Fachbereich Informatik und Mathematik
ISMI - Institut für Stochastik
& Mathematische Informatik

Mathematik der Zeitreihen

WS 2012/13

H. Dinges

4. Februar 2013

Inhaltsverzeichnis

1	Konstruktionen strikt stationärer Prozesse	3
1.1	Der Ornstein-Uhlenbeck-Prozess	3
1.2	Der vom Zufall getriebene harmonische Oszillator	6
1.3	Komplexe gaußsche Prozesse	8
2	Analysis der Kovarianzfolgen und Spektralmaße	12
2.1	Der Satz von Herglotz	13
2.2	Lineare Filter und ihre Transfer-Funktionen	17
2.3	Spezielle Transferfunktionen	21
2.4	Kausale Filter, ARMA-Konstruktionen.	27
3	Stationäre Zeitreihen im L^2-Sinn	33
3.1	Finite lineare Prädiktion	33
3.2	Unendliche Vergangenheit, reguläre Zeitreihen	39
3.3	Wiener-Hopf-Faktorisierungen	48
3.4	Filtern gegen Rauschen	55
3.5	Maße mit orthogonalen Werten	59

Stationäre stochastische Prozesse

Die Zeitreihenanalyse wird meistens als eine angewandte Disziplin gelehrt und geübt. Dabei bleibt oft unklar, wie die inhaltlichen Gegebenheiten zu denjenigen Voraussetzungen passen, welche präzise mathematische Schlüsse erlauben. Der Glaube, dass man die mathematische Sätze aus der Theorie der stationären stochastischen Prozesse auf empirisch gegebene Zeitreihen anwenden kann, gründet sich anscheinend vor allem darauf, dass die Praktiker Analogien zu mathematisch gut verstandenen stochastischen Prozessen vermuten, (die man auch in Simulationen erfahrbar machen kann). Mit Simulationen und Datenanalyse wollen wir uns hier nicht befassen. Wir sehen uns auch leider nicht in der Lage, auf die eindrucksvollen Anwendungen in der Signalbearbeitung einzugehen. Eben- sowenig können wir die statistischen Probleme der Zeitreihenanalyse näher erörtern. Es soll uns also nicht um Rechentechniken gehen, sondern vielmehr um die in der Tat recht breiten mathematischen Zusammenhänge, in welchen die stationären Prozesse stehen.. Der mathematische Dreh- und Angelpunkt sind die stationären stochastischen Prozesse. Dabei sind für den Wahrscheinlichkeitstheoretiker sowohl die strikt stationären Prozesse interessant, als auch die 'lineare Theorie' der schwach stationären Prozesse.

Definition (Strikte Stationarität). Ein S -wertiger strikt stationärer Prozess über dem Zeitbereich T ist eine Schar von S -wertigen Zufallsgrößen $\{X_t : t \in T\}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \{P_\theta : \theta \in \Theta\})$, sodass für alle 'Hypothesen' $\theta \in \Theta$ gilt

$$\mathfrak{L}_\theta(X_{t_1}, X_{t_2}, \dots, X_{t_n}) = \mathfrak{L}_\theta(X_{t_1+s}, X_{t_2+s}, \dots, X_{t_n+s})$$

für alle Tupel $(t_1, t_2, \dots, t_n)$ und alle $s \in T$.

Der Zeitbereich T ist (bei uns hier) eine der Mengen $\mathbb{R}, \mathbb{R}_+, h \cdot \mathbb{Z}, h \cdot \mathbb{Z}_+$ mit einer 'Schrittweite' $h > 0$. Der 'Zustandsraum' S ist üblicherweise ein polnischer Raum es existiert also eine Metrik, welche S zu einem vollständigen metrischen Raum mit abzählbarer Basis macht. Häufig betrachten wir die Schar der Zufallsgrößen unter einer festen Hypothese; Θ ist dann also einpunktig. Häufig ist $(t, \omega) \mapsto X_t(\omega)$ messbar. Für jedes ω haben wir dann einen messbaren 'Pfad' $X(\omega)$.

Definition (Schwache Stationarität). Ein reell- d -dimensionaler schwach stationärer Prozess ist eine Schar $\mathbb{R}^d$ -wertiger quadratintegrierbarer Zufallsgrößen $\{X_t : t \in T\}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \{P_\theta : \theta \in \Theta\})$, sodass unter jeder der Hypothesen θ der Erwartungswert $\mathfrak{E}(X_t)$ zeitlich konstant ist, und die Covarianzmatrix $\text{cov}(X_{t+h}, X_t)$ eine stetige Funktion, die nur vom zeitlichen Abstand h abhängt.

Bemerkung Wenn $\{X_t : t \in T\}$ ein strikt stationärer Prozess ist und $f^1, f^2, \dots, f^d$ ein d -Tupel quadratintegrierbarer Funktionen auf dem Zustandsraum, dann ist die Familie $(Y_t)_{t \in T}$ der d -Tupel $(f^1(X_t), \dots, f^d(X_t))_{t \in T}$ ein stark stationärer vektorwertiger Prozess, der obendrein schwachstationär ist.

Definition (Gaussische stationäre Prozesse). Ein d -dimensionaler schwachstationärer Prozess, in welchem alle Tupel $(X_{t_1}^{j_1}, X_{t_2}^{j_2}, \dots, X_{t_n}^{j_n})$, $(j_k \in 1, 2, \dots, d, t_k \in T)$ gemeinsam gaussisch verteilt sind, heist ein stationärer gaussischer Prozess.

(Es zeigt sich, dass ein stationärer gaussischer Prozess auch strikt stationär ist.)

Diese Begriffe wollen wir nun durch einige konkrete Beispiele beleuchten, die in vielen Zusammenhängen von Interesse sind. Es geht noch nicht darum, komplizierte Zusammenhänge streng mathematisch abzuleiten.

1 Konstruktionen strikt stationärer Prozesse

1.1 Der Ornstein-Uhlenbeck-Prozess

Es gibt viele Wege, den skalaren Ornstein-Uhlenbeck-Prozess $\{U_t : t \in \mathbb{R}\}$ zu konstruieren. Der folgende ist besonders einfach, weil er sich direkt auf den elementaren Prozess der Brownschen Bewegung $\{W_t : t \in \mathbb{R}_+\}$ bezieht. Wir betrachten für ein $\alpha > 0$

$$U_t = e^{-\alpha t} \cdot W(e^{2\alpha t})$$

Diese U_t sind allesamt gemeinsam normalverteilt mit $\mathfrak{E}(U_t) = 0$ und

$$\text{cov}(U_t, U_{t+s}) = e^{-\alpha t} \cdot e^{-\alpha(t+s)} \cdot \text{cov}(W(e^{2\alpha t}), W(e^{2\alpha(t+s)})) = e^{-2\alpha s}.$$

Der Prozess $(W_t)_t$ hat unabhängige Zuwächse. Für alle $s > 0$ ist $W(e^{2\alpha(t+s)}) - W(e^{2\alpha t})$ unabhängig von den U_r mit $r \leq t$. Wegen

$$\Delta := U_{t+s} - e^{-\alpha s} U_t = e^{-\alpha(t+s)} \cdot (W(e^{2\alpha(t+s)}) - W(e^{2\alpha t}))$$

haben wir

$$U_{t+s} = e^{-\alpha s} U_t + \Delta([t, t+s])$$

mit einer normalverteilten Zufallsgrößen $\Delta([t, t+s])$, die unabhängig ist von der Vorgeschichte $\mathfrak{A}_t^{-\infty} = \{U_r : r \leq t\}^\sigma$. $\mathfrak{E}(\Delta) = 0$, $\text{var}(\Delta) = 1 - e^{-2\alpha s}$.

Der Ornstein-Uhlenbeck-Prozess ist also ein Markov-Prozess bezgl. der von den U_t erzeugten Filtrierung.

Wir wollen in intuitiver Weise ohne mathematische Strenge weitere Aspekte des OU-Prozesses herausstellen. Die genaue Bedeutung der Konstruktionen erfährt man in der stochastischen Analysis. Man lernt dort den Umgang mit infinitesimalen Zuwächsen und mit stochastischen Integralen. Für $s = dt$ haben wir mit einer Standard-Brown'schen Bewegung (B_t)

$$U_{t+dt} - U_t = (e^{-\alpha dt} - 1) \cdot U_t + \sqrt{\frac{1}{dt}(1 - e^{-2\alpha dt})} \cdot (B_{t+dt} - B_t).$$

Dies schreibt man als stochastische Differentialgleichung

$$dU_t = -\alpha U_t \cdot dt + \sqrt{2\alpha} \cdot dB_t.$$

Man versteht den Prozess als eine Diffusion mit einer Drift hin zum Nullpunkt, die linear anwächst mit dem Abstand vom Nullpunkt. Die stochastische Störung hat in allen Punkten dieselbe Größe; der Diffusionskoeffizient ist ortsunabhängig. Die stationäre Verteilung ist die Standardnormalverteilung.

Wir betrachten den Prozess eingeschränkt auf den diskreten Zeitbereich $T_h = h \cdot \mathbb{Z}$. Mit der Abkürzung $\beta = \beta_h = e^{-\alpha h}$ haben wir für $t \in T_h$

$$\begin{aligned} U_t &= \beta \cdot U_{t-\delta} + \Delta([t - \delta, t]) \\ &= \beta(\beta \cdot U_{t-2\delta} + \Delta([t - 2\delta, t - \delta])) + \Delta([t - \delta, t]) \\ &= \dots \\ &= \beta^N \cdot U_{t-N\delta} + \beta^{N-1} \Delta([t - N\delta, t - (N-1)\delta]) + \dots + \Delta([t - \delta, t]) \\ &= \sum_{n=0}^{\infty} \beta^n \cdot \Delta([t - (n+1)\delta, t - n\delta]). \end{aligned}$$

Die Summanden sind unabhängig normalverteilt mit den Varianzen $\beta^{2n}(1 - e^{-2\alpha h}) = e^{-2\alpha hn}(1 - e^{-2\alpha h})$, die sich zum Wert 1 summieren. Wenn wir das Gitter verfeinern, $h \rightarrow 0$, dann wir aus der Summe das Integral

$$U_t = \int_0^{\infty} e^{-\alpha s} dB(t-s) = \int_{-\infty}^t e^{-\alpha(t-u)} dB(u).$$

Hinweis: Wenn die Variablen in einem stochastischen Prozess $(V_t) - t$ eine Darstellung der folgenden Form besitzen

$$V_t = \int_{-\infty}^t a(t-u) dY(u)$$

wo $(Y_u)_u$ ein Prozess mit unkorrelierten Zuwächsen ist, dann nennt man die Zuwächse des Y -Prozesses die Innovationen und die Funktion $a(\cdot)$ die Influenzfunktion. In dieser Sprache erscheint der OU-Prozess als stationärer Prozess mit gaussischen Innovationen und einer exponentiell abfallenden Influenzfunktion.

Ein weiterer Hinweis: In der Theorie der schwach stationären Prozesse $(X_t)_t$ spielt die Kovarianzfunktion $\gamma(\cdot)$ eine entscheidende Rolle, $\gamma(h) = \text{cov}(X_{t+h}, X_t)$. Für gaussische stationäre Prozesse ist die gemeinsame Verteilung der X_t allein durch den Erwartungswert und die Kovarianzfunktion bestimmt.

Die Kovarianzfunktion für den OU-Prozess ist die nach beiden Seiten exponentiell abfallende Funktion $\gamma(h) = e^{-\alpha|h|}$. Die Konstruktion eines stationären gaussischen Prozesses zu einer Kovarianzfunktion, die wir später kennenlernen werden, gibt (auch bei sehr schönen Kovarianzfunktionen) zunächst einmal keinen Anhaltspunkt, dass eine Modifikation zu einem Prozess mit stetigen Pfaden existiert. Bei unserer Konstruktion ergibt sich Stetigkeit der Pfade des OU-Prozesses aus der wohlbekanntem Stetigkeit der Brown'schen Pfade.

Ein diskretes Analogon zum OU-Prozess ergibt sich aus dem bekannten Modell von Paul und Tatjana Ehrenfest: N Teilchen werden unabhängig auf zwei Kammern verteilt, jedes Teilchen mit Wahrscheinlichkeit $\frac{1}{2}$ in die erste Kammer. Dort landet also zur Zeit 0 eine binomialverteilte Anzahl N_0 ($\mathbb{E}N_0 = \frac{1}{2}N$, $\text{var}N_0 = \frac{1}{4}N$). In jedem Zeitschritt

wechselt ein zufällig ausgewähltes Teilchen die Kammer. Die Anzahl der Teilchen in der ersten Kammer bildet einen stationären stochastischen Prozess $N_0, N_1, N_2, \dots$. Es handelt sich um einen reversiblen Markoff-Prozess. Für die Zuwächse $N_{n+1} - N_n$ gilt

$$\mathfrak{E}(N_{n+1} - N_n | \mathfrak{A}_n) = 1 - 2\frac{N_n}{N} = -2\left(\frac{N_n}{N} - \frac{1}{2}\right), \quad \mathfrak{E}((N_{n+1} - N_n)^2 | \mathfrak{A}_n) = 1.$$

Der Prozess schwankt um die Nulllage bei $\frac{1}{2}N$. Für große Anzahlen N ist $\sqrt{4N}\left(\frac{N_n}{N} - \frac{1}{2}\right)$ nahezu standardnormalverteilt. Der Erwartungswert der 'rücktreibenden Kraft' steigt linear an mit der Auslenkung aus der Nulllage. Andererseits ist die bedingte Varianz des Zuwachses nahezu konstant in der Nähe der Nulllage. Das erklärt die Ähnlichkeit des Prozesses $\sqrt{4N}\left(\frac{N_n}{N} - \frac{1}{2}\right)$ mit dem OU-Prozess.

1.2 Der vom Zufall getriebene harmonische Oszillator

Es geht uns in diesem Unterabschnitt nicht um mathematisch solide Konstruktionen. Wir erinnern an Bekanntes aus der Schwingungslehre und deuten an, wie die bekannten Fakten in die Welt der stationären Prozesse passen.

In einem elementaren elektrischen Schwingkreis sei R der Ohm'sche Widerstand, L die Induktivität und C die Kapazität. Wenn wir die zeitlich variable Spannung ('elektromotorische Kraft') $U(t)$ anlegen, dann genügt die Ladung auf dem Kondensator $Q(t)$ der folgenden inhomogenen linearen Differentialgleichung zweiter Ordnung:

$$L \cdot \ddot{Q}(t) + R \cdot \dot{Q}(t) + \frac{1}{C}Q(t) = U(t).$$

Mit der Notation $\gamma = \frac{R}{L}$, $\omega_0^2 = \frac{1}{LC}$, $b(\cdot) = \frac{1}{C}U(\cdot)$ wird daraus die Gleichung

$$\ddot{x} + \gamma \cdot \dot{x} + \omega_0^2 \cdot x = b.$$

Wir wollen uns auf den sog. schwach gedämpften Fall $\omega_0 > \frac{1}{2}\gamma$ konzentrieren und definieren $\tilde{\omega} = \sqrt{\omega_0^2 - \frac{1}{4}\gamma^2}$. Die reellwertigen Lösungen der homogenen Gleichung sind dann bekanntlich die Linearkombinationen der folgenden Funktionen

$$x_g(t) = \exp(-\frac{1}{2}\gamma \cdot t) \cdot \cos(\tilde{\omega}t), \quad x_u(t) = \exp(-\frac{1}{2}\gamma \cdot t) \cdot \sin(\tilde{\omega}t).$$

Eine beliebige Präsentationsform einer Lösung der homogenen Gleichung ist übrigens auch

$$x(t) = \exp(-\frac{1}{2}\gamma \cdot t) \cdot |A| \cdot \cos(\tilde{\omega}(t - t_0)).$$

Auf die Präsentation mit Hilfe der komplexen Amplitude $A = |A| \cdot \exp(-i\tilde{\omega}t_0)$ kommen wir später zu sprechen. Die einzige Lösung, die für $t \rightarrow -\infty$ beschränkt bleibt, ist die Nulllösung.

Wenn $b(\cdot)$ auf jedem Intervall $(-\infty, T]$ beschränkt ist, dann besitzt die inhomogene Gleichung genau eine auf jedem Intervall $(-\infty, T]$ beschränkte Lösung. Man nennt sie die Gleichgewichtslösung, und man gewinnt eine Darstellung durch Integrale nach der Methode der Variation der Konstanten.

Es wird übersichtlicher, wenn man von der skalaren Gleichung zweiter Ordnung zur Gleichung erster Ordnung für die Spalte $\begin{pmatrix} x^1(\cdot) \\ x^2(\cdot) \end{pmatrix} = \begin{pmatrix} x(\cdot) \\ \dot{x}(\cdot) \end{pmatrix}$ übergeht.

$$\frac{d}{dt} \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\omega_0^2 & -\gamma \end{pmatrix} \cdot \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} + \begin{pmatrix} 0 \\ b \end{pmatrix} = A \cdot \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} + \begin{pmatrix} 0 \\ b \end{pmatrix}.$$

Die Gleichgewichtslösung hat die Form

$$\begin{pmatrix} x^1(t) \\ x^2(t) \end{pmatrix} = \exp(t \cdot A) \cdot \begin{pmatrix} c^1(t) \\ c^2(t) \end{pmatrix} = \int_{-\infty}^t \exp((t-s)A) \cdot \begin{pmatrix} 0 \\ b(s) \end{pmatrix} ds.$$

Wir bemerken, dass die Matrix die Eigenwerte $-\frac{1}{2}\gamma \pm i\tilde{\omega}$ hat; sie kann diagonalisiert werden. Im aktuellen Fall ergibt sich

$$x^1(t) = \int_{-\infty}^t \exp(-\frac{1}{2}\gamma(t-s)) \cdot \frac{1}{\tilde{\omega}} \sin((t-s)\tilde{\omega}) \cdot b(s) ds.$$

Man kann nun auch allgemeinere Integratoren dB_s statt $b(s) ds$ zulassen. Die 'Influenz-funktion' ist in jedem Fall die Funktion

$$a(h) = \exp(-\frac{1}{2}\gamma h) \cdot \frac{1}{\tilde{\omega}} \sin(h\tilde{\omega}) = \frac{1}{2i\tilde{\omega}} \left(e^{(-\frac{1}{2}\gamma + i\tilde{\omega})h} - e^{(-\frac{1}{2}\gamma - i\tilde{\omega})h} \right).$$

Besonders interessant ist das stochastische Integral $\int_{-\infty}^t \exp((t-s)A) \cdot d\begin{pmatrix} 0 \\ B(s) \end{pmatrix}$, wo $\{B(s) : s \in (-\infty, \infty)\}$ eine Brownsche Bewegung ist. Der resultierende Prozess $\begin{pmatrix} X^1(t) \\ X^2(t) \end{pmatrix}$ ist dann ein stationärer gaussianer Prozess.

Das stochastische Integral bzgl. eines Poisson-Prozesses ist bekannt unter dem Namen Yule's Pendel. Man denkt an einen mechanischen harmonischen Oszillator ('Pendel'), der zu zufälligen Zeitpunkten beschossen wird, sodass auf kurze Zeitspannen konzentrierte Impulsübertragungen stattfinden. Bei der 'Covarianz-Funktion' $C(h)$ kommt es nur auf die Intensität des anregenden 'weissen Rauschens' an. Wenn man den Oszillator mit allgemeineren stationären Prozessen anregt, dann gestaltet sich die Berechnung von $C(h)$ natürlich etwas schwieriger; wir verschieben das auf später. Man versteht den Oszillator als einen zeitinvarianten linearen Filter. Man sagt: Der stationäre Prozess der Eingabe wird durch den Filter in den stationären Prozess der Ausgabe transformiert.

Man kann hier natürlich auch den zweidimensionalen Prozess $\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix}$ betrachten. (Wenn der anregende Prozess der Poisson-Prozess ist, dann hat der resultierende Prozess Sprünge.) An die Stelle der Covarianzfunktion tritt hier die matrixwertige Funktion

$$\Gamma(h) = \text{cov} \left(\begin{pmatrix} X_1(t+h) \\ X_2(t+h) \end{pmatrix}, \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} \right)$$

Die Funktion ausserhalb der Diagonale nennt man die Kreuz-Covarianzfunktion ('cross-covariance') der Prozesse $X_1(t)$ und $X_2(t)$. Die Funktionen in der Diagonale heissen auch die Autocovarianzfunktionen der skalaren Prozesse $X_1(t)$ bzw. $X_2(t)$, zur Unterscheidung von den Kreuz-Covarianzen.

1.3 Komplexgaussische Prozesse

Ein n -Tupel quadratintegrabler komplexwertiger Zufallsgrößen bezeichnet man auch als einen n -dimensionalen komplexen Zufallsvektor (oder als Zufallsvektor der Länge n). Man notiert den Zufallsvektor Z als Spalte und betrachtet die paarweisen Kovarianzen als die Einträge der Kovarianzmatrix $\text{cov}(Z, Z)$

$$c_{kl} = \text{cov}(Y^k, Y^l) = \mathfrak{E}(Z^k \cdot \bar{Z}^l) - \mathfrak{E}Z^k \mathfrak{E}\bar{Z}^l.$$

$$\text{cov}(Z, Z) = \mathfrak{E}((Z - \mathfrak{E}Z) \cdot (Z - \mathfrak{E}Z)^*)$$

Wenn $\mathfrak{E}Z = 0$, dann spricht man von einem zentrierten Zufallsvektor. Sind U und V quadratintegrale Zufallsvektoren, U von der Dimension m und V von der Dimension n , so ist ihre Kovarianz die $m \times n$ -Matrix

$$\text{cov}(U, V) = \mathfrak{E}(U \cdot V^*) - \mathfrak{E}U \mathfrak{E}V^*.$$

Es gilt offenbar $\text{cov}(V, U) = \text{cov}(U, V)^*$. Ist A eine $m' \times m$ -Matrix und B eine $n' \times n$ -Matrix, so ist $A \cdot U$ ein Zufallsvektor der Länge m' , $B \cdot V$ ein Zufallsvektor der Länge n' , und es gilt

$$\text{cov}(A \cdot U, B \cdot V) = A \cdot \text{cov}(U, V) \cdot B^*.$$

Die Bildung der Kovarianzmatrizen ist biadditiv, d. h.

$$\text{cov}(U_1 + U_2, V) = \text{cov}(U_1, V) + \text{cov}(U_2, V); \quad \text{cov}(U, V_1 + V_2) = \text{cov}(U, V_1) + \text{cov}(U, V_2).$$

Für zusammengefügte Spalten gilt

$$\text{cov} \left(\begin{pmatrix} U^{\parallel} \\ U^{\perp} \end{pmatrix}, \begin{pmatrix} V^{\parallel} \\ V^{\perp} \end{pmatrix} \right) = \begin{pmatrix} \text{cov}(U^{\parallel}, V^{\parallel}) & \text{cov}(U^{\parallel}, V^{\perp}) \\ \text{cov}(U^{\perp}, V^{\parallel}) & \text{cov}(U^{\perp}, V^{\perp}) \end{pmatrix}.$$

Die Sprache der Hilberträume:

Die Gesamtheit aller zentrierten quadratintegrablen Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ wird zu einem Hilbertraum $(\mathcal{H}, \|\cdot\|)$, wenn man die Quadratwurzel der Varianz zur Norm erklärt. Die dazugehörige Sesquilinearform, das 'innere Produkt' ist die Kovarianz. (Man beachte allerdings, dass die Kovarianz im ersten Argument linear ist und im zweiten antilinear, entgegen der Konvention in der mathematischen Physik.) Unkorrelierte Zufallsgrößen heißen in der Sprache der Hilberträume zueinander orthogonale Vektoren. Die Kovarianzmatrix eines n -Tupels heisst in der Sprache der Hilberträume die Gram-Matrix dieses n -Tupels. Jede Gram-Matrix ist positiv semidefinit. Sie ist genau dann positiv definit, wenn das n -Tupel linear unabhängig ist. Man sagt von einem Zufallsvektor Y , dass er den Rang r hat, wenn die zentrierten Komponenten $Y^k - \mathfrak{E}Y^k$ einen r -dimensionalen Vektorraum aufspannen; dies ist genau dann der Fall, wenn die Kovarianzmatrix den Rang r hat. Eine fundamentale Eigenschaft der Hilberträume ist bekanntlich die, dass es zu jedem abgeschlossenen Teilraum $\mathcal{M}$ eine wohlbestimmte orthogonale Projektion auf $\mathcal{M}$ gibt; wir werden diese meistens mit $\pi_{\mathcal{M}}$ bezeichnen.

Die Stochastiker denken bei den quadratintegralen Zufallsgrößen zunächst einmal an reellwertige Zufallsgrößen. Für die Betrachtungen im Hilbertraum sind aber die komplexen Zufallsvektoren sehr praktisch und nützlich. Vor allem für die Betrachtung der gaussischen Vektoren müssen wir uns etwas näher mit Fragen der Komplexifikation und Dekomplexifikation beschäftigen.

Reelle und komplexe gaussische Vektoren

Es ist wohlbekannt, was eine normalverteilte (oder gaussische) reelle Zufallsgröße ist. Ein reeller Zufallsvektor heisst gaussisch verteilt, wenn alle reellen Linearkombinationen der Komponenten normalverteilt sind; man sagt in diesem Fall auch, dass die Komponenten gemeinsam gaussisch verteilt sind.— Es gibt noch weitere wichtige Charakterisierungen der reellen Gauss-Vektoren; wir werden sie bei passender Gelegenheit formulieren.

Definition. Eine komplexwertige Zufallsgröße $Z = X + iY$ heisst komplex gaussisch verteilt, wenn der Realteil $X = \Re Z$ und der Imaginärteil $Y = \Im Z$ unabhängig normalverteilt sind mit derselben Varianz.

Ein komplexer Zufallsvektor der Länge n heisst komplexgaussisch verteilt, wenn das $2n$ -Tupel der Real- und Imaginärteile gemeinsam gaussisch verteilt und jede komplexe Linearkombination der Komponenten komplexgaussisch verteilt ist. Man sagt in diesem Falle auch, dass die Komponenten gemeinsam komplexnormalverteilt sind.

Satz.

a) Wenn $Z^1, \dots, Z^n$ unabhängige komplex gaussische Zufallsgrößen sind, dann ist der Vektor mit diesen Komponenten ein komplex gaussischer Vektor.

b) Ist Z ein komplex gaussisch verteilter Zufallsvektor der Dimension m und A eine komplexe $m' \times m$ -Matrix, dann ist $A \cdot Z$ komplex gaussisch verteilt.

Der Satz ist selbstverständlich. Für alternative Kennzeichnungen der komplex gaussischen Vektoren holen wir weiter aus. Wir definieren zu jedem komplexen Zufallsvektor $Z = X + iY$ den konjugiert komplexen Vektor $\bar{Z} = X - iY$ und bemerken

Lemma.

