\pi kt}{T}\right) dt$$

sehen wir, dass

$$a_0 = \frac{2}{T} \int_0^T h(t) dt$$

und somit, dass der erste Term der Reihenentwicklung für $h(t)$ dem Mittelwert von $h(t)$ entspricht:

$$h_0(t) = \frac{a_0}{2} = \overline{h(t)}.$$

Für $k = 1$ ist

$$h_1(t) = a_1 \cos\left(\frac{2\pi t}{T}\right) + b_1 \sin\left(\frac{2\pi t}{T}\right).$$

Aus der Tatsache, dass $0 \leq t \leq T$, folgt in diesem Fall, dass das Argument der trigonometrischen Funktionen von 0 bis 2π geht. Somit entspricht der Term $h_1(t)$ genau einer harmonischen Schwingung mit der Periode $T_1 = T$, bzw. der Frequenz $f_1 = 1/T$. In gleicher Weise kann man zeigen, dass der Term $h_2(t)$ genau zwei Schwingungen mit der Periode $T/2$, bzw. der Frequenz $2/T$ entspricht. Daraus folgt, dass die anfangs vorausgesetzte Periodizität bzw. die Grundperiode T sowohl die kleinste im Spektrum enthaltene Frequenz f_1 , als auch den Abstand zwischen zwei benachbarten Spektralwerten, also die Auflösung Δf , bedingt:

$$f_1 = \Delta f = f_{i+1} - f_i = 1/T.$$

Wenn T gegen ∞ geht, dann geht f_1 gegen 0 und Δf geht in ein infinitesimales df über. Dies führt uns zwanglos auf das Fourier Integral.

Das Fourier Integral

Definition

Lässt man die Periode T gegen unendlich gehen, dann wird aus der Fourier Reihe ein Integral über ein kontinuierliches Frequenzspektrum:

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega) e^{j\omega t} d\omega,$$

$$H(\omega) = \int_{-\infty}^{\infty} h(t) e^{-j\omega t} dt.$$

Eine hinreichende und in der Praxis sehr brauchbare Bedingung für die Existenz des kontinuierlichen Fourier Integrals ist, dass die Funktion $h(t)$ eine endliche Energie aufweist:

$$\int_{-\infty}^{\infty} |h(t)|^2 dt < \infty.$$

Die zwei Integrale für $h(t)$ und $H(\omega)$ werden als kontinuierliche Fourier Transformation bezeichnet: eine Funktion der Zeit (oder einer beliebigen anderen Variable, z.Bsp. des Raumes) wird in eine Funktion der Frequenz übertragen und umgekehrt. Diese Beziehung wird oft durch die Korrespondenz $h(t) \leftrightarrow H(\omega)$ dargestellt.

Wenn t die Zeit ist, dann ist in obiger Formulierung $\omega = 2\pi f$ die Kreisfrequenz (Bogenmass/sekunden, engl. radians/s). Einfacher ist es, wenn man direkt über die Frequenz f (Hz) integriert, weil dann der Faktor $1/2\pi$ vor dem Integral wegfällt. Dann lautet die Fourier-(Hin)Transformation,

$$\mathcal{F}\{h(t)\} = H(f) = \int_{-\infty}^{\infty} h(t)e^{-j2\pi ft} dt,$$

und die Rücktransformation,

$$\mathcal{F}^{-1}\{H(f)\} = h(t) = \int_{-\infty}^{\infty} H(f)e^{+j2\pi ft} df.$$

Die Fourier Transformierte $H(f)$ ist im Allgemeinen eine komplexe Grösse, welche aus der Summe eines Real- und eines Imaginärteils besteht:

$$H(f) = \Re\{H(f)\} + j\Im\{H(f)\} = \int_{-\infty}^{\infty} h(t) \cos(2\pi ft) dt - j \int_{-\infty}^{\infty} h(t) \sin(2\pi ft) dt.$$

Also kann sie auch als ein Produkt des Betrages und des Argumentes dargestellt werden:

$$H(f) = |H(f)|e^{j\phi(f)} = |H(f)|[\cos \phi(f) + j \sin \phi(f)].$$

Der Betrag (engl. modulus) wird als Amplituden-Spektrum (oder Amplitudendichte) bezeichnet,

$$|H(f)| = \sqrt{\Re\{H(f)\}^2 + \Im\{H(f)\}^2},$$

und das Argument ist das Phasen-Spektrum,

$$\phi(f) = \arctan \frac{\Im\{H(f)\}}{\Re\{H(f)\}}.$$

In der Literatur findet man ebenso oft auch die Definition der Fourier Transformation mit vertauschten Vorzeichen im Exponent. Wie im Zusammenhang mit der komplexen Darstellung der Fourier Reihe schon gesagt, sind beide Formulierungen absolut gleichwertig. Es ist einzig darauf zu achten, dass sich dann auch in den Beziehungen, welche die Phase betreffen, andere Vorzeichen ergeben.

Wichtige Eigenschaften der Fourier Transformation

Sei im Folgenden $H(f) = \mathcal{F}\{h(t)\}$ die Fourier Transformation von $h(t)$. Dann lassen sich durch Anwendung der Definition der Fourier Transformation, bzw. ihrer Rücktransformation, folgende Beziehungen beweisen:

Reziprozität

$$\mathcal{F}\{H(t)\} = h(-f),$$

was bedeutet, dass $h(-f)$ die Fourier Transformierte von $H(t)$ ist.

Beweis: Ausgehend von der Definition der Fourier Rücktransformation

$$h(t) = \int_{-\infty}^{\infty} H(f)e^{+j2\pi ft} df$$

substituiert man erst f durch x (die Bezeichnung der Integrationsvariabel ist willkürlich), dann t durch $-f$, und schliesslich x durch t , dann erhält man

$$h(-f) = \int_{-\infty}^{\infty} H(t)e^{-j2\pi ft} dt$$

was der Definition der Fourier Transformation von $H(t)$ entspricht.

Verwendet man die Kreisfrequenz ω anstelle von f , dann lautet letztere Beziehung,

$$\mathcal{F}\{H(t)\} = 2\pi h(-\omega)$$

Linearität

$$\mathcal{F}\{c_1 h(t) + c_2 g(t)\} = c_1 H(f) + c_2 G(f)$$

Der Beweis wurde schon in Kapitel 2 erbracht.

Nullpunktverschiebung in der Zeit

$$\mathcal{F}\{h(t + t_0)\} = e^{+j2\pi t_0 f} H(f)$$

Beweis: Die Definition der Fourier Transformation auf $h(t + t_0)$ angewendet gibt

$$\mathcal{F}\{h(t + t_0)\} = \int_{-\infty}^{\infty} h(t + t_0)e^{-j2\pi ft} dt$$

Definiere $\tau = t + t_0$ und ersetze t durch $\tau - t_0$ und dt durch $d\tau$, dann kann man

$$e^{+j2\pi t_0 f}$$

vor das Integral ziehen, und das resultierende Integral mit der Integrationsvariabel τ an Stelle von t entspricht genau $H(f)$.

Man beachte, dass der Betrag des Exponentialterms vor dem Integral gleich 1 ist. Also hat die zeitliche Verschiebung einer Funktion keinen Einfluss auf die Amplitude der Fourier Transformierten sondern nur auf

die Phase. Sei $\phi(f)$ die Phase der unverschobenen Funktion und $\theta(f)$ die der um $\pm t_0$ verschobenen Funktion, dann folgt daraus

$$\theta(f) = \phi(f) \pm 2\pi t_0 f$$

Die Phase verschiebt sich also um einen Betrag der proportional zur Frequenz ist, und der Proportionalitätsfaktor ist gleich $2\pi t_0$.

Nullpunktverschiebung in der Frequenz

$$\mathcal{F}\{e^{+j2\pi f_0 t} h(t)\} = H(f - f_0)$$

Der Beweis ergibt sich durch Anwendung der Definition der Fourier Transformation auf die linke Seite und anschliessendem Ausmultiplizieren der zwei resultierenden Exponentialfunktionen; als Ergebnis erhält man dann die Fourier Transformation von $h(t)$ mit einer neuen Frequenz $f - f_0$.

Skalenänderung

$$\mathcal{F}\{h(ct)\} = \frac{1}{c} H\left(\frac{f}{c}\right)$$

Beweis: Ausgehend von

$$\mathcal{F}\{h(ct)\} = \int_{-\infty}^{\infty} h(ct) e^{-j2\pi ft} dt$$

definiere man $\tau = ct$ und ersetze t durch τ/c und dt durch $\frac{1}{c} d\tau$. Dann kann man $1/c$ vor das Integral ziehen und das resultierende Integral ist die Fourier Transformation von $h(\tau)$ mit einer neuen Frequenz f/c .

Das dies auch anschaulich Sinn macht, kann man sich folgendermassen überlegen: wenn die Zeiteinheit geändert wird, dann muss sich auch die Frequenzeinheit ändern. Als Beispiel: wenn wir die Zeit in Sekunden haben, dann ist die Frequenz in Hz; wenn wir stattdessen die Zeit in Millisekunden haben, dann haben wir die Zeiteinheit durch 1000 dividiert, und entsprechend müssen wir die Frequenzeinheit mit 1000 multiplizieren und erhalten kHz.

Ableitung

$$\mathcal{F}\{h'(t)\} = j2\pi f H(f)$$

Beweis: Die Fourier Transformation der Ableitung $h'(t)$ ist

$$\mathcal{F}\{h'(t)\} = \int_{-\infty}^{\infty} h'(t) e^{-j2\pi ft} dt$$

Nun wende man die Regel für die partielle Integration an:

$$\int u'v = uv - \int uv'$$

Das ergibt

$$\int_{-\infty}^{\infty} h'(t)e^{-j2\pi ft} dt = h(t)e^{-j2\pi ft} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} h(t) \frac{d[e^{-j2\pi ft}]}{dt} dt$$

Die Bedingung dafür, dass die Fourier Transformation existiert, impliziert, dass $h(t)$ gegen 0 gehen muss, wenn t gegen $\pm\infty$ geht. Somit verschwindet der erste Summand auf der rechten Seite der Gleichung. Die Ableitung im zweiten Summand trägt einen zusätzlichen Term $-j2\pi f$ bei, der vor das entsprechende Integral gezogen werden kann, wobei sich dann die zwei Minus-Zeichen aufheben. Das verbleibende Integral entspricht schliesslich $H(f)$.

Man beachte, dass durch die Multiplikation mit dem Term $j2\pi f$ die Ableitung einer Funktion eine Verstärkung der hohen Frequenzanteile bewirkt und wegen der Multiplikation mit j zu einer Phasendrehung von $-\pi/2$ führt.

Integral

$$\mathcal{F}\left\{\int h(t)dt\right\} = \frac{1}{j2\pi f} H(f)$$

Der Beweis dieser Eigenschaft folgt aus der Umkehrung des Satzes über die Ableitung. Dem entsprechend hat das Integrieren einer Funktion auch den gegenteiligen Effekt einer Ableitung: statt einer Erhöhung der hohen Frequenzanteile bewirkt die Division durch $j2\pi f$ eine Erhöhung der tiefen Frequenzanteile sowie eine Phasendrehung von $+\pi/2$.

Man beachte, dass wenn die Fourier Transformation mit positivem Vorzeichen im komplexen Exponenten definiert ist (und die Rücktransformation mit negativem Vorzeichen), dann bei der Nullpunktverschiebung, der Ableitung und dem Integral das j zu einem $-j$ wird (das kommt daher, dass dies Beziehungen sind, welche eine Phasenverschiebung verursachen).

Faltung Die Faltung (Engl. *convolution*) von zwei Funktionen $h(t)$ und $g(t)$ ist definiert als,

$$h(t) * g(t) = \int_{-\infty}^{\infty} h(\tau)g(t - \tau)d\tau.$$

Es gilt dann die folgende Beziehung:

$$\mathcal{F}\{h(t) * g(t)\} = H(f)G(f).$$

Der Beweis dieses Satzes folgt aus der Definition der Fourier Transformation, durch Vertauschen der Reihenfolge der Integration und Anwendung des Verschiebungssatzes.

Es gilt auch die Umkehrung dieses Satzes:

$$\mathcal{F}\{h(t)g(t)\} = H(f) * G(f).$$

Das heisst, die Fourier Transformation des Produktes von zwei Funktionen der Zeit ist gleich der Faltung der einzelnen Fourier Transformaten:

$$H(f) * G(f) = \int_{-\infty}^{\infty} H(\sigma)G(f - \sigma)d\sigma.$$

Die Faltung ist eine der wichtigsten Operationen in der Signalverarbeitung. Zum Beispiel, ergibt sich das Ausgangssignal eines Filters aus der Faltung des Eingangssignals mit der Filterfunktion, und ein Seismogramm kann als Faltung der Quellfunktion mit der Green'schen Funktion des Uebertragungsmediums und mit der Impulsantwort des Seismometers dargestellt werden. Was im Zeitbereich eine komplizierte und rechenintensive Operation ist, reduziert sich im Frequenzbereich auf eine einfache Multiplikation der entsprechenden Spektren.