Für einen komplexen Zufallsvektor $Z = X + iY$ mit der Covarianzmatrix $\text{cov}(Z, Z) = S + iT$ gilt $\text{cov}(\bar{Z}, \bar{Z}) = S - iT$ und

$$S = \text{cov}(X, X) + \text{cov}(Y, Y), \quad T = \text{cov}(Y, X) - \text{cov}(X, Y),$$

$$\text{cov}(Z, \bar{Z}) = (\text{cov}(X, X) - \text{cov}(Y, Y)) + i(\text{cov}(Y, X) + \text{cov}(X, Y))$$

Daraus ergibt sich (siehe Übungen)

Satz. Es sei $Z = X + iY$ ein Zufallsvektor mit der Covarianzmatrix $C = S + iT$, dessen Real- und Imaginärteile gemeinsam gaussisch verteilt sind. Wenn $\text{cov}(Z, \bar{Z}) = 0$, dann sind alle komplexen Linearkombinationen der Komponenten $\sum c_k \cdot Z_k$ komplex gaussisch verteilt. Für die Real- und Imaginärteile gilt

$$\text{cov} \left(\begin{pmatrix} X \\ Y \end{pmatrix}, \begin{pmatrix} X \\ Y \end{pmatrix} \right) = \frac{1}{2} \begin{pmatrix} S & -T \\ T & S \end{pmatrix}.$$

Aus der Theorie der Normalverteilungen im Reellen ergibt sich somit: Zu jeder positiv semidefiniten Matrix $C = S + iT$ existiert ein zentrierter komplexgaussisch verteilter Zufallsvektor mit dieser Kovarianz; und die Verteilung ist eindeutig bestimmt. Das Resultat kann man auch folgendermaßen ausdrücken. Wenn $C = S + iT$ eine positiv semidefinite $n \times n$ -Matrix ist, dann existiert genau ein Wahrscheinlichkeitsmaß auf dem Produktraum $\mathbb{C}^n$, sodass die Koordinaten gemeinsam komplexnormalverteilt sind mit der gegebenen Kovarianz. Das $2n$ -Tupel der Real- und Imaginärteile ist reellnormalverteilt mit der Kovarianzmatrix $\frac{1}{2} \begin{pmatrix} S & -T \\ T & S \end{pmatrix}$. Wenn C positiv definit ist, $C^{-1} = Q + iR$, dann besitzt die Verteilung eine Dichte im $\mathbb{R}^{2n}$, und zwar

$$p(x, y) dx dy = \text{const} \exp \left(- (x^T, y^T) \begin{pmatrix} Q & -R \\ R & Q \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \right) dx dy.$$

Komplexgaussische Zeitreihen

Wir wollen jetzt ein Maß auf dem unendlichen Produktraum $\Omega = \mathbb{C}^{\mathbb{Z}}$ zu einer $\mathbb{Z} \times \mathbb{Z}$ -Matrix $C = \{c_{kl} : k, l \in \mathbb{Z}\}$ konstruieren, sodass jedes endliche Tupel von Koordinatenfunktionen $(Z^m)_{m \in M}$ zentriert und gemeinsam komplexnormalverteilt ist mit den Kovarianzen in der entsprechenden Teilmatrix. Von C ist zu fordern, dass jede endliche quadratische Teilmatrix $\{c_{kl} : k, l \in M\}$ positiv semidefinit ist. Wenn dies der Fall ist, dann existiert für jedes endliche M genau ein Maß auf der von den Koordinaten in M erzeugten Produkt-Borelalgebra $\mathfrak{B}_M = \bigotimes_{m \in M} \mathfrak{B}_m$. Diese Maße auf den σ -Algebren $\mathfrak{B}_M$ sind verträglich. Nach einem berühmten Satz von Kolmogorov besitzen sie eine gemeinsame Fortsetzung auf die Produktalgebra des unendlichen Produktraums $\Omega = \mathbb{C}^{\mathbb{Z}}$. Damit sind wir am Ziel: Zu jeder positiv semidefiniten $\mathbb{Z} \times \mathbb{Z}$ -Matrix existiert ein zentrierte komplexgaussische Folge $(Z_n)_n$ mit den vorgegebenen Kovarianzen. Die gemeinsame Verteilung der Z_n ist durch C eindeutig bestimmt.

Hinweis: Der Satz von Kolmogorov über verträgliche Maße, die sich zu einem Maß auf dem unendlichen Produktraum fortsetzen lassen, gilt auch für Indexmengen T , die nicht abzählbar sind, z. B. für ein Zeitintervall $T = [t^l, t^h]$. Man findet zu jeder Kovarianz-Struktur $\{c(s, t) : s, t \in T\}$ einen gaussischen Prozess $(Z_t)_{t \in T}$. Man könnte so z. B. versuchen, den OU-Prozess aus der Kovarianzstruktur $\{c(s, t) = \exp |t - s| : s, t \in T\}$ zu konstruieren. Die Konstruktion nach dem Satz von Kolmogorov leistet aber nicht dasselbe wie unsere direkte Konstruktion aus der Brown'schen Bewegung. Die Produkt σ -Algebra auf dem Produktraum $\Omega = \mathbb{R}^{\mathbb{R}}$ ist nicht groß. Eine Menge wie beispielsweise $\{\omega : x_t \leq h \text{ für alle } t \in [t^l, t^h]\}$ ist nicht produktmessbar. Die Menge der stetigen 'Pfade' ist ebensowenig produktmessbar. Wenn man Prozesse mit guten Pfaden konstruieren will muss man anders vorgehen.

Beispiel 1.3.1 (Fastperiodische Prozesse).

Es seien $Z^1, \dots, Z^N$ unabhängige komplexgaussische Variable mit $\text{var} Z = 1$, $\sigma_1, \dots, \sigma_N$ Zahlen mit $\sum |\sigma_k|^2 = 1$ und $\omega_1, \dots, \omega_N \in \mathbb{R}$ paarweise verschieden. Dazu konstruieren wir den Prozess

$$X^t = \sum_{k=1}^N \sigma_k \cdot Z^k e^{it\omega_k}.$$

Es handelt sich um einen gaussischen Prozess, denn die X^t sind gemeinsam gaussisch verteilt. Die Covarianzfunktion ist

$$\begin{aligned}\gamma(h) &= \text{cov}(X^{t+h}, X^t) = \text{cov}\left(\sum_k \sigma_k \cdot Z^k e^{i(t+h)\omega_k}, \sum_l \sigma_l \cdot Z^l e^{it\omega_l}\right) = \\ &= \sum_1^N |\sigma_k|^2 \cdot e^{ih\omega_k} = \int e^{it\omega} dF(\omega),\end{aligned}$$

wo $F(\cdot)$ die Verteilungsfunktion ist, die in den Positionen ω_k Sprünge der Höhe $|\sigma_k|^2$ besitzt. Es handelt sich um eine sog. fastperiodische Funktion, d. h. eine Überlagerung periodischer Funktionen.

Nicht nur die Covarianzfunktion ist fastperiodisch. Jeder Pfad ist fastperiodisch. Wenn der Zufall die Werte der Zufallsgrößen $\sigma_k Z^k$ bestimmt hat, dann liegt der Pfad fest.

Man kann übrigens die gesamte Zukunft (und die gesamte Vergangenheit) rekonstruieren, wenn man den Prozess zu N verschiedenen Zeitpunkten beobachtet hat. Die Matrix des linearen Gleichungssystems, welches $X^{t_1}, \dots, X^{t_N}$ aus den $\sigma_k Z^k$ bestimmt ist invertierbar. Diese Tatsache dürfte in einem Spezialfall aus der Algebra-Vorlesung bekannt sein: wenn $t_1, \dots, t_N$ die Zeitpunkte $1, \dots, N$ sind, dann ist die Determinante des Gleichungssystems eine sog. Vandermonde-Determinante.

Beispiel 1.3.2.

Es sei $F(\cdot)$ eine Verteilungsfunktion auf dem Intervall $(0, 2\pi]$ und $Z(\omega)$ eine komplexe Brown'sche Bewegung auf $[0, 1]$. Wir betrachten den Prozess $Y(\omega) = Z(F(\omega))$ für $\omega \in (0, 2\pi]$. Die Zuwächse in disjunkten Intervallen sind bei diesem Prozess unabhängig gaussisch verteilt mit $\text{var}(Y(\omega^{\text{II}}) - Y(\omega^{\text{I}})) = F(\omega^{\text{II}}) - F(\omega^{\text{I}})$ für $\omega^{\text{I}} < \omega^{\text{II}}$. Wir betrachten nun für $n \in \mathbb{Z}$ die stochastischen Integrale

$$X^n = \int e^{in\omega} dY(\omega).$$

Es handelt sich um einen gaussischen Prozess mit der Covarianzfolge

$$\gamma(h) = \text{cov}(X^{n+h}, X^n) = \text{cov}\left(\int e^{i(n+h)\omega'} dY(\omega'), \int e^{in\omega''} dY(\omega'')\right) = \int e^{ih\omega} dF(\omega).$$

Dies und das weitere sollte vom weniger erfahrenen Stochastiker als ein Ansporn gelesen werden, sich näher mit stochastischer Integration zu befassen.

Man nennt die Zeitreihe $(X^n)_{n \in \mathbb{Z}}$ den gaussischen Prozess zum Spektralmaß $\mu(d\omega) = dF(\omega)$. Interessant ist der Spezialfall $dF(\omega) = \frac{1}{2\pi} d\omega$. Wir haben dann $\gamma(h) = 0$ für alle $h \neq 0$. Die Zeitreihe ist also ein gaussisches weisses Rauschen.

2 Analysis der Covarianzfolgen und Spektralmaße

Wir wollen klären, welche Funktionen $\gamma(\cdot)$ die Covarianzstruktur $\{c_{rs} = \gamma(r-s) : r, s \in T\}$ eines komplexen schwach stationären Prozesses $\{Z_t : t \in T\}$ sein können. Beim OU-Prozesses haben wir $\gamma_{OU}(h) = e^{-\alpha|h|}$ gefunden. Zu Yule's Pendel gehört die Covarianzfunktion $\gamma_Y(h) = \sigma^2 \cdot \exp(-\frac{1}{2}\gamma|h|) \cdot \cos(\tilde{\omega}h)$, wie wir sehen werden.

Es ist jedenfalls notwendig, dass für beliebige $t_1, \dots, t_m$ die Matrix mit den Einträgen $\text{cov}(Z_{t_k}, Z_{t_l}) = \gamma(t_k - t_l)$ positiv semidefinit ist. Diese Eigenschaft ist auch hinreichend, wie wir im vorigen Unterabschnitt gezeigt haben. Es existiert sogar ein gaussischer Prozess.

Wir nehmen zunächst Abstand von der Welt der Zufallsgrößen. Aus der Analysis übernehmen wir die

Sprechweise. Eine komplexwertige Funktion $\gamma(h)$ auf $\mathbb{R}$ heisst eine positiv semidefinite Funktion, wenn sie stetig und beschränkt ist mit $\gamma(h) = \overline{\gamma(-h)}$ und

$$\sum_{k,l=1}^m a_k \gamma(t_k - t_l) \overline{a_l} \geq 0 \quad \text{für alle } t_1, \dots, t_m \in \mathbb{R} \quad \text{und alle } (a_1, \dots, a_m) \in \mathbb{C}^m.$$

Wir bemerken, dass der Fall $m = 2$ insbesondere $|\gamma(h)| \leq \gamma(0)$ liefert. Die positiv semidefinite Funktion $\gamma(\cdot)$ heisst normiert wenn $\gamma(0) = 1$. Die einfachsten normierten positiv semidefiniten Funktionen sind die Funktionen $\gamma_\omega(\cdot) = e^{i\omega(\cdot)}$ mit $\omega \in \mathbb{R}$; sie heissen die Charaktere der Gruppe $\mathbb{R}$. Sie sind die multiplikative beschränkten Funktionen auf $\mathbb{R}$, $\gamma_\omega(h_1 + h_2) = \gamma_\omega(h_1) \cdot \gamma_\omega(h_2)$. Die Menge aller normierten positiv semidefiniten Funktionen ist offenbar konvex. Für jedes Wahrscheinlichkeitsmaß $\mu(d\omega)$ ist

$$\gamma_\mu(\cdot) = \int \gamma_\omega(\cdot) \mu(d\omega) = \int e^{i\omega(\cdot)} \mu(d\omega)$$

eine normierte positiv semidefinite Funktion. Es gilt

$$\begin{aligned} \sum_{kl} a_k \gamma_\mu(t_k - t_l) \overline{a_l} &= \int \sum_k a_k \exp(i\omega t_k) \cdot \overline{\sum_l a_l \exp(i\omega t_l)} d\mu(\omega) \\ &= \int \left| \sum_k a_k \exp(i\omega t_k) \right|^2 d\mu(\omega). \end{aligned}$$

In der Stochastik heisst die Funktion $\gamma_\mu(\cdot)$ bekanntlich die charakteristische Funktion des W-Maßes μ auf dem Raum $\Omega = \mathbb{R}$.

Wir wollen nun zeigen, dass die charakteristischen Funktionen die einzigen normierten positiv semidefiniten Funktionen sind. Zu jeder normierten positiv semidefiniten Funktion $\gamma(\cdot)$ existiert genau ein W-Maß μ auf $\mathbb{R}$, sodass $\gamma(\cdot) = \gamma_\mu(\cdot)$. Wir behandeln zuerst den zeitdiskreten Fall.

2.1 Der Satz von Herglotz

Satz (Satz von Herglotz).

Es sei $(\gamma_n)_{n \in \mathbb{Z}}$ eine beschränkte Folge komplexer Zahlen mit $\gamma(0) = 1$, $\gamma(n) = \overline{\gamma(-n)}$ und

$$\sum_{kl} a_k \gamma(k-l) \bar{a}_l \geq 0 \quad \text{für alle finiten komplexen Folgen } (a_n)_n.$$

Es existiert dann genau ein W -Maß μ auf $\mathbb{R}/2\pi$ sodass

$$\gamma(n) = \int e^{in\omega} \mu(d\omega) \quad \text{für alle } n$$

Beweis. 1) Man nennt die $\gamma(n)$ die charakteristischen Koeffizienten des W -Maßes μ . Man spricht den Satz folgendermaßen aus: Jede normierte positiv semidefinite Folge ist die Folge der charakteristischen Koeffizienten eines wohlbestimmten W -Maßes auf $\mathbb{R}/2\pi$. Es ist klar, dass es höchstens ein Maß auf dem kompakten Raum $\mathbb{R}/2\pi$ mit vorgegebenen $\gamma(k) = \int e^{ik\omega} d\mu(\omega)$ geben kann; denn jede stetige Funktion $h(\omega)$ lässt sich (nach dem Satz von Stone-Weierstraß) gleichmäßig durch trigonometrische Polynome approximieren.

2) Wir betrachten nun eine summable semidefinite Folge $\gamma(k)_k$ und dazu die Funktion $f(\omega) = \sum \gamma(k) e^{-ik\omega}$. Wir zeigen, dass diese 2π -periodische Funktion nichtnegativ ist. Dazu benützen wir, dass $\sum_{ij} a_j \gamma(j-i) \bar{a}_i \geq 0$ insbesondere für jedes n -Tupel der Form $\{a_j = \zeta^j : j = 1, \dots, n\}$ mit $\zeta = e^{i\omega}$. Die Doppelsumme mit n^2 Summanden lässt sich in diesem Fall als eine Einfachsumme schreiben

$$\sum_{ij} \zeta^j \gamma(j-i) \bar{\zeta}^i = \sum_{ij} \zeta^{j-i} \gamma(j-i) = \sum_k (n-|k|)^+ \gamma(k) \zeta^k.$$

Es gibt in der Tat n Paare (i, j) mit $i-j=0$, $n-1$ Paare mit $i-j=1$ (oder $=-1$), und allgemein $n-|k|$ Paare mit $i-j=k$ (oder $=-k$) für $|k| < n$. Wir haben also

$$f^{(n)}(\omega) = \sum_k (1 - \frac{|k|}{n})^+ \gamma(k) e^{ik\omega} \geq 0$$

für alle n . Im Limes erhalten wir $f(\omega) \geq 0$. Für das Maß $d\mu(\omega) = \frac{1}{2\pi} f(\omega) d\omega$ gilt $\int e^{ik\omega} \mu(d\omega) = \gamma(k)$.

3) Die Rechnung ist auch nützlich, wenn $\gamma(\cdot)$ irgendeine semidefinite Folge ist. Sie ergibt nämlich für das Maß $d\mu^{(n)}(\omega) = \frac{1}{2\pi} f^{(n)}(\omega) d\omega$ die charakteristischen Koeffizienten $(\gamma^{(n)}(k))_k = (1 - \frac{|k|}{n})^+ \gamma(k)$. Wir zeigen, dass die Maße $d\mu^{(n)}(\omega)$ schwach gegen ein Maß $d\mu(\omega)$ auf dem kompakten Raum $\mathbb{R}/2\pi$ konvergiert:

$$\int h(\omega) d\mu^{(n)}(\omega) \rightarrow \int h(\omega) d\mu(\omega) \quad \text{für alle stetigen } h(\cdot)$$

In der Tat haben wir Konvergenz für alle $h_k(\omega) = e^{ik\omega}$ und somit für alle trigonometrischen Polynome wegen

$$\gamma(k) = \int h_k(\omega) d\mu(\omega) = \lim \int h_k(\omega) d\mu^{(n)}(\omega) = \lim (1 - \frac{|k|}{n})^+ \gamma(k).$$

Ein Kriterium für Semidefinitheit. Die semidefiniten Folgen $(c_n)_n$ sind hermitisch in dem Sinn $c(n) = \overline{c(-n)}$. Der Beweis zeigt, dass eine summable hermitische Folge $(c_n)_n$ genau dann semidefinit ist, wenn die trigonometrische Reihe $f(\omega) = \sum c(n)e^{-in\omega}$ überall nichtnegative Werte hat.

Der Beweis zeigt auch, dass eine hermitische Folge $(c_n)_n$ genau dann semidefinit ist, wenn die Folge $c^{(N)}(n) = c(n) \cdot (1 - \frac{|n|}{N})^+$ für jedes $N \in \mathbb{N}$ definit ist. Die Funktionen $(1 - \frac{|n|}{N})^+$ selbst sind übrigens semidefinit. Sie sind erzeugt von den Maßen $d\mu^{(N)}(\omega) = \frac{1}{2\pi} |\sum_1^N e^{ik\omega}|^2 d\omega$; die Dichten sind als die 2π -periodischen Fejér-Kerne bekannt.

Wir bemerken: Wenn $(c_n)_n$ und $(d_n)_n$ semidefinite Folgen sind, dann auch $(c_n \cdot d_n)_n$; das darstellende Maß ergibt sich durch Faltung der darstellenden Maße.

Beispiel 2.1.1.

Es sei $c(0) = 1, c(1) = c(-1) = \rho$ und $c(n) = 0$ für $|n| > 1$. Die Folge ist genau dann positiv semidefinit, wenn $|\rho| \leq \frac{1}{2}$. Es gilt nämlich

$$f(\omega) = 1 + \rho e^{-i\omega} + \rho e^{i\omega} = 1 + 2\rho \cos \omega.$$

Es ist auch leicht, einen stationären gaussischen Prozess mit dieser Kovarianzfunktion anzugeben. Für den Fall $\rho = 0$ leistet eine Folge unabhängiger standardnormalverteilter Zufallsgrößen $(Z_n)_n \in \mathbb{Z}$ das Verlangte; man spricht von einem normierten gaussischen weissen Rauschen. Sei (Z_n) ein solcher Prozess. Der Prozess $W_n = \frac{1}{\sqrt{1+|\alpha|^2}} \cdot (Z_n + \alpha Z_{n-1})$ leistet das Verlangte für $\rho = \frac{\alpha}{1+|\alpha|^2}$.

Beispiel 2.1.2. Der diskret abgetastete OU-Prozess hat die Kovarianzfolge $c(n) = |\rho|^{|n|}$ mit einem $\rho < 1$. Das darstellende Maß hat die Dichte

$$\begin{aligned} f(\omega) &= 1 + \sum_1^\infty \rho^n e^{-in\omega} + \sum_1^\infty \rho^n e^{in\omega} = 1 + \frac{\rho e^{-i\omega}}{1 - \rho e^{-i\omega}} + \frac{\rho e^{i\omega}}{1 - \rho e^{i\omega}} = \\ &= 1 + \frac{\rho}{e^{i\omega} - \rho} + \frac{\rho}{e^{-i\omega} - \rho} = \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos \omega} = \frac{1 - \rho^2}{|e^{i\omega} - \rho|^2}. \end{aligned}$$

Sprechweise. In manchen Anwendung der schwach stationären Prozesse werden die Pfade einer Zeitreihe $(Z_n)_n$ als 'verrauschte' Signale interpretiert, die zu den diskreten Zeitpunkten n abgetastet werden. Das $\omega \in \mathbb{R}/2\pi$, welches aus mathematischer Sicht lediglich eine Integrationsvariable ist, wird als eine Frequenz (genauer Kreisfrequenz) gedeutet. In der Tat erinnert die Folge $(e^{in\omega})_n$ an die Abtastwerte der harmonischen Schwingung mit der Kreisfrequenz ω . (Wenn sich ω' nur um ein ganzzahliges Vielfaches von 2π unterscheidet, dann ist die Folge der Abtastwerte die gleiche.)

Warnung: Es ist höchst problematisch, wenn man versucht, die Realisierungen des Prozesses als Überlagerungen von (diskret abgetasteten) harmonischen Schwingungen zu deuten. Eine solche Deutung ist nur bei den Prozessen mit einem diskreten Leistungsspektrum zulässig; und diese Prozesse sind aus der Sicht der Stochastik ein ausgeartet und so untypisch, dass man sie keiner Weise als Beispiel ansehen kann.

Beispiel 2.1.3 (Fastperiodische Pfade).

Es sei $\mu(d\omega)$ eine konvexe Kombination von Punktmaßen $\delta_{\omega_j}(d\omega)$ mit den Gewichten $|a_j|^2$. ($\sum |a_j|^2 = 1$) Die Kovarianzfunktion $(\gamma_\mu(h))_{h \in \mathbb{R}}$ ist dann die fastperiodische Funktion

$$\gamma_\mu(h) = \sum_j |a_j|^2 e^{ih\omega_j}$$

Einen gaussischen Prozess (mit stetigen Pfaden) kann man dazu leicht konstruieren: Sind nämlich W^j unabhängige standardnormalverteilte Zufallsgrößen und $Z^t = \sum a_j W^j \cdot e^{it\omega_j}$ so gilt für alle t und h $\text{var}(Z^t) = 1$, $\text{cov}(Z^{t+h}, Z^t) = \gamma_\mu(h)$. Die Z_t sind gemeinsam komplexgaussisch verteilt und alle Pfade sind fastperiodisch.

Trotz der Interpretationsprobleme bei allgemeineren Prozessen benützt man weithin den Ausdruck 'Spektralmaß' (oder 'Leistungsspektrum' (zum stationären Prozess) an Stelle des oben eingeführten Begriffs des die Kovarianzfolge 'erzeugenden Maßes'. Dass es sehr gewagt wäre, sich die Pfade eines stationären Prozesses als Überlagerungen von (zeitlich unendlich ausgedehnten) Sinusschwingungen vorzustellen, sollte klar sein, wenn man z. B. an den OU-Prozess denkt. Hier ist auch die Idee der Fourier-Integrale fehl am Platz. Die Pfade $(Z^t(\omega))_{t \in \mathbb{R}}$ sind keine quadratintegrablen Funktionen der Zeit; und nur die quadratintegrablen $f(t)$ als Fourier-Integrale darstellen.

Der zeitkontinuierlichen Fall Es sei $(Z^t)_t \in \mathbb{R}$ ein schwach stationärer Prozess mit der stetigen Kovarianzfunktion $\gamma(t)$ (o. B. d. A. $\text{var} Z^t = \gamma(0) = 1$). Wenn wir ihn zu ganzzahligen Zeiten $n \in \mathbb{Z}$ abtasten, dann erhalten wir eine Kovarianzfolge $(\gamma(n))_{n \in \mathbb{Z}}$ und dazu ein Wahrscheinlichkeitsmaß $d\mu_1(\omega)$, welches wir als W-Maß auf dem Intervall $[0, 2\pi)$ auffassen können. Wenn wir die Abtastfrequenz verdoppeln, dann erhalten wir die semi-definite Folge $(\gamma(n/2))_{n \in \mathbb{Z}}$. Diese können wir darstellen mit einem (eindeutig bestimmten) W-Maß $d\mu_2(\omega)$ auf dem verdoppelten Intervall $[0, 4\pi)$. Es ist gekennzeichnet durch die Eigenschaft

$$\int_0^{4\pi} e^{i\frac{1}{2}n\omega} d\mu_2(\omega) = \gamma\left(\frac{n}{2}\right).$$

Das Maß $d\mu_1(\omega)$ erhalten wir aus $d\mu_2(\omega)$ als Bildmaß bei der Abbildung, welche das Intervall $[0, 2\pi)$ festlässt und das Intervall $[2\pi, 4\pi)$ durch Subtraktion von 2π auf $[0, 2\pi)$ abbildet. Wir konstruieren weiter $d\mu_4(\omega)$ auf dem Intervall $[0, 8\pi)$, welche die Folge $\gamma(n/4)$ richtig beschreibt:

$$\int_0^{8\pi} e^{i\frac{1}{4}n\omega} d\mu_4(\omega) = \gamma\left(\frac{n}{4}\right).$$

So fahren wir fort. Im n -ten Schritt wird aus dem Maß auf $[0, 2^n\pi)$ ein Maß auf $[0, 2^{n+1}\pi)$, indem man einen Teil des Maßes $d\mu_n$ um den Betrag $2^n\pi$ verschiebt. Wenn die Gewichte

der zu verschiebenden Maße nach Null konvergieren, dann haben wir eine schwach konvergente Folge, deren Limes $d\mu_\infty$ (wegen der Stetigkeit von $\gamma(\cdot)$) die Werte in allen Punkten t beschreibt.

$$\int_0^\infty e^{it\omega} d\mu_\infty(\omega) = \gamma(t).$$

Die Konvergenz der Folge μ_{2^n} ergibt sich leicht aus der Straffheit dieser Folge von Maßen, welche wiederum aus der Stetigkeit der Covarianzfunktion folgt.— Wir wollen das nicht im Detail ausführen.

Im Übrigen ist fast alles wie im zeitdiskreten Fall: Wenn die semidefinite Funktion $\gamma(\cdot)$ integrierbar ist, dann hat das darstellende Maß eine Dichte und zwar

$$d\mu(\omega) = \frac{1}{2\pi} f(\omega) d\omega \quad \text{mit} \quad f(\omega) = \int e^{-it\omega} \gamma(t) dt.$$

Die Rechnung ist aus der Theorie der Fourierintegrale wohlbekannt. Wir bemerken: Die Eigenschaft eines Maßes, dass seine charakteristische Funktion integrierbar ist, hat mit der Glätte der Dichte zu tun. Beispielsweise ist bekanntlich die charakteristische Funktion der uniformen Verteilung (auf irgendeinem Intervall) nicht integrierbar. Von der charakteristischen Funktion eines Maßes mit Dichte kann man allerdings sagen, dass sie für $|\omega| \rightarrow \infty$ nach 0 strebt ('Lemma von Riemann-Lebesgue').