Symmetrie-Eigenschaften

Eine Funktion $h(t)$ wird als *gerade* bezeichnet wenn $h(-t) = h(t)$ und als *ungerade* wenn $h(-t) = -h(t)$. Somit ist $\cos \omega t$ eine gerade Funktion und $\sin \omega t$ ungerade. Wenn $h(t)$ reell ist, dann gilt $H(-f) = H^*(f)$ (* = konjugiert komplex). Folgende Symmetrie-Eigenschaften erleichtern oft die Berechnung der Fourier Transformation:

$h(t)$	$H(f)$
gerade	gerade
ungerade	ungerade
reell und gerade	reell und gerade
reell und ungerade	imaginär und ungerade

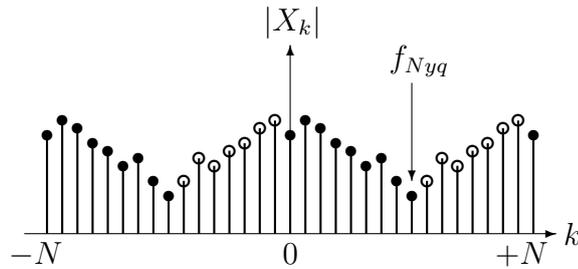


Abbildung 3.2: Betrag des diskreten Fourier Spektrums einer reellen diskreten Funktion mit $N = 16$. Die leeren Kreise entsprechen den konjugiert komplexen Werten der vollen Kreise.

Die diskrete Fourier Transformation

Betrachten wir nun eine Funktion (in unserem Fall ein Signal als eine Funktion der Zeit) dessen Werte x_n nur über ein beschränktes Intervall T und zu diskreten Zeitpunkten $n\Delta t$ gegeben sind (Δt ist das Abtastintervall und $n = 0, \dots, N - 1$). Die Abtastrate ist dann $f_s = 1/\Delta t$. Die Diskretisierung der Integrale in der kontinuierlichen Fourier Transformation führt auf die *Zeitdiskrete Fourier-Reihe*, welche den Uebergang zwischen diskretem Signal und diskretem Spektrum darstellt:

$$\begin{aligned} dt &\longrightarrow \Delta t \\ t &\longrightarrow t_n = n\Delta t \\ h(t) &\longrightarrow h(t_n) = x_n \\ df &\longrightarrow \Delta f = \frac{1}{N\Delta t} = \frac{1}{T} \\ f &\longrightarrow f_k = k\Delta f = \frac{k}{N\Delta t} \\ H(f) &\longrightarrow H(f_k) = X_k \end{aligned}$$

$$H(f) = \int_{-\infty}^{\infty} h(t)e^{-j2\pi ft} dt \longrightarrow X_k = \Delta t \sum_{n=0}^{N-1} x_n e^{-j\frac{2\pi kn}{N}} \quad (k = 0, \dots, N-1)$$

$$h(t) = \int_{-\infty}^{\infty} H(f)e^{+j2\pi ft} df \longrightarrow x_n = \frac{1}{N\Delta t} \sum_{k=0}^{N-1} X_k e^{+j\frac{2\pi kn}{N}} \quad (n = 0, \dots, N-1).$$

Die Folge X_k wird als das diskrete *Fourier Spektrum* der diskreten Zeitreihe x_n bezeichnet. Dabei ist Δf nicht die Abtastrate f_s sondern das Frequenzintervall zwischen zwei benachbarten Frequenzwerten im Spektrum. Ausserdem ist N die Anzahl Punkte in der Zeitreihe x_n , welche als Folge $\{x_n\}$ zu verstehen ist, und dementsprechend ist auch X_k als eine komplexe Folge $\{X_k\}$ zu betrachten.

In der Literatur wird die diskrete Fourier Transformation, DFT, oft ohne den Faktor Δt angegeben. Tatsächlich ist die DFT auch gültig ohne diesen Skalierungsfaktor. Dieser ist aber unerlässlich falls es auf die richtigen Einheiten ankommt. Nur dann entspricht X_n dem kontinuierlichen Spektrum und

nur dann hat $|X_n|$ wirklich die Einheit einer spektralen Amplituden-Dichte (z.Bsp. m/Hz).

Die Hintransformation stellt eine Analyse dar, welche eine gegebene Zeitreihe x_n in ihre Frequenzanteile X_k zerlegt. X_k wird auch als *diskretes Fourier Frequenzspektrum* bezeichnet. Die Rücktransformation entspricht hingegen der Synthese einer Zeitreihe aus einzelnen harmonischen Schwingungen.

Man beachte dreierlei:

- Die komplexe Folge $\{X_k\}$ ist periodisch mit der Periode N .
- Für reelle Zeitreihen $\{x_n\}$ gilt $X_{-k} = X_k^*$ für $(0 \leq k \leq N - 1)$.
- Für $k = 0$ sowie für $k = \pm N/2$ ist X_k rein reell.

Zum Beweis der ersten Aussage betrachten wir die DFT für einen einzigen beliebigen Wert von k :

$$X_k = \Delta t \sum_{n=0}^{N-1} x_n e^{-j \frac{2\pi kn}{N}} = \Delta t \sum_{n=0}^{N-1} x_n \left\{ \cos\left[\frac{2\pi kn}{N}\right] - j \sin\left[\frac{2\pi kn}{N}\right] \right\}$$

Entsprechend gilt für $k + N$:

$$X_{k+N} = \Delta t \sum_{n=0}^{N-1} x_n \left\{ \cos\left[\frac{2\pi(k+N)n}{N}\right] - j \sin\left[\frac{2\pi(k+N)n}{N}\right] \right\}$$

Durch anwenden der Summationsformeln für Cosinus und Sinus und unter Berücksichtigung, dass für ganzzahlige n , $\cos(2\pi n) = 1$ und $\sin(2\pi n) = 0$, folgt:

$$\cos\left[\frac{2\pi(k+N)n}{N}\right] = \cos\left[\frac{2\pi kn}{N} + 2\pi n\right] = \cos\left[\frac{2\pi kn}{N}\right]$$

und

$$\sin\left[\frac{2\pi(k+N)n}{N}\right] = \sin\left[\frac{2\pi kn}{N} + 2\pi n\right] = \sin\left[\frac{2\pi kn}{N}\right]$$

und somit ist

$$X_{k+N} = X_k.$$

Die zweite Aussage lässt sich auf ähnliche Art beweisen:

$$X_{-k} = \Delta t \sum_{n=0}^{N-1} x_n \left\{ \cos\left[-\frac{2\pi kn}{N}\right] - j \sin\left[-\frac{2\pi kn}{N}\right] \right\}$$

und

$$X_{-k} = \Delta t \sum_{n=0}^{N-1} x_n \left\{ \cos\left[\frac{2\pi kn}{N}\right] + j \sin\left[\frac{2\pi kn}{N}\right] \right\}$$

und somit ist

$$X_{-k} = X_k^*.$$

Die dritte Aussage folgt aus der Tatsache, dass, für $k = 0$ sowie für $k = N/2$ bei ganzzahligen Werten von n , $\sin(2\pi kn/N) = 0$ ist.

Aus diesen drei Aussagen folgt ausserdem, dass

$$X_{N-k} = X_{-k}$$

und

$$X_{N-k} = X_k^*.$$

Somit wird in der Praxis die DFT für reelle Signale auf die $N/2 + 1$ unabhängigen Spektralkoeffizienten mit den Werten $k = 0, \dots, N/2$ beschränkt. Die höchste im diskreten Spektrum enthaltene Frequenz folgt aus $f_k = k/(N\Delta t)$ mit $k = N/2$ und wird als *Nyquist Frequenz*, f_{Nyq} bezeichnet (siehe Abbildung 3.2):

$$f_{Nyq} = \frac{N}{2N\Delta t} = \frac{1}{2\Delta t} = \frac{f_s}{2}.$$

Dies führt zu einem der wichtigsten Ergebnisse der digitalen Signalanalyse: *Die höchste in einem diskreten Signal enthaltene Frequenz ist gleich der halben Abtastrate.*

Wir sprechen in diesem Fall von einem Signal mit beschränktem Frequenzumfang (engl. band-limited signal). Wegen der Beschränkung des Signals sowohl in der Zeit als auch in der Frequenz, entspricht einer diskreten Zeitserie endlicher Länge auch ein diskretes Frequenzspektrum endlicher Länge.

Übung

Zeigen Sie, dass es für die inverse DFT einer reellen Zeitsreihe, x_n , genügt über k von 0 bis $N/2$ zu summieren, wobei die Beiträge für $1 \leq k \leq (N/2 - 1)$ mit 2 zu multiplizieren sind. Anders ausgedrückt

$$x_n = \frac{1}{N\Delta t} [X_0 + X_{N/2}e^{+j\pi n} + 2 \sum_{k=1}^{N/2-1} X_k e^{+j\frac{2\pi kn}{N}}] \quad (n = 0, \dots, N-1).$$

Tipp: Zeigen Sie für ein beliebiges k im Bereich $k = 1, \dots, (N/2 - 1)$, dass die jeweiligen Cosinus und Sinus Anteile der entsprechenden Schwingung für k und $N - k$ identisch sind. Da x_n eine reelle Zeitsreihe ist, interessiert uns nur der reelle Beitrag aus obiger Gleichung. Beachten Sie dabei, dass die Terme X_k komplex sind und dass bei der Multiplikation von X_k mit der komplexen Exponentialfunktion die Produkte der zwei Imaginärteile wieder reell sind.

	Diskrete FT	Fourier Reihe	Fourier Integral
Auflösung Δt	Diskret (x_n) Δt	Kontinuierlich dt	Kontinuierlich dt
Auflösung Δf	Diskret (c_k) $\frac{1}{N\Delta t}$	Diskret (c_k) $\frac{1}{T}$	Kontinuierlich df
Signallänge	Endlich Periodizität = $N\Delta t$	Endlich Periodizität = T	Unendlich
Max. Frequenz	Endlich $F_{Nyq} = \frac{1}{2\Delta t}$	Unendlich	Unendlich

Beziehung zwischen diskreter Fourier Transformation, Fourier Reihe und Fourier Integral

Ausgehend von der diskreten Fourier Transformation, welche auch als *zeitdiskrete Fourier Reihe* bezeichnet wird, lässt man Δt gegen 0 gehen, dann geht f_N gegen ∞ und man erhält die Fourier Reihe (genauer, die *zeitkontinuierliche Fourier Reihe*). Lässt man nun die Signallänge T gegen ∞ gehen, dann geht Δf gegen 0, bzw. gegen df , und man erhält das Fourier Integral (oder genauer, das *zeitkontinuierliche Fourier Integral*).

Wie man sich leicht überlegen kann, kann man den Uebergang von der diskreten Fourier Transformation zum Fourier Integral auch über das sogenannte *zeitdiskrete Fourier Integral* ausführen, indem man das endliche Δt beibehält aber die Signallänge gegen ∞ und somit das endliche Δf gegen das infinitesimale df gehen lässt.

Wichtig ist die Erkenntnis, dass die Diskretisierung der Zeit mit der Abtastperiode Δt zu einer Beschränkung der maximalen Frequenz $f_{max} = f_{Nyq} = 1/(2\Delta t)$ führt und dass die Begrenzung der Zeitdauer T eine Diskretisierung der Frequenz mit einer Auflösung von $\Delta f = 1/T$ bedingt.

Die FFT

Die schnelle Fourier Transformation (engl. Fast Fourier Transform, FFT) ist nicht eine weitere Transformation, sondern nur ein Algorithmus um die diskrete Fourier Transformation mit relativ geringem Rechenaufwand durchzuführen. Um die gewünschte hohe Rechengeschwindigkeit zu erreichen, arbeiten die meisten FFT Routinen mit einer Signallänge von $N = 2^n$ Punkten. Dabei wird das gegebene Signal mit Nullen bis zur nächsten Zweierpotenz verlängert (zero padding). Dadurch wird der Informationsgehalt des resultierenden Spektrums nicht erhöht. Das Spektrum wird zwar feiner abgetastet, aber die Erhöhung der Auflösung ist nur optisch. Insbesondere sind die Amplitudenwerte des Spektrums nur von der Dauer des tatsächlich für die Transformation ausgewählten Signals abhängig und nicht von der Anzahl angehängter Nullen.

Sei Δt das Abtastintervall, dann ist $1/\Delta t$ die Abtastrate. Unabhängig von der Signallänge ist der Frequenzumfang der FFT nur von der Abtastrate abhängig und umfasst immer den Bereich von der Frequenz 0 bis zur Nyquist-Frequenz, $f_{Nyq} = 1/(2\Delta t)$. Bei einer Signallänge von $N = 2^n$ diskreten Abtastwerten besteht das Spektrum aus $N/2 + 1 = 2^{n-1} + 1$ diskreten Spektralwerten. Das heisst, der Frequenzbereich zwischen 0 und der Nyquist-Frequenz ist in $N/2$ gleichlange Frequenzintervalle Δf eingeteilt. Dabei gilt:

$$\Delta f = f_{Nyq}/2^{n-1} = \frac{f_{Nyq}}{N/2} = \frac{1}{N\Delta t}.$$

Man beachte, dass, sowohl hier als auch in der vorhergehenden Tabelle, N die Anzahl Datenpunkte nach Verlängerung mit Nullen bis zur nächsten Zweierpotenz ist!

Wegen der Nichtexistenz des $\log 0$ ist in einer logarithmischen Darstellung der Frequenzachse die kleinste Frequenz gleich Δf . Ausserdem ist der Bereich tiefer Frequenzen stark gedehnt (geringe Punktdichte) während der Bereich hoher Frequenzen stark gedrängt ist (hohe Punktdichte). Das Spektrum kann natürlich auch als Funktion der Periode ($1/f$) dargestellt werden, wobei dann auch wieder die Nichtlinearität der Skala beachtet werden muss.

Führt man die FFT für eine reelle Zeitserie durch, besteht das Resultat der komplexen FFT aus Paaren von Cosinus- und Sinus-Termen, die jeweils paarweise einem Frequenzwert entsprechen (Cosinus = reell, Sinus = imaginär). Die Sinus-Terme zur Frequenz 0 und zur Nyquist-Frequenz sind immer gleich 0, so dass zur Speicherplatz-Optimierung oft der Cosinus-Term der Nyquist-Frequenz an Stelle des Sinus-Terms der Frequenz 0 abgespeichert wird.

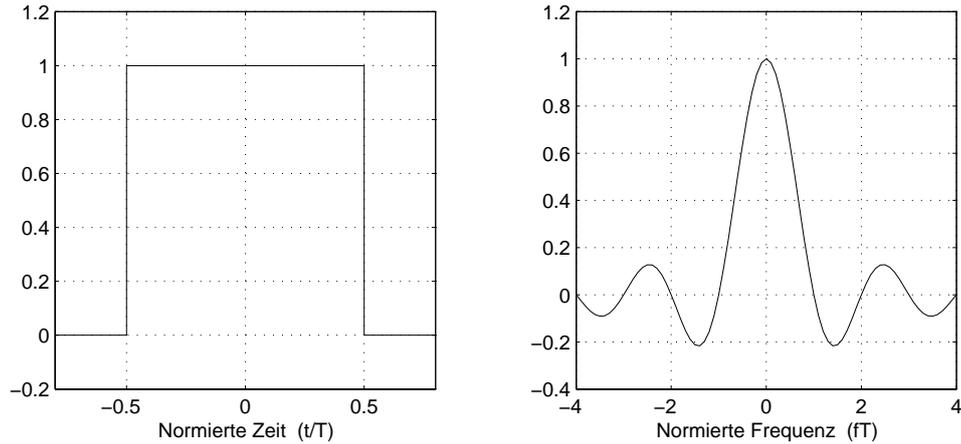


Abbildung 3.3: Die Rechteckfunktion und der Realteil ihrer Fourier-Transformierten. Da die so definierte Rechteckfunktion eine reelle und gerade Funktion ist, verschwindet der Imaginärteil der FT für alle Frequenzen und somit stellt der Realteil die vollständige Transformierte dar.

Beispiele

Fourier Transformation einer Rechteckfunktion

Eine Rechteckfunktion der Höhe 1 und der Breite $T = 2\tau$ sei wie folgt definiert:

$$h(t) = \begin{cases} 1 & |t| \leq \tau = T/2 \\ 0 & |t| > \tau = T/2 \end{cases}$$

Die Fourier Transformation ist dann:

$$H(\omega) = \int_{-\infty}^{\infty} h(t)e^{-j\omega t} dt = \int_{-\tau}^{\tau} 1e^{-j\omega t} dt,$$

und da h reell und gerade ist folgt,

$$H(\omega) = \int_{-\tau}^{\tau} \cos \omega t dt = \frac{1}{\omega} \sin \omega t \Big|_{-\tau}^{\tau} = 2 \frac{\sin \omega \tau}{\omega}.$$

Durch Multiplikation mit τ/τ und mit den Substitutionen $\tau = T/2$ und $\omega = 2\pi f$ erhält man das Ergebnis in Form einer sogenannten *sinc*-Funktion:

$$H(f) = T \frac{\sin(\pi T f)}{\pi T f} = T \operatorname{sinc}(\pi T f).$$

Unter Anwendung der Regel von L'Hospital erhält man auch den Wert für $f = 0$:

$$H(0) = \lim_{f \rightarrow 0} T \frac{\sin(\pi T f)}{\pi T f} = \lim_{f \rightarrow 0} T \frac{\pi T \cos(\pi T f)}{\pi T} = T.$$

Beachte, dass $H(0)$ gleich der Fläche des Rechteckes (also gleich dem Integral der Funktion) ist. Dies ist eine allgemeine Eigenschaft der FT!

Die Nulldurchgänge von $H(f)$ befinden sich bei den Werten

$$\pi T f_n = n\pi, \quad |n| = 1, 2, 3, \dots$$

woraus folgt

$$f_n = \frac{n}{T}.$$

Mit $n = 1$ sieht man, dass je schmaler (zeitlich kürzer) das Rechteck ist, desto breiter (frequenzmässig unschärfer) ist das Spektrum. Dies ist ein Beispiel der allgemein gültigen Unschärferelation der Fourier Transformation.

Statt die Höhe des Rechteckes gleich 1 zu setzen, soll sie im Folgenden gleich $1/2\tau$ sein:

$$h(t) = \begin{cases} \frac{1}{2\tau} & |t| \leq \tau \\ 0 & |t| > \tau \end{cases}$$

Dann ist die Fläche:

$$\int_{-\infty}^{\infty} h(t) dt = 2\tau \frac{1}{2\tau} = 1.$$

Die Fourier Transformierte ist dann:

$$H(\omega) = \frac{\sin \omega\tau}{\omega\tau}.$$

Lassen wir τ gegen 0 gehen, so dass der Impuls immer schmaler und höher wird, dann erhalten wir für alle ω :

$$\lim_{\tau \rightarrow 0} H(\omega) = \lim_{\tau \rightarrow 0} \frac{\sin \omega\tau}{\omega\tau} = 1.$$

Der Grenzwert $\tau \rightarrow 0$ in obiger Definition von $h(t)$ führt genau auf die Definition des *Dirac Delta Impulses* (Delta-Funktion, Dirac-Stoss):

$$\delta(t) = \begin{cases} \infty & t = 0 \\ 0 & t \neq 0 \end{cases}$$

mit der Eigenschaft,

$$\int_{-\infty}^{\infty} \delta(t) dt = 1.$$

Somit erhält man für die Fourier Transformation des Delta-Impulses:

$$\int_{-\infty}^{\infty} \delta(t) e^{-j\omega t} dt = 1.$$

Beachte, mit $g(\tau) = e^{-j\omega\tau}$ hat diese Beziehung für die Fourier Transformation des Delta-Impulses die Form:

$$\int_{-\infty}^{\infty} \delta(0 - \tau) g(\tau) d\tau = g(0).$$

Dies ist eine allgemeine Eigenschaft der δ -Funktion, welche auch als Definition verwendet wird.

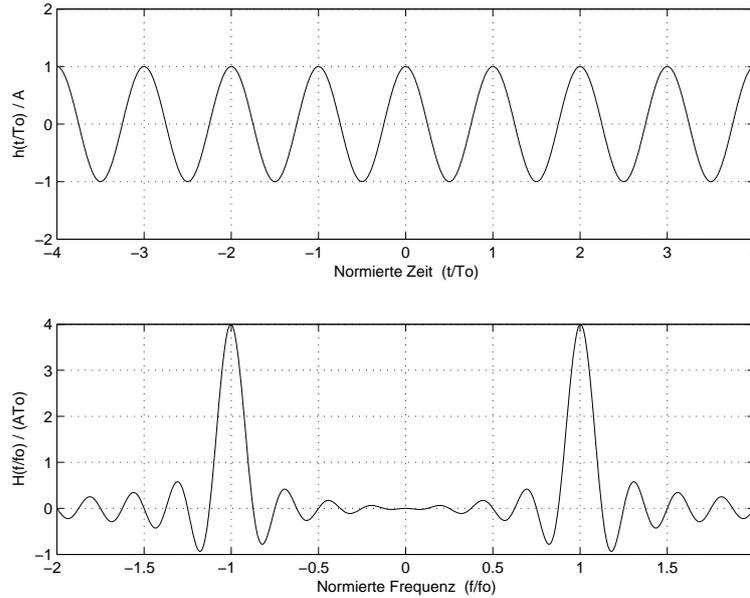


Abbildung 3.4: Eine zeitlich beschränkte Cosinus-Funktion und der Realteil ihrer Fourier-Transformierten. Da die so definierte harmonische Schwingung eine reelle und gerade Funktion ist, verschwindet der Imaginärteil der FT für alle Frequenzen und somit stellt der Realteil die vollständige Transformierte dar.

Fourier Transformation einer harmonischen Schwingung

Das Fourier Integral einer unendlich langen harmonischen Schwingung existiert nicht, da die Energie eines solchen Signals nicht endlich ist. Da sie aber periodisch ist, kann man sie natürlich in eine Fourier Reihe zerlegen und erhält zu der gegebenen Frequenz den entsprechenden Spektralwert. In der Praxis wendet man auch für solche Signale mit nicht endlicher Energie die Fourier Transformation an, indem man sich ein mehr oder weniger willkürliches Segment davon herausschneidet. Im Folgenden wollen wir die Fourier Transformation eines in dieser Art künstlich beschränkten Signals untersuchen.

Sei $h(t)$ ein Cosinus-Signal mit der Amplitude A und der Periode T_0 , bzw. mit der Eigenfrequenz $f_0 = 1/T_0$. Das betrachtete Zeitintervall T erstrecke sich über M Perioden, so dass $T = MT_0$. Um die Symmetrie-Eigenschaft der FT im Falle einer geraden Funktion ausnützen zu können, definieren wir das betrachtete Zeitintervall, so dass $-MT_0/2 < t < MT_0/2$. Mit $\omega_0 = 2\pi f_0 = 2\pi/T_0$ gilt:

$$h(t) = \begin{cases} A \cos(\omega_0 t) & |t| \leq \frac{MT_0}{2} = \frac{M\pi}{\omega_0} \\ 0 & |t| > \frac{MT_0}{2} = \frac{M\pi}{\omega_0} \end{cases}$$

Die Fourier Transformation ist dann:

$$H(\omega) = \int_{-\infty}^{\infty} h(t)e^{-j\omega t} dt = 2 \int_0^{\frac{M\pi}{\omega_0}} A \cos(\omega_0 t) e^{-j\omega t} dt,$$

und da h reell und gerade ist folgt,

$$H(\omega) = 2 \int_0^{\frac{M\pi}{\omega_0}} A \cos(\omega_0 t) \cos(\omega t) dt.$$

Die Lösung des Integrals ist:

$$H(\omega) = 2 \left\{ \frac{A \sin[(\omega_0 - \omega)t]}{2(\omega_0 - \omega)} + \frac{A \sin[(\omega_0 + \omega)t]}{2(\omega_0 + \omega)} \right\} \Big|_0^{\frac{M\pi}{\omega_0}}$$

$$H(\omega) = \frac{A \sin[(\omega_0 - \omega) \frac{M\pi}{\omega_0}]}{(\omega_0 - \omega)} + \frac{A \sin[(\omega_0 + \omega) \frac{M\pi}{\omega_0}]}{(\omega_0 + \omega)}.$$

Wiederum mit $T = MT_0$ und mit $\omega_0 = 2\pi f_0 = 2\pi/T_0$ gilt

$$H(f) = \frac{AMT_0}{2} \left\{ \frac{\sin[\pi MT_0(f_0 - f)]}{\pi MT_0(f_0 - f)} + \frac{\sin[\pi MT_0(f_0 + f)]}{\pi MT_0(f_0 + f)} \right\}.$$

Das Spektrum besteht also aus einer Summe von zwei *sinc*-Funktionen, die je an den Werten $\pm f_0$ zentriert sind.