Beispiel 2.1.4. Bekannte positivdefinite Funktionen sind u. a.

$$\begin{aligned} \gamma(t) = e^{-|t|} & \quad \text{zu} \quad \frac{1}{\pi} \cdot \frac{1}{1 + \omega^2} d\omega \\ \gamma(t) = \left(1 - \frac{|t|}{T}\right)^+ & \quad \text{zu} \quad \frac{1}{2\pi} \left(\frac{1}{T\omega} \cdot \sin\left(\frac{T}{2}\omega\right)\right)^2 T d\omega. \end{aligned}$$

Wir bemerken, dass die Funktionen $\gamma_\omega(\cdot) = \exp(i(\cdot)\omega)$ die beschränkten stetigen multiplikativen Funktionen auf der Gruppe $\mathbb{R}$ sind. Sie sind die Gruppencharaktere. Die semidefiniten Funktionen mit $\gamma(0) = 1$ sind die konvexen Mittel dieser Charaktere.— Wir erwähnen das, um nochmals die Distanz zu den Fourier-Integralen zu betonen.

2.2 Lineare Filter und ihre Transfer-Funktionen

Sprechweise 2.2.1.

Es sei $\mathbf{a} = a(h)_{h \in \mathbb{Z}}$ eine summable Folge. Wir benützen sie zunächst einmal, um einer beliebigen beschränkten Folge $\mathbf{x} = (x(n))_{n \in \mathbb{Z}}$ eine beschränkte Folge $\mathbf{y} = (y(n))_{n \in \mathbb{Z}}$ zuzuordnen

$$y(n) = \sum_h a_h \cdot x(n-h) = \sum_k a(n-k) \cdot x_k$$

Wir notieren $\mathbf{y} = T_{\mathbf{a}}(\mathbf{x})$ und nennen die Abbildung $T_{\mathbf{a}}$ den zeitinvarianten linearen Filter zur Koeffizientenfolge $\mathbf{a} = (a_h)_{h \in \mathbb{Z}}$. (TLF für 'timeinvariant linear filter') Man spricht auch von einem gleitenden Mittel (Moving average, MA im Englischen)

Der TLF macht aus dem 'Eingabesignal' $\mathbf{x}$ das 'Ausgabesignal' $\mathbf{y}$.

Die komplexe Zahl $a(h)$ heisst die Impulsantwort zur Zeitverzögerung h . Die Folge $\mathbf{a}$ heisst die Influenzfolge des TLF, oder die Folge der Impulsantworten.

Wenn $a(h) = 0$ für alle $h < 0$, dann nennt man $T_{\mathbf{a}}$ einen kausalen Filter; der Wert $y(n)$ hängt hier nur von den x -Werten in der Vergangenheit ab. Wenn nur endlich viele $a(h)$ nicht verschwinden, dann nennen wir $T_{\mathbf{a}}$ einen TLF mit endlicher Reichweite oder auch einen finiten Filter.

Einen TLF kann man auch dazu benützen, um eine schwachstationären Zeitreihe $\mathbf{X}$ eine schwachstationären Zeitreihe $\mathbf{Y} = T_{\mathbf{a}}\mathbf{X}$ zuzuordnen. Die Summabilität der Influenzfolge sichert die Konvergenz der Reihe $Y^t = \sum a(h) X^{t-h}$ sowohl fastsicher als auch im quadratischen Mittel.

Offensichtlich gilt

Satz 2.2.1. *Zeitinvariante lineare Filter kann man hintereinanderschalten, sowohl als Abbildungen beschränkter Funktionen als auch als Abbildungen schwachstationärer Zeitreihen. Es gilt $T_{\mathbf{a}} \circ T_{\mathbf{b}} = T_{\mathbf{c}}$ mit $\mathbf{c} = \mathbf{a} * \mathbf{b}$ (Faltungsprodukt)*

Beispiel 2.2.1.

Der TLF zur Folge mit $a_0 = 1$ und $a(h) = 0$ für alle $h \neq 0$ ist die Identitätsabbildung.

Der TLF zur Folge mit $a_1 = 1$ und $a(h) = 0$ für alle $h \neq 1$ heisst der Rückwärtsshift und wird mit B bezeichnet. $B : (x_n)_{n \in \mathbb{Z}} \mapsto (x_{n-1})_{n \in \mathbb{Z}}$.

Der TLF $(I - B)$ bildet die Folge $(x_n)_n$ in die Folge $(x_n - x_{n-1})_n$ ab. Wir werden sehen, dass er für Folgen eine ähnliche Rolle spielt wie die Differentiation $\frac{d}{dt}$ für (genügend glatte) Funktionen $(x_t)_{t \in \mathbb{R}}$.

Der TLF $(I - B) \circ (I - B) = I - 2B + B^2$ ist analog zu zweiten Ableitung.

Wir können für beliebige summable $\mathbf{a} = (a_h)_h$ schreiben $T_{\mathbf{a}} = \sum_h a_h B^h$. Für einen kausalen Filter haben wir $T_{\mathbf{a}} = p(B)$ mit einer formalen Potenzreihe $p(\cdot)$. Für die Unbestimmte in der formalen Potenzreihe wählen wir meistens die Bezeichnung z , wie das in der Analysis üblich ist. Da die Koeffizienten summabel sind, können wir $p(z)$ auch als eine im Einheitskreis holomorphe Funktion auffassen. Sind $p(z)$ und $q(z)$ formale Potenzreihen mit summablen Koeffizienten, so gilt $p(B) \circ q(B) = r(B)$. mit $r(z) = p(z) \cdot q(z)$.

Bekanntlich kann man jedes Polynom vom Grad p als Produkt von p Linearfaktoren schreiben: $1 - \varphi_1 z - \varphi_2 z^2 - \dots - \varphi_p z^p = (1 - \zeta_1^{-1} z) \cdot \dots \cdot (1 - \zeta_p^{-1} z)$. Daraus folgt

Satz 2.2.2. Für jeden finiten TLF existieren (bis auf die Reihenfolge eindeutig bestimmte) komplexe Zahlen ζ_j^{-1} , sodass mit einem $N \in \mathbb{Z}$ und einem $c \in \mathbb{C}$ gilt

$$T_{\mathbf{a}} = cB^{-N}(I - \zeta_1^{-1}B) \cdot \dots \cdot (I - \zeta_p^{-1}B).$$

Satz 2.2.3.

Gegeben sei der TLF $T_{\mathbf{a}}$. Wenn das Eingangssignal $(x(n))_n$ eine reine Sinusschwingung ist, d. h. wenn $x(n) = C e^{in\omega}$ mit einer 'komplexen Amplitude' C und einer 'Kreisfrequenz' $\omega \in \mathbb{R}/2\pi$, dann ist auch das Ausgangssignal eine reine Sinusschwingung, und zwar

$$y(n) = A(e^{-i\omega}) \cdot x(n) \quad \text{mit} \quad A(z) = \sum a_h z^h$$

Beweis. $y(n) = \sum a_h x(n-h) = C \cdot \sum a_h e^{i(n-h)\omega} = C e^{in\omega} \cdot \sum a_h e^{-ih\omega} = x(n) \cdot A(e^{-i\omega})$.

Sprechweise 2.2.2.

Die komplexwertige Funktion $A(e^{-i\omega}) = \sum_h a_h e^{-ih\omega}$ heisst die **Transfer-Funktion** des TLF $T_{\mathbf{a}}$. Das Absolutquadrat $|A(e^{-i\omega})|^2$ heisst die **Leistungstransferfunktion** des Filters. ('power transfer function' im Englischen).

Beispiel 2.2.2. Sei $\alpha = |\alpha| \cdot e^{i\omega_0}$.

Der TLF $(I - \alpha B)$ hat die Transferfunktion $A(e^{-i\omega}) = 1 - |\alpha| e^{-i(\omega - \omega_0)}$ und die Leistungstransferfunktion $|A(e^{-i\omega})|^2 = |1 - |\alpha| e^{-i(\omega - \omega_0)}|^2 = 1 + |\alpha|^2 - 2|\alpha| \cos(\omega - \omega_0)$. Für $|\alpha| < 1$ hat der TLF $(I - \alpha B)^{-1} = \sum_0^\infty \alpha^k B^k$ die Leistungstransferfunktion

$$|A(e^{-i\omega})|^2 = |1 - \alpha e^{-i\omega}|^{-2} = \frac{1}{1 + |\alpha|^2 - 2|\alpha| \cos(\omega - \omega_0)}.$$

Satz 2.2.4. Es sei $\mathbf{Z} = (Z_n)_n$ ein schwach stationärer Prozess zur Covarianzfolge $\gamma_h = \int e^{ih\omega} d\mu(\omega)$. Der Prozess $\mathbf{W} = T_{\mathbf{a}}(\mathbf{Z})$ hat dann die Covarianzfolge

$$(c_h)_h \quad \text{mit} \quad c_h = \int e^{ih\omega} |A(e^{-i\omega})|^2 d\mu(\omega).$$

Beweis. Wegen $W_{n+h} = \sum_j a_j Z_{n+h-j}$ und $W_n = \sum_k a_k Z_{n-k}$ haben wir

$$\begin{aligned} \text{cov}(W_{n+h}, W_n) &= \sum_{jk} a_j \overline{a_k} \gamma(h-j+k) = \sum_{jk} a_j \overline{a_k} \int e^{i(h-j+k)\omega} d\mu(\omega) = \\ &= \int e^{ih\omega} \left(\sum_j a_j e^{-ij\omega} \right) \cdot \left(\sum_k a_k e^{-ik\omega} \right)^* d\mu(\omega) = \int e^{ih\omega} |A(e^{-i\omega})|^2 d\mu(\omega). \end{aligned}$$

Der Satz erklärt die Sprechweise vom Leistungstransfer. Während im Eingabeprozess die 'Leistung' gemäß $d\mu_X(\omega)$ über die Kreisfrequenzen verteilt war, ist sie im Ausgabeprozess gemäß dem mit der Leistungstransferfunktion multiplizierten Maß verteilt: $d\mu_Y(\omega) = |A(e^{-i\omega})|^2 d\mu_X(\omega)$.

Vorausschau: Es zeigt sich: Wenn der zentrierte schwach stationäre Prozess $(W_n)_n$ eine Spektraldichte der Form $\frac{1}{2\pi} f(\omega) d\omega = \frac{1}{2\pi} |A(e^{-i\omega})|^2 d\omega$ besitzt mit $A(z) = \sum_0^\infty a_h z^h$, dann kann man ihn unter Umständen durch einen kausalen Filter aus einem weissen Rauschen $(Z_n)_n$ gewinnen.

Wie man gegebenenfalls die unkorrelierten 'Innovationen' Z_n findet, werden wir sehen. (Satz von Szegö-Krein-Kolmogorov). Unproblematische Beispiele sind die Spektraldichten der Form

$f(\omega) = \left| \frac{\theta(e^{-i\omega})}{\varphi(e^{-i\omega})} \right|^2$, wo $\theta(z)$, $\varphi(z)$ Polynome sind und $\varphi(z)$ keine Nullstellen im Einheitskreis besitzt. Diese werden wir im Abschnitt über ARMA-Prozesse diskutieren..

Der zeitkontinuierliche Fall

Wir erinnern hier in informeller und hoffentlich intuitiver Sprache an die Gleichung des schwach gedämpften harmonischen Oszillators und seine Gleichgewichtslösung

$$\int_{t'}^{t''} (\ddot{X}(t) + \gamma \dot{X}(t)) dt = -\omega_0^2 \int_{t'}^{t''} X(t) dt + \int_{t'}^{t''} b(t) dt.$$

Wir stören jetzt aber nicht mit dem Zuwachs einer deterministischen Funktion, die auch noch eine Dichte besitzt, $dB_t = b(t) dt$, sondern mit den Zuwächsen eines Prozesses wie der Brown'schen Bewegung. Wir fordern, dass die Zuwächse zentriert, quadratintegabel und unabhängig sind mit $\text{var}(B(t'') - B(t')) = \sigma^2 |t'' - t'|$. Wir denken z. B. an die quadratintegablen Levy-Prozesse. Die stochastische Differentialgleichung lautet dann

$$d_t(\dot{X} + \gamma X) = -\omega_0^2 X dt + dB_t.$$

Die Gleichgewichtslösung ergibt sich mit der oben berechneten Influenzfunktion $(a(h))_{h \in \mathbb{R}_+}$, die man übrigens auch die Green's-Funktion der Differentialgleichung nennt.

$$X(t) = \int_{-\infty}^t a(t-s) dB_s.$$

Die Konstruktion, die einem weissen Rauschen in kontinuierlicher Zeit (dB_s) einen stationären Prozess zuordnet, heisst wie im zeitdiskreten Fall ein zeitinvarianter linearer Filter. Es gibt da technische Probleme mit den stochastischen Integralen, die wir hier natürlich nicht behandeln können. Wir stellen fest: $(X_t)_t$ ist ein schwach stationärer Prozess mit

$$\begin{aligned} \text{cov}(X_{t+h}, X_t) &= \text{cov}\left(\int_{-\infty}^{t+h} a(t+h-s) dB_s, \int_{-\infty}^t a(t-s) dB_s\right) \\ &= \sigma^2 \int_{-\infty}^t a(t+h-s) \cdot \overline{a(t-s)} ds. \end{aligned}$$

Die Influenzfunktion ist $A(-\omega) = \int_0^\infty a(h) e^{-ih\omega} dh$ und das Spektralmaß des gefilterten Prozesse ist $d\nu(\omega) = |A(-\omega)|^2 \cdot \frac{\sigma^2}{2\pi} d\omega$.

Hinweis auf funktionalanalytische Aspekte

In der sog. linearen Theorie der zentrierten schwachstationären Prozesse gründen sich die Konstruktionen und Argumente allein auf die Covarianzfunktion. Alle sonstigen Annahmen und Kenntnisse über gemeinsame Verteilungen bleiben aus dem Spiel. Aus der Sicht der linearen Theorie ist ein zentrierte schwachstationäre Folge zur Covarianzfolge $(\gamma_h)_{h \in \mathbb{R}}$ nichts weiter als eine Folge von Vektoren in einem Hilbertraum $(v_n)_n$ mit $\langle v_{n+h}, v_n \rangle = \gamma_h$. Eine besonders einfache Folge ist die Folge $e_n(\omega) = (e^{in\omega})_n$ auf dem endlichen Maßraum $(\mathbb{R}/2\pi, \mathfrak{B}, d\mu(\omega))$. Die inneren Produkte sind

$$\langle e_{n+h}, e_n \rangle = \int e_{n+h}(\omega) \cdot \overline{e_n(\omega)} d\mu(\omega) = \int e^{ih\omega} d\mu(\omega) = \gamma(h).$$

In der Hilbertraumwelt macht die zeitkontinuierliche Version keine Probleme. Es sei $d\mu(\omega)$ ein endliches Maß auf $\mathbb{R}$ und $\gamma_h = \int e^{ih\omega} d\mu(\omega)$. Ein Prozess zur semidefiniten Funktion $(\gamma_h)_{h \in \mathbb{R}}$ ist eine im Hilbertraum stetige Schar $(v_t)_{t \in \mathbb{R}}$ mit $\langle v_{t+h}, v_t \rangle = \gamma_h$ für alle $h \in \mathbb{R}$. Solche 'Kurven' im Hilbertraum ergeben sich z. B. folgendermaßen: Es sei $(U_t)_{t \in \mathbb{R}}$ eine stark stetige Gruppe von unitären Operatoren. Für jeden Vektor v_0 hat dann die 'Bahn' $v_t = U_t v_0$ die Eigenschaft, dass das innere Produkt $\langle v_{n+h}, v_n \rangle$ nur von h abhängt. Wenn $a(h)_{h \in \mathbb{R}}$ eine integrable Funktion ist und $(v_t)_{t \in \mathbb{R}}$ ein (im Hilbertraum-Sinn) stationärer Prozess, dann ergeben die Integrale (im Hilbertraum-Sinn) $w_t = \int_{-\infty}^t a(t-s) v_s ds$ einen stationären Prozess. Die Konstruktion $(v_s)_s \mapsto (w_t)_t$ ist der lineare Filter zur Influenzfunktion $a(\cdot)$.

Stark stetige einparametrische Gruppen unitärer Operatoren sind in der Hilbertraumtheorie ein bestens ausgearbeitetes Thema. Ein zentrales Resultat ist die Spektraldarstellung. Die technischen Probleme mit der stochastischen Integration werden in dieser Theorie aber nicht erfasst. Wir bleiben zunächst noch in der klassischen Analysis. Die Betrachtungen um Integration werden wir aufnehmen, wenn wir orthogonale stochastische Maße diskutieren. (Im Englischen 'Orthogonal stochastic measures').

2.3 Spezielle Transferfunktionen

Einige interessante TLF haben Transferfunktionen bzw. Impulsantworten, die man aus der Analysis oder aus der elementaren Stochastik gut kennt. Wir benützen die Gelegenheit, an einige spezielle Folgen bzw. 2π -periodische Funktionen zu erinnern, die gerne für 'Beispiele' herangezogen werden.

Es sei T eine nichtnegative ganzzahlige Zufallsgröße mit $\text{Ws}(T = k) = p_k$ für $k = 0, 1, \dots$. Die Transferfunktion zum kausalen Filter $x(\cdot) \mapsto y(n) = \sum_0^\infty p_k x(n - k)$ kann man als einen Erwartungswert schreiben

$$A(e^{-i\omega}) = \sum_0^\infty p_k e^{-ik\omega} = \mathfrak{E}(e^{-i\omega T})$$

Sind T_1, T_2 unabhängig nach $(p_k)_k$ verteilt, so gilt für die Leistungstransferfunktion

$$|A(e^{-i\omega})|^2 = \mathfrak{E}(e^{i\omega(T_1 - T_2)}) = \sum_{-\infty}^\infty q_n e^{-in\omega} \quad \text{mit} \quad q_n = \sum p_{n+k} p_k.$$

Im Einzelfall kann die eine oder die andere Formel bequemer benützt werden.

Beispiel 2.3.1 (Geometrisch abfallende Impulsantworten). Ein Versuch mit der Erfolgswahrscheinlichkeit $(1-q)$ wird unabhängig wiederholt. Die Wartezeit bis zum ersten Erfolg ist bekanntlich geometrisch verteilt mit dem Erwartungswert $\frac{1}{1-q}$:

$$\text{Ws}(T = k) = p_k = (1 - q) \cdot q^{k-1} \quad \text{für } k = 1, 2, \dots$$

Die dazugehörigen Transfer- und Leistungstransferfunktionen sind

$$A(e^{-i\omega}) = e^{-i\omega}(1 - q) \cdot (1 - qe^{-i\omega})^{-1},$$

$$|A(e^{-i\omega})|^2 = (1 - q)^2 \cdot (1 - 2q \cos \omega + q^2)^{-1} = \frac{1 - q}{1 + q} \sum_{-\infty}^\infty e^{-in\omega} q^{|n|}.$$

denn für $n \geq 0$ gilt $\sum_1^\infty p_{k+n} p_k = (1 - q)^2 q^n \sum_1^\infty q^{2k-2} = (1 - q)^2 q^n (1 - q^2)^{-1}$.

Wenn der Filter auf ein weisses Rauschen wirkt, dann ist das Ergebnis eine Zeitreihe mit der Kovarianzfolge $\sigma^2 \frac{1-q}{1+q} \cdot q^{|n|}$. Wenn das weisse Rauschen gaussisch ist, dann ergibt sich der zu ganzzahligen Zeiten beobachtete Ornstein-Uhlenbeck-Prozess.

Ebenso bekannt ist die etwas anders normierte Transferfunktion zu einem komplexen $q = |q|e^{i\omega_0}$ mit $0 < |q| < 1$

$$P_q(e^{-i\omega}) = \sqrt{1 - |q|^2} (1 - qe^{-i\omega})^{-1} = \sqrt{1 - |q|^2} \left(1 + \sum_1^\infty q^h e^{-ih\omega} \right).$$

Es handelt sich um einen kausalen Filter. Seine Leistungstransferfunktion (für festes $|q|$) ist die Dichte des Poisson-Kerns. Diesen werden wir unten genauer diskutieren. Hier erwähnen wir nur seine charakteristischen Koeffizienten $\gamma_q(h) = |q|^{|h|} \cdot e^{-ih\omega_0}$.

Beispiel 2.3.2. Zur zentrierten Zeitreihe $(X^t)_{t \in \mathbb{Z}}$ betrachten wir das (nichtkausale) gleitende Mittel $Y^t = \sum_{-n}^n X^{t-h}$. Die Covarianzfolge $\gamma_Y(\cdot)$ ergibt sich aus der Covarianzfolge $\gamma_X(\cdot)$ in der bekannten Weise

$$\gamma_Y(h) = \text{cov} \left(\sum_{-n}^n X^{h-i}, \sum_{-n}^n X^{-j} \right) = (2n+1) \sum_k \left(1 - \frac{|k|}{2n+1} \right)^+ \gamma_X(n-k)$$

Die dazugehörigen Transferfunktion ist der n -te Dirichlet-Kern

$$A(e^{-i\omega}) = \sum_{h=-n}^n e^{-ih\omega} = \frac{\sin(n + \frac{1}{2})\omega}{\sin \frac{1}{2}\omega} = D_n(\omega)$$

Die Leistungstransferfunktion ist der Fejér-Kern der Ordnung n $|A(e^{-i\omega})|^2 = F_n(\omega)$. Da im Beispiel die (nichtkausalen!) Impulsantworten symmetrisch um den Nullpunkt sind, ist die Transferfunktion reellwertig. Ebenfalls nullsymmetrisch (aber nicht reellwertig) sind die Impulsantworten in folgenden Verallgemeinerung: Für ein reelles ω_0 sei $a(h) = e^{ih\omega_0}$ für und $a(h) = 0$ für $|h| > n$. Die Transferfunktion ist $D_n(\omega - \omega_0)$, reellwertig aber nicht nullsymmetrisch.

Wenn ein Input-Prozess $\mathbf{X}$ das Spektralmaß $d\mu_X(\omega)$ hat, dann hat der Y -Prozess das Spektralmaß $d\mu_Y(\omega) = |A(e^{-i\omega})|^2 d\mu_X(\omega)$, und dieses ist in unserem Fall (für große n) eng auf die Umgebung von ω_0 konzentriert. Der TLF bevorzugt also die Frequenzen in der Nähe von λ .

Im zeitkontinuierlichen Fall sind die Verhältnisse ähnlich. Es sei T eine nichtnegative Zufallsvariable mit der Dichte $a(h) dh$. Wir gewinnen aus dem Prozess $(X^t)_{t \in \mathbb{R}}$ mit dem Spektralmaß $d\mu_X(\omega)$ den Prozess $Y^t = \int_0^\infty X^{t-h} a(h) dh$. Die Spektralmaß des Y -Prozesses ist dann $d\mu_Y(\omega) = |A(-\omega)|^2 \cdot d\mu(\omega)$ mit der Transferfunktion $A(-\omega) = \int_0^\infty a(h) e^{-ih\omega} dh$. Dies ist die charakteristische Funktion der Wahrscheinlichkeitsdichte $a(h) dh$.

Beispiel 2.3.3 (Exponentiell abfallende Impulsantworten). Es sei T exponentiell verteilt mit Erwartungswert $\frac{1}{\alpha}$. Die Dichte ist $a(h) dh = \alpha e^{-\alpha h} dh$ auf $\mathbb{R}_+$. Die dazugehörigen Transfer- und Leistungstransfer-funktionen sind

$$A(-\omega) = \mathfrak{E} (e^{-i\omega T}) = \frac{\alpha}{\alpha + i\omega}$$

$$|A(-\omega)|^2 = \frac{\alpha^2}{\alpha^2 + \omega^2} = \mathfrak{E} (e^{-i\omega(T_1 - T_2)})$$

wo T_1, T_2 unabhängig exponentiell verteilt sind mit dem Erwartungswert $\frac{1}{\alpha}$. Die Verteilung ist die doppeltextponentielle Verteilung mit der Dichte $\frac{1}{2}\alpha e^{-\alpha|h|} dh$. Man bemerke, dass die charakteristische Funktion bis auf einen Faktor eine Cauchy-Dichte ist.

Beispiel 2.3.4. Die Techniken der gekoppelten Schwingkreise liefert reichhaltige Möglichkeiten, lineare Filter technisch zu realisieren. Betrachten wir als einfachsten Fall den

gedämpften harmonischen Oszillator $\ddot{x} + \gamma \cdot \dot{x} + \omega_0^2 \cdot x = b(t)$. Wir wollen uns auf den sog. schwach gedämpften Fall $\omega_0 > \frac{1}{2}\gamma$ konzentrieren und definieren $\tilde{\omega} = \sqrt{\omega_0^2 - \frac{1}{4}\gamma^2}$. Die komplexen Lösungen der homogenen Gleichung sind dann bekanntlich die Linearkombinationen der beiden Funktionen $x_{\pm}(t) = \exp\left(-\left(\frac{1}{2}\gamma \pm i\tilde{\omega}\right)t\right)$. Die einzige Lösung, die für $t \rightarrow -\infty$ beschränkt bleibt, ist die Nulllösung. Wenn $b(\cdot)$ auf jedem Intervall $(-\infty, T]$ beschränkt ist, dann besitzt die inhomogene Gleichung genau eine auf jedem Intervall $(-\infty, T]$ beschränkte Lösung. Man nennt sie die Gleichgewichtslösung, und man gewinnt eine Darstellung durch Integrale nach der Methode der Variation der Konstanten. Es existiert eine 'Impulsantwortfunktion' $a(h)$ auf $\mathbb{R}_+$, sodass $y(t) = \int_0^\infty a(h)b(t-h) dh$ die Gleichgewichtslösung zu $b(\cdot)$ ist. Es stellt sich heraus

$$\begin{aligned} a(h) &= \exp\left(-\frac{1}{2}\gamma h\right) \cdot \frac{1}{\tilde{\omega}} \sin h\tilde{\omega} = \frac{1}{2i\tilde{\omega}} \left[e^{(-\frac{1}{2}\gamma + i\tilde{\omega})h} - e^{(-\frac{1}{2}\gamma - i\tilde{\omega})h} \right] \\ A(-\omega) &= \int_0^\infty a(h)e^{-i\omega h} dh = \int_0^\infty e^{-i\omega h} \frac{1}{2i\tilde{\omega}} \left(e^{(-\frac{1}{2}\gamma + i\tilde{\omega})h} - e^{(-\frac{1}{2}\gamma - i\tilde{\omega})h} \right) dh = \\ &= \frac{1}{2i\tilde{\omega}} \left[-\left(-\frac{1}{2}\gamma + i\tilde{\omega} - i\omega\right)^{-1} + \left(-\frac{1}{2}\gamma - i\tilde{\omega} - i\omega\right)^{-1} \right]. \\ &= \left(-\frac{1}{2}\gamma + i\tilde{\omega} - i\omega\right)^{-1} \cdot \left(-\frac{1}{2}\gamma - i\tilde{\omega} - i\omega\right)^{-1} = \left(\omega_0^2 - \omega^2 + \gamma i\omega\right)^{-1} \end{aligned}$$

Die Formel für die Transferfunktion $A(-\omega)$ hätten wir auch einfacher erhalten können. Wir können nämlich die Gleichgewichtslösungen zu allen 'reinen Sinusschwingungen' direkt ausrechnen. Die Lösung $y(\cdot)$ zu $b_\omega(\cdot) = e^{i(\cdot)\omega}$ ist

$$y(t) = A(-\omega) \cdot e^{it\omega} \quad \text{mit} \quad A(-\omega) = \left(\omega_0^2 - \omega^2 + \gamma i\omega\right)^{-1},$$

denn für diese Funktion gilt in der Tat

$$\ddot{y} + \gamma \dot{y} + \omega_0^2 y = A(-\omega) \left[-\omega^2 + \gamma i\omega + \omega_0^2\right] e^{it\omega} = e^{it\omega}$$

Interessant ist auch noch ein Blick auf die Leistungstransferfunktion

$$|A(-\omega)|^2 = \left((\omega_0^2 - \omega^2)^2 + \gamma^2 \omega^2\right)^{-1}$$

Für die ω in der Nähe von ω_0 können wir approximieren $\omega_0^2 - \omega^2 = (\omega_0 - \omega)(\omega_0 + \omega) \sim 2\omega_0(\omega_0 - \omega)$ und erhalten somit

$$|A(-\omega)|^2 \sim \frac{1}{4\omega_0^2} \cdot \frac{1}{(\omega_0 - \omega)^2 + \frac{1}{4}\gamma^2}$$

Diese Funktion hat ihr Maximum in ω_0 und ist größer als die Hälfte dieses Maximalwerts im Bereich $\omega_0 \pm \frac{1}{2}\gamma$. Dies gibt eine gute Vorstellung, wie ein gedämpfter Oszillator das Leistungsspektrum der Eingabe modifiziert. Wenn wir dem gedämpften Oszillator einen stationären Input-Prozess $\mathbf{X}$ mit dem Spektralmaß eingeben, dann hat der Output-Prozess $\mathbf{Y}$ das Spektralmaß $d\mu_Y(\omega) = |A(-\omega)|^2 d\mu(\omega)$ mit dem eben berechneten $A(-\omega)$.