Untersuchen wir nun den Betrag des Spektrums an der Stelle f_0 :

$$|H(f_0)| = \lim_{f \rightarrow f_0} \frac{AMT_0}{2} \left\{ \frac{\sin[\pi MT_0(f_0 - f)]}{\pi MT_0(f_0 - f)} + \frac{\sin[\pi MT_0(f_0 + f)]}{\pi MT_0(f_0 + f)} \right\}.$$

Wie wir im vorigen Beispiel gesehen haben, ist der erste Summand zwischen den geschweiften Klammern, unter Anwendung der Regel von L'Hospital, gleich 1. Der zweite Summand wird zu

$$\frac{\sin[2\pi M]}{2\pi M}$$

und verschwindet für alle ganzzahligen M . Der Betrag des Spektrums einer einzelnen harmonischen Schwingung der Amplitude A ist also proportional zur Länge T des betrachteten Zeitintervalls:

$$|H_{f_0}| = \frac{AMT_0}{2} = \frac{A}{2}T.$$

Statt der intuitiv erwarteten diskreten Spektralwerte bei den Frequenzen $\pm f_0$ (bzw. $\pm \omega_0$) erhalten wir ein Spektrum, welches der Faltung dieser diskreten Spektralwerte mit der *sinc*-Function entspricht. Diese Faltung im

Spektralbereich folgt auch aus der Tatsache, dass wir die zeitliche Begrenzung des Signals durch eine Multiplikation der Cosinus-Funktion mit einer Rechteck-Funktion im Zeitbereich erhalten haben. Die Breite des Hauptmaximums, die sich aus dem ersten Nulldurchgang ergibt, ist eine Funktion der Länge des betrachteten Zeitfensters. Wenn $M \gg 1$ dann kann man in erster Näherung den Beitrag von $H(-f)$ auf das Verhalten von $H(+f)$ in der Nähe von $+f_0$ vernachlässigen. Die Nullstellen von $H(f)$ ergeben sich dann aus,

$$\pi T(f_0 - f_n) = n\pi, \quad |n| = 1, 2, 3, \dots$$

oder

$$f_0 - f_n = n/T$$

und für die Halbbreite des Hauptmaximums

$$|f_0 - f_1| = 1/T.$$

Die allgemein gültige Folgerung daraus ist, je länger das Zeitfenster, also je mehr Daten zur Verfügung stehen, desto genauer kann der Spektralgehalt eines Signals aufgelöst werden.

Das Abtasttheorem

Nun sind wir in der Lage eine formale Antwort auf die grundsätzliche Frage zu geben, *was ist die minimale Abtastrate, um eine gegebene kontinuierliche Funktion vollständig und unverfälscht zu diskretisieren?*

Wir definieren zuerst die verallgemeinerte Funktion $\delta(t)$ (Deltafunktion, Dirac–Stoss):

$$\delta(t) = \begin{cases} 0, & \text{für } t \neq 0 \\ \infty, & \text{für } t = 0 \end{cases}$$

so dass

$$\int_{-\infty}^{+\infty} \delta(t) dt = 1.$$

Diese “Funktion” besitzt folgende Eigenschaft:

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

Das heisst, dass die Faltung einer Funktion $x(\tau)$ mit dem δ -Impuls eine Selektion des Wertes von x an der Stelle $\tau = t$ bewirkt.

Formal führen wir nun die Abtastung durch, indem wir $x(t)$ mit einer Folge von Deltafunktionen $\delta(t)$ multiplizieren:

$$x(t) \Rightarrow x_k(t) = \sum_{k=-\infty}^{+\infty} x(t) \delta(t - k\Delta t),$$

wobei Δt hier die Abtastperiode darstellt.

Wie müssen wir nun beim Diskretisierungsprozess vorgehen, um sicher zu sein, dass wesentliche Merkmale dieser Funktion nicht verloren gehen?

In den Jahren 1948–1949 veröffentlichte C.E. Shannon Arbeiten, mit denen er die Informationstheorie begründete. In diesen Arbeiten formulierte er das sogenannte **Abtasttheorem** (in der hier verwendete Notation) wie folgt:

.... Let $x(t)$ contain no frequencies over f_{Nyq} , then:

$$x(t) = \sum_{k=-\infty}^{+\infty} x_k \frac{\sin \pi(2f_{Nyq}t - k)}{\pi(2f_{Nyq}t - k)},$$

where

$$x_k = x\left(\frac{k}{2f_{Nyq}}\right) = x(k\Delta t).$$

Obwohl einige Mathematiker dieses Theorem schon vor Shannon exakt formulierten und auf Probleme der Nachrichtentechnik anwendeten, wird es heute gewöhnlich als “Shannon’s Abtasttheorem” bezeichnet. Dieses Theorem sagt, dass man eine beliebige Funktion $x(t)$, die sich aus einer diskreten Anzahl Frequenzen von 0 bis f_{Nyq} zusammensetzt, durch die obige Reihe darstellen kann. Umgekehrt wiederum besteht eine beliebige Funktion $f(t)$, die

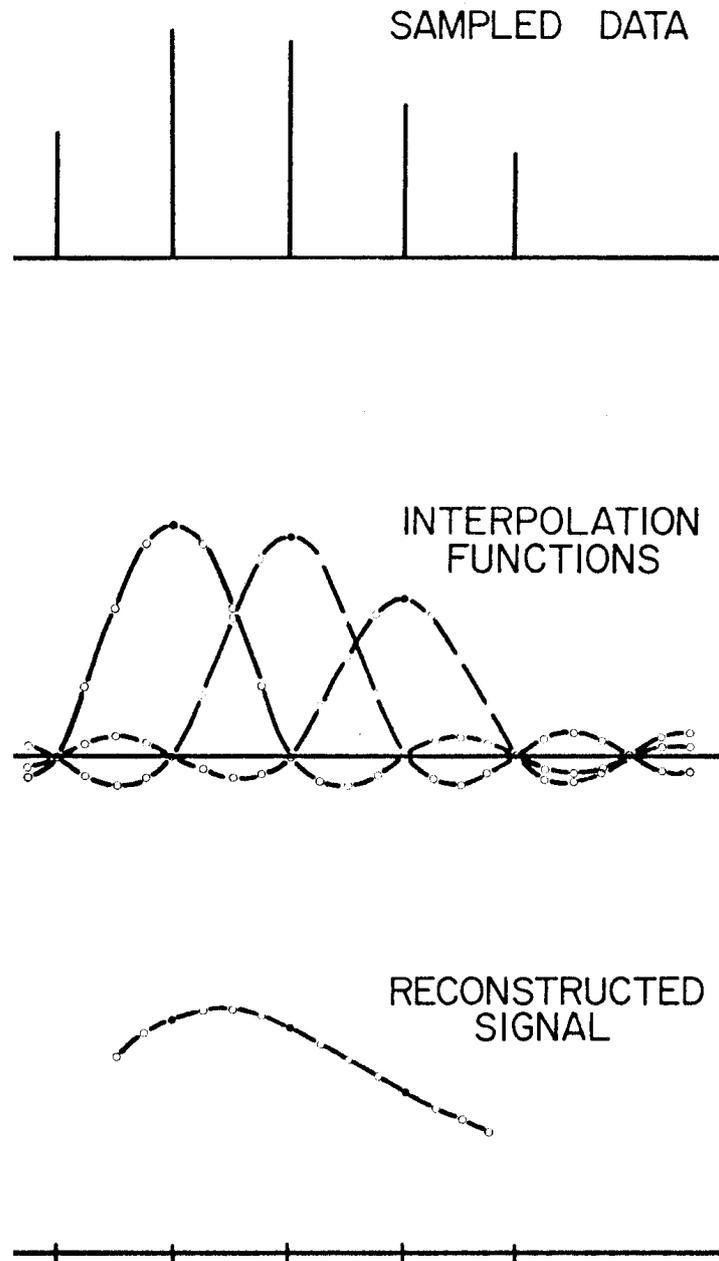


Abbildung 3.5: Graphische Darstellung des auf Shannons Abtasttheorem basierenden Interpolationsformalismus. (Aus: A Pictorial Digital Atlas, Bendix United Geophysical Corporation, 1966)

durch die obige Reihe dargestellt wird, nur aus Frequenzen von 0 bis f_{Nyq} . In der Formulierung von Shannon liefert das Abtasttheorem auch einen Formalismus, der es erlaubt, eine diskrete Funktion beliebig fein zu interpolieren (siehe Abbildung 3.5).

Wir können das Abtasttheorem auch folgendermassen formulieren: Die Abtastung mit der Periode Δt ist nur dann sinnvoll, wenn die zu diskretisierende Funktion keine Energie bei den Frequenzen $f > 1/(2\Delta t)$ enthält. Die halbe Abtastfrequenz wird Nyquist-Frequenz genannt:

$$f_{Nyquist} = f_{Nyq} = \frac{f_s}{2} = \frac{1}{2\Delta t}.$$

Auf Englisch wird letztere auch *folding frequency* genannt, weil Frequenzen grösser als f_{Nyq} auf kleinere Frequenzen zurückgefaltet werden: $f = f_{Nyq} + y$ wird zu $f_a = f_{Nyq} - y$ (dies hat aber nichts mit dem oben definierten Faltungintegral zu tun!) (siehe Abbildung 3.6). Dieser Effekt wird mit dem englischen Wort *Aliasing* bezeichnet und f_a ist die *Alias Frequenz* von f (*Alias* heisst Doppelname). Mit anderen Worten, zu hohe Frequenzen werden bei ungenügender Abtastrate durch die Analog/Digital Wandlung als falsche tieferfrequente Signalanteile wiedergegeben. Daten, welche durch Aliasing verfälscht wurden, sind unbrauchbar! Um Aliasing zu vermeiden, muss entweder mit einer genügend hohen Frequenz abgetastet werden, oder mit einem Tiefpass-Filter (*Anti-Alias Filter*) vor der A/D Wandlung dafür gesorgt werden, dass keine Frequenzen höher als f_{Nyq} im Signal vorhanden sind. Aliasing ist ein allgemeines Phänomen diskreter Daten und kann sowohl bei Zeitserien, wie Seismogramme oder Temperaturmessungen, als auch bei räumlichen Daten, wie über ein Gebiet verteilte Schweremessungen oder Seismometer-Arrays, auftreten. Letzteres wird im Englischen als *spatial aliasing* bezeichnet.

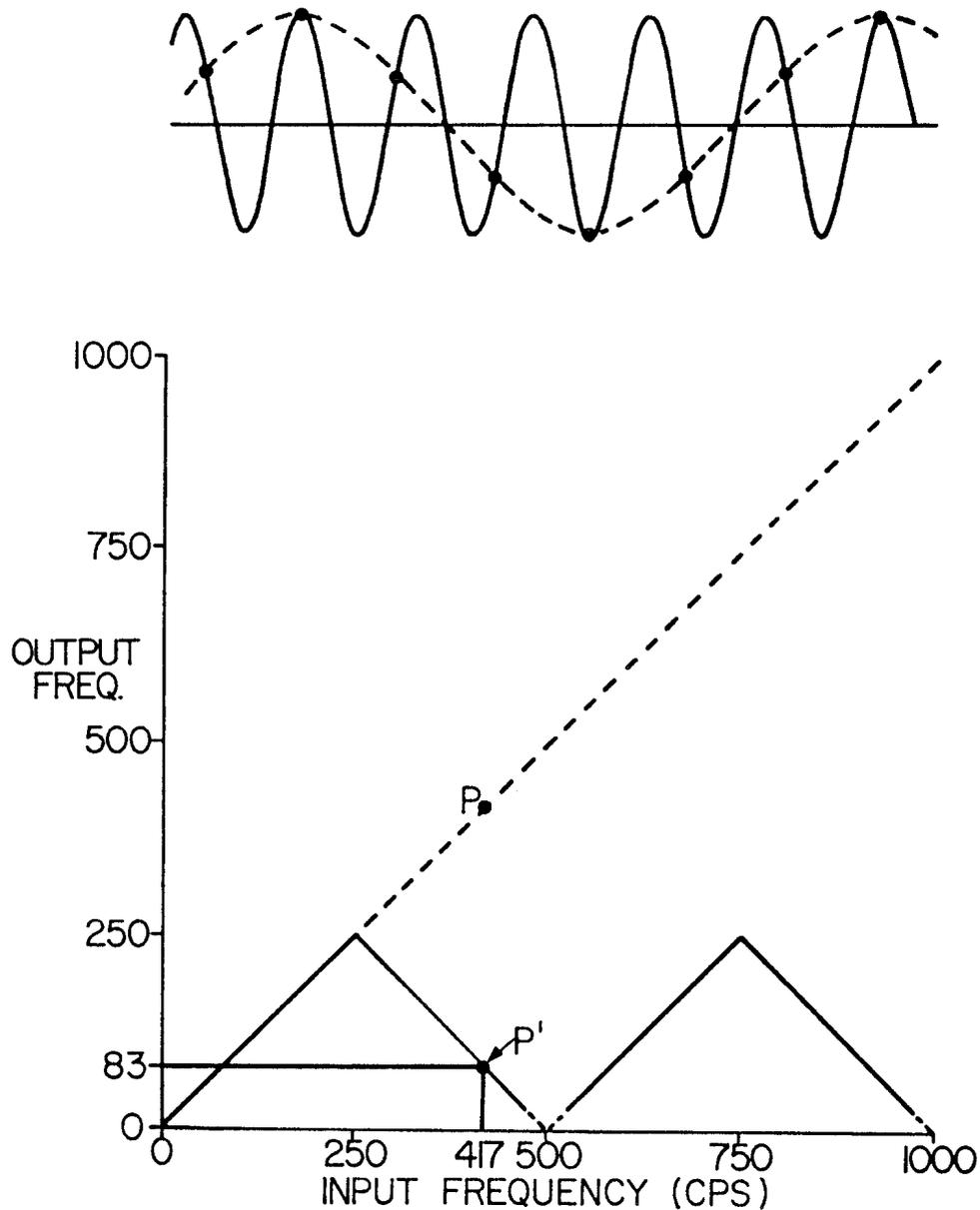


Abbildung 3.6: Schematische Darstellung des Aliasing Phänomens: Gegeben sei eine harmonische Schwingung mit einer Frequenz von $f = 417 \text{ Hz}$, welche mit 500 Hz abgetastet wurde ($f_{Nyquist} = 250 \text{ Hz}$). Das Ergebnis ist, dass die $f = 250 + 167 = 417 \text{ Hz}$ auf $f_a = 250 - 167 = 83 \text{ Hz}$ zurückgefallen werden. (Aus: A Pictorial Digital Atlas, Bendix United Geophysical Corporation, 1966)