Beispiel 2.3.5 (Ausblick auf stochastische Integration). Wir haben eben Integrale betrachtet der Art

$$Y^t = \int_0^\infty X^{t-h} a(h) dh = \int_{-\infty}^t a(t-s) X^s ds.$$

Wir können nun aber auch allgemeinere stochastische Integratoren von der Gestalt zulassen, etwa

$$Y^t = \int_{-\infty}^t a(t-s) dB_s$$

wo B_s eine Brown'sche Bewegung ist. Die Zuwächse dB_s gelten als das kontinuierliche Analogon zum daussischen weissen Rauschen in diskreter Zeit. Der gedämpfte Oszillator macht aus dem weissen Rauschen einen gaussischen Prozess mit der Spektraldichte $\text{const}|A(-\omega)|^2 d\omega$. Wir können den Oszillator auch mit einem anderen Prozess mit unabhängigen Zuwächsen (und endlicher Varianz) anregen, etwa mit einem Poisson-Prozess ('Yule's Pendel'). Wir erhalten einen schwachstationären Prozess mit derselben Spektraldichte. Die Pfade eines solchen nicht-gaussischen Prozesses sehen aber natürlich ganz anders aus.

Sprechweise (Tapern und Glätten). Ein Verfahren, welches beschränkte Folgen dazu bringt, für $|n| \rightarrow \infty$ ausreichend schnell nach 0 zu konvergieren, heisst ein Tapering-Verfahren oder kurz ein Taper. Ein Verfahren, welches eine lokalintegrale Funktion glatt macht, heisst ein Glättungsverfahren ('smoothing' im Englischen).

Die Fejér-Kerne werden gerne zum Glätten von 2π -periodischen Funktionen benützt oder auch zur Glättung eines Maßes auf $\mathbb{R}/2\pi$. Die Glättung einer Spektraldichte mit dem Fejér-Kern bewirkt sehr radikales Tapern der Folge der Fourier-Koeffizienten; die Fourier-Koeffizienten des geglätteten Spektralmaßes verschwinden für große $|n|$; die geglätteten Funktionen sind trigonometrische Polynome.

In unserem Beweis des Satzes von Herglotz nutzten wir dieses radikale Tapern um die Kovarianzfolge summabel zu machen. Das Ziel dieses Taperns war es, glatte Dichten im Frequenzbereich zu erzeugen, welche das Spektralmaß approximieren.

Eine ähnliche Rolle können die Poisson-Kerne spielen. Wenn man das Maß zur Kovarianzfolge $\gamma(h)$ mit dem Poisson-Kern zu einem $q = |q|e^{i\omega_0}$ glättet, dann erhält man als Kovarianzfolge die getaperte Folge $|q|^{|h|}e^{-ih\omega_0}\gamma(h)$. Das Tapern der Kovarianzfolge mit geometrisch abfallenden Gewichten entspricht dem Glätten des Spektralmaßes mit einem Poisson-Kern.

Beachte: Wenn wir von den Fejér-Kernen (zu einer festen Ordnung n) und den Poisson-Kernen (zu einem festen $|q|$) sprechen, dann meinen wir Familien von Wahrscheinlichkeitsdichten, die auseinander durch Verschiebungen im $\mathbb{R}/2\pi$ hervorgehen. Wenn wir um ω_0 verschieben, dann nimmt die Kovarianzfolge den Faktor $e^{-ih\omega_0}$ auf.

Lassen wir die Verschiebungen erst einmal aus dem Spiel, und denken wir an einen einzelnen Akt des Glättens bzw. Taperns. Unsere Beispiele beleuchten die folgende

Faustregel

Ein Tapern von Kovarianzfolgen durch die Multiplikation mit einer (für $|n| \rightarrow \infty$) abfallenden Folge erreicht man dadurch, dass man die Spektraldichte mit einer glatten Wahrscheinlichkeitsdichte faltet.

Ein Glätten von Kovarianzfolgen ergibt sich, wenn man den Prozess durch einen Filter schiebt, dessen Leistungstransferfunktion die niedrigen Frequenzen weniger dämpft als die hohen. (Ein Filter dieser Art heisst ein Tiefpass.) Auch anderes 'Präparieren' der Kovarianzfolgen kann man durch geeignete TLF erreichen. Die Leistungstransferfunktionen werden (punktweise auf $\mathbb{R}/2\pi$) multipliziert.

Die Faustregeln sind nicht überraschend, wenn man bedenkt, dass die Fourier-Transformation (in guten Fällen) aus dem Faltungsprodukt ein punktweises Produkt macht und umgekehrt. Der Haken ist der, dass Kovarianzfolgen nicht notwendigerweise summabel sind und Spektralmaße nicht notwendigerweise eine glatte Dichte besitzen. Ausserdem sind die Leistungstransferfunktionen nicht einfach Fouriertransformierte sondern Absolutquadrate von Fouriertransformierten. Man möchte für das Präparieren des Spektralmaße auch nicht immer alle Filter in Betracht ziehen, sondern eventuell nur die kausalen. Der Weg von den Leistungstransferfunktion zu geeigneten Transferfunktionen wird uns in späteren Abschnitten intensiv beschäftigen; das Stichwort ist die Wiener-Hopf-Faktorisierung.

Wir wollen noch zwei spezielle Filter im Kontinuierliche ansprechen, und schliesslich auch noch ein beliebtes Tapering.

Beispiel 2.3.6 (Hart gemittelte Beobachtungen). Das gleitende Mittel über einen Abschnitt einer Zeitreihe mit gleichen Gewichten hat uns oben auf die Dirichlet-Kerne (bei der Transferfunktion) und auf die Fejér-Kerne (bei der Leistungstransferfunktion) geführt.

Das kontinuierliche Analogon gleitenden Mittels über Beobachtungswerte im zeitkontinuierlichen Fall ist mathematisch leicht zu behandeln, praktisch allerdings wenig bedeutsam; weichere Mittelungen sind in der Regel vorzuziehen. Man beachte: Auch im Beweis des Satz von Herglotz haben sich die Fejér-Kerne bewährt, aber in einem ganz anderen Sinn. Es ging da ums Tapern (zur Glättung des Spektralmaßes) und nicht um Filter (zur 'Glättung' von Kovarianzfolge.)

Im kontinuierlichen Fall liefert das gleitende Mittel

$$Y^t = \frac{1}{2\Delta} \int_{-\Delta}^{\Delta} X^{t-s} ds$$

die Transferfunktion

$$A(-\omega) = \frac{1}{2\Delta} \int_{-\Delta}^{\Delta} e^{-is\omega} ds = \frac{\sin(\Delta\omega)}{\Delta\omega}.$$

Aus dem Spektralmaß $d\mu_X(\omega)$ wird das Spektralmaß $d\mu_Y(\omega) = \left(\frac{\sin(\Delta\omega)}{\Delta\omega}\right)^2 d\mu_X(\omega)$.

Beispiel 2.3.7 (Weich gemittelte Beobachtungen).

Es sei $a(h) dh = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{1}{2\sigma^2} h^2\right) dh$ mit kleinem σ ; und $Y^t = \int X^{t-h} a(h) dh$.

Die Transferfunktion dieses glättenden TLF ist

$$A(-\omega) = \int a(h)e^{ih\omega} dh = \exp\left(-\frac{\sigma^2}{2}\omega^2\right).$$

Der Filter lässt (für kleine σ) das Spektralmaß bei den kleinen Frequenzen nahezu unverändert; im Bereich hoher Frequenzen wird die Leistung entscheidend geschwächt.

Beispiel 2.3.8 (Tapern mit Hilfe der Poisson-Kerne).

Wir zeigen, dass das Tapern einer Covarianzfolge mit beidseitig geometrisch abfallenden Gewichten zusammengehört mit dem Glätten des Spektralmaßes durch die Poisson-Kerne.

Es sei μ ein Maß auf $\mathbb{R}/2\pi$ (manchmal $d\mu(\omega) = \frac{1}{2\pi}f(\omega) d\omega$), und $a(n)$ die Folge seiner charakteristischen Koeffizienten, $a(n) = \int e^{-in\omega} d\mu(\omega)$. Wenn $|a(n)|$ genügend schnell abfällt für $n \rightarrow \infty$, dann liefert bekanntlich $\sum_{-\infty}^{\infty} a(n)e^{in\omega}$ die Dichte von $d\mu$. Dies wollen wir nicht voraussetzen.

Die Folge $a^{(r)}(n) = a(n) \cdot r^{|n|}$ ist summabel für jedes $0 < r < 1$ und ebenfalls positiv definit. Sie ist, wie wir werden werden, die Covarianzfolge zum geglätteten Maß $\mu * \rho^{(r)}(d\omega)$, wo $\rho^{(r)}$ der Poissonkern zum Parameter r ist.

$$\rho^{(r)}(d\omega) = \frac{1-r^2}{1+r^2-2r\cos\omega} \frac{1}{2\pi} d\omega.$$

Wenn man ein Maß μ mit einem Maß ρ faltet, welches eine (stetige) Dichte besitzt, $\rho(d\omega) = p(\omega) \frac{1}{2\pi} d\omega$, dann hat das Ergebnis eine (stetige) Dichte

$$\mu * \rho(d\omega) = g(\omega) \frac{1}{2\pi} d\omega \quad \text{mit} \quad g(\omega) = \int d\mu(\eta) p(\omega - \eta).$$

Für festes ω ist $p^{(r)}(\omega - \eta) = \frac{1-r^2}{1+r^2-2r\cos(\omega-\eta)}$ eine 2π -periodische Funktion. Sie nimmt ihren Maximalwert $\frac{1-r^2}{(1-r)^2} = \frac{1+r}{1-r}$ in den Punkten $\omega + 2k\pi$ an, und den Minimalwert $\frac{1+r}{1-r}$ in den Punkten dazwischen. In jedem Intervall $[\omega - \pi, \omega + \pi]$ ist sie monoton. Für r nahe bei 1 ist sind die Spitzen sehr ausgeprägt. Die Bereiche, in welchem $p(\cdot)$ größer als der halbe Maximalwert ist, sind kleine Intervalle $[\omega - \Delta_r, \omega + \Delta_r\pi]$, denn $1+r^2-2r\cos\eta < 2(1+r^2-2r) \iff 1-\cos\eta < \frac{1}{2r}(1-r)^2 \iff 2\sin(\eta/2) < \frac{1}{\sqrt{r}}(1-r)$. Die Fourier-Darstellung der Funktion $p^{(r)}(\omega)$ ergibt sich aus

$$\frac{1-r^2}{1+r^2-2r\cos\omega} = \frac{1-r^2}{|1-re^{-i\omega}|^2} = \frac{1}{1-re^{-i\omega}} + \frac{re^{i\omega}}{1-re^{+i\omega}} = \sum_{-\infty}^{\infty} r^{|n|} e^{in\omega}.$$

$$\int e^{in\omega} (\mu * \rho^{(r)})(d\omega) = \int e^{in\omega} g^{(r)}(\omega) \frac{1}{2\pi} d\omega = a(n) \cdot r^{|n|}.$$

Wenn wir ein trigonometrisches Polynom $b(\omega) = \sum_j b_j e^{ij\omega}$ mit dem geglätteten Maß integrieren, dann erhalten wir $\int b(\omega) (\mu * \rho^{(r)})(d\omega) = \sum b_j r^{|j|} a_j$.

2.4 Kausale Filter, ARMA-Konstruktionen.

Eine Folge unkorrelierter Zufallsgrößen mit derselben Varianz $\mathbf{Z} = (Z^t)_{t \in \mathbb{Z}}$ nennt man ein weisses Rauschen mit der Intensität σ^2 ('White noise', abgekürzt $\text{WN}(0, \sigma^2)$). Wenn die Z^t unabhängig (komplex)gaussisch sind, dann haben wir ein (komplex)gaussisches weisses Rauschen $\text{GWN}(0, \sigma^2)$.

Definition 2.1. Es sei $\mathbf{Z}$ ein weisses Rauschen und $a(z) = a_0 + a_1 z + a_2 z^2 + \dots$ eine Potenzreihe mit $\sum |a_h| < \infty$. Der Prozess $X^t = \sum_0^\infty a_h Z^{t-h}$ heisst dann ein (bzgl. $\mathbf{Z}$) kausaler MA-Prozess (causal moving average process) oder eine kausale MA-Zeitreihe.

Satz 2.4.1. Ist $\mathbf{Z}$ ein weisses Rauschen. Wenn $|\varphi_1| < 1$, so existiert genau eine bzgl. $\mathbf{Z}$ kausaler MA-Zeitreihe mit

$$X^t = \varphi_1 \cdot X^{t-1} + Z^t \quad \text{für alle } t \in \mathbb{Z}$$

Man nennt ihn einen autoregressiven Prozess (AR(1)-Prozess) zum Polynom $1 - \varphi_1 z$.

Beweis. Es ist hier für alle m gefordert

$$X_t = Z_t + \varphi_1 Z_{t-1} + \varphi_1^2 Z_{t-2} + \dots + \varphi_1^m Z_{t-m} + \varphi_1^{m+1} X_{t-m-1}.$$

Für $m \rightarrow \infty$ ergibt sich für X_t eine (im fastsicheren und im L^2 -Sinn) konvergente Reihe.

Für $|\varphi_1| \geq 1$ existiert offenbar kein bzgl. $\mathbf{Z}$ kausaler AR(1)-Prozess. Im Fall $|\varphi_1| > 1$ können wir folgendermaßen eine Zeitreihe $\mathbf{X}$ konstruieren: Wir konstruieren zunächst zum weissen Rauschen $W^s = (Z^{-s+t})$ eine Schar $\mathbf{Y}$, sodass $Y^t = \frac{1}{\varphi_1} Y^{t-1} + W^t$ für alle $t \in \mathbb{Z}$. Die Schar $X^t = \varphi_1^{-1} \cdot Y^{-t}$ erfüllt dann die Gleichungen $X^t = \varphi_1 \cdot X^{t-1} + Z^t$ mit dem weissen Rauschen der Intensität σ^2 , nämlich zum weissen Rauschen $Z^t = -W^{-t+1}$. Die so konstruierte Zeitreihe $\mathbf{X}$ ist nicht kausal bzgl. des gegebenen $\text{WN } \mathbf{Z}$.

Notation. Wir schreiben die Bestimmungsgleichung des AR-Prozesses $\mathbf{X}$ und die Darstellung durch den Prozess $\mathbf{Z}$ mit Hilfe des Rückwärtsshiftoperators B ('backward shift').

$$(I - \varphi_1 B) \mathbf{X} = \mathbf{Z}, \quad \mathbf{X} = (I + \varphi_1 B + \varphi_1^2 B^2 + \dots) \mathbf{Z} = (I - \varphi_1 B)^{-1} \mathbf{Z}.$$

Man kann sagen, dass $\mathbf{X} = (I - B)^{-1} \mathbf{Z}$ der kausale MA-Prozess ist, der mit geometrisch abfallenden Gewichten aus $\mathbf{Z}$ entsteht.

Eine Analogie: Man kann die Autoregressionsgleichung auch folgendermaßen schreiben: $(X_t - X_{t-1}) = (\varphi_1 - 1)X_{t-1} + Z_t$. Ein zeitkontinuierliches Analogon ist die stochastische Differentialgleichung $dX_t = \alpha X dt + dZ_t$. Sie hat eine kausale stationäre Lösung, wenn $\Re \alpha < 0$, nämlich $X_t = \int_{-\infty}^t e^{\alpha(t-s)} dZ_s$.

Satz 2.4.2.

Es sei $\mathbf{Z}$ ein weisses Rauschen und $\varphi(z) = 1 - \varphi_1 z - \dots - \varphi_p z^p$ ein Polynom, dessen Nullstellen ausserhalb des Einheitskreises liegen. Es existiert dann genau ein bzgl. $\mathbf{Z}$ kausaler Prozess mit

$$X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} - \dots - \varphi_p X_{t-p} = Z_t \quad \text{für alle } t.$$

Man nennt ihn einen $AR(p)$ -Prozess zum Polynom $\varphi(z)$.

Beweis. Wir schreiben die Autoregressionsgleichung mit dem Rückwärtshiftoperator B

$$(I - \varphi_1 B - \dots - \varphi_p B^p) \mathbf{X} = \mathbf{Z} \quad \text{oder kurz} \quad \varphi(B) \mathbf{X} = \mathbf{Z}.$$

Sind $\zeta_1, \dots, \zeta_p$ die Nullstellen des Polynoms $\varphi(z) = 1 - \varphi_1 z - \dots - \varphi_p z^p$, so gilt

$$(I - \zeta_1^{-1} B) \circ \dots \circ (I - \zeta_p^{-1} B) \mathbf{X} = \mathbf{Z}.$$

Die Konstruktion eines kausalen Prozesses $\mathbf{X}$ gemäß $\mathbf{X} = (I - \varphi B)^{-1} \mathbf{Y}$ kann man nicht nur für ein weisses Rauschen durchführen, sondern für beliebige Prozesse $\mathbf{Y}$, insbesondere für solche, die aus $\mathbf{Z}$ durch einen kausalen MA hervorgegangen sind. Es entsteht ein bzgl. $\mathbf{Z}$ kausaler Prozess.

Unsere Gleichung kann als die $AR(1)$ -Gleichung $(I - \zeta_1^{-1} B) \mathbf{Y} = \mathbf{Z}$ verstanden werden für den Prozess $\mathbf{Y} = (I - \zeta_2^{-1} B) \circ \dots \circ (I - \zeta_p^{-1} B) \mathbf{X}$. Die eindeutige Lösung ist $\mathbf{Y} = (I - \zeta_1^{-1} B)^{-1} \mathbf{Z}$. Wenn wir so fortfahren, erhalten wir die gewünschte Konstruktion von $\mathbf{X}$ durch ein Hintereinanderschalten von $AR(1)$ -Konstruktionen. $\mathbf{X}$ ist also der kausale MA-Prozess mit den Gewichten aus der Potenzreihe $\psi(z) = \frac{1}{\varphi(z)}$.

Analogie: Der Satz erinnert an den folgenden Satz, welcher lineare Differentialgleichungen mit konstanten Koeffizienten betrifft: Wenn alle Nullstellen des Polynom $\varphi(z)$ negativen Realteil haben, dann besitzt die Gleichung

$$x^{(p)} - \varphi_1 x^{(p-1)} - \dots - \varphi_{p-1} \dot{x} - \varphi_p x = b(t)$$

genau eine für $t \rightarrow -\infty$ beschränkte Lösung für jede beschränkte Funktion $b(\cdot)$. Die Lösungen der homogenen Gleichung sind nämlich Linearkombinationen von Funktionen der Gestalt $t^k \cdot e^{\lambda t}$, wo λ eine Nullstelle des Polynoms $\varphi(z)$ ist. Sie alle (ausser der Nullfunktion) sind unbeschränkt für $t \rightarrow -\infty$.

Beachte: Der Begriff der Kausalität (bzgl. eines Prozesses) gibt es natürlich nicht in der Welt der deterministischen Differentialgleichungen.

Definition 2.2 (Invertierbare ARMA-Prozesse). Ein zentrierter schwachstationärer Prozess $\mathbf{X}$ heisst ein $ARMA(p, q)$ -Prozess, wenn ein weisses Rauschen $\mathbf{Z}$ existiert und Polynome

$$\varphi(z) = 1 - \varphi_1 z - \dots - \varphi_p z^p, \quad \theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q, \quad \text{sodass}$$

$$\varphi(B) \mathbf{X} = \theta(B) \mathbf{Z}.$$

Hierbei ist zu fordern, dass $\varphi(z)$ und $\theta(z)$ keine gemeinsamen Nullstellen besitzen und $\varphi(z)$ keine Nullstellen im Einheitskreis besitzt. Wenn auch $\theta(z)$ keine Nullstellen im Einheitskreis besitzt, dann spricht man von einem invertierbaren $ARMA(p, q)$ -Prozess.

Die Bezeichnung 'invertierbar' erklärt sich daraus, dass man in dem Fall, wo das Polynom $\theta(z)$ keine Nullstellen im Einheitskreis besitzt, das weisse Rauschen $\mathbf{Z}$ aus dem ARMA-Prozess zurückgewinnen kann. Und das geht sogar mit einer ARMA-Konstruktion. $\mathbf{Z} = \beta(B) \mathbf{X}$ mit $\beta(z) = \frac{\varphi(z)}{\theta(z)}$.

Definition 2.3. Allgemein nennt man $\mathbf{X}$ einen invertierbaren kausalen Prozess, wenn Potenzreihen $\alpha(z), \beta(z)$ mit summablen Koeffizientenfolgen und ein weisses Rauschen $\mathbf{Z}$ existieren, sodass

$$\mathbf{X} = \alpha(B)\mathbf{X}, \quad \mathbf{Z} = \beta(B)\mathbf{X}.$$

Das Spektralmaß zu einem MA-Prozess haben wir oben allgemein (nicht nur im kausalen Fall) bestimmt. Für die ARMA-Prozesse ergibt sich

Satz 2.4.3. *Das Spektralmaß zum ARMA-Prozess $\mathbf{X}$ besitzt eine Dichte $f(\omega)$, die eine gebrochenrationale Funktion von $e^{-i\omega}$ ist.*

$$d\mu(\omega) = \frac{1}{2\pi} f(\omega) d\omega = \frac{|\tau|^2}{2\pi} \left| \frac{\theta(e^{-i\omega})}{\varphi(e^{-i\omega})} \right|^2 d\omega.$$

Wirklich interessant ist nun das Umgekehrte

Satz 2.4.4. *Sind $p(\omega)$ und $q(\omega)$ trigonometrische Polynome ohne Nullstellen, so existiert zu jedem weissen Rauschen ein invertierbarer kausaler ARMA-Prozesse mit der Spektraldichte*

$$f(\omega) = \left| \frac{p(\omega)}{q(\omega)} \right|^2.$$

Beweis. Wir schreiben $p(\omega) = \sum p_k e^{-ik\omega}$, $q(\omega) = \sum q_k e^{-ik\omega}$ und betrachten einen Spezialfall: Es sei $p_0 \neq 0$ und $p_k = 0$ für $k < 0$ sowie $q_0 \neq 0$ und $q_k = 0$ für $k < 0$; ausserdem nehmen wir an, dass die Polynome $p(z) = \sum p_k z^k$, $q(z) = \sum q_k z^k$ keine Nullstellen im Einheitskreis besitzen. In diesem Spezialfall gewinnt man den gewünschten Prozess $\mathbf{X}$ als die Lösung der Gleichung

$$q(B) \mathbf{X} = p(B) \mathbf{Z}.$$

Wir schreiben in diesem Fall mit $A(z) = \frac{p(z)}{q(z)}$

$$\mathbf{X} = A(B)\mathbf{Z} = p(B) \circ q(B)^{-1}\mathbf{Z}.$$

Das folgende Lemma zeigt, wie man im allgemeinen Fall vorzugehen hat. Man muss die trigonometrischen Polynome durch solche trigonometrische Polynome $\theta(e^{-i\omega})$, $\varphi(e^{-i\omega})$, ersetzen, die keine Nullstellen im Einheitskreis besitzen ohne das Absolutquadrat $\left| \frac{p(\omega)}{q(\omega)} \right|^2$ zu ändern.

Lemma. Zu jedem Polynom $p(z)$ mit $p(e^{-i\omega}) \neq 0$ für alle $\omega \in \mathbb{R}/2\pi$ existiert ein Polynom $\tilde{p}(z)$ ohne Nullstellen im Einheitskreis, sodass

$$|\tilde{p}(e^{-i\omega})|^2 = |p(e^{-i\omega})|^2 \quad \text{für } \omega \in \mathbb{R}/2\pi$$

Beweis. Wenn $\check{p}(z) = c \cdot s^M p(z)$, dann gilt $|\check{p}(e^{-i\omega})|^2 = |c|^2 |p(e^{-i\omega})|^2$. Wir können daher annehmen, dass $p(\cdot)$ die Form hat

$$p(z) = (1 - \zeta_1^{-1}z) \cdots (1 - \zeta_n^{-1}z)$$

mit $|\zeta_j| < 1$ für $j = 1, \dots, r$ und $|\zeta_j| > 1$ für $j = r + 1, \dots, p$.

Wir setzen $\eta_j = \bar{\zeta}_j^{-1}$ für $j = 1, \dots, r$ und $\eta_j = \zeta_j$ für $j = r + 1, \dots, p$. Das Polynom $\tilde{p}(z) = \prod_1^r |\zeta_j|^{-2} \cdot \prod_1^p (1 - \eta_j^{-1}z)$ leistet das Verlangte, denn

$$|1 - \eta_j^{-1}e^{-i\omega}|^2 = |e^{i\omega} - \eta_j^{-1}|^2 = |e^{-i\omega} - \bar{\eta}_j^{-1}|^2 = |e^{-i\omega} - \zeta_j|^2 = |\zeta_j|^2 \cdot |1 - \zeta_j^{-1}e^{-i\omega}|^2$$

Exkurs in die Algebra Mit den hier als ausgezeichnet erschienenen gebrochenrationalen Funktionen wollen wir uns noch etwas näher befassen. Es sei $\mathcal{R}$ die Menge Funktionen von der Gestalt

$$A(z) = \frac{\theta(z)}{\varphi(z)} \quad \text{mit } \theta(z) \cdot \varphi(z) \neq 0 \quad \text{für } |z| \leq 1$$

Mit $A(\cdot)$ liegt auch die reziproke Funktion $\frac{1}{A(\cdot)}$ in $\mathcal{R}$. Mit $A(\cdot), B(\cdot) \in \mathcal{R}$ liegt auch das Produkt $A \cdot B(\cdot)$ in $\mathcal{R}$; eine Linearkombination $\alpha_1 A_1(\cdot) + \dots + \alpha_m A_m(\cdot)$ gehört genau dann zu $\mathcal{R}$, wenn sie in keinem Punkt des Einheitskreises verschwindet.