Kapitel 4

Das Leistungsspektrum

Energie und Leistungsspektrum

Eine wichtige Beziehung zwischen einer Funktion und ihrer Fourier Transformierten ist der Energieerhaltungssatz (*Theorem von Parseval*):

$$\int_{-\infty}^{\infty} |h(t)|^2 dt = \int_{-\infty}^{\infty} |H(f)|^2 df.$$

Betrachtet man $|h(t)|^2$ als ein Mass der momentanen Leistung, dann ergibt das Integral über die Zeit die Energie. Folglich wird die Grösse $|H(f)|^2 = \Re(f)^2 + \Im(f)^2$ als *spektrale Energiedichte* bezeichnet, dessen Integral über alle Frequenzen wiederum gleich der Energie ist. Wenn z.Bsp. die Funktion $h(t)$ eine Verschiebung mit Einheit m ist, dann hat $|H(f)|^2$ die Einheit $\text{m}^2/\text{Hz}^2 = \text{m}^2\text{s}/\text{Hz}$, nämlich Leistung mal Zeit pro Frequenz.

Wenn $H(f)$ die Fourier Transformierte von $h(t)$ ist, dann ist also das Quadrat des Amplitudenspektrums $|H(f)|^2$ die spektrale Energiedichte. Für ein zeitlich begrenztes Signal der Dauer T ergibt sich dann das Leistungsspektrum oder genauer ausgedrückt die spektrale Leistungsdichte $P(f)$ (*power spectral density*) aus

$$P(f) = \frac{1}{T} |H(f)|^2.$$

Dieses direkt aus der Fourier Transformation ermittelte Leistungsspektrum ist als *Periodogramm* bekannt.

Das diskrete Leistungsspektrum

In Analogie zum kontinuierlichen Fall bezeichnet man das Betragsquadrat der DFT $|X_k|^2$ als spektrale Energiedichte (dies folgt daraus, dass x_n^2 als die Leistung des Signals betrachtet wird, und wenn z.Bsp. $[x_n] = \text{m}$ ist, dann ist

$[x_n^2] = \text{m}^2$ und $[|X_k|^2] = \text{m}^2/\text{Hz}^2 = \text{m}^2\text{s}/\text{Hz}$, was einer Energie pro Frequenz entspricht):

$$|X_k|^2 = \left| \Delta t \sum_{n=0}^{N-1} x_n e^{-j \frac{2\pi kn}{N}} \right|^2.$$

Dann ergibt sich der Energieerhaltungssatz (*Theorem von Parseval*) im Diskreten Fall wie folgt:

$$\Delta t \sum_{n=0}^{N-1} |x_n|^2 = \Delta f \sum_{k=0}^{N-1} |X_k|^2 = \frac{1}{N\Delta t} \sum_{k=0}^{N-1} |X_k|^2.$$

Da die Leistung als Energie pro Zeiteinheit definiert ist, erhalten wir die spektrale Leistungsdichte, $P(f_k)$, (engl. power spectral density, PSD) aus der spektralen Energiedichte mittels Division durch die Zeit:

$$P(f_k) = \frac{1}{N\Delta t} |X_k|^2 = \frac{\Delta t}{N} \left| \sum_{n=0}^{N-1} x_n e^{-j \frac{2\pi kn}{N}} \right|^2.$$

So normiert hat das Leistungsspektrum $P(f_k)$ die korrekten Einheiten von Leistung pro Frequenz (z.Bsp. m^2/Hz).

Eine andere oft verwendete Normierung des Leistungsspektrums lautet:

$$P'(f_k) = \frac{1}{N^2} \left| \sum_{n=0}^{N-1} x_n e^{-j \frac{2\pi kn}{N}} \right|^2$$

oder anders geschrieben:

$$P'(f_k) = \frac{1}{N\Delta t} \frac{\Delta t}{N} \left| \sum_{n=0}^{N-1} x_n e^{-j \frac{2\pi kn}{N}} \right|^2 = \frac{1}{N\Delta t} P(f_k).$$

Daraus sieht man, dass $P'(f_k)$ einer spektralen Leistungsdichte pro Zeiteinheit entspricht, mit einer Dimension der Leistung. Bei einem komplexen Signal hat diese Normierung den Vorteil, dass die so gewonnenen Koeffizienten genau der Leistung der jeweiligen harmonischen Schwingung entsprechen. Im Gegensatz zum Fall der Amplituden- und Leistungsdichte, ist ausserdem der Betrag von $P'(f_k)$ unabhängig von der Dauer des analysierten Zeitfensters, was in der Praxis von Vorteil ist, wenn Leistungsspektren von verschiedenen langen Signalen miteinander verglichen werden sollen.

Wenn man beide Seiten der Parseval'schen Beziehung durch die Signallänge $N\Delta t$ dividiert, erhält man eine Beziehung für die mittlere Leistung,

$$\frac{1}{N} \sum_{n=0}^{N-1} |x_n|^2 = \frac{1}{(N\Delta t)^2} \sum_{k=0}^{N-1} |X_k|^2 = \frac{1}{N\Delta t} \sum_{k=0}^{N-1} P(f_k) = \sum_{k=0}^{N-1} P'(f_k).$$

Für eine Zeitreihe deren Mittelwert 0 ist, stellt der linke Term dieser Gleichung die Varianz dar. In diesem Fall ist somit die Summe des Leistungsspektrums über alle Frequenzen gleich der Varianz von x_n :

$$\sum_{k=0}^{N-1} P'(f_k) = \text{Var}(x_n).$$

Kapitel 5

Korrelation

Kreuzkovarianz und Korrelation

In der Praxis stellt sich oft das Problem, die Aehnlichkeit zwischen zwei Zeitreihen oder die Verschiebung zwischen zwei Zeitreihen, bei der die Aehnlichkeit maximal ist, zu bestimmen. Oder es stellt sich die Frage nach der möglichen Abhängigkeit zwischen zwei Zeitreihen. In diesem Zusammenhang sind die Begriffe der Kovarianz und Korrelation von grosser Wichtigkeit.

Kovarianz

Die Kreuzkovarianzfunktion zweier kontinuierlicher reeller Funktionen $x(t)$ und $y(t)$ ist wie folgt definiert:

$$R_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T [x(t + \tau) - \overline{x(t)}][y(t) - \overline{y(t)}] dt.$$

Hier sind \overline{x} und \overline{y} die entsprechenden Mittelwerte

$$\overline{x(t)} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) dt,$$

und analog für \overline{y} . Wir wollen im folgenden davon ausgehen, dass die Mittelwerte Null sind oder von unseren Zeitreihen abgezogen worden sind.

Im Fall diskreter Zeitreihen lautet dann die Definition für die diskrete Kreuzkovarianzfunktion

$$R_{xy}(k) = \lim_{N \rightarrow \infty} \frac{1}{2N + 1} \sum_{n=-N}^{n=+N} x_{n+k} y_n.$$

Die Grösse τ bzw. der Index k werden als Verschiebung oder *lag* zwischen den zwei Zeitreihen x und y bezeichnet. Von im Prinzip unendlich langen stochastischen Zeitreihen stehen uns in der Praxis natürlich nur Abschnitte

endlicher Länge zur Verfügung ($x_n, y_n; n = 0, \dots, N-1$), wobei angenommen wird, dass die Zeitreihen für $n < 0$ und $n > N-1$ Null sind. Wir sprechen dann von einer Schätzung für die Kreuzkovarianz:

$$R_{xy}(-k) = \frac{1}{N} \sum_{n=0}^{N-1-k} x_n y_{n+k}, \quad (k = N-1, \dots, 1)$$

$$R_{xy}(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} x_{n+k} y_n, \quad (k = 0, \dots, N-1)$$

Beachte, dass das Δt im Zähler vor dem Summenzeichen, welches dem dt im Integral entspricht, stillschweigend weggelassen wurde, weil es sich mit dem Δt von $T = N\Delta t$ im Nenner aufhebt. Oft wird auch der Normierungsfaktor $1/N$ weggelassen. Die Rechenprozedur sieht dann wie folgt aus:

$$\begin{aligned} \hat{R}_{xy}(-N+1) &= x_0 y_{N-1} \\ &\vdots \\ \hat{R}_{xy}(-1) &= x_0 y_1 + x_1 y_2 + \dots + x_{N-2} y_{N-1} \\ \hat{R}_{xy}(0) &= x_0 y_0 + x_1 y_1 + \dots + x_{N-1} y_{N-1} \\ \hat{R}_{xy}(1) &= x_1 y_0 + x_2 y_1 + \dots + x_{N-1} y_{N-2} \\ \hat{R}_{xy}(2) &= x_2 y_0 + x_3 y_1 + \dots + x_{N-1} y_{N-3} \\ &\vdots \\ \hat{R}_{xy}(N-1) &= x_{N-1} y_0 \end{aligned}$$

Stellt man sich die Zeitreihen x und y untereinander geschrieben vor, dann wird die Zeitreihe y jeweils um k Stützwerte nach links ($k < 0$) oder nach rechts ($k > 0$) verschoben und die entsprechenden Werte von x und y werden miteinander multipliziert und aufaddiert.

In der folgenden Abbildung wird das Resultat gezeigt für die zwei Folgen

$$x_n = 0, 0, 0, 0, 0, 3, 0, -3, 0, 0$$

und

$$y_n = 0, -1, 0, 2, 0, -1, 0, 0, 0, 0$$

mit der Kovarianz

$$\hat{R}_{xy} = 0, 0, 0, 0, 0, 0, 0, 0, 0, -3, 0, 9, 0, -9, 0, 3, 0, 0, 0.$$

Das Maximum der Kreuzkovarianzfunktion wird bei einer Verschiebung (*Lag*) von $k = 2$ Stützwerten erreicht. Das entspricht dem Versatz zwischen der Maxima von x und y und gibt somit an, wo die grösstmögliche Uebereinstimmung zwischen den zwei Zeitreihen herrscht.

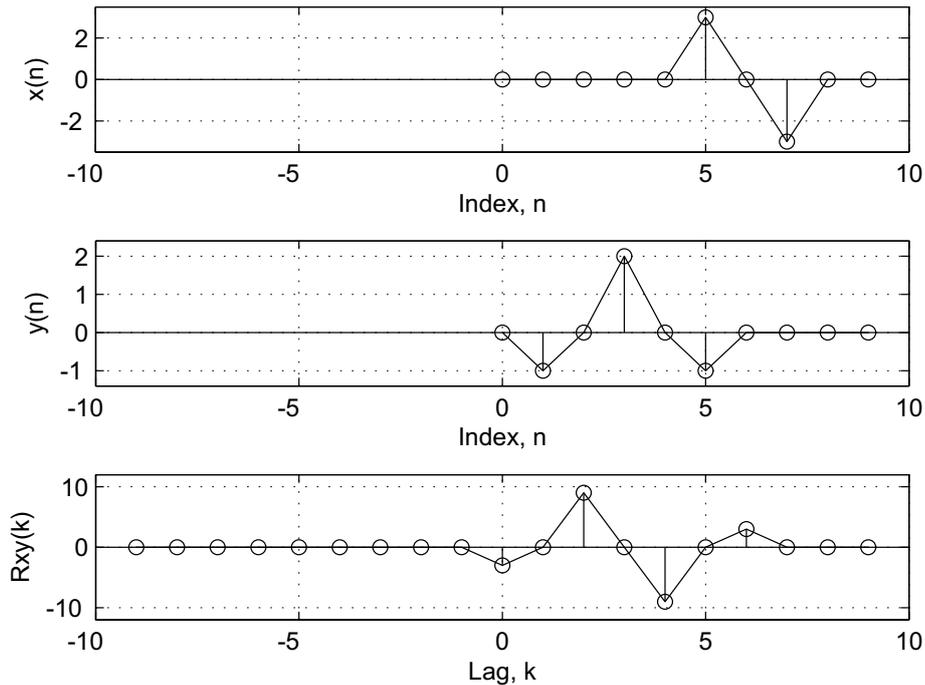


Abbildung 5.1: Diskrete unnormierte Kovarianzfunktion $\hat{R}_{xy}(k)$ der Zeitreihen $x(n)$ und $y(n)$. Diese wurde in MATLAB mit der Funktion `xcorr(x,y)` berechnet. $\text{Max}(\hat{R}_{xy}(k))$ ist bei $k > 0$ weil $x(n)$ gegenüber $y(n)$ verzögert ist.