Wenn $A(0) = 1$, dann gibt es bis auf die Reihenfolge der Faktoren eindeutig bestimmte Produktdarstellungen

$$\begin{aligned} \varphi(z) &= (1 - \zeta_1^{-1}z) \cdots (1 - \zeta_p^{-1}z); & \varphi(z) &= (1 - \eta_1^{-1}z) \cdots (1 - \eta_q^{-1}z) \\ &\text{mit } |\zeta_1|, \dots, |\zeta_p| > 1, & &\text{und } |\eta_1|, \dots, |\eta_q| > 1. \end{aligned}$$

Jedes $A(\cdot) \in \mathcal{R}$ besitzt eine Partialbruchzerlegung. Im Falle, wo der Zählergrad kleiner ist als der Nennergrad, $p < q$, und die ζ_k paarweise verschieden sind, hat sie die Gestalt

$$A(z) = \frac{\alpha_1}{1 - \zeta_1^{-1}z} + \dots + \frac{\alpha_p}{1 - \zeta_p^{-1}z} \quad \text{mit eindeutig bestimmten } \alpha_k.$$

(Wir nehmen uns hier nicht die Mühe, die Formel auch in dem Fall aufzuschreiben, wo das Polynom $\varphi(z)$ mehrfache Nullstellen hat. Das alles sollte hinlänglich bekannt sein.)

Es sei $\mathbf{Z}$ ein weisses Rauschen. Jedem $A(\cdot) \in \mathcal{R}$ können wir einen invertierbaren kausalen Prozess $A(B)\mathbf{Z}$ zuordnen. Die Impulsantworten des Filters $A(B)$ erhält man meistens am bequemsten aus der Partialbruchzerlegung $A(z) = P(z) + \hat{A}(z)$ mit einem Polynom

und einem $\tilde{A}(z)$ mit Zählergrad kleiner als Nennergrad, dann definiert $P(B)$ einen finiten Moving Average und $A(B) = P(B) + \tilde{A}(B)$ einen invertierbaren kausalen Filter mit

$$\begin{aligned} TLF_A = A(B) &= \tau \cdot (I - \zeta_1 B)^{-1} \circ \cdots \circ (I - \zeta_p B)^{-1} \circ (I - \eta_1 B) \circ \cdots \circ (I - \eta_q B) = \\ &= P(B) + \frac{\theta(B)}{\varphi(B)} = P(B) + c_1 (I - \zeta_1 B)^{-1} + \cdots + c_p (I - \zeta_p B)^{-1}. \end{aligned}$$

Die Filter zu den $A(z) \in \mathcal{R}$: Unsere Filter $A(B)$ können natürlich auch als Potenzreihen dargestellt werden. Wenn $A(0) = 1$, dann ist der führende Koeffizient = 1 und wir haben

$$A(B) = I + \alpha_1 B + \alpha_2 B^2 + \cdots \quad \text{mit} \quad \sum |\alpha_j| < \infty$$

Es handelt sich für große h um endliche Linearkombinationen von geometrisch abfallenden Folgen. Wir halten fest, dass jedenfalls solche Koeffizientenfolgen als die Impulsantworten von kausalen invertierbaren Filtern auftreten können.

Es gibt natürlich noch mehr Folgen, die als Influenzfolgen von kausalen invertierbaren Filtern auftreten können. Die $A(B)$ mit $A \in \mathcal{R}$ sind aber eine reiche und bedeutsame bedeutsame Klasse von kausalen invertierbaren TLF. Sie ist (nach Meinung der Fachleute) im Hinblick auf die Anwendungen ausreichend groß. Und sie erlaubt dem Theoretiker, eine realistische Vorstellung zu entwickeln, welche konkreten Möglichkeiten der Konstruktion die Theorie bereitstellt.

Der folgende Satz ist ein Beispiel zur Berechnung der 'guten' Darstellung einer rationalen Leistungstransferfunktion.

Satz 2.4.5. *Gegeben sind Paare komplexer Zahlen (a, b) , (c, d) mit $\frac{|b|}{|a|} < 1$, $\frac{|d|}{|c|} < 1$.*

Es existiert dann ein Paar (A, B) mit $\frac{|B|}{|A|} < 1$, sodass

$$|a + b e^{-i\omega}|^2 + |c + d e^{-i\omega}|^2 = |A + B e^{-i\omega}|^2 = |A|^2 |1 + \frac{B}{A} e^{-i\omega}|^2.$$

Die Zahlen $|A|^2, \frac{B}{A}$ (und damit auch $A \cdot \bar{B}$ und $|A|^2 + |B|^2$) sind eindeutig bestimmt.

Beweis 2.4.1.

$$\begin{aligned} |a + b e^{-i\omega}|^2 &= (a + b e^{-i\omega})(\bar{a} + \bar{b} e^{i\omega}) = |a|^2 + |b|^2 + 2\Re(\bar{a} b e^{i\omega}) \\ |c + d e^{-i\omega}|^2 &= |c|^2 + |d|^2 + 2\Re(\bar{c} d e^{i\omega}) \\ |A + B e^{-i\omega}|^2 &= |A|^2 + |B|^2 + 2\Re(\bar{A} B e^{i\omega}) \end{aligned}$$

Wir bestimmen zuerst $|A|$, $|B|$ aus den Gleichungen

$$\begin{aligned} |A|^2 + |B|^2 &= |a|^2 + |b|^2 + |c|^2 + |d|^2 \\ \frac{1}{2} \left(\frac{|A|}{|B|} + \frac{|B|}{|A|} \right) &= \frac{|A|^2 + |B|^2}{2|A||B|} = \frac{|a|^2 + |b|^2 + |c|^2 + |d|^2}{|2\bar{a}b + 2c\bar{d}|} \end{aligned}$$

Es gibt genau eine Lösung mit $\frac{|B|}{|A|} < 1$ wegen $|2\bar{a}b + 2c\bar{d}| < |a|^2 + |b|^2 + |c|^2 + |d|^2$.

Schliesslich finden wir $A \cdot \bar{B} = \bar{a}b + c\bar{d}$.

Beispiel 2.4.1. Was im Kontinuierlichen das Szenario des gedämpften Oszillators ist, das ist im Diskreten der TLF_A mit

$$\begin{aligned} A(B) &= (I - (\alpha + \beta)B + \alpha\beta B^2)^{-1} = (I - \alpha B)^{-1} \circ (I - \beta B)^{-1} = \\ &= \frac{\alpha}{\alpha - \beta} (I - \alpha B)^{-1} - \frac{\beta}{\alpha - \beta} (I - \beta B)^{-1} = \\ &= \frac{\alpha}{\alpha - \beta} (I + \alpha B + \alpha^2 B^2 + \dots) - \frac{\beta}{\alpha - \beta} (I + \beta B + \beta^2 B^2 + \dots). \end{aligned}$$

Die Transferfunktion $A(e^{-i\omega})$ des Filters TLF_A erhält man, indem wir das Argument B durch die Variable $e^{-i\omega}$ ersetzen. Die Koeffizienten ψ_j , d. h. die Folge der Impulsantworten, erhält man am bequemsten aus der Partialbruchzerlegung. Es ergibt sich

$$A(e^{-i\omega}) = 1 + \frac{1}{\alpha - \beta} (\alpha^2 - \beta^2) e^{-i\omega} + \frac{1}{\alpha - \beta} (\alpha^3 - \beta^3) e^{-i2\omega} + \dots$$

Bei all dem war natürlich vorausgesetzt $\alpha \neq \beta$, $|\alpha| < 1$, $|\beta| < 1$. Der Fall, der dem Oszillator mit schwacher Dämpfung entspricht, ist der, wo α und β konjugiert komplex sind. Im Falle $\alpha = |\alpha| e^{i\lambda} = \bar{\beta}$ haben die Impulsantworten die Gestalt

$$a(h) = \frac{1}{\alpha - \beta} (\alpha^{h+1} - \beta^{h+1}) = |\alpha|^h \frac{\sin((h+1)\lambda)}{\sin \lambda} \quad \text{für } h = 0, 1, \dots$$

ARMA-Filter

Es ist richtig, dass in vielen Bereichen der Theorie der Zeitreihen das weisse Rauschen eine ausgezeichnete Rolle spielt. Bei den Überlegungen zum Begriff des invertierbaren kausalen Filters ist das nicht so. Unseren Hauptsatz können wir auch folgendermaßen formulieren:

Satz 2.4.6.

Wenn eine überall auf $\mathbb{R}/2\pi$ positive Funktion $f(\omega)$ die Gestalt hat $f(\omega) = \left| \frac{p(\omega)}{q(\omega)} \right|^2$ mit trigonometrischen Polynomen $p(\cdot), q(\cdot)$, dann existiert ein invertierbarer kausaler Filter mit der Leistungstransferfunktion $f(\omega)$. Das Spektralmaß des Ausgabeprozesses ergibt sich aus dem Spektralmaß des Eingabeprozesses durch Multiplikation mit $f(\omega)$. Die Transferfunktion ist bis auf einen Faktor der Quotient einseitiger trigonometrischer Polynome mit führendem Koeffizienten 1, die keine Nullstellen im Einheitskreis haben: $A(e^{-i\omega}) = \tau \frac{\theta(\omega)}{\phi(\omega)}$ mit $\theta(\omega) = 1 + \theta_1 e^{-i\omega} + \theta_2 e^{-i2\omega} + \dots$, $\phi(\omega) = 1 - \phi_1 e^{-i\omega} - \phi_2 e^{-i2\omega} + \dots$.

Die so normierten Polynome und die Zahl $|\tau|^2$ sind durch $f(\omega)$ eindeutig bestimmt. Man nennt die Faktorisierung

$$f(\omega) = |\tau|^2 \cdot \Upsilon(e^{-i\omega}) \cdot \overline{\Upsilon(e^{-i\omega})}$$

mit $\Upsilon(e^{-i\omega}) = \frac{\theta(e^{-i\omega})}{\phi(e^{-i\omega})}$ die Wiener-Hopf-Faktorisierung von $f(\omega)$ mit dem kausalen Faktor $\Upsilon(e^{-i\omega}) = \frac{\theta(e^{-i\omega})}{\phi(e^{-i\omega})}$ und dem antikausalen Faktor $\overline{\Upsilon(e^{-i\omega})}$.

Schlussbemerkung:

Wenn $\mathbf{Z}$ weisses Rauschen ist und $(a_n)_n$ eine quadratsummable Zahlenfolge, dann ist für alle n $\sum_{j=0}^{\infty} a_j Z^{n-j}$ eine Zufallsgröße mit endlicher Varianz. Die Konvergenz ist sowohl punktweise (fast-überall) gesichert, als auch im L^2 -Sinn. Auch mit solchen Summen werden wir uns gelegentlich zu befassen haben. Die Sprache der Filter ist da aber nicht angemessen, weil es in diesem Rahmen mit dem Hintereinanderschalten von Operationen Probleme geben kann. Das weisse Rauschen spielt da dann doch eine besondere Rolle.

3 Stationäre Zeitreihen im L^2 -Sinn

Ideen und Konstruktionen aus der Welt der Hilberträume spielen eine wichtige Rolle in der Theorie der zentrierten schwachstationären Zeitreihen. Eine stationäre Zeitreihe $\mathbf{X} = (X^t)_{t \in \mathbb{Z}}$ betrachtet man als eine Folge von Elementen in einem Hilbertraum mit der Eigenschaft, dass das Skalarprodukt nur von der Differenz der Indizes abhängt. Das ist im Sinne einer wichtigen Arbeit von A. N. Kolmogorov: 'Curves in Hilbertspace which are invariant with respect to a one-parameter group of motions' (Doklady 1940).

Wir beginnen mit allgemein bekannten Überlegungen zur Orthogonalisierung im endlichdimensionalen Raum, benützen dabei aber immer wieder auch die Sprache der Zufallsgrößen.

3.1 Finite lineare Prädiktion

Es sei $(X^t)_{t \in \mathbb{Z}}$ eine zentrierte stationäre Zeitreihe zur positivdefiniten Kovarianzfolge $(\gamma(h))_{h \in \mathbb{Z}}$. Wir stellen uns vor, wir hätten die Werte $X^1, \dots, X^{n-1}$ beobachtet und möchten daraus eine Vorhersage von X^{n+m} gewinnen. Dabei beschränken wir uns auf lineare Prädiktoren, und wir versuchen den 'Vorhersagefehler' $\text{var}(X^{n+m} - \sum_1^{n-1} a_j X^{n-j})$ zu minimieren. Der optimale $m+1$ -Schritt Prädiktor ist der Fußpunkt des Lots von X^{n+m} auf den Vektorraum $V_{[1, n-1]}$, der von den Zufallsgrößen $X^1, \dots, X^{n-1}$ aufgespannt wird. Wir bezeichnen das Lot mit $X^{n+m} - \pi_{[1, n-1]} X^{n+m}$.

Im Fall $m=0$ interessieren wir uns für den besten 'Einschritt'-Prädiktor von X^n aufgrund der vorherigen $n-1$ Beobachtungen und die Varianz des Vorhersagefehlers $\sigma_n^2 = \text{var}(X^n - \pi_{[1, n-1]} X^n)$. Um diesen Einschritt-Prädiktor zu berechnen benötigen wir die Gram-Matrix $C^{(n)}$ mit den Einträgen $c_{kl} = \gamma(k-l)$, $1 \leq k, l \leq n$. Es wird sich oft als günstig erweisen, wenn wir $\pi_{[1, n-1]} X^n$ nicht als Linearkombination von $X^1, \dots, X^{n-1}$ darstellen sondern als Linearkombination der Zufallsgrößen $W^1, \dots, W^{n-1}$, die wir mit dem Schmidt'schen Orthogonalisierungsverfahren aus der Folge $X^1, X^2, \dots$ gewinnen. Wir wollen die W^j nicht normieren; das Verfahren führt uns auf W^n mit $\text{var}(W^n) = \sigma_n^2$.

$$\begin{aligned} X^n &= \pi_{[i, n-1]} X^n + W^n, & \ell_{n-1} X^1 + \dots + \ell_{n-1} X^{n-1} + X^n &= W^n \\ \pi_{[i, n-1]} X^n &= -\ell_{n-1} X^{n-1} - \ell_{n-2} X^{n-2} - \dots - \ell_1 X^1 \\ &= +\chi_1^{(n)} W^{n-1} + \chi_2^{(n)} W^{n-2} + \dots + \chi_{n-1}^{(n)} W^1 \end{aligned}$$

Die Idee ist, dass die Koeffizienten für $n \rightarrow \infty$ konvergieren, und zwar $\chi_j^{(n)} \rightarrow \chi_j$ schneller als $-\ell_{n-j} \rightarrow \psi_j$ für festes j . Mit den χ_j haben wir eine brauchbare Lösung für die Aufgabe der zeitnahen Prädiktion. Darunter verstehen wir die Aufgabe:

$$\text{Schätze } X^t \text{ zur Zeit } t-1, \quad X^{t+1} \text{ zur Zeit } t, \quad X^{t+2} \text{ zur Zeit } t+1, \dots$$

Man produziere also die Folge der Einschrittprädiktoren. Für diese Aufgabe ist es hilfreich, wenn man die Prädiktionsfehler aus früheren Zeiten gespeichert hat, $W^n = \pi_{[1, n-1]} X^n$;

denn ein Anfangsstück der Reihe $\sum_{j=1} \chi_j W^{t-j}$ sollte für nicht zu kleine t eine gute Approximation von $\pi_{[1,t-1]} X^t$ sein. Die Folge der optimalen m -Schrittprädiktoren ergibt sich aus der abkürzten Reihe $\pi_{[1,t-m]} X^t \sim \sum_{j=m} \chi_j W^{t-j}$.

Diese Aussagen zu erhärten ist das Ziel dieses Abschnitts über finite Prädiktion.

Exkurs zur Erinnerung: Das sog. Gauss'sche Eliminationsverfahren aus der Anfängervorlesung kann u. a. dazu verwendet werden, eine nichtsinguläre $n \times n$ -Matrix C (nach einer geeigneten Umordnung der Reihen) als Produkt von Dreiecksmatrizen zu schreiben. $C = L^{-1}M$. (LU-Zerlegung, oder 'lower-upper factorization') Die Faktorisierung ist eindeutig, wenn wir fordern, dass die Diagonalelemente der unteren Dreiecksmatrix L allesamt $= 1$ sind. Die Eindeutigkeit ist schnell bewiesen:

$L_1^{-1}M_1 = C = L_2^{-1}M_2$ impliziert $L_2L_1 = M_2M_1^{-1}$. Diese Matrix ist einerseits eine obere Dreiecksmatrix und andererseits eine untere Dreiecksmatrix mit Einsen in der Diagonale; sie ist also die Einheitsmatrix.

Die obere Dreiecksmatrix M berechnet man durch eine Folge von sog. elementaren Zeilenumformungen aus der Matrix C . Von jeder Zeile werden Linearkombinationen früherer Zeilen subtrahiert, sodass nach und nach Nullen unterhalb der Diagonalen entstehen. In der Anfängervorlesung erzeugt man (meistens) zuerst Nullen in der ersten Spalte, dann in der zweiten usw. Das wird für uns nicht günstig sein; wir wollen lieber Zeile für Zeile abarbeiten:

Im ersten Schritt wird von der zweiten Zeile ein Vielfaches der ersten subtrahiert; und wenn wir im $(k-1)$ -ten Schritt eine Matrix gewonnen haben, die in den ersten k Zeilen links von der Diagonalen Nullen hat, dann subtrahieren wir im k -ten Schritt die passende Linearkombination dieser Zeilen von der $(k+1)$ -ten Zeile. Dieser k -te Schritt wird offenbar durch eine untere Dreiecksmatrix E_k bewerkstelligt, die ausserhalb der Diagonalen nur in der k -ten Zeile Einträge $\neq 0$ haben kann. Das Produkt $E_{n-1} \cdots E_2 \cdot E_1$ ist die gesuchte untere Dreiecksmatrix L . $L \cdot C = M$ ist die damit erzeugte obere Dreiecksmatrix.

Ist C eine positivdefinite Matrix, so sind keine Zeilenvertauschungen nötig. Die Diagonalelemente der oberen Dreiecksmatrix M seien $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$. Wenn Σ_n^2 die Diagonalmatrix mit diesen Einträgen bezeichnet, dann gilt $M = \Sigma_n^2 \cdot (L^{-1})^*$. Das ergibt sich aus $C = C^*$ und der Eindeutigkeit der LU-Faktorisierung. $C = L^{-1} \cdot \Sigma_n^2 \cdot (L^{-1})^* = M^* \cdot \Sigma_n^{-2} \cdot M$.

Die Ausdrucksweise der linearen Prädiktion. Wir beginnen mit einem n -Tupel von Zufallsgrößen, welches wir als eine Spalte schreiben $\mathbf{X} = (X^1, X^2, X^3, \dots, X^n)^T$. Wir setzen $W^1 = X^1$ und gewinnen im ersten Schritt das Tupel $(W^1, W^2, X^3, \dots, X^n)^T$, indem wir von X^2 dasjenige Vielfache von W^1 subtrahieren, welches $W^2 = X^2 - cW^1$ orthogonal zu W^1 macht. Im k -ten Schritt wird aus dem Tupel $(W^1, \dots, W^k, X^{k+1}, \dots, X^n)^T$ das Tupel $(W^1, \dots, W^k, W^{k+1}, \dots, X^n)^T$. Im $(n-1)$ -ten Schritt gewinnen wir schliesslich das orthogonalisierte Tupel der W . Die Zufallsgröße W^k heisst die **Innovation** zur Zeit k . Es gilt $\text{span}\{W^1, \dots, W^k\} = \text{span}\{X^1, \dots, X^k\}$ für alle k .

$$\begin{pmatrix} W_1 \\ \vdots \\ W^{n-1} \\ W^n \end{pmatrix} = E_{n-1} \begin{pmatrix} W^1 \\ \vdots \\ W^{n-1} \\ X^n \end{pmatrix} = E_{n-1} \cdot E_{n-2} \begin{pmatrix} W^1 \\ \vdots \\ X^{n-1} \\ X^n \end{pmatrix} = \cdots = L_n \begin{pmatrix} X^1 \\ \vdots \\ X^{n-1} \\ X^n \end{pmatrix}$$

Wir bemerken: Die $n \times n$ - Matrix $E_{k-1} \cdots \cdots E_2 \cdot E_1$ hat die Gestalt

$$E_{k-1} \cdots \cdots E_2 \cdot E_1 = \begin{pmatrix} L_k & 0 \\ 0 & I_{n-k} \end{pmatrix}$$

wo L_k eine normierte untere Dreiecksmatrix vom Format $k \times k$ ist und I_{n-k} die Einheitsmatrix vom Format $(n-k) \times (n-k)$. (Von einer normierten Dreiecksmatrix sprechen wir, wenn die Diagonaleinträge = 1 sind.)

Es sei C_n die Kovarianzmatrix des n -Tupels X . Die Matrix $L_n = E_{n-1} \cdots \cdots E_1$ ist die normierte untere Dreiecksmatrix, welche das Produkt $L \cdot C$ zu einer oberen Dreiecksmatrix macht. $L \cdot C = M = \Sigma^2 \cdot L^{-1*}$. Verfolgen wir die Entwicklung der beteiligten Matrizen im Laufe der Diagonalisierungsschritte. Die Kovarianzmatrix des n -Tupels $(W^1, \dots, W^k, X^{k+1}, \dots, X^n)^T$ ist

$$\begin{pmatrix} L_k & 0 \\ 0 & I_{n-k} \end{pmatrix} \cdot C_n \cdot \begin{pmatrix} L_k^* & 0 \\ 0 & I_{n-k} \end{pmatrix}$$

Wir bemerken, dass die ersten k Zeilen der Dreiecksmatrizen L und $L \cdot C = M$ nach dem k ten Orthogonalisierungsschritt nicht mehr verändert werden.

In der Sprache der linearen Prädiktion können wir sagen: Der beste lineare Prädiktor für X^n ist der Fußpunkt des Lots auf den Raum $\text{span}\{X^1, \dots, X^{n-1}\}$. Der Prädiktionsfehler für X^n aufgrund der $(n-1)$ -Vergangenheit ist die Innovation $W^n = X^n - \pi_{\{1,2,\dots,n-1\}} X^n$. Den Prädiktor können wir sowohl als Linearkombination der X^j , $1 \leq j \leq n-1$ als auch als Linearkombination der W^k , $1 \leq k \leq n-1$ darstellen. Wir brauchen dazu die n -te Zeile von L und die n -te Spalte von M sowie die Diagonalelemente $\sigma_1^2, \dots, \sigma_{n-1}^2$.

$$X^n - W^n = \pi_{[1,2,\dots,n-1]} X^n = - \sum_1^{n-1} \ell^n_j X^j = \sum_1^{n-1} \bar{m}_n^k \sigma_k^{-2} W^k.$$

Dabei ist $\{W^1, W^2, \dots, W^m\}$ für jedes $m \leq n$ eine Orthogonalbasis von $\text{span}\{X^1, \dots, X^m\}$. Es gilt $\text{cov}(W, W) = L \cdot C \cdot L^* = \Sigma^2$.

Wir denken jetzt an eine lange Folge $X^1, \dots, X^n, X^{n+1}, \dots, X^N$ und studieren Mehrschritt-Prädiktoren aufgrund der n -Vergangenheit $\text{span}\{X^1, \dots, X^n\}$. Für die Prädiktion von X^k für die $k \in \{n+1, \dots, N\}$ brauchen wir die $n \times N$ -Matrix $C_{n \times N}$, die aus den ersten n Zeilen der großen Kovarianzmatrix besteht und die gemäß der Zeilenoperationen nach L_n umgeformte Matrix $M_{n \times N} = L_n \cdot C_{n \times N}$. Der vordere Teil von M ist eine obere Dreiecksmatrix mit Diagonaleinträgen $\sigma_1^2, \dots, \sigma_n^2$. Die weiteren Einträge sind $m_{lk} = \text{cov}(W^l, X^k)$

$$\text{cov} \left(\begin{pmatrix} W^1 \\ \vdots \\ W^n \end{pmatrix}, \begin{pmatrix} X^1 \\ \vdots \\ X^N \end{pmatrix} \right) = \text{cov} \left(L_n \begin{pmatrix} X^1 \\ \vdots \\ X^N \end{pmatrix}, \begin{pmatrix} X^1 \\ \vdots \\ X^N \end{pmatrix} \right) = L \cdot C_{n \times N} = M_{n \times N}.$$

Wir halten fest:

Satz 3.1.1. *Es sei $\mathbf{X}$ ein Tupel linear unabhängiger zentrierter Zufallsgrößen mit der Kovarianzmatrix $\text{cov}(\mathbf{X}, \mathbf{X}) = C$. Es sei L die normierte untere Dreiecksmatrix, für welche $M = L \cdot C$ eine obere Dreiecksmatrix ist. Σ^2 sei die Diagonalmatrix mit den Diagonaleinträgen aus M . Dann ist $L \cdot \mathbf{X} = \mathbf{W}$ ein Tupel unkorrelierter Zufallsgrößen mit $\text{cov}(\mathbf{W}, \mathbf{W}) = \Sigma^2$ und es gilt $\mathbf{X} = M^* \cdot \Sigma^{-2} \cdot \mathbf{W}$.*

Wir wollen das Vorhersagen von X^N aufgrund der länger zurückliegenden Vergangenheiten $\text{span}\{X^1, \dots, X^n\}$ nochmals als dynamisches Problem formulieren. Zu den Zeitpunkten $1, 2, \dots$ kommen Daten X^n , die sich vom eben vorhergesagten Wert $\pi_{[1, \dots, n-1]} X^n$ unterscheiden, und zwar um $W^n = X^n - \pi_{[1, \dots, n-1]} X^n$. Die Werte dieser W^n sollten wir festhalten; Linearkombinationen dieser W^n eignen sich nämlich deswegen in besonderer Weise zur Darstellung der Prädiktoren späterer X^N , weil die W^n eine Folge unkorrelierter Zufallsgrößen bilden. Für $n < N$ haben wir $\pi_{[1, \dots, n]} X^N = b_1^N W_1 + b_2^N W_2 + \dots + b_n^N W_n$ mit $b_l^N = \sigma_l^{-2} \text{cov}(X^N, W^l)$. Den Koeffizienten b_n^N können wir bereits nach dem n -ten Umformungsschritt der Kovarianzmatrix ablesen.

Alle diese Überlegungen gelten für jede Folge von linear unabhängigen Beobachtungen. Im dem Fall, dass die Beobachtungen aus einer stationären Zeitreihe kommen, sind die Kovarianzmatrizen sog. **Töplitzmatrizen**; in jeder Nebendiagonalen stehen identische Einträge $c_{kl} = c(k-l)$. Die Folge $\sigma_n^2 = \text{var } W^n$ ist abnehmend und man darf in guten Fällen erwarten, dass die Einträge in jeder Nebendiagonalen konvergieren. Im nächsten Unterabschnitt werden wir sehen, dass das tatsächlich so ist in den für die Stochastik interessanten Situationen, bei den sog. regulären Zeitreihen. Die σ_n^2 fallen gegen einen Wert $\sigma_\infty^2 > 0$. (Man könnte diese Zahl die (asymptotische) Varianz der Einschnitt-Innovation nennen.) Wir werden lernen, wie sich der Limes aus der Spektraldichte berechnet: $\ln \sigma_\infty^2 = \frac{1}{2\pi} \int \ln f(\omega) d\omega$. Wir bleiben hier aber noch beim finiten Fall. Wir wollen kurz diskutieren, wieviel verloren geht, wenn man die Prädiktion von X^N nur auf die Vergangenheit $\text{span}\{X^{N-1}, X^{N-2}, \dots, X^{N-k}\}$ gründen will. Die entscheidende Größe ist der sog. **partielle Autokorrelationskoeffizient** zum Zeitabstand k : $\rho_1 = \text{Corr}(X^2, X^1) = \frac{\text{cov}(X^2, X^1)}{\sqrt{\text{var} X^2 \cdot \text{var} X^1}}$ und für $k = 2, 3, \dots$

$$\rho_k = \text{Corr}(X^{n+1} - \pi_{\{2,3,\dots,n\}} X^{n+1}, X^1 - \pi_{\{2,3,\dots,n\}} X^1).$$

Der Abstand von X^{n+1} zum Raum $\text{span}\{X^2, \dots, X^n\}$ ist größer als der Abstand zum größeren Raum $\text{span}\{X^1, X^2, \dots, X^n\}$. Die Abstandskvadrat sind σ_n^2 bzw. $\sigma_{n+1}^2 = \text{var} W^{n+1}$, die Varianz der Innovation zur Zeit $n+1$. Wir zeigen jetzt $\sigma_{n+1}^2 = \sigma_n^2 (1 - |\rho_n|^2)$.

Die Fußpunkte der Lote $W^{n+1} = X^{n+1} - \pi_{[1,2,\dots,n]}$ und $X^{n+1} - \pi_{[2,\dots,n]}$ unterscheiden sich um ein Vielfaches von $X^1 - \pi_{[2,\dots,n]} X^1$.

$$(X^{n+1} - \pi_{[1,2,\dots,n]} X^{n+1}) = (X^{n+1} - \pi_{[2,\dots,n]} X^{n+1}) - \alpha \cdot (X^1 - \pi_{[2,\dots,n]} X^1)$$

$$\text{mit } \alpha \text{ so, dass } (X^{n+1} - \pi_{[2,\dots,n]} X^{n+1}) - \alpha \cdot (X^1 - \pi_{[2,\dots,n]} X^1) \perp (X^1 - \pi_{[2,\dots,n]} X^1).$$

Das ergibt $\alpha = \rho_n$ und $\sigma_{n+1}^2 = \sigma_n^2 - |\alpha|^2 \sigma_n^2$.

Bei den regulären Prozesse gilt $\lim \sigma_n^2 > 0$ und die Prädiktoren aufgrund finiter Vergangenheit konvergierengenegen die Prädiktoren aufgrund der vollen Vergangenheit. Zur Bestimmung der partiellen Autokorrelationen bemerken wir noch

Lemma. Es seien $X^0, X^1, \dots, X^n, X^{n+1}$ irgendwelche Zufallsgrößen und W^{n+1}, W^0 die Lote auf $\text{span}\{X^1, \dots, X^n\}$, also $W^{n+1} = X^{n+1} - \pi_{[1,n]}X^{n+1}$, $W^0 = X^0 - \pi_{[1,n]}X^0$. Dann gilt für die 'partielle Kovarianz zur Zeitverzögerung' n

$$\text{cov}_{[1,n]}(X^{n+1}, X^0) = \text{cov}(W^{n+1}, X^0)$$

In der Tat steht W^{n+1} senkrecht auf $\pi_{[1,n]}X^0$.

Beispiel 3.1.1. (Autoregressive Prozesse) Nehmen wir an, dass der beste lineare Prädiktor eines X^n bzgl. der gesamten Vergangenheit $\text{span}\{X^{n-1}, X^{n-2}, \dots\}$ im p -dimensionalen Vektorraum $\text{span}\{X^{n-1}, \dots, X^{n-p}\}$ liegt. Es existieren dann also Koeffizienten $\phi_1, \dots, \phi_p$, sodass

$$W^n = (X^n - \phi_1 X^{n-1} - \dots - \phi_p X^{n-p}) \perp \text{span}\{X^{n-1}, X^{n-2}, \dots\}$$

Wenn das für ein n gilt, dann für alle. Die W^n sind also weisses Rauschen. Die Zeitreihe $\mathbf{X}$ ist ein AR(p)-Prozess.

Die allgemeine Formel $W^t = X^t - \pi_{[1,t-1]}X^t = X^t + \sum_1^{t-1} \ell_{t-k}^t X^{t-k}$ nimmt hier für $t > p$ die Form an $W^t = X^t - \sum_1^p \phi_k X^{t-k}$ mit $\phi_k = -\ell_{t-k}^t$.

Man könnte sagen: Die Matrix L ist eine untere Dreiecksmatrix der Dicke p ; denn es gilt $\ell_l^n = 0$ für $n - l > p$. Wenn man die ersten p Zeilen weglässt, dann entsteht eine normierte Töplitz-Matrix mit den Einträgen $\phi_1, \dots, \phi_p$ in den unteren Nebendiagonalen.

Beispiel 3.1.2 (Der MA(1)-Prozess). Es sei $\mathbf{Z}$ ein weisses Rauschen mit der Intensität 1 und $\mathbf{X} = (I + \theta B)\mathbf{Z}$, d. h. $X^n = Z^n + \theta Z^{n-1}$. Die Kovarianzfolge ist

$$\gamma_0 = \text{cov}(X^n, X^n) = 1 + |\theta|^2, \quad \gamma_1 = \theta = \overline{\gamma_{-1}}, \quad \gamma_h = 0 \quad \text{für } |h| > 1$$

Die Kovarianzmatrix hat also in der Hauptdiagonalen den Eintrag $1 + |\theta|^2$, in der ersten unteren Nebendiagonalen den Eintrag θ , in der ersten oberen Nebendiagonalen den Wert $\bar{\theta}$ und sonst überall nur Nullen. Die Diagonalisierungsschritte gestalten sich sehr einfach; im n -ten Schritt wird von der $(n+1)$ -ten Zeile ein Vielfaches der n -ten Zeile subtrahiert. Die Einträge in den oberen Nebendiagonalen bleiben unverändert. In der Diagonalen entsteht eine nach 1 absteigende Folge $(\sigma_n^2)_n$,

$$\sigma_1^2 = 1 + |\theta|^2, \quad \sigma_2^2 = 1 + |\theta|^2 - \sigma_1^{-2}|\theta|^2, \quad \dots \quad \sigma_{n+1}^2 = 1 + |\theta|^2 - \sigma_n^{-2}|\theta|^2.$$

Für die Prädiktion von X^{n+1} interessiert die $(n+1)$ -te Spalte. Es ergibt sich

$$\pi_{[1,n]}X^{n+1} = \sigma_n^{-2} \theta W^n = \sigma_n^{-2} \theta (X^n - \hat{X}^n).$$

Die Konvergenz $\sigma_n^2 \searrow 1$ ist recht schnell. Nach kurzem Anlauf schätzen wir also

$$\hat{X}^{n+1} \sim \theta (X^n - \hat{X}^n).$$

Es wäre nicht so schön, wenn man versuchte, den optimalen linearen Schätzer als Linearkombination der X_t mit $t \leq n$ darzustellen: $\pi_{[1,n]}X^{n+1} = \phi_1^{(n)} X^n + \phi_2^{(n)} X^{n-1} + \dots$.

Wir diskutieren noch kurz das finite Prädiktionsproblem für MA(q)-Prozesse.

Beispiel 3.1.3. Es sei $\mathbf{Z}$ ein normiertes ($\sigma^2 = 1$) weisses Rauschen und

$$X^t = Z^t + \theta_1 Z^{t-1} + \dots + \theta_q Z^{t-q}$$

$\mathbf{X} = \Theta(B)\mathbf{Z}$ ist also ein kausaler MA(q)-Prozess mit $\text{var}X^t = 1 + |\theta_1|^2 + \dots + |\theta_q|^2$. Es sollen die θ_j im Laufe des Orthogonalisierungsverfahrens approximativ bestimmt werden. Wir betrachten die Kovarianzmatrix C der Folge $(X^1, X^2, \dots)$. Es handelt sich um eine hermitesche Töplitz-Matrix vom Format $\mathbb{N} \times \mathbb{N}$ mit der 'Dicke' $1 + 2q$; denn es gilt $\text{cov}(X^s, X^t) = 0$ für $|s - t| > q$. Uns interessieren die Matrizen M_n die bei der Diagonalisierung nach und nach entstehen. Die Nebendiagonalen im Abstand $> q$ bleiben leer; $m_N^n = 0$ für $N > n + q$. Aus der allgemeinen Formel $X^t - W^t = \sum_1^{t-1} \bar{m}^{t-j}_t \sigma_{t-j}^{-2} W^{t-j}$ wird mit dem normierten weissen Rauschen $Y^t = \frac{1}{\sigma_t} W^t$ die Darstellung

$$\frac{1}{\sigma_t} X^t = Y^t + \sum_1^q \theta_j^{(t)} Y^{t-j}$$

Die endliche Summe $\sum_1^q \theta_j^{(t)} Y^{t-j}$ ist der beste lineare Prädiktor für $\frac{1}{\sigma_t} X^t$. Y^t ist das auf Länge 1 normierte Lot von X^t auf den Raum $\text{span}\{X^{t-1}, X^{t-2}, \dots, X^2, X^1\}$

Es unterscheidet sich für große t wenig vom normierten Lot auf die unendliche Vergangenheit $\text{span}\{X^{t-1}, X^{t-2}, \dots\}$. Es gilt $\text{var}\left(\frac{1}{\sigma_t} X^t\right) = 1 + \sum |\theta_j^{(t)}|^2$ und $\sigma_t^2 \searrow 1$. Die Koeffizienten

$$\theta_j^{(t)} = \text{cov}\left(\frac{1}{\sigma_t} X^t, Y^{t-j}\right) = \frac{1}{\sigma_{t-j} \cdot \sigma_t} \cdot \bar{m}^{t-j}_t$$

die wir aus den transformierten Matrizen M_t ablesen können, konvergieren. $\theta_j^{(t)} \rightarrow \theta_j$. Unsere Darstellung von $\frac{1}{\sigma_t} X^t$ durch das weisse Rauschen $\mathbf{Y}$ wird also für große t immer ähnlicher zur ursprünglichen Konstruktion der Zeitreihe $X^t = Z^t + \theta_1 Z^{t-1} + \dots + \theta_q Z^{t-q}$.

Man beachte aber: Die Werte der Zufallsgrößen $Y^t = \frac{1}{\sigma_t} W^t$ für $t = 1, 2, \dots$ ergeben sich aus den Beobachtungen $X^1, X^2, \dots$, während die Werte der Zufallsgrößen Z^t , die den Prozess bestimmt haben, nicht aus den beobachteten $X^1, X^2, \dots$ ermittelt werden können. Die Koeffizienten σ_t^2 und $\theta_j^{(t)}$ können aus der Kovarianzfolge berechnet werden. Sie liefern im Grenzwert $t \rightarrow \infty$ die θ_j aus dem 'Filter' $\Theta(B)$.

Verallgemeinerung: Wie im Beispiel sollte man sich das Verfahren auch bei allgemeineren (im Sinne der Wold-Zerlegung) regulären Zeitreihen vorstellen: Wenn die Zeitreihe $\mathbf{X}$ durch ein unendliches kausales gleitendes Mittel entstanden ist,

$$X^t = Z^t + \sum_1^{\infty} \chi_j Z^{t-j} \quad \text{mit einem normierten weissen Rauschen } \mathbf{Z},$$

dann ergeben sich die Z^t im Wesentlichen als die auf Länge 1 normierten Lote von X^t auf die Vergangenheit bis $t - 1$. So kann man sagen, dass der Orthogonalisierungsalgorithmus aus der Kovarianzfolge approximativ die Koeffizienten χ_j errechnet.— Im nächsten Abschnitt behandeln wir solchen Fragen in einer grundsätzlichen Weise. Dort werden wir auch diskutieren, wie man die Funktion $\Upsilon(e^{-i\omega}) = 1 + \sum_1^{\infty} \chi_j e^{-ij\omega}$ durch eine sog. Wiener-Hopf-Faktorisierung aus der Spektraldichte einer regulären Zeitreihe $\mathbf{X}$ gewinnt: $|\Upsilon(e^{-i\omega})|^2 = f(\omega)$.

3.2 Unendliche Vergangenheit, reguläre Zeitreihen

Die stationäre Zeitreihe $\mathbf{X} = (X^t)_{t \in \mathbb{Z}}$ betrachten wir als eine Folge von Elementen in einem Hilbertraum. Den von den X^t aufgespannten Teil-Hilbertraum bezeichnen wir mit $\mathcal{H}(\mathbf{X})$. Den von $\{X^t : t \leq n\}$ aufgespannten Raum bezeichnen wir $\mathcal{H}_n^{-\infty}(\mathbf{X})$. Das ist die lineare(!) Vergangenheit bis zur Zeit n . Weiter sei $\mathcal{S}(\mathbf{X}) = \bigcap_n \mathcal{H}_n^{-\infty}(\mathbf{X})$ und $\mathcal{R}(\mathbf{X})$ das orthogonale Komplement. Die orthogonale Projektion auf $\mathcal{H}_n^{-\infty}(\mathbf{X})$ bezeichnen wir π_n ; $\pi_{-\infty}$ sei die orthogonale Projektion auf $\mathcal{S}(\mathbf{X})$. Für jedes Element $Y \in \mathcal{H}(\mathbf{X})$ gilt offenbar $\pi_n(Y) \rightarrow \pi_{-\infty}(Y)$ wenn $n \rightarrow -\infty$. Der Einschnitt-Prädiktor für X^0 ist der Fußpunkt $\pi_{-1}(X^0)$ des Lots von X^0 auf $\mathcal{H}_{-1}^{-\infty}(\mathbf{X})$. Die Zufallsgröße $Z^t = X^t - \pi_{t-1}X^t$ nennen wir die Innovation zur Zeit t .

Für jede stationäre Zeitreihe haben wir die direkte Zerlegung $\mathcal{H}(\mathbf{X}) = \mathcal{S}(\mathbf{X}) \oplus \mathcal{R}(\mathbf{X})$. Eine stationäre Folge $\mathbf{X}$ heisst regulär, wenn $\mathcal{H}(\mathbf{X}) = \mathcal{R}(\mathbf{X})$ und singulär, wenn $\mathcal{H}(\mathbf{X}) = \mathcal{S}(\mathbf{X})$. Man kann leicht zeigen, dass für jede stationäre Zeitreihe $\mathbf{X}$ die Zeitreihe $\mathbf{X}_{sing} = (\pi_{-\infty}X^t)_{t \in \mathbb{Z}}$ singulär und die Zeitreihe $\mathbf{X}_{reg} = (X^t - \pi_{-\infty}X^t)_{t \in \mathbb{Z}}$ regulär ist. Dem Stochastiker erscheint es als ein ausgearteter Fall, wenn $\pi_{-1}(X^0) = X^0$. Wenn man X^0 durch Linearkombinationen der X^t mit $t \leq -1$ beliebig gut approximieren kann, dann kann man (wegen der Stationarität) auch X^{-1} durch Linearkombinationen der X^t mit $t \leq -2$ beliebig gut approximieren. Somit kann man X^0 durch Linearkombinationen der X^t mit $t \leq -2$ beliebig gut approximieren, usw. X^0 liegt in jedem $\mathcal{H}_n^{-\infty}(\mathbf{X})$ und somit in $\mathcal{S}(\mathbf{X})$. Die Zeitreihe ist singulär. Eine reguläre Zeitreihe ist dadurch gekennzeichnet, dass sie keinen singulären Anteil besitzt. Etwas mehr über die (üblicherweise nach H. Wold benannten) orthogonale Zerlegung

$$\mathbf{X} = \mathbf{X}_{reg} \oplus \mathbf{X}_{sing}$$

findet man z. B. im Lehrbuch *A. N. Shiriyayev: Probability, Springer-Verlag 1995*

Beispiel. Es sei Z eine komplexgaussische Zufallsgröße mit $\mathfrak{E}(Z) = 0$, $\text{var}Z = 1$, $\lambda \in \mathbb{R}/2\pi$, $X^n = Z \cdot e^{in\lambda}$. Die Zeitreihe $\mathbf{X}$ ist singulär. Das Spektralmaß ist das δ -Maß im Punkt λ .

Wir wollen uns im Folgenden nur mit regulären Zeitreihen befassen. Wir werden sehen, dass das Spektralmaß einer regulären Zeitreihe absolutstetig ist bzgl. des Lebesgue-Maßes. $d\mu(\omega) = f(\omega) \frac{1}{2\pi} d\omega$. Genauer gilt: Eine Zeitreihe mit der Spektraldichte $f(\omega)$ ist genau dann regulär, wenn die Spektraldichte $\ln f(\omega)$ integrabel ist. — Dies ist ein relativ schwieriger Satz, der üblicherweise mit Hilfe von relativ diffizilen Aussagen aus der klassischen Funktionentheorie bewiesen wird, den wir hier aber direkt beweisen wollen.

Satz 3.2.1. *Wenn $\mathbf{X}$ regulär ist, dann ist die Folge der Innovationen $(Z^t)_t$ ein weißes Rauschen mit einer gewissen Intensität $\sigma_\infty^2 > 0$. Es existiert eine quadratsummable Folge $(1, \chi_1, \chi_2, \dots)$, sodass $X^t = Z^t + \sum_1^\infty \chi_j Z^{t-j}$. Für die von der Folge der Innovationen aufgespannten Hilberträume $\mathcal{H}_n^{-\infty}(\mathbf{Z}) = \text{span}\{Z^n, Z^{n-1}, \dots\}$ gilt $\mathcal{H}_n^{-\infty}(\mathbf{Z}) = \mathcal{H}_n^{-\infty}(\mathbf{X})$. Die Spektraldichte besitzt die Darstellung*

$$f(\omega) = \sigma_\infty^2 \cdot |Y(e^{-i\omega})|^2 \quad \text{mit} \quad Y(e^{-i\omega}) = 1 + \sum_1^\infty \chi_j e^{-ij\omega}$$

Beweis. Wenn $[Z^n]$ den von Z^n aufgespannten eindimensionalen Vektorraum bezeichnet, dann haben wir für jedes $k > 0$

$$\mathcal{H}_n^{-\infty}(\mathbf{X}) = \mathcal{H}_{n-k}^{-\infty}(\mathbf{X}) \oplus [Z^{n-k+1}] \oplus \dots \oplus [Z^{n-1}] \oplus [Z^n].$$

$$X^n = Z^n + \sum_1^{k-1} \chi_j Z^{n-j} + \pi_{n-k}(X^n) \quad \text{mit } \chi_j = \sigma_\infty^{-2} \cdot \langle Z^{n-j}, X^n \rangle \text{ für } j \geq 1.$$

Aus der Regularität ergibt sich $\lim_k \|\pi_{n-k}(X^n)\|^2 = 0$. Die Innovationen $\{Z^t : t \leq n\}$ spannen den Hilbertraum $\mathcal{H}_n^{-\infty}(\mathbf{X})$ auf.

Die Folge $(\chi_1, \chi_2, \dots)$ nennen wir die Folge der Nachwirkungen einer Innovation. Manchmal notieren wir auch $\chi_0 = 1$ und $\chi_k = 0$ für $k < 0$, sowie $Y(e^{-i\omega}) = \sum \chi_k e^{-ik\omega} = 1 + \sum_1^\infty \chi_j e^{-ij\omega}$. Die quadratintegrale Funktion $Y(e^{-i\omega})$ nennen wir die Wold-Funktion der regulären Zeitreihe. Es gilt

$$\gamma_X(h) = \langle X^h, X^0 \rangle = \left\langle \sum_j \chi_j Z^{h-j}, \sum_k \chi_k Z^k \right\rangle = \sigma_\infty^2 \sum_k \chi_{h+k} \bar{\chi}_k,$$

$$\frac{1}{2\pi} \int e^{ih\omega} \sigma_\infty^2 Y(e^{-i\omega}) \overline{Y(e^{-i\omega})} d\omega = \sigma_\infty^2 \frac{1}{2\pi} \int \sum_j e^{i(h-j)\omega} \chi_j \cdot \sum_k e^{ik\omega} \bar{\chi}_k d\omega = \gamma_X(h).$$

Man kann leicht zeigen (siehe Shiryaev)

Satz 3.2.2. Eine Zeitreihe $\mathbf{X}$ mit $\text{var}X^t = \sigma^2 > 0$ ist genau dann regulär, wenn ein weisses Rauschen $\mathbf{W}$ und eine quadratsummable Folge (χ_j) existieren, sodass für alle n gilt $\mathcal{H}_n^{-\infty}(\mathbf{W}) = \mathcal{H}_n^{-\infty}(\mathbf{X})$ und $X^n = Z^n + \sum_0^\infty \chi_j W^{n-j}$.

Warnung: Wenn $\mathbf{W}$ ein weisses Rauschen ist und $(\chi_0, \chi_1, \dots)$ eine quadratsummable Folge, dann bedeutet das noch nicht, dass $\mathbf{X} = (\chi_0 I + \chi_1 B + \chi_2 B^2 + \dots) \mathbf{W}$ eine reguläre Zeitreihe ist. Es kann sehr wohl sein, dass die so gebildete Zeitreihe einen singulären Anteil besitzt. Wichtig ist, dass der Einschritt-Prädiktionsfehler positive Varianz $\sigma_\infty^2 > 0$ besitzt. Wenn das gesichert ist, dann kann man das weisse Rauschen leicht so modifizieren, dass $\mathcal{H}_n^{-\infty}(\mathbf{W}) = \mathcal{H}_n^{-\infty}(\mathbf{X})$ für alle n .

Zur Einordnung: Mehrere bedeutende Mathematiker haben sich in der Mitte des 20.ten Jahrhunderts für die Frage interessiert, wie man diejenigen Spektralmaße charakterisieren kann, die auf eine regulären Zeitreihe führen, und wie man gegebenenfalls aus der Spektraldichte die Varianz des Einschrittprädiktors σ_∞^2 findet. Sie haben schliesslich die folgende überzeugende Antwort gefunden:

Eine notwendige Bedingung ist die, dass das Spektralmaß totalstetig ist, $d\mu(\omega) = f(\omega) \frac{1}{2\pi} d\omega$. Es kommen aber nicht alle Spektraldichten $f(\omega)$ in Betracht. Es ist zu fordern, dass der Logarithmus der Spektraldichte Lebesgue-integrierbar ist. Die Spektraldichte darf sich also nirgends allzu stark an die Null anschmiegen. Wenn die Bedingungen an das Spektralmaß erfüllt sind, dann ist die dazugehörige Zeitreihe regulär und die Varianz des Einschritt-Prädiktors ist $\sigma_\infty^2 = \exp\left(\frac{1}{2\pi} \int \ln f(\omega) d\omega\right)$. ('Formel von Szegő-Krein-Kolmogorov')

Der flüchtige Leser könnte die Aussage der Wold-Entwicklung dahingehend (miss)verstehen, dass er sich vorstellt, jede reguläre Zeitreihe (und somit der reguläre Anteil jeder Zeitreihe) ergäbe sich aus einem weissen Rauschen durch einen kausalen TLF. So ist es aber nicht! Bei einem TLF sind (nach unserer Definition) die 'Influenzkoeffizienten' summabel und nicht lediglich quadratsummabel. Die Transferfunktion eines TLF des Prozesses, der durch den TLF aus einem weissen Rauschen entsteht, ist bei uns durch eine absolutkonvergente Fourier-Reihe dargestellt. Im Unterschied dazu ist die Folge der Nachwirkungen bei einer regulären Zeitreihe i. Allg. nur quadratsummabel. Man beachte insbesondere: Die 'Wold-Funktion' einer regulären Zeitreihe $Y(e^{-i\omega}) = 1 + \sum_j \chi_j e^{-ij\omega}$ ist keine Transferfunktion; sie ist wohl quadratintegrierbar aber nicht notwendigerweise stetig.

Wir werden später diskutieren, was die Regularität der Zeitreihe für seine Spektraldichte bedeutet. Zunächst wollen wir aber noch unter der Annahme der Regularität einige Rechnungen durchführen.

Der Ansatz mit trigonometrischen Reihen. Wir wollen die Betrachtungen über Projektionen auf unendliche Vergangenheiten in die Welt der Spektralmaße übertragen und dort konkretisieren. Dazu benützen wir eine isometrische Abbildung des Hilbertraums $\mathcal{H}(X)$ auf einen Hilbertraum 2π -periodischer Funktionen. Dieser wird von dem Raum der trigonometrischen Polynome erzeugt. Der Linearkombination $\sum a_n X^n$ wird das trigonometrische Polynom $\sum a_n e^{in\omega}$ zugeordnet. Die Norm ist aber nicht die übliche Norm aus der Theorie der Fourierreihen. Bei uns macht das Spektralmaß den Vektorraum der trigonometrischen Polynome zu einem Prä-Hilbertraum, indem wir für $a(\cdot) = \sum a_n e^{in\cdot}$, $b(\cdot) = \sum b_m e^{im\cdot}$ definieren

$$\langle a, b \rangle_\mu = \int a(\omega) \cdot \overline{b(\omega)} d\mu(\omega) = \text{cov}\left(\sum a_n X^n, \sum b_m X^m\right)$$

Mit $\gamma_X(h) = \text{cov}(X^{n+h}, X^n) = \int e^{ih\omega} d\mu_X(\omega)$ gilt also $\langle a, b \rangle_\mu = \sum_{nm} a_n \bar{b}_m \gamma_X(n-m)$. Insbesondere haben wir für das trigonometrische Polynom $a(\omega)$ das Normquadrat

$$\|a\|_\mu^2 = \int |a(\omega)|^2 d\mu(\omega) = \text{var}\left(\sum a_n X^n\right) = \sum a_n \bar{a}_m \gamma(n-m).$$

Die Sesquilinearform $\langle \cdot, \cdot \rangle_\mu$ nennt man manchmal die **Töplitz-Form** zu $\gamma(\cdot)$.

Der erzeugte Hilbertraum ist der Raum $L^2(\mu)$, wo $\mu(d\omega)$ das Spektralmaß der Zeitreihe ist. Genauer gesagt, geht es bei den Elementen natürlich um Äquivalenzklassen 2π -periodischer Funktionen; Funktionen, die sich nur auf einer $\mu_{\mathbf{X}}$ -Nullmenge unterscheiden, sind zu identifizieren.