Das genau gleiche Resultat erhält man, wenn man die Reihenfolge der zwei Zeitreihen vertauscht und dann die untere Zeitreihe relativ zur oberen nach links verschiebt. Daraus ergibt sich das allgemeine Resultat

$$R_{xy}(k) = R_{yx}(-k).$$

Korrelation

Die Kreuzkovarianzfunktion oder Kovarianzfunktion wird auch oft als Kreuzkorrelation oder einfach Korrelation bezeichnet. Streng genommen sollten die letzten zwei Bezeichnungen der durch die Varianzen normierten Kreuzkovarianzfunktion vorbehalten sein. Für eine kontinuierliche Funktion ist die Varianz definiert als

$$\text{Var}\{x(t)\} = \sigma_x^2 = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T [x(t) - \bar{x}]^2 dt.$$

Wiederum unter der Annahme, dass der Mittelwert Null ist, erhalten wir für die Varianz einer diskreten Zeitreihe

$$\text{Var}(x_n) = \sigma_x^2 = \lim_{N \rightarrow \infty} \frac{1}{2N + 1} \sum_{n=-N}^{n=+N} x_n^2,$$

oder für eine Schätzung der Varianz aus Ausschnitten endlicher Länge

$$\text{Var}(x_n) = \sigma_x^2 = \frac{1}{N} \sum_{n=0}^{N-1} x_n^2.$$

Die normierte Kreuzkorrelation, auch Korrelationskoeffizient genannt, ist dann

$$C_{xy}(k) = \frac{R_{xy}(k)}{\sqrt{\text{Var}(x)\text{Var}(y)}}.$$

Der Maximalwert der Kreuzkorrelation ist ein Mass für die Aehnlichkeit zweier Zeitreihen. Für zwei Zeitreihen, die vollkommen unkorreliert sind und somit nichts Gemeinsames haben, ist $C_{xy} = 0$. Die Grösstmögliche Aehnlichkeit haben natürlich zwei identische Zeitreihen bei der Verschiebung $k = 0$. Wie sich aus den Definitionen der Kovarianz und der Varianz leicht verifizieren lässt, ist dann $c_{xy}(0) = 1$. Ist hingegen $y_n = -x_n$ dann ist $C_{xy}(0) = -1$.

Autokovarianz und Autokorrelation

Autokovarianz

Analog zur Kovarianz von zwei verschiedenen Zeitreihen koennen wir auch die Kovarianz einer Zeitreihe mit sich selber bilden. Diese wird Autokovarianz genannt. Für eine diskrete Zeitreihe x_n mit Mittelwert gleich Null ist sie definiert als

$$R_{xx}(k) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^{n=+N} x_{n+k}x_n.$$

Aus der für die Kreuzkovarianz gültigen Beziehung $R_{xy}(k) = R_{yx}(-k)$ folgt $R_{xx}(k) = R_{xx}(-k)$. Somit ist die Autokovarianz eine gerade Funktion und es genügt, sie für positive Werte von k zu berechnen. In der Praxis können wir die Autokovarianz einer unendlich langen Zeitreihe nicht berechnen, sondern nur aus Abschnitten endlicher Länge abschätzen:

$$R_{xx}(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} x_{n+k}x_n, \quad (k = 0, \dots, M),$$

wobei $M \leq N - 1$.

Es zeigt sich jedoch, dass obige Schätzung mit einem systematischen Fehler behaftet ist (*biased estimate*). Eine Fehlerfreie Schätzung (*unbiased estimate*) erhält man wenn man den Faktor $1/N$ vor dem Summenzeichen mit $1/(N - |k|)$ ersetzt. Letztere Normierung hat aber den Nachteil, dass die Varianz der Schätzung für grosse Werte von M sehr gross wird. Da ausserdem für $N \rightarrow \infty$ sowohl der Fehler als auch die Varianz bei beiden Normierungsarten gegen Null gehen, wird in der Praxis meistens die erste Form verwendet.

Wie man sofort sieht, ist der Wert der Autokovarianz bei der Verschiebung $k = 0$ nichts anderes als die Varianz des Signals:

$$R_{xx}(0) = \frac{1}{N} \sum_{n=0}^{N-1} x_n^2 = \sigma_x^2 = \text{Var}(x).$$

Für den Fall eines Leistungssignals entspricht diese Definition der mittleren Leistung des Signals. Ausserdem gilt $R_{xx}(k) \leq R_{xx}(0)$.

Autokorrelation

Die Autokorrelation eines Signals ist einfach die mit der Varianz des Signals normierte Autokovarianz:

$$C_{xx}(k) = \frac{R_{xx}(k)}{R_{xx}(0)}.$$

Der Maximalwert der Autokorrelation ist somit $C_{xx}(0) = 1$.

Für determinierte Signale endlicher Länge (*transient signals* oder *wavelets*), deren Energie endlich ist, geht die mittlere Leistung mit $N \rightarrow \infty$ gegen Null. Für solche Signale wird die Autokovarianz ohne Division durch N definiert:

$$R_{xx}(k) = \sum_{n=0}^{N-1-k} x_{n+k}x_n.$$

Die Autokovarianz für $k = 0$ entspricht dann der Energie des Signals:

$$R_{xx}(0) = \sum_{n=0}^{N-1} x_n^2.$$

Die Definition der Autokorrelation bleibt aber für diesen Fall unverändert.

Aus den Eigenschaften der Autokovarianz lässt sich einiges über die interne Gesetzmässigkeit oder umgekehrt über den Grad der Regellosigkeit eines Signals aussagen. Die Autokovarianz einer harmonische Schwingung ist wiederum eine harmonische Schwingung der gleichen Frequenz (aber ohne Phaseninformation, da ja die Autokovarianz eine gerade Funktion ist). Je regelloser ein stochastisches Signal ist, umso schmaler wird das Maximum der Autokovarianzfunktion. Ein vollkommen regelloses Signal (sogenanntes weisses Rauschen) ist von einem Zeitpunkt zu einem anderen maximal unkorreliert. Somit gilt im Idealfall für die Autokovarianz eines regellosen Prozesses $R_{xx}(k) = 0$ ($k \neq 0$) und $R_{xx}(0) = \sigma_x^2$.

Eine ausführliche Diskussion über die Aussagekraft der Autokovarianz sowie einige konkrete Beispiele sind in Kapitel 8 des Buches von Buttkus¹ zu finden.

¹B. Buttkus, Spektralanalyse und Filtertheorie in der angewandten Geophysik, Springer Verlag, Berlin (1991)

Autokovarianz und Leistungsspektrum

Eine direkte Anwendung der Autokovarianz besteht im Zusammenhang mit der Schätzung der spektralen Leistungsdichte von stochastischen Signalen.

Wenn $X(f)$ die Fourier Transformierte eines zeitlich begrenzten Signals $x(t)$ der Dauer T ist, dann ist bekanntlich das Leistungsspektrum oder genauer ausgedrückt die spektrale Leistungsdichte $P(f)$ (*power spectral density*)

$$P(f) = \frac{1}{T}|X(f)|^2.$$

Dieses direkt aus der Fourier Transformation ermittelte Leistungsspektrum ist als *Periodogramm* bekannt.

Ein alternatives Vorgehen geht von der Autokovarianz des Signals aus. Nach dem Theorem von Wiener und Khintchine bilden nämlich die Autokovarianzfunktion und die quadratische Spektraldichtefunktion ein Fourier-Transformationspaar:

$$R_{xx}(\tau) = \int_{-\infty}^{\infty} P(f)e^{j2\pi f\tau} df$$

und

$$P(f) = \int_{-\infty}^{\infty} R_{xx}(\tau)e^{-j2\pi f\tau} d\tau.$$

Ein Beweis dieses Satzes ist z.Bsp. im Buch von Buttkus (Kap. 8) zu finden.

Dieser Sachverhalt kann genutzt werden, um die spektrale Leistungsdichte zu schätzen. Die Methode ist als *Korrelogramm* Verfahren oder als Blackman-Tukey Schätzung bekannt. Das praktische Vorgehen sowie eine Diskussion über Auflösung und Varianz sowohl des Korrelogramm- als auch des Periodogramm-Verfahrens sind wiederum bei Buttkus zu finden (Kap 9).

Die Autokovarianz sowie die spektrale Leistungsdichte, zusammen mit dem Mittelwert und der Varianz, werden verwendet, um statistische Eigenschaften von mehr oder weniger regellosen Prozessen von im Prinzip unbegrenzter Dauer zu erfassen. Eine Eigenschaft regelloser Prozesse besteht darin, dass die Phasenlage nicht determiniert ist, sondern zufälligen Schwankungen unterliegt. Somit ist es auch sinnvoll, dass in diesen Größen keine Information über die Phase der analysierten Signale enthalten ist.

Kapitel 6

Stochastische Zeitreihen

Definitionen

Im Folgenden sollen einige wichtige Begriffe im Zusammenhang mit der Analyse von stochastischen Prozessen erläutert werden. Weiterführende und ausführliche Darstellungen dieses sehr weitläufigen Gebietes der Zeitreihenanalyse sind zum Beispiel in den Lehrbüchern von Schlittgen und Streitberg sowie von Shumway und Stoffer zu finden ¹

Stochastische Zeitreihen $\{x_n\}$ sind als Realisierungen von stochastischen Prozessen zu verstehen, d.h. von dynamischen Vorgängen mit Zufallscharakter. Ein stochastischer Prozess ist eine Folge $\{X_t\}$ von Zufallsvariablen X_t , wobei der Index t Element einer höchstens abzählbaren Menge T ist. In der Analyse von stochastischen Zeitreihen geht es meistens darum, gewisse statistische Gesetzmässigkeiten zu finden, welche den zugrundeliegenden stochastischen Prozess beschreiben. Dabei ist es wichtig, zu unterscheiden zwischen den beobachteten oder berechenbaren Eigenschaften einer oder mehrerer Realisierungen des Prozesses und diesen zugrundeliegenden Eigenschaften, die wir nicht direkt beobachten können sondern aus den Beobachtungen schätzen möchten. Somit unterscheidet die Notation von Schlittgen und Streitberg durch Verwendung von Gross- und Kleinbuchstaben zwischen dem stochastischen Prozess einerseits und der gegebenen Realisierung als Zeitreihe andererseits. Wir werden wie Shumway und Stoffer im Folgenden aber in beiden Fällen Kleinbuchstaben verwenden, in der Meinung, dass der Sinn aus dem Zusammenhang klar wird.

In der Notation muss man jedoch unterscheiden zwischen den entsprechenden statistischen Grössen. So sind die in Kapitel 2 eingeführten Mittelwerte, Varianzen und Autokovarianzen empirisch bestimmte Grössen einer

¹R. Schlittgen & B. Streitberg: Zeitreihenanalyse. Oldenburg Verlag, München, 8.Aufl., 1999.

R. H. Shumway & D. S. Stoffer: Time Series Analysis and its Applications. Springer Verlag, New York, 2000.

gegebenen Zeitreihe x_n :

$$\bar{x} = \frac{1}{N} \sum_{n=0}^{N-1} x_n$$

$$\text{Var}(x_n) = \sigma_x^2 = \frac{1}{N} \sum_{n=0}^{N-1} (x_n - \bar{x})^2$$

$$R_{xx}(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} (x_{n+k} - \bar{x})(x_n - \bar{x}), \text{ mit } k = 0, \dots, M \text{ und } M \leq N - 1.$$

Der Mittelwert des entsprechenden Zufallsprozesses ist hingegen gemäss der klassischen Statistik als Erwartungswert für einen bestimmten Index t über alle möglichen Realisierungen des Prozesses definiert:

$$\mu_t = \mu_{tx} = E[x_t] = \int_{-\infty}^{\infty} x p_t(x) dx,$$

wobei $p_t(x)$ die entsprechende Wahrscheinlichkeitsdichte ist. In gleicher Art ist auch die theoretische Autokovarianz als Erwartungswert über alle möglichen Realisierungen der Werte x_s und x_t definiert:

$$\gamma(s, t) = \gamma_x(s, t) = \text{Cov}[x_s, x_t] = E[(x_s - \mu_s)(x_t - \mu_t)].$$

Die theoretische Autokorrelation ist die durch die Varianzen normierte Autokovarianz:

$$\rho(s, t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}}.$$

Zufallsprozesse, für die es möglich ist, den Ensemble-Mittelwert μ und die Ensemble-Autokovarianzfunktion γ durch Bildung des zeitlichen Mittelwertes über eine einzige Zeitreihe bzw. durch Berechnung der zeitlichen Autokovarianz zu schätzen, nennt man **ergodisch**. Eine exakte Darlegung der Bedingungen für Ergodizität ist z.Bsp. im Kapitel 5 von Schlittgen und Streitberg zu finden. Es zeigt sich, dass die meisten in der Praxis vorkommenden Zufallsprozesse diesen Bedingungen zumindest annähernd genügen.