Es sollte nicht zu Missverständnissen führen, wenn wir gelegentlich hin und her springen zwischen der Sprechweisen mit Zufallsgrößen und der mit Funktionen aus L^2_μ . Wir identifizieren also den Raum L^2_μ mit $\mathcal{H}(\mathbf{X})$ und den von $\{e^{in\cdot}, e^{i(n-1)\cdot}, e^{i(n-2)\cdot}, \dots\}$ erzeugten Teilvektorraum mit $\mathcal{H}_n^{-\infty}(\mathbf{X})$. Seine Elemente sind diejenigen 2π -periodischen Funktionen, die sich in der L^2_μ -Norm durch trigonometrische Polynome der Form $\sum_{m \leq n} a_m e^{im\cdot}$ approximieren lassen. Besonders häufig werden wir den Raum $\mathcal{H}_{-1}^{-\infty}(\mathbf{X})$ benötigen, den

wir dann auch mit $L^2_-(\mu)$ bezeichnen. Seine Elemente sind diejenigen bzgl. μ quadrat-integrablen Funktionen, die sich durch trigonometrische Polynome der Form $\sum_{j=1}^N b_j e^{-ij\omega}$ approximieren lassen. Ein Element in $L^2_-(\mu)$ ist beispielsweise die Funktion zum Einschnitt-Prädiktor $\pi_{-1}^\infty X^0$.

Diese Funktion zur Innovation $Z^0 = X^0 - \pi_{-1} X^0$ wollen wir mit $e_0(\omega)$ bezeichnen. Die Funktion zur Innovation $Z^n = X^n - \pi_{n-1} X^n$ ist offenbar $e_n(\omega) = e^{in\omega} \cdot e_0(\omega)$.

Satz 3.2.3. *Es sei $\mathbf{X}$ eine zentrierte schwachstationäre Zeitreihe mit dem absolutstetigen Spektralmaß $d\mu(\omega) = f(\omega) \frac{1}{2\pi} d\omega$. Für die Funktion e_0 zur Innovation $Z^0 = X^0 - \pi_{-1} X^0$ gilt dann*

$$(1) \quad |e_0(\omega)|^2 f(\omega) = \sigma_\infty^2 = \text{var}(Z^0).$$

Wenn $\sigma_\infty^2 > 0$, dann gilt für die zu $e_0(\omega)$ reziproke Funktion $y(\omega)$

$$y = e_0^{-1} = \frac{1}{\sigma_\infty^2} \bar{e}_0 \cdot f = 1 + \sum_1^\infty \chi_j e^{-ij\omega} = Y(e^{-i\omega})$$

Beweis. Die erste Formel folgt aus der paarweisen Orthogonalität der Z^n : Diese bedeutet

$$0 = \text{cov}(Z^h, Z^0) = \frac{1}{2\pi} \int e^{ih\omega} |e_0(\omega)|^2 f(\omega) d\omega = 0 \quad \text{für alle } h \neq 0.$$

Da die integrable Funktion $g(\omega) = |e_0(\omega)|^2 f(\omega)$ verschwindende charakteristische Koeffizienten besitzt $\int e^{ih\omega} g(\omega) d\omega$ für alle $h \neq 0$, ist sie eine Konstante. Sie ist genau dann von Null verschieden, wenn die Zeitreihe regulär ist. Aus $(1 - e_0) \perp e_0$ ergibt sich

$$0 = \frac{1}{2\pi} \int (1 - e_0) \cdot \bar{e}_0 \cdot f d\omega = -\sigma_\infty^2 + \frac{1}{2\pi} \int \bar{e}_0 \cdot f d\omega.$$

Weiter gilt $\frac{1}{2\pi} \int |y|^2 d\omega = \frac{1}{2\pi} \int e_0^{-1} \cdot \frac{1}{\sigma_\infty^2} e_0 \cdot f d\omega = \frac{1}{\sigma_\infty^2} \cdot \sigma^2$ mit $\sigma^2 = \frac{1}{2\pi} \int f d\omega$. Die Funktion e_0 wird durch einseitige trigonometrische Polynome $A(e^{-i\omega}) = \prod (1 - \zeta_k^{-1} e^{-i\omega})$ (mit $|\zeta_k| > 1$) approximiert. Für die Reziproken gilt $(A(e^{-i\omega}))^{-1} - 1 \in L^2_-(1)$. Daher die Darstellung $y = 1 + \sum_1^\infty \chi_j e^{-ij\omega}$. Wir schliessen $X^0 = Z^0 + \sum_1^\infty \chi_j Z^{-j}$ aus der Gleichung

$$(2) \quad 1 = e_0(\omega) \cdot \left(1 + \sum_{j=1}^\infty \chi_j e^{-ij\omega}\right) = e_0(\omega) \cdot Y(e^{-i\omega}).$$

Interessant sind auch die Mehrschritt-Prädiktoren: Wir kennen schon $\pi_{-1}(1) = 1 - e_0(\omega)$, die Funktion zum Fußpunkt des Lots, den Einschnitt-Prädiktor. Für die Fußpunkte der Lote von 1 auf die kleineren Räume $L^2_{-2} \supset L^2_{-3} \supset \dots$ gilt offenbar

$$\begin{aligned} \pi_{-2}(1) &= e_0 \cdot [\chi_2 e^{-i2\omega} + \chi_3 e^{-i3\omega} + \dots] \\ \pi_{-3}(1) &= e_0 \cdot [\chi_3 e^{-i3\omega} + \chi_4 e^{-i4\omega} + \dots] \\ &\dots \end{aligned}$$

Wir haben gesehen: Für die Spektraldichte einer regulären Zeitreihe existiert eine Zahl σ_∞^2 und eine im Einheitskreis holomorphe Funktion $Y(z) = 1 + \sum_1^\infty \chi_j z^j$ mit $\sum |\chi_j|^2 < \infty$ sodass $f(\omega) = \sigma_\infty^2 \cdot |Y(e^{-i\omega})|^2$. Es gibt viele Faktorisierungen $f(\omega) = \tau^2 \cdot |B(e^{-i\omega})|^2$ mit $B(z)$ holomorph und $B(0) = 1$. Es zeigt sich, dass für die Prädiktion interessante Faktorisierung die ist, wo die Funktion $B(z)$ im Einheitskreis keine Nullstellen besitzt; für sie ist der Faktor τ^2 minimal. Das sieht man so: Wenn $f(\omega) = \sigma_\infty^2 \cdot |Y(e^{-i\omega})|^2$ und $B(z) = \frac{1}{\zeta} \frac{z-\zeta}{z-1} Y(z) = 1 + \sum_1 b_j z^j$ mit einem $|\zeta| < 1$, dann ist $B(z)$ eine im Einheitskreis holomorphe Funktion mit $B(0) = 1$ und $f(\omega) = |\zeta|^2 \sigma_\infty^2 \cdot |B(e^{-i\omega})|^2$. Ist $\mathbf{W}$ ein weisses Rauschen mit der Intensität $\tau^2 = |\zeta|^{-2} \sigma_\infty^2$ und $Y^t = W^t + \sum_1 b_j W^{t-j}$, so ist $\mathbf{Y}$ eine stationäre Zeitreihe mit der Spektraldichte $f(\omega)$. Die Intensität τ^2 des anregenden weissen Rauschens ist nicht die Prädiktionsvarianz. Die Prädiktionsvarianz ($> 0!$) ist die minimale Intensität eines Rauschens, aus welchem man mit einem (verallgemeinerten) kausalen Moving Average $TLF = I + \sum_1 b_j B^j$ eine Zeitreihe mit der Spektraldichte $f(\omega)$ herstellen kann. Unser Argument zeigt auch: Wenn wir zu einer regulären Zeitreihe die Funktion $y(\omega)$ wie oben konstruieren, d. h. $y(\omega) = Y(e^{-i\omega}) = e_0(\omega)^{-1}$ oder aus den Nachwirkungen χ_j der Innovationen, dann hat die dazugehörige Funktion $Y(z)$ im Einheitskreis keine Nullstellen. Ist nämlich $f(\omega) = \tau^2 \cdot |B(e^{-i\omega})|^2$ eine Faktorisierung mit $B(\zeta) = 0$, dann liefert $C(z) = \zeta \frac{z-1}{z-\zeta} B(z)$ eine Verbesserung: $f(\omega) = |\zeta|^{-2} \tau^2 \cdot |C(e^{-i\omega})|^2$.

Wir halten das fest als

Satz 3.2.4 (Eindeutige W-H-Faktorisierung regulärer Spektraldichten).

Ist $f(\omega)$ die Spektraldichte zu einem regulären Prozess, so existiert genau eine Zahl $\sigma_\infty^2 > 0$ und genau eine holomorphe Funktion $Y(z) = 1 + \sum_1^\infty \chi_j z^j$ ohne Nullstellen im Einheitskreis, sodass

$$f(\omega) = \tau^2 \cdot |B(e^{-i\omega})|^2.$$

Es gilt und $\sum_j |b_j|^2 < \infty$. Die Funktion $e_0(\omega) = B(e^{-i\omega})$ ist die Funktion zum Einschnitt-Prädiktor; σ_∞^2 ist seine Varianz.

Wir wollen jetzt diejenigen Spektraldichten $f(\omega)$ analytisch kennzeichnen, zu welchen eine (im Sinne der Wold-Zerlegung) reguläre Zeitreihe existiert. Der zentrale Punkt ist die untere Abschätzung der Länge der Einschnitt-Prädiktors

Satz 3.2.5. Ist $\mathbf{X}$ eine schwachstationäre Zeitreihe mit einer Spektraldichte f , deren Logarithmus integrierbar ist, dann gilt

$$(3) \quad \text{var}(X^0 - \pi_{-1}^- X^0) \geq \exp \left(\frac{1}{2\pi} \int \ln f(\omega) d(\omega) \right).$$

Beweis. Die Suche nach dem Einschnitt-Prädiktor von X^0 kann als ein Problem für trigonometrische Polynome (im Raum $L^2_-(\mu)$) formuliert werden. Gesucht sind $a_1, a_2, \dots, a_N$, sodass

$$\text{var} \left(X^0 - \sum_{j=1}^N a_j X^{-j} \right) = \frac{1}{2\pi} \int \left| 1 - \sum_{j=1}^N a_j e^{-ij\omega} \right|^2 f(\omega) d\omega$$

nahe beim Infimum σ_∞^2 liegt. Wir suchen also zu unserem Prozess $\mathbf{X}$ nach Polynomen der Gestalt $A(z) = 1 - \sum_1 a_j z^j$ für welche $\int |A(e^{-i\omega})|^2 \mu(d\omega)$ klein ist.

Wir zeigen zuerst, dass wir uns auf solche Polynome $A(z)$ beschränken können, die keine Nullstellen im Einheitskreis besitzen. Dazu schreiben wir ein gegebenes Polynom als Produkt von Linearfaktoren $A(z) = \prod (1 - \zeta_k^{-1} z)$. Wenn eine der Nullstellen ζ im Einheitskreis liegt, dann liegt $\bar{\zeta}^{-1}$ ausserhalb des Einheitskreises und es gilt

$$|1 - \bar{\zeta} e^{-i\omega}|^2 = |1 - \zeta e^{i\omega}|^2 = |\zeta|^2 |\zeta^{-1} - e^{i\omega}|^2 = |\zeta|^2 |1 - \zeta^{-1} e^{-i\omega}|^2.$$

Die Linearfaktoren zu ζ und $\bar{\zeta}^{-1}$ liefern bis auf eine Konstante dasselbe Absolutquadrat auf der Peripherie des Einheitskreises. Die Filter $I - \zeta^{-1} B$ mit $|\zeta| > 1$ sind vorzuziehen.

Es wird sich als nützlich erweisen, die Bedingung $A(0) = 1$ umzuformulieren. Dazu die Vorbemerkung:

$$\frac{1}{2\pi} \int \ln |1 - c e^{-i\omega}| d\omega = \begin{cases} \ln |c| & \text{für } |c| > 1 \\ 0 & \text{für } |c| < 1 \end{cases}$$

Unser Polynom $A(z)$ mit den Nullstellen $|\zeta_k| > 1$ schreiben wir

$$A(z) = \left(\prod_k \zeta_k^{-1} \right) \cdot \left(\prod_k (\zeta_k - z) \right) = \left(\prod_k \zeta_k^{-1} \right) \cdot B(z)$$

$$|A(e^{-i\omega})|^2 = |B(e^{-i\omega})|^2 \cdot \prod_k |\zeta_k|^{-2} = |B(e^{-i\omega})|^2 \cdot \exp\left(-\sum \ln |\zeta_k|^2\right)$$

$$\int \ln |A(e^{-i\omega})|^2 d\omega = 0 \quad \text{ergibt} \quad \frac{1}{2\pi} \int \ln |B(e^{-i\omega})|^2 d\omega = \sum_k \ln |\zeta_k|^2.$$

Somit können wir schreiben

$$\frac{1}{2\pi} \int |A(e^{-i\omega})|^2 f(\omega) d\omega = \frac{1}{2\pi} \int |B(e^{-i\omega})|^2 f(\omega) d\omega \cdot \exp\left(-\frac{1}{2\pi} \int \ln |B(e^{-i\omega})|^2 d\omega\right).$$

Mit den Abkürzungen $\|B\|^2 = \frac{1}{2\pi} \int |B(e^{-i\omega})|^2 f(\omega) d\omega$, $\mathcal{L}(B) = \frac{1}{2\pi} \int \ln |B(e^{-i\omega})|^2 d\omega$ heisst das

$$\frac{1}{2\pi} \int |A(e^{-i\omega})|^2 f(\omega) d\omega = \|B\|^2 \exp(-\mathcal{L}(B))$$

Jedes Polynom $B(z) = \text{const} \prod_k (\zeta_k - z)$ liefert eine Approximation $\tilde{W}^0$ an das Lot $Z^0 = X^0 - \pi_{-1} X^0$. Wir können die Betrachtung aber auf die $B(\cdot)$ mit $\|B\|^2 = 1$ einengen, weil die Multiplikation von $B(\cdot)$ mit einer Konstanten die rechte Seite nicht ändert.

Wir fassen zusammen: Es sei $A(z) = 1 - \sum_1^N a_j z^j$ ein Polynom ohne Nullstellen im Einheitskreis und $B(z)$ das dazugehörige Polynom mit $\|B\|^2 = 1$. Für das approximative Lot dazu gilt dann

$$\text{var} \left(X^0 - \sum_1^N a_j X^{-j} \right) = e^{-\mathcal{L}(B)}$$

Wir zeigen jetzt: Wenn $\ln f$ integrierbar ist, dann ist das Funktional $\mathcal{L}(\cdot)$ auf der Einheitskugel nach oben beschränkt. Es gilt

$$\mathcal{L}(B) \leq -\frac{1}{2\pi} \int \ln f(\omega) d\omega \quad \text{für alle } B \text{ mit } \|B\|^2 = 1.$$

Der Beweis dieser Abschätzung ergibt sich aus dem folgenden wohlbekanntem

Lemma. *Auf einem Maßraum $(\Omega, \mathfrak{A}, d\rho(\omega))$ seien $g(\omega), h(\omega) \geq 0$ mit $\int g(\omega) d\rho = 1 = \int h(\omega) d\rho$. Es gilt dann*

$$\int h(\omega) \ln \frac{h(\omega)}{g(\omega)} d\rho \geq 0$$

Dieser sog. Informationsabstand verschwindet nur, wenn $g = h$ ρ -fastüberall.

Wir benötigen den folgenden Spezialfall: Auf dem Wahrscheinlichkeitsraum $(\mathbb{R}/2\pi, \frac{1}{2\pi} d\omega)$ ist $g(\omega) = |B(e^{-i\omega})|^2 \cdot f(\omega)$ eine Wahrscheinlichkeitsdichte für jedes $B(\cdot)$ mit $\|B\|^2 = 1$.

Nach dem Lemma gilt also $\frac{1}{2\pi} \int \ln \left(|B(e^{-i\omega})|^2 \cdot f(\omega) \right) d\omega \leq 0$ für alle B mit $\|B\|^2 = 1$. Dies liefert die gewünschte untere Abschätzung der Varianz des Prädiktionsfehlers σ_∞^2 .

$$\text{var } \tilde{W}^0 \cdot \exp \left(-\frac{1}{2\pi} \int \ln f(\omega) d\omega \right) \geq 1, \quad \ln \sigma_\infty^2 - \frac{1}{2\pi} \int \ln f(\omega) d\omega \geq 0.$$

Mit der Abschätzung $\sigma_\infty^2 > 0$ ist der Weg frei für die obigen Konstruktionen, die zur entscheidenden Faktorisierung führten: $f(\omega) = \sigma_\infty^2 |Y(e^{-i\omega})|^2$ mit $Y(\zeta) \neq 0$ für $|\zeta| < 1$. Wir haben lückenlos bewiesen: Zu jedem f mit integrierbarem Logarithmus existiert eine reguläre Zeitreihe. Und wir werden sehen, dass wir so viel über diese Zeitreihen gelernt haben, dass wir in Reihe von konkreten Fällen die Prädiktionsvarianz und die Nachwirkungen der Innovationen explizit bestimmen können. Diese speziellen Fälle stellen wir zurück. In diesem Abschnitt bleiben wir noch beim allgemeinsten Fall einer regulären Zeitreihe.

Die Formel $\ln \sigma_\infty^2 = \frac{1}{2\pi} \int \ln f(\omega) d\omega$ haben wir nicht bewiesen; sie erscheint plausibel, wenn man bedenkt, dass für die Funktion $b(\omega) = \frac{1}{\sigma_\infty} e_0$ gilt

$$\|b(\omega)\|^2 = 1, \quad \int \ln (|b(\omega)|^2 \cdot f(\omega)) d\omega = 0.$$

Leider haben wir keinen stichhaltigen Beweis der Formel gefunden, welcher nahe bei den Vorstellungsweisen der Zeitreihen bleibt. Wir benötigen ein wenig Funktionentheorie.

Anhang: Einige funktionentheoretische Argumente

Ein wichtiger Typ von Funktionen sind bekanntlich die harmonischen Funktionen. Eine reellwertige Funktion ist genau dann harmonisch, wenn der Mittelwert über den Rand jedes Kreises (im Definitionsbereich) gleich dem Wert im Mittelpunkt ist. Für jede holomorphe Funktion sind Real- und Imaginärteil harmonische Funktionen. Zu jeder harmonischen Funktion $a(x, y)$ in einem einfach zusammenhängenden Gebiet existiert eine harmonische Funktion $b(x, y)$ sodass $a(x, y) + i \cdot b(x, y)$ eine holomorphe Funktion ist; sie ist bis auf eine additive Konstante eindeutig bestimmt.

Mit Hilfe dieser Begriffe können wir unsere Überlegungen um die reguläre Spektraldichten ergänzen. Rekapitulieren wir zuerst: Mit unseren Überlegungen mit approximierenden Polynomen haben wir aus der Integrierbarkeit von $\ln f$ eine untere Abschätzung für σ_∞^2 gewonnen. Andererseits haben wir durch unsere Hilbertraum-Konstruktionen gezeigt

$$-\ln \sigma_\infty^2 + \ln f(\omega) = \ln |Y(e^{-i\omega})|^2 = 2\Re \left(1 + \sum_1 \chi_j e^{-ij\omega} \right),$$

wobei $Y(z)$ im Einheitskreis keine Nullstellen hat. Hier setzt nun die Funktionentheorie ein. Die Funktion $H(z) = \ln Y(z) = 1 + \sum_j \chi_j z^j$ ist holomorph im Einheitskreis; ihr Realteil ist harmonisch; $H(0) = \ln Y(0) = 0$. Der Mittelwert über die Kreislinie $\zeta = r \cdot e^{i\omega}$ verschwindet für jedes $r < 1$. Ausführlich geschrieben: Es gilt $\frac{1}{2\pi} \int h_r(e^{i\omega}) d\omega = 0$ für die Funktion

$$h_r(\omega) = 2\Re \ln (Y(re^{i\omega})) = \ln |Y(re^{i\omega})|^2.$$

Der Beweis der Formel von Szegő-Krein-Kolmogorov ist erbracht, wenn wir zeigen, dass $\lim_{r \uparrow 1} h_r(\omega) = h(\omega) = \ln f(\omega) - \ln \sigma_\infty^2$ gilt im Sinne einer Konvergenz, welche die Konvergenz der Integrale impliziert.

Es ist lohnend, die Funktionen h_r etwas genauer zu betrachten. Wir zeigen, dass sich h_r aus h durch Glättung mit dem Poisson-Kern ergibt:

$$h_r(\varphi) = \frac{1}{2\pi} \int h(\omega) p_r(\varphi - \omega) d\omega \quad \text{mit} \quad p_r(\varphi - \omega) = \frac{1 - r^2}{1 - 2r \cos(\varphi - \omega) + r^2}.$$

Für jedes h im Banachraum $L^1(\mathbb{R}/2\pi, \frac{1}{2\pi}d\omega)$ gilt in der Tat $\|h * p_r - h\|_1 \rightarrow 0$; die Aussage gilt für alle elementaren Treppenfunktionen und somit für alle Funktionen, die man in der l^1 -Norm durch solche Funktionen approximieren kann.

Glättung mit den Poisson-Kernen

Es sei $h(\zeta) = h(e^{i\omega})$ integrabel und $h_n = \frac{1}{2\pi} \int h(e^{i\omega}) \cdot e^{-in\omega} d\omega$. Das Umlaufintegral über den Einheitskreis ist eine in $\{|z| < 1\}$ holomorphe Funktion

$$H(z) = \frac{1}{2\pi i} \int h(\zeta) \frac{1}{\zeta - z} d\zeta = \frac{1}{2\pi} \int h(e^{i\omega}) \frac{1}{1 - ze^{-i\omega}} d\omega = \sum_0^\infty h_n z^n.$$

Wenn h reellwertig ist, dann gilt $h_{-n} = \bar{h}_n$ und $-h_0 + 2\Re H(re^{i\varphi}) = \sum_{-\infty}^\infty h_n r^{|n|} e^{in\varphi}$. Man bemerke: Wenn h (mit $h_0 = 0$) eine absolutkonvergente Fourier-Reihe besitzt, dann ist $H(z)$ offensichtlich eine holomorphe Funktion mit $H(0) = 0$ und $2\Re H(\cdot) = h(\cdot)$ am

Rand. Es gibt nur eine solche Funktion zu h . Der folgende Aufwand muss getrieben werden für Randfunktionen h , die nur als integrel angenommen werden.

Für die Funktion $h_r(\varphi) = -h_0 + 2\Re H(re^{i\varphi})$ ergibt sich wegen

$$1 + 2\Re \frac{1}{1-a} = 1 + \frac{a}{1-a} + \frac{\bar{a}}{1-\bar{a}} = \frac{1-|a|^2}{|1-a|^2}$$

$$\begin{aligned} h_r(\varphi) &= \frac{1}{2\pi} \int h(e^{i\omega}) \frac{1-|z|^2}{|1-ze^{-i\omega}|^2} d\omega = \\ &= \frac{1}{2\pi} \int h(e^{i\omega}) \frac{1-r^2}{1-2r\cos(\varphi-\omega)+r^2} d\omega = h * p_r(\varphi). \end{aligned}$$

und das zeigt, dass die glatten h_r bei radialer Annäherung an den Rand in einem guten Sinn gegen die (möglicherweise unstetige) Funktion h konvergieren.

Den analytischen Kern dieses Abschnitts, welcher die (im Sinne der Wold-Zerlegung) regulären Zeitreihen (in voller Allgemeinheit!) behandelt hat, kann man in analytischen Termen folgendermaßen ausdrücken.

Satz 3.2.6. *Es sei $f(\omega) > 0$ fastüberall auf $\mathbb{R}/2\pi$ mit $\ln \sigma_\infty^2 = \frac{1}{2\pi} \int \ln f(\omega) d\omega > -\infty$. Es sei $h_n = \int e^{-in\omega} \ln f(\omega) d\omega$, $H(z) = \sum_1 h_n z^n$, $Y(z) = \exp H(z) = 1 - \sum \chi_j z^j$. Es gilt dann $\sum |\chi_j|^2 < \infty$ und $\sigma_\infty^2 \cdot |Y(e^{-i\omega})|^2 = f(\omega)$.*

Im nächsten Abschnitt wollen wir uns mit spezielleren Spektraldichten regulärer Zeitreihen befassen.

3.3 Wiener-Hopf-Faktorisierungen

In den Anwendungen der Zeitreihen interessiert man sich weniger für die allerallgemeinsten regulären Spektraldichten und mehr für gewisse spezielle Typen wie z. B. die Dichten zu invertierbaren ARMA-Prozessen. Eine wichtige Rolle spielt jedenfalls die spezielle Faktorisierung $f(\omega) = \sigma_\infty^2 Y(e^{-i\omega}) \cdot \overline{Y(e^{-i\omega})}$, die wir oben konstruiert haben. Die Nachrichten-Ingenieure nennen gelegentlich $Y(e^{-i\omega})$ den kausalen Faktor in der Wiener-Hopf-Faktorisierung von $f(\cdot)$ und $\overline{Y(e^{-i\omega})}$ den antikausalen Faktor.

Doppeltreguläre Zeitreihen Wir klären, was es für eine Spektraldichte bedeutet, dass sowohl zu f als auch zur reziproken Funktion f^{-1} eine reguläre Zeitreihe gehört.

Satz 3.3.1 (Einpunkt-Interpolation). *Es sei $f(\omega)$ die Spektraldichte zu einem Prozess $\mathbf{X}$ mit positivem Prädiktionsfehler $\sigma_\infty^2 = \text{var}(X^0 - \pi_{\mathcal{H}_{-1}^\infty} X^0) > 0$. Das Lot Y^0 von X^0 auf den Raum $\text{span}\{X^n : n \neq 0\}$ hat genau dann positive Varianz $\tau^2 > 0$, wenn $f^{-1}(\omega)$ Lebesgue-integrierbar ist. Es gilt $\tau^{-2} = \left(\frac{1}{2\pi} \int \frac{1}{f(\omega)} d\omega\right)$.*

Beweis. $L^2(f)$ sei der Raum der Funktionen $g(\omega)$ mit $\|g\|^2 = \frac{1}{2\pi} \int |g(\omega)|^2 f(\omega) d\omega < \infty$. Für jedes $g(\omega) \in L^2(f)$ ist gf Lebesgue-integrierbar; die Schwarz'sche Ungleichung liefert für das Quadrat des Integrals die Abschätzung

$$\left| \int f \cdot g d\omega \right|^2 \leq \int (g\sqrt{f})^2 d\omega \cdot \int (\sqrt{f})^2 d\omega.$$

Das innere Produkt mit der Funktion $e^{ik\omega}$ ist $\langle g(\omega), e^{ik\omega} \rangle_f = \frac{1}{2\pi} \int gf(\omega) e^{-ik\omega} d\omega$. Ein $g \in L^2(f)$ steht genau dann auf allen $e^{ik\omega}$ mit $k \neq 0$ senkrecht, wenn für die integrierbare Funktion $h = g \cdot f$ gilt $\int h(\omega) e^{-ik\omega} d\omega = 0$ für alle $k \neq 0$. Die einzigen Lebesgue-integrierbaren Funktionen $h(\omega)$, die das leisten, sind die Konstanten c , $g \cdot f = c$. Zum Lot Y^0 gehört eine Funktion $g \in L^2(f)$, welche auf $e^{ik\omega}$ mit $k \neq 0$ senkrecht steht. Wenn $f(\omega)^{-1}$ nicht Lebesgue-integrierbar ist, dann kommt nur $c = 0$ in Betracht.