Eine weitere Bedingung, die für die Modellierung eines stochastischen Prozesses vorausgesetzt wird, ist dass er **stationär** ist. Konkret, bezeichnet man einen stochastischen Prozess als:

mittelwertstationär wenn $\mu_t = \mu$ konstant ist,

varianzstationär wenn $\sigma_t^2 = \sigma^2$ konstant ist und

kovarianzstationär wenn die $\gamma(s, t) = \gamma(\tau)$ ist, d.h. wenn die Kovarianz nur vom Abstand (oder *lag*) $\tau = s - t$ abhängt. In diesem Fall ist auch die Autokorrelation $\rho(\tau) = \gamma(\tau)/\gamma(0)$.

Ein stochastischer Prozess, der sowohl mittelwertstationär als auch kovarianzstationär ist, wird als **schwach stationär** bezeichnet. (Frage: ist ein schwach stationärer Prozess auch Varianzstationär?) In der Praxis wird meistens der Begriff stationär an Stelle von schwach stationär verwendet.

Beispiele

White-Noise Prozesse

Ein **White-Noise Prozess** ist ein reiner Zufallsprozess der aus einer Folge (w_t) von identisch verteilten und unabhängigen Zufallsvariablen w_t besteht. Oft wird für ein solches **weisses Rauschen** eine Normalverteilung mit Mittelwert 0 und gegebener Varianz σ_w^2 vorausgesetzt. Aus der Unabhängigkeit der einzelnen Werte w_t folgt für die Autokovarianz:

$$\gamma_w(s, t) = \begin{cases} \sigma_w^2, & s = t \\ 0, & s \neq t \end{cases}$$

Daraus ist ersichtlich, dass White-Noise Prozesse stationär sind. White-Noise Prozesse sind zwar an sich nicht besonders interessant, sie dienen aber, wie im Folgenden gezeigt wird, als Input für eine Vielzahl von Modellen von stochastischen Prozessen.

MA Prozesse

Ein MA oder Moving-Average Prozess erhält man aus einer Linearkombination von White-Noise Variablen:

$$x_t = a_0 w_t + a_1 w_{t-1} + a_2 w_{t-2} + \dots + a_q w_{t-q}.$$

Dies entspricht der Konvolution von weißem Rauschen w_t mit einem Moving-Average Operator (a_n) . Man spricht in diesem Fall von einem Moving-Average Prozess der Ordnung q oder kurz von einem $MA(q)$ Prozess.

AR Prozesse

Ein AR oder autoregressiver Prozess x_t wird durch eine Linearkombination von früheren Werten von x und jeweils einer White-Noise Variable erzeugt:

$$x_t = a_0 w_t - b_1 x_{t-1} - b_2 x_{t-2} - \dots - b_p x_{t-p}.$$

Man spricht in diesem Fall von einem autoregressiven Prozess der Ordnung p oder kurz von einem $AR(p)$ Prozess. (Beachte: In den erwähnten Lehrbüchern sind die Koeffizienten b_n mit anderem Vorzeichen definiert.)

ARMA Prozesse

Die Kombination eines MA und eines AR Prozesses führt zwangslos zu einem sogenannten $ARMA(p, q)$ Prozess:

$$x_t = a_0 w_t + a_1 w_{t-1} + \dots + a_q w_{t-q} - b_1 x_{t-1} - \dots - b_p x_{t-p}.$$

ARIMA Prozesse

In der Praxis trifft man oft Zeitreihen an, die man als stationären ARMA Prozess, x_t , modellieren kann, dem aber noch ein linearer Trend überlagert ist. Eine solche Zeitreihe y_t kann man wie folgt beschreiben:

$$y_t = \beta_0 + \beta_1 t + x_t$$

Da der Mittelwert einer solchen Zeitreihe nicht für alle t gleich ist, ist die Zeitreihe y_t nichtstationär. Mit dem Differenzen-Operator Δ erhalten wir:

$$\Delta y_t = y_t - y_{t-1} = \beta_1 + x_t - x_{t-1} = \beta_1 + \Delta x_t.$$

Die neue Zeitreihe Δy_t ist somit wieder stationär und kann als stationärer ARMA Prozess modelliert werden. Dieses Vorgehen kann natürlich beliebig oft wiederholt werden und erlaubt somit auch Trends höherer Ordnung zu entfernen. Die ursprüngliche Zeitreihe y_t kann daher als eine Realisierung eines integrierten ARMA Prozesses angesehen werden. Verallgemeinert spricht man in diesen Fällen von $ARIMA(p, d, q)$ Prozessen, wobei d die Ordnung der Integration, bzw. des Trends bezeichnet.

Kapitel 7

Anhang

7.1	Die FFT in MATLAB	58
7.2	Das Leistungsspektrum in MATLAB	61
7.3	Kovarianz und Korrelation in MATLAB	64
7.4	ARMA Filter in MATLAB	65

Die FFT in MATLAB

Die FFT einer Zeitreihe x_n , die N Werte lang ist, wird in MATLAB mit dem Befehl `fft(x,M)` ausgeführt. Wenn $M < N$ ist, dann wird die Zeitreihe gekürzt, wenn $M > N$ ist, dann werden entsprechend viele 0 angehängt. Ohne Angabe von M , also mit `fft(x)`, wird die FFT mit N Werten ausgeführt. Weder N noch M müssen einer Zweierpotenz entsprechen, aber für den Rechenaufwand ist es von Vorteil: Im ungünstigsten Fall wird aus der FFT eine langsame DFT. Sei im Folgenden $M = N$, dann ist

$$\text{fft}(x,N) = \sum_{n=0}^{N-1} x_n e^{-j \frac{2\pi kn}{N}}$$

Wird ein Spektrum mit korrekten Einheiten gewünscht, muss das Resultat noch mit der Länge des Abtastintervalls Δt multipliziert werden.

Mit `Delta = Δt` ist dann

$$X = \text{Delta} * \text{fft}(x,N) = \Delta t \sum_{n=0}^{N-1} x_n e^{-j \frac{2\pi kn}{N}} = X_k$$

Der komplexe Vektor X hat N Werte. Zur Darstellung des vollständigen Spektrums von der Frequenz 0 bis zur Nyquist-Frequenz genügen aber die ersten $N/2 + 1$ Werte (man erinnere sich, in MATLAB hat die Frequenz 0 den Index 1). Mit der Abtastrate f_s , sei `Fnyq = $F_{Nyq} = f_s/2 = 1/(2\Delta t)$` die Nyquist Frequenz und `Nf = $N/2 + 1$` die Anzahl Frequenzwerte von 0 bis F_{Nyq} , dann kann man zur Darstellung des Resultates der FFT den Vektor mit den entsprechenden Frequenzwerten wie folgt erstellen:

$$f = \text{linspace}(0, \text{Fnyq}, \text{Nf});$$

oder mit dem Frequenzintervall `df = $\Delta f = 1/(N\Delta t)$` ,

$$f = 0:\text{df}:\text{Fnyq};$$

Das Amplitudenspektrum

Das Amplitudenspektrum kann man dann darstellen zum Beispiel mit

$$\text{plot}(f, \text{abs}(X(1:\text{Nf})))$$

oder für logarithmische Achsen mit

$$\text{loglog}(f, \text{abs}(X(1:\text{Nf})))$$

Das Phasenspektrum

Das Phasenspektrum kann man wie folgt berechnen

```
phase = unwrap(angle(X(1:nf))) ;
```

Die Funktion `angle(X)` berechnet den $\arctan(\Im(X_k)/\Re(X_k))$ und die Funktion `unwrap` beseitigt die durch die Unstetigkeiten der Tangens-Funktion verursachten Sprünge des Arcustangens. Darstellen kann man die Phase schliesslich mit den Befehlen

```
plot(f,phase/pi)
```

oder für eine logarithmische Frequenz-Achse mit

```
semilogx(f,phase/pi)
```

Dabei sorgt die Division durch π dafür, dass die Phase als Vielfaches von π dargestellt wird.

Die inverse FFT

Die inverse FFT wird mit `ifft(X)` ausgeführt, wobei

$$y = \text{ifft}(X) = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{+j \frac{2\pi kn}{N}}$$

Das Resultat `y` ist ein komplexer Vektor und die ursprüngliche reelle Zeitreihe x_n erhält man mit `x = real(y)`, oder mit `x = real(y) / Delta`, falls `X` korrekt mit Δt normiert worden ist. Falls X_k das Resultat von `X = fft(x,M)` und $M \neq N$ war, ist die Länge von x_n als Resultat von `ifft(X)` nun aber gleich M statt N .

Bemerkungen zur Auflösung der FFT

Bei einer gegebenen Abtastrate f_s bzw. einem gegebenen Abtastintervall Δt ist die Auflösung der diskreten Fourier Transformation Δf und somit die Fähigkeit, zwei benachbarte Frequenzen zu unterscheiden, einzig und alleine ein Funktion der zur Verfügung stehenden Datenmenge N , denn $\Delta f = 1/(N\Delta t)$ (Zur Erinnerung, $N\Delta t = T$ ist die Dauer der gegebenen Zeitreihe in Einheiten der Zeit).

Wird eine bestimmte minimale Auflösung Δf_{min} durch die gegebene Problemstellung gefordert, lässt sich die dazu notwendige minimale Länge der Zeitreihe leicht berechnen:

$$N_{min} = f_s / \Delta f_{min}$$

Durch *zero-padding* kann man zwar die optische Auflösung verbessern, indem dies der FFT mehr Stützwerte (Punkte) zur Verfügung stellt, für welche die FFT berechnet wird, aber am grundsätzlichen Informationsgehalt ändert dies nichts.

Ob ein Wert der diskreten Fourier Transformation für eine bestimmte Frequenz f_n im Spektrum tatsächlich berechnet wird, d.h. ob dieser Frequenzwert im Spektrum tatsächlich abgebildet wird, hängt davon ab, ob diese Frequenz ein ganzzahliges Vielfaches von Δf ist: $f_n = n\Delta f$. Aus der Tatsache, dass $f_s = 1/\Delta t$ und $\Delta f = (1/N\Delta t)$ sind, folgt somit die Bedingung

$$\frac{N}{n} = \frac{f_s}{f_n}$$

wobei N und n ganze Zahlen sein müssen. Bei einer vorgegebenen Länge der Zeitreihe kann in diesem Fall die Anzahl N durch *zero-padding* auf eine geeignete Menge vergrößert werden. In MATLAB gilt es dann noch zu berücksichtigen, dass der Index des entsprechenden Spektralwertes $k = n + 1$ ist.

Für im Prinzip unendlich lange Zeitreihen oder im Falle der Anwendung der FFT auf einen Ausschnitt einer gegebenen Zeitreihe ist immer zu berücksichtigen, dass dieses Ausschneiden einer Multiplikation der gegebenen Zeitreihe mit einer Fensterfunktion entspricht, und dass somit das Resultat eine Faltung der eigentlich gesuchten Fourier Transformation mit der FFT der Fensterfunktion ist. Wenn die Fensterfunktion nicht speziell definiert wird, dann hat man tatsächlich ein Rechteck-Fenster verwendet, und das Resultat enthält die Faltung mit der *sinc*-Funktion, deren Nebenmaxima unter Umständen andere im Spektrum vorhandene Frequenzanteile überlagern oder vortäuschen können. Die Breite des Hauptmaximums der *sinc*-Funktion ist $1/(N\Delta t) = \Delta f$, wobei hier N immer die Länge der Zeitreihe ohne *zero-padding* ist. Das Rechteckfenster ist daher dasjenige Fenster, welches die höchste spektrale Auflösung ergibt. Es ist aber auch das Fenster mit den grössten und dichtesten Nebenmaxima im Spektralbereich. Die Nulldurchgänge der *sinc*-Funktion befinden sich ebenfalls bei Frequenzen, die ganzzahlige Vielfache von Δf sind. Somit lassen sich zumindest optisch die Nebenmaxima der *sinc*-Funktion im Spektrum für gegebene Frequenzen durch geeignete Wahl von N manchmal unterdrücken. Bei der FFT über die gesamte Länge einer endlichen Zeitreihe mit klar definiertem Anfang und Ende spielen aber die Fenster-Effekte keine Rolle, und es sollte auch keines verwendet werden.

Das Leistungsspektrum in MATLAB

Das gebräuchlichste Verfahren das Leistungsspektrum einer Zeitreihe zu berechnen, beruht auf der diskreten Fourier Transformation und wird als *Periodogramm* bezeichnet. Da in der Literatur und Praxis sehr oft die verschiedenen Normierungsarten des Leistungsspektrums unterschiedlich definiert sind, ist es wichtig, sich bei der Anwendung von gegebenen Computer-Routinen zu vergewissern, dass das Resultat den Erwartungen entsprechend normiert ist.

Das Leistungsspektrum kann natürlich auch direkt aus der FFT berechnet werden:

```
X = fft(x,N);  
nf = N/2 + 1;  
pp = abs(X(1:nf)) * abs(X(1:nf));  
So ist
```

$$pp = \left| \sum_{n=0}^{N-1} x_n e^{-j \frac{2\pi kn}{N}} \right|^2$$

In dieser Form entspricht aber **pp** weder der Definition der spektralen Leistungsdichte noch der spektralen Leistung.

Für die spektrale Leistungsdichte muss **pp** mit $\Delta t/N$ multipliziert werden (wobei Δt das Abtastintervall ist)

$$pd = (\text{Deltat}/N) * pp = \frac{\Delta t}{N} \left| \sum_{n=0}^{N-1} x_n e^{-j \frac{2\pi kn}{N}} \right|^2$$

und für die spektrale Leistung, die unabhängig von der Länge von x_n ist, muss **pp** mit $1/N^2$ multipliziert werden

$$ps = (1/(N*N)) * pp = \frac{1}{N^2} \left| \sum_{n=0}^{N-1} x_n e^{-j \frac{2\pi kn}{N}} \right|^2$$

In der Regel stellt x_n einen Ausschnitt aus einem im Prinzip unbegrenzten Prozess dar und entspricht einer Multiplikation dieser unbegrenzten Zeitreihe mit einem Rechteckfenster der entsprechenden Länge. Wird eine andere Fensterfunktion **w** als das Rechteckfenster verwendet, spricht man von einem *modifiziertem Periodogramm*, denn diese Fensterfunktion (z.Bsp. ein Dreiecksfenster oder ein Cosinusfenster) hat eine Auswirkung sowohl auf die Auflösung des Leistungsspektrums als auch auf den Betrag. Um die Auswirkung auf den Betrag rückgängig zu machen, muss der Faktor $1/N$ in den obigen Gleichungen für **pd** und **ps** durch den folgenden Normierungsfaktor

ersetzt werden:

$$(\text{norm}(\mathbf{w}))^2 / (\text{sum}(\mathbf{w}))^2 = \sum_{n=1}^N |w_n|^2 / \left(\sum_{n=1}^N w_n\right)^2$$

Dass dieser Faktor auch für das Rechteckfenster gilt, lässt sich leicht zeigen:

Rechteckfenster: $w_n = 1, (n = 1, \dots, N)$

$$\sum_{n=1}^N |w_n|^2 / \left(\sum_{n=1}^N w_n\right)^2 = N/N^2 = 1/N.$$

Da uns keine unendliche Datenmenge zur Verfügung steht, können wir mit den uns zur Verfügung stehenden Methoden bestenfalls einen Schätzwert des tatsächlichen Leistungsspektrums berechnen. Es lässt sich zeigen, dass die Varianz des wie oben berechneten Leistungsspektrums unabhängig von der Länge des Zeitintervalls und in der Grössenordnung des jeweiligen Erwartungswertes des Leistungsspektrums selber ist. Ein längeres Zeitfenster verbessert nur die Auflösung des Spektrums aber nicht die Varianz. Wie bei anderen Messverfahren, lässt sich auch bei der Abschätzung des Leistungsspektrums die Varianz durch Mittelwertbildung über mehrere Beobachtungen reduzieren. Hierzu wird die zu analysierende Zeitreihe in einzelne kürzere Zeitfenster zerlegt und die Leistungsspektren der einzelnen Segmente gemittelt. Wegen der kürzeren Zeitfenster geht dies aber auf Kosten der spektralen Auflösung. Als Kompromiss wird die zu analysierende Zeitreihe häufig in überlappende Datenabschnitte zerlegt. Dabei geht aber die statistische Unabhängigkeit der einzelnen Datenabschnitte verloren, was dem Prinzip der Mittelwertbildung zuwider läuft und daher nicht so sinnvoll ist.

Hat man ein auf diese Art gemittelttes Leistungsspektrum `ps` berechnet kann man, unter der Annahme, dass diese Leistungsspektren einer *chi-quadrat* Verteilung gehorchen, in MATLAB mit der Funktion `chi2inv` relativ einfach die entsprechenden Konfidenzintervalle berechnen (vorausgesetzt man hat Zugang zur STATISTICS TOOLBOX):

```
p = 0.95;  
k = fix(length(x)/N);  
alfa = 1-p;  
c1 = 2*k*ps*(1./chi2inv(1-alfa/2,2*k));  
c2 = 2*k*ps*(1./chi2inv(alfa/2,2*k));
```

Dabei sind `p` die vorgegebene Wahrscheinlichkeit für die zu berechnenden Konfidenzintervallen, `k` die Anzahl Segmente der Länge `N` die aus der Zerlegung der Zeitreihe `x` gebildet wurden, und `ps` das berechnete Leistungsspektrum. Das Resultat sind zwei Vektoren `c1` und `c2`, die die untere und obere Grenzen des Bereiches bezeichnen, innerhalb dessen der wahre Wert von `ps` mit der vorgegebenen Wahrscheinlichkeit `p` zu erwarten ist. Wie im Beispiel

in der folgenden Abbildung gezeigt ist, je mehr Segmente für die Mittelwertbildung zur Verfügung stehen, umso kleiner wird die Varianz der Spektral-Amplituden und umso mehr verbessert sich somit die Vertrauenswürdigkeit des Resultates. Ausführliche Behandlungen der Berechnung der Konfidenzintervalle sind in den Büchern von Buttkus und von Press et al. (*Numerical Recipes*) zu finden.

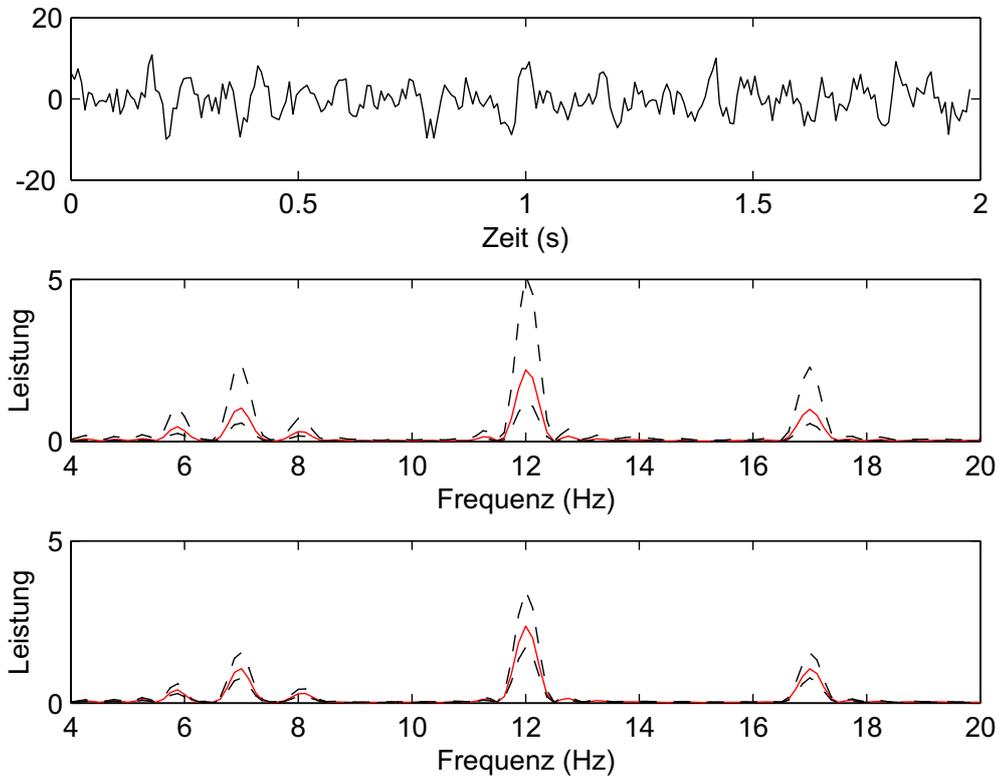


Abbildung 7.1: Oben: 2 Sekunden langer Ausschnitt aus einer synthetischen Zeitreihe. Gemittelte Leistungsspektren aus 8 (mitte) und 32 (unten) je 2 Sekunden langen Datensegmenten (rote ausgezogene Kurven) und die entsprechenden Konfidenzintervalle (schwarze gestrichelte Kurven) für eine Wahrscheinlichkeit von 0.95. Bei der Berechnung der Spektren wurden die einzelnen Zeitfenster mit *zero-padding* auf 8 Sekunden verlängert. Mit zunehmender Anzahl Datensegmenten verkleinert sich die Varianz der Amplituden, aber die Frequenz-Auflösung bleibt wegen der konstanten Fensterlänge gleich.

Kovarianz und Korrelation in MATLAB

Für die Berechnung von Kovarianz und Korrelation in MATLAB stehen die zwei Funktionen `xcorr` und `xcov` zur Verfügung:

$$\text{xcorr}(\mathbf{x}, \mathbf{y}) = \hat{R}_{xy}(k) = \begin{cases} \sum_{n=0}^{N-1-k} x_{n+k} y_n, & (k = 0, \dots, N-1) \\ \hat{R}_{yx}(-k), & (k = -(N-1), \dots, -1) \end{cases}$$

Dies wird auch als *rohe Korrelation* bezeichnet.

$$\text{xcorr}(\mathbf{x}, \mathbf{y}, \text{'biased'}) = R_{xy}(k) = \frac{1}{N} \hat{R}_{xy}(k)$$

$$\text{xcorr}(\mathbf{x}, \mathbf{y}, \text{'unbiased'}) = \frac{1}{N - |k|} \hat{R}_{xy}(k)$$

$$\text{xcorr}(\mathbf{x}, \mathbf{y}, \text{'coeff'}) = C_{xy}(k) = \frac{R_{xy}(k)}{\sqrt{\text{Var}(x)\text{Var}(y)}}$$

Das heißt, die ersten drei Befehle liefern die Kreuzkovarianz und der vierte (`xcorr(x,y,'coeff')`) die Kreuzkorrelation, wobei letztere mit $1/N$ normiert und somit *biased* ist. Um eine Korrelation zu erhalten, die *unbiased* ist, muss man

$$\text{xcorr}(\mathbf{x}, \mathbf{y}, \text{'unbiased'}) / (\text{std}(\mathbf{x}) * \text{std}(\mathbf{y}))$$

eingeben.

Die Funktion `xcov` unterscheidet sich von `xcorr` dadurch, dass vor der Korrelation die Mittelwerte von x und y abgezogen werden. Bezüglich Normierung verhalten sich die beiden Funktionen gleich.

Die Autokovarianz erhält man mit `xcorr(x,x)` oder `xcorr(x)` bzw. `xcov(x,x)` oder `xcov(x)` und den oben angegebenen Normierungen.

Bei einer Länge N der Zeitreihen (Vektoren) x und y , ist die Länge des Resultates gleich $2N - 1$ und der N -te Wert entspricht $lag = 0$. Ausserdem kann der maximale *lag* mit einem zusätzlichen Parameter `maxlag` festgelegt werden, z. Bsp. `xcov(x,y,maxlag,'unbiased')`.

Zu beachten: Wenn x und y korreliert sind und das Maximum von $\hat{R}_{xy}(k)$ bei $k > 0$ ist, dann ist x gegenüber y verzögert; ist hingegen y gegenüber x verzögert, dann ist das Maximum von $\hat{R}_{xy}(k)$ bei $k < 0$.

ARMA Filter in MATLAB

Die Realisierung x_t eines $ARMA(p, q)$ Prozesses als Linearkombination von White Noise und früheren Werten von x entspricht der allgemeinsten Form einer linearen Filteroperation. Die a_n sind die Koeffizienten des Moving-Average Filters und die b_n sind die Koeffizienten des autoregressiven Filters. Letzterer wird auch als rekursives Filter bezeichnet. Dabei wird die Zeitreihe w_t (Input) durch die Gesamtwirkung aller Koeffizienten in eine gefilterte Zeitreihe x_t (Output) verwandelt.

In MATLAB steht dafür die Funktion `filter` zur Verfügung:

$$\mathbf{x} = \text{filter}(\mathbf{ma}, \mathbf{ar}, \mathbf{w}).$$

Der Vektor `ma` entspricht den MA Koeffizienten a_n und der Vektor `ar` den AR Koeffizienten b_n . Dabei gilt es aber zu berücksichtigen, dass MATLAB keinen Index 0 kennt. Somit sind

$$\mathbf{ma}(1) = a_0 \text{ und } \mathbf{ma}(\mathbf{q}+1) = a_q$$

$$\mathbf{ar}(1) = b_0 \text{ und } \mathbf{ar}(\mathbf{p}+1) = b_p$$

Unserer Notation entsprechend muss für alle Filter $\mathbf{ar}(1) = b_0 = 1$ sein. Ausserdem gilt für ein reines AR Filter $\mathbf{ma} = [a_0]$ und für ein reines MA Filter $\mathbf{ar} = [1]$.