Es sei nun $f(\omega)^{-1}$ Lebesgue-integrierbar. Dies bedeutet gleichzeitig $f(\omega)^{-1} \in L^2(f)$ mit $\|\frac{1}{f}\|_\mu^2 = \frac{1}{2\pi} \int |\frac{1}{f}(\omega)|^2 f(\omega) d\omega = \frac{1}{2\pi} \int \frac{1}{f}(\omega) d\omega (= \tau^{-2})$. Nun gilt offenbar $\frac{1}{f} \perp e^{ik\omega}$ im $L^2(f)$ für alle $k \neq 0$. Das Lot der Funktion 1 (entsprechend der Zufallsgrößen X^0) auf den Raum $\text{span}\{e^{ik\omega} : k \neq 0\}$ ist ein Vielfaches von $\frac{1}{f}$. Der orthogonalen Zerlegung $X^0 = Y^0 + (X^0 - Y^0)$ entspricht im Raum $L^2(f)$ die orthogonale Zerlegung $1 = \alpha \frac{1}{f} + (1 - \alpha \frac{1}{f})$. Die Orthogonalität bedeutet $0 = \int \left(1 - \frac{\alpha}{f}\right) \cdot \frac{\alpha}{f} f d\omega$. Daraus ergibt sich die Konstante $\alpha = \tau^2$. Die Funktion zum Lot Y^0 ist also $\tau^2 \frac{1}{f}(\omega)$.

Satz 3.3.2 (Existenz und Eindeutigkeit der Faktorisierung im doppelt regulären Fall). *Es sei $f(\omega)$ eine f. ü. positive Funktion auf $\mathbb{R}/2\pi$ mit $f, \frac{1}{f}$ integrierbar. Es existieren dann quadratsummable linksseitige Fourier-Reihen*

$$y(\omega) = 1 + \sum_1^\infty \chi_j e^{-ij\omega}, \quad e(\omega) = 1 - \sum_1^\infty \psi_j e^{-ij\omega},$$

sodass mit $\frac{1}{2\pi} \int \ln f(\omega) d\omega = \ln \sigma_\infty^2$ gilt

$$e(\omega) \cdot f(\omega) = \sigma_\infty^2 \cdot \overline{y(\omega)}, \quad |e(\omega)|^2 \cdot f(\omega) = \sigma_\infty^2.$$

Sind $a(\omega)$, $b(\omega)$ durch linksseitige Fourier-Reihen darstellbar mit $a \cdot f = \bar{b}$, so existiert eine Konstante c , sodass

$$a = c \cdot e, \quad \bar{b} = c \cdot \bar{y}.$$

Vorbemerkungen: Es bezeichne $L^2(f)$ den Hilbertraum der bzgl. dem Maß $f(\omega) \frac{1}{2\pi} d\omega$ quadratintegriblen Funktionen. $L_-^2(f)$ sei der Teilraum, der von $\{1, e^{-i\omega}, e^{-i2\omega}, \dots\}$ aufgespannt wird. Die Räume $L^2(\frac{1}{f})$, $L_-^2(\frac{1}{f})$, $L^2(1)$, $L_-^2(1)$ seien entsprechend definiert. $L_-^2(1)$ ist also der Raum der einseitigen Fourier-Reihen $c(\omega) = \sum_1^\infty c_n e^{-in\omega}$ mit der üblichen Norm $\|c\| = (\sum |c_n|^2)^{1/2}$. Mit $L_-^1(1)$ bezeichnen wir den Raum der Lebesgue-integriblen Funktionen, die sich in der L^1 -Norm durch einseitige trigonometrische Polynome approximieren lassen.

Lemma. Wenn $e \in L_-^2(f)$, $b \in L_-^2(\frac{1}{f})$, dann gilt $e \cdot b \in L_-^1(1)$

Beweis. Wenn e und b einseitige trigonometrische Polynome sind, dann ist alles klar. Wir approximieren nun e durch ein einseitiges trigonometrisches Polynom e' , sodass gilt $\int |e - e'|^2 f(\omega) d\omega < \varepsilon$. Entsprechend wählen wir b' mit $\int |b - b'|^2 \frac{1}{f}(\omega) d\omega < \varepsilon$. Wir haben dann $e \cdot b = e' \cdot b' + (e - e') \cdot b' + e' \cdot (b - b') + (e - e') \cdot (b - b')$. Wir zeigen mit der Schwarz'schen Ungleichung, dass der zweite Term ein kleines Lebesgue-Integral hat

$$\int |(e - e') \cdot b'| d\omega = \int \left(|e - e'| \cdot \sqrt{f} \right) \cdot \left(b' \frac{1}{\sqrt{f}} \right) d\omega \leq \sqrt{\varepsilon} \cdot \left(\int |b'|^2 \cdot \frac{1}{f} d\omega \right)^{1/2}.$$

Ebenso werden die weiteren Terme abgeschätzt. So kann also $e \cdot b$ durch ein einseitiges trigonometrisches Polynom in der L_-^1 -Norm approximiert.

Beim folgenden Existenzbeweis der Wiener-Hopf-Faktorisierung stützen wir uns auf die Begriffsbildungen der regulären Zeitreihen. Der Eindeutigkeitsbeweis wird dann reine Algebra sein.

Beweis.

Wir beweisen, dass sich die Funktion $e(\omega)$ zur Innovation durch eine quadratsummable trigonometrische Reihe dargestellt werden kann, d. h. dass e quadratintegribel ist. Die Einschnittsprädiktoren für Zeitreihen mit den Spektralmaßen $f(\omega) \frac{1}{2\pi} d\omega$ und $\frac{1}{f}(\omega) \frac{1}{2\pi} d\omega$ führen auf Funktionen $e \in L_-^2(f)$, $y \in L_-^2(1)$, $b \in L_-^2(\frac{1}{f})$, $a \in L_-^2(1)$ mit

$$\begin{aligned} e(\omega) \cdot f(\omega) &= \sigma_\infty^2 \cdot \overline{y(\omega)}, & e(\omega) \cdot y(\omega) &= 1, \\ b(\omega) \cdot \frac{1}{f(\omega)} &= \frac{1}{\sigma_\infty^2} \cdot \overline{a(\omega)}, & b(\omega) \cdot a(\omega) &= 1, \\ a(\omega) \cdot f(\omega) &= \sigma_\infty^2 \cdot \overline{b(\omega)}, & b(\omega) \cdot a(\omega) &= 1. \end{aligned}$$

Wir bemerken $\int \ln f d\omega = -\int \ln \frac{1}{f} d\omega$; es geht daher wirklich um das gleiche σ_∞^2 . Wir zeigen $a(\omega) = e(\omega)$, $b(\omega) = y(\omega)$. Die beiden Darstellungen von f führen auf

$$\frac{b}{\bar{a}} = \frac{f}{\sigma_\infty^2} = \frac{\bar{y}}{e}, \quad b \cdot e = \bar{a} \cdot \bar{y}.$$

$a \cdot y$ ist das Produkt zweier Funktionen in $L_-^2(1)$, gehört also zu $L_-^1(1)$; nach dem Lemma von oben gehört auch $b \cdot e$ zu $L_-^1(1)$.

$$\int e^{ik\omega} \cdot (b \cdot e)(\omega) d\omega = 0 \quad \text{für } k = 1, 2, \dots$$

$$\int e^{ik\omega} \cdot (\bar{a} \cdot \bar{y})(\omega) d\omega = 0 \quad \text{für } k = -1, -2, \dots$$

Da $b \cdot e = \bar{a} \cdot \bar{y}$, haben wir hier eine integrierbare Funktion, deren sämtliche Fourier-Koeffizienten ausser den nullten verschwinden. Es handelt sich also um eine Konstante c . Da alle unsere Reihen mit dem nullten Koeffizienten = 1 beginnen, haben wir $c = 1$. Somit $b = e^{-1} = y$, $a = y^{-1} = e$.

Wenn $a \cdot f = \sigma_\infty^2 \overline{b(\omega)}$ mit irgendwelchen $a(\omega), b(\omega) \in L_-^1(1)$ so ist $b \cdot e = \bar{a} \cdot \bar{y}$ eine Konstante. Wegen $f = \sigma_\infty^2 \frac{\bar{y}}{e}$ gilt $ay\sigma_\infty^2 = c = \bar{b}\bar{e}$, $a\sigma_\infty^2 = ce$, $\bar{b} = c\bar{y}$.

Man bemerke: Im vorigen Abschnitt haben wir die Eindeutigkeit der Darstellung $f(\omega) = \sigma_\infty^2 |Y(e^{-i\omega})|^2$ dadurch sicher gestellt, dass wir (neben $Y(0) = 1$) von $Y(\cdot)$ verlangten, dass keine Nullstellen im Einheitskreis liegen. Hier ist die Eindeutigkeit dadurch gesichert, dass $Y(e^{-i\omega})^{-1}$ eine einseitige Fourier-Reihe besitzt. Zur Erinnerung: Bei einer regulären Zeitreihe $\mathbf{X}$ ist der Einschnitt-Prädiktor derjenige (verallgemeinerte) kausale TLF, welcher den Prozess in ein weisses Rauschen (mit der Intensität σ_∞^2) transformiert. Die zusätzliche Annahme der Integrierbarkeit von $\frac{1}{f}$ garantiert, dass der Einschnitt-Prädiktor mit quadratsummablen Koeffizienten realisiert werden kann.

$$e(\omega) = 1 - \sum \psi_j e^{-ij\omega}, \quad Z^t = X^t - \sum \psi_j X^{t-j}.$$

Die Fourier-Koeffizienten der reziproken Funktion $y(\omega) = e(\omega)^{-1}$ zeigen, wie der Prozess aus dem weissen Rauschen $\mathbf{Z}$ der Innovationen (mit der Intensität σ_∞^2) gewonnen wird.

$$y(\omega) = 1 + \sum \chi_j e^{-ij\omega}, \quad X^t = Z^t + \sum \chi_j Z^{t-j}$$

Kurz geschrieben: $e(B)\mathbf{X} = \mathbf{Z}$, $y(B)\mathbf{Z} = \mathbf{X}$.

Wenn $\mathbf{Y}$ ein Prozess zur Spektraldichte $\frac{1}{f}$ ist, dann $y(B)\mathbf{Y} = \mathbf{W}$ ein weisses Rauschen mit der Intensität σ_∞^{-2} , und es gilt $e(B)\mathbf{W} = \mathbf{Y}$.

Den Existenzsatz können wir auch folgendermaßen formulieren: Sind $f, \frac{1}{f}$ integrierbar, so existiert ein ('verallgemeinerte') kausaler 'invertierbarer' TLF mit den Leistungstransferfunktionen $f(\omega)$. Seine Transferfunktion ergibt sich aus dem kausalen Faktor in der

(eindeutig bestimmten!) Wiener-Hopf-Faktorisierung. $\text{TLF} = \sigma_\infty \cdot (I + \sum \chi_j B^j)$. Die Impulsantworten sind die Koeffizienten $(\sigma_\infty \chi_j)_{j \geq 0}$. ('Verallgemeinert' müssen wir sagen, weil die Folge der Impulsantworten auch im 'doppeltregulären' Fall nicht notwendigerweise summabel ist ; sie ist auch im Fall $(f, \frac{1}{f}$ integrabel) i. Allg. nur quadratsummabel. Man beachte, dass man den 'verallgemeinerten' TLF nur auf solche Zeitreihen $\mathbf{X}$ 'anwenden' kann, für deren Spektralmaß $d\mu_{\mathbf{X}}$ gilt $\int f(\omega) d\mu_{\mathbf{X}}(\omega) < \infty$.)

Beispiele für Faktorisierungen Wir werden nun sehen, dass sich für manche Typen von Funktionen $f(\omega)$ ein direkter rein analytischer Zugang zu den Funktionen y und e_0 gibt. Entscheidend dafür die Tatsache, dass die Faktorisierung bis auf die Normierungskonstanten eindeutig bestimmt sind.

Die folgende Einsicht geht anscheinend auf L. Fejér zurück: Jedes positivwertige trigonometrische Polynom vom Grad q ist das Absolutquadrat eines trigonometrischen Polynoms vom Grad q . Der klassische Beweis macht zuerst aus der Koeffizientenfolge ein Polynom vom Grad $2q + 1$. Wenn ζ eine Nullstelle ist, dann auch $\bar{\zeta}^{-1}$. Die Nullstellen mit $|\zeta| > 1$ werden herausgesucht; mit ihnen konstruiert man dann leicht das gesuchte einseitige trigonometrische Polynom vom Grad q . Diesen Beweissgang im Geiste der klassischen Analysis wollen wir hier nicht weiter ausführen. Wir wollen hier stattdessen üben, uns in der Gedankenwelt der MA(q)-Prozesse zu bewegen.

Beispiel 3.3.1 (Der kausale MA(q)-Prozess). Ist $f(\omega)$ ein trigonometrisches Polynom vom Grad q mit $f(\omega) > 0$, so existiert eine Zahl σ_∞^2 und einseitiges trigonometrisches Polynom vom Grad q $Y(e^{-i\omega}) = 1 + \sum_1^q \chi_j e^{-ij\omega}$, sodass $f(\omega) = \sigma_\infty^2 \cdot |Y(e^{-i\omega})|^2$.

Wir haben die Bezeichnungen passend zum obigen Hauptsatz gewählt. $f(\omega)$ führt offenbar zu einer doppelt regulären Zeitreihe. Es existiert daher eine Wiener-Hopf-Faktorisierung $f(\omega) = \sigma_\infty^2 \cdot |Y(e^{-i\omega})|^2$. Es gilt $\text{cov}(X^t, X^{t-h}) = \text{cov}(Z^t + \sum \chi_j Z^{t-j}, Z^{t-h}) = \chi_h \cdot \text{var}Z$.

Die Kovarianzen $\gamma(h) = \text{cov}(X^t, X^{t-h})$ sind die Koeffizienten des Polynoms $f(\omega)$; sie verschwinden für $|h| > q$. Also verschwinden auch die Kovarianzen $\text{cov}(X^t, Z^{t-h})$ für alle $h > q$; und das bedeutet $\chi_h = 0$ für alle $h > q$. Y ist ein Polynom vom Grad q .

Ein möglicher Weg, das trigonometrische Polynom $Y(\cdot)$ zum Polynom $f(\cdot)$ mit den Koeffizienten $\gamma(\cdot)$ schnell (approximativ) zu finden, ist wohl der, dass man auf die Töplitz-Matrix zu $\gamma(\cdot)$ das Orthogonalisierungsverfahren (im Sinne der finiten Prädiktion) anwendet. Das mag bequemer sein als das Sortieren der Nullstellen des gegebenen Polynoms.

Beispiel 3.3.2 (Kausale ARMA-Prozesse). Ein kausaler ARMA(p, q)-Prozess ergibt sich bekanntlich aus einem Paar von Polynomen und einem weissen Rauschen als die eindeutig bestimmte Lösung von $\phi(B)\mathbf{X} = \theta(B)\mathbf{Z}$ mit einem Polynom $\phi(z)$ ohne Nullstellen im Einheitskreis. Wir wollen auch vom Polynom $\theta(z)$ annehmen, dass es keine Nullstellen im Einheitskreis hat. Wir haben einen invertierbaren Prozess. Da hier $\mathcal{H}_t^{-\infty}(X) = \mathcal{H}_t^{-\infty}(Z)$ für alle t , handelt es sich um einen doppelt regulären Prozess. Wenn $\text{var}\mathbf{Z} = 1$, dann können wir $\theta(0) = 1 = \phi(0)$ annehmen. Die Spektraldichte ist

$$f(\omega) = \left| \frac{\theta(e^{-i\omega})}{\phi(e^{-i\omega})} \right|^2,$$

Den Prozess der Einschnitt-Prädiktoren kann man als einen kausalen $MA(\infty)$ schreiben:

$$\hat{X}^t = \pi_{[-\infty, t-1]} X^t = \sum_1^{\infty} \psi_j X^{t-j} \quad \text{mit} \quad 1 - \sum \psi_j B^j = \frac{\phi(B)}{\theta(B)}.$$

Man ist aber nicht gezwungen, die Potenzreihendarstellungen ins Spiel zu bringen. Die Einschnitt-Prädiktoren $\hat{X}^t$ können durch endliche Summen dargestellt werden. Es gilt

$$\hat{X}^t = \phi_1 X^{t-1} + \dots + \phi_p X^{t-p} + \theta_1 Z^{t-1} + \dots + \theta_q Z^{t-q} = X^t - Z^t.$$

Den Beweis führen wir dadurch, dass wir verifizieren: $X^t - \hat{X}^t$ steht senkrecht auf allen Z^{t-j} , $j > 0$ und damit auf $\mathcal{H}_{t-1}^{-\infty}$. Dies ergibt sich durch einfaches Umschreiben der definierenden Gleichung

$$\begin{aligned} X^t - \phi_1 X^{t-1} - \dots - \phi_p X^{t-p} &= Z^t + \theta_1 Z^{t-1} + \dots + \theta_q Z^{t-q}. \\ X^t - \hat{X}^t &= (X^t - \phi_1 X^{t-1} - \dots - \phi_p X^{t-p}) - (\theta_1 Z^{t-1} + \dots + \theta_q Z^{t-q}) = Z^t. \end{aligned}$$

Die Rechnung zeigt übrigens, dass dank der Normierung $\theta_0 = 1 = \phi_0$ die $(Z^t)_t$ die Innovationen des gegebenen ARMA-Prozesses sind.

Anmerkung: Das Resultat sollte nützlich sein für die Praxis der (zeitnahen) Vorhersage einer Zeitreihe. Es sagt uns: Wenn man den ARMA(p, q)-Prozess $\mathbf{X}$ bis zur Zeit $t - 1$ beobachtet hat, und sowohl die letzten p Werte X^t auch die letzten q Prädiktionsfehler $X^t - \hat{X}^t$ registriert hat, dann gewinnt man den Prädiktor für den t -ten Beobachtungswert X^t als Linearkombination dieser $p + q$ Zahlen. Man muß hier also nicht unendlich viele Beobachtungswerte aus der Vergangenheit (mit den Gewichten ψ_j) aufsummieren.

Beispiel 3.3.3 (Die Yule-Walker-Gleichungen zu einem AR(p)-Prozess).

Es sei $\mathbf{Z}$ ein weisses Rauschen und $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ ein Polynom ohne Nullstellen im Einheitskreis. $\mathbf{X}$ sei der eindeutig bestimmte AR-Prozess mit $\phi(B)\mathbf{X} = \mathbf{Z}$. Es handelt sich um einen doppeltregulären Prozess mit der Spektraldichte $f(\omega) = |\phi(e^{-i\omega})|^{-2}$. Der kausale Faktor der Wiener-Hopf-Faktorisierung ist offensichtlich $e_0(\omega) = \phi(e^{-i\omega})$.

Bemerke: Die Kovarianzfolge ist bekanntlich für jede Zeitreihe die Folge der Fourier-Koeffizienten $\gamma(h) = \frac{1}{2\pi} \int e^{-ih\omega} f(\omega) d\omega$. Bei einem AR(p)-Prozess gewinnt man die Koeffizienten ϕ_j aus den sog. Yule-Walker-Gleichungen. Das p -Tupel der ϕ_j ist die eindeutig bestimmte Lösung des inhomogenen Gleichungssystems

$$\gamma(k) - \sum_1^p \phi_j \gamma(k-j) = 0, \quad \text{für } k = 1, 2, \dots, p$$

Das sieht man folgendermaßen:

- 1) Die Matrix dieses Gleichungssystems ist die $p \times p$ -Töplitzmatrix zur Kovarianzfolge; sie ist also nichtsingulär. Es gibt genau eine Lösung.
- 2) Die Koeffizienten des Polynoms erfüllen die Yule-Walker-Gleichungen; denn die Zufallsgrößen $W^t = X^t - \sum \phi_j X^{t-j}$ sind in der Tat die Innovationen Z^t des AR-Prozesses; für $k = 1, 2, \dots$ gilt $0 = \text{cov}(Z^t, X^{t-k}) = \text{cov}(X^t - \sum \phi_j X^{t-j}, X^{t-k}) = \gamma(k) - \sum_1^p \phi_j \gamma(k-j)$.

Bei einer Yule-Walker-Gleichung der Länge p hat die Lösung $(\psi_1, \dots, \psi_p)$ die Eigenschaft, dass $\gamma(k) - \sum_1^\infty \psi_j \gamma(k-j)$ für alle $k > 0$ verschwindet. Das liegt an der speziellen Struktur der Töplitz-Matrix. Man betrachtet auch allgemeinere Töplitz-Matrizen. Sei γ_n die Covarianzfolge zu einem doppeltregulären Prozess und $1 - e(\omega) = \sum_1^\infty \psi_j e^{-ij\omega}$ die Funktion zum Einschrittprädiktor. Wenn γ_n summabel ist, also $f(\omega) = \sum_{-\infty}^\infty \gamma_n e^{in\omega}$, dann hat die Eigenschaft $e \perp e^{-ik\omega}$ für alle $k > 0$ die Form

$$\frac{1}{2\pi} \int \left(\sum_1^\infty \psi_j e^{-ij\omega} \right) \left(\sum_{-\infty}^\infty \gamma_n e^{in\omega} \right) e^{ik\omega} d\omega = 0 \quad \text{für } k = 1, 2, \dots,$$

und das heisst $\gamma(k) - \sum_1^\infty \gamma(k-j) = 0$ für alle $k > 0$. Wir haben oben bewiesen: Es existiert genau eine quadratsummable Lösung $(\psi_1, \psi_2, \dots)$ mit $(1 - \sum_1^\infty \psi_j e^{-ij\omega})^{-1} \in L_-^2$. **Ausblick:** Man hat auch bei anderen Folgen Γ_n das Problem gestellt: Gesucht ist eine einseitige Folge $(\psi_1, \psi_2, \dots)$, sodass

$$\Gamma(k) - \sum_1^\infty \psi_j \Gamma(k-j) = 0 \quad \text{für } k = 1, 2, \dots$$

Diese unendliche Gleichungssystem hat eine Struktur, welche an die aus anderen Zusammenhängen stammenden sog. Wiener-Hopf-Gleichungen erinnert:

Problem: Gegeben ist ein 'Kern' $\Gamma(n)$, d. h. eine komplexwertige Funktionen auf $\mathbb{Z}$, und eine einseitige Folge $(g_k)_{k \geq 0}$. Gesucht ist eine einseitige Folge $(\psi_n)_{n \geq 0}$, sodass gilt

$$\psi(k) - \sum_1^\infty \psi_j \Gamma(k-j) = g(k) \quad \text{für } k = 0, 1, 2, \dots$$

Besonders interessant ist die dazugehörige 'homogene' Gleichung

$$\psi(k) - \sum_1^\infty \psi_j \gamma(k-j) = 0 \quad \text{für } k = 1, 2, \dots$$

Um das Problem zu präzisieren, muss man noch sagen, welche Art von Folgen $(\psi_1, \psi_2, \dots)$ als Lösungen in Betracht gezogen werden sollen.

Der bekannteste Typ von Wiener-Hopf-Gleichungen im Kontinuierlichen hat die Form

$$K(t) = \int_0^\infty K(t-s) \psi(s) ds \quad \text{für alle } t > 0,$$

Beispiel 3.3.4. In den 'guten' Fällen besitzt die Funktion $1 - e_0 \in L_-^2(f)$ eine Darstellung als trigonometrische Reihe und man kann die Bedingung $e_0 \perp \text{span}\{e^{-ik\omega} : k < 0\}$ als eine Bedingung an die Koeffizientenfolge formulieren: Gesucht sind $\psi_1, \psi_2, \dots$, sodass

$$\begin{aligned} e_0(\omega) &= 1 - \sum_1^\infty \psi_j e^{-ij\omega} && \text{(konvergent im } L_\mu^2\text{-Sinn)} \\ \text{cov}(Z^0, X^{-k}) &= 0, && \text{cov}(X^0 - \sum_1^\infty \psi_j X^{-j}, X^{-k}) = 0, \\ \gamma(k) &= \sum_1^\infty \psi_j \cdot \gamma(k-j) && \text{für } k = 1, 2, \dots \end{aligned}$$

Wenn $f(\omega)$ die Spektraldichte zu einer doppeltregulären Zeitreihe ist, dann kann man die Funktion zum Einschnitt-Prädiktor durch eine trigonometrische Reihe darstellen: $e(\omega) = 1 - \sum \psi_j e^{-ij\omega}$. Wenn auch $f(\omega)$ durch eine trigonometrische Reihe dargestellt ist, dann kann man die W-H-Faktorisierung in 'guten' Fällen als ein Gleichungssystem für die Koeffizienten-Folgen verstehen: wir lassen alle Konvergenzfragen ausser Betracht und schreiben

$$\left(1 - \sum_1 \psi_j e^{-ij\omega}\right) \left(\sum_{-\infty}^{\infty} f_n e^{in\omega}\right) = \sigma_{\infty}^2 + r^+(\omega) \quad \text{wo } r^+(\omega) \in \text{span}\{e^{i\omega}, e^{i2\omega}, e^{i3\omega}, \dots\}$$

In Worten: Eine linksseitige trigonometrische Reihe, multipliziert mit einer beidseitigen ergibt eine rechtsseitige trigonometrische Reihe

$$\begin{aligned} e_0(\omega) &= 1 - \sum_1^{\infty} \psi_j e^{-ij\omega} && \text{(konvergent im } L_{\mu}^2\text{-Sinn)} \\ \text{cov}(Z^0, X^{-k}) &= 0, && \text{cov}(X^0 - \sum_1^{\infty} \psi_j X^{-j}, X^{-k}) = 0, \\ \gamma(k) &= \sum_1^{\infty} \psi_j \cdot \gamma(k-j) && \text{für } k = 1, 2, \dots \end{aligned}$$

Die letzte Gleichung ist eine Gleichung vom Wiener-Hopf-Typ.

3.4 Filtern gegen Rauschen

Stellen wir uns vor, dass ein Sender in gewissen (zufälligen!) Zeitabschnitten ein Signal sendet, welches uns durch einen gestörten Kanal übermittelt wird. Das Signal ist eine Funktion $S(t)$ oder auch ein d -Tupel von Funktionen. Wir beschränken uns aber zunächst auf den skalaren komplexen Fall. Auf der Empfängerseite entsteht das Problem, das Signal (einigermaßen genau) zu rekonstruieren oder wenigstens zu identifizieren. Die Identifikation ist dann aussichtsreich, wenn man davon ausgehen kann, dass das Signal aus einer bekannten Kollektion möglicher Signal stammt.

In der Zeitreihentheorie studiert man den Fall, wo die Signale S^t die Realisierungen eines stationären Prozesses $\mathbf{S}$ sind, dessen Spektraldichte $d\mu_S(\omega) = f_S(\omega) \frac{1}{2\pi} d\omega$ man (mehr oder weniger) kennt. Wir empfangen $X^t = S^t + N^t$ (N steht für 'noise'), wobei wir (der Übersichtlichkeit halber) annehmen, dass $\mathbf{N} = (N^t)_t \in \mathbb{Z}$ ein von $\mathbf{S}$ unabhängiger zentrierter stationärer Prozess ist, dessen Spektraldichte $d\mu_N(\omega) = f_N(\omega) \frac{1}{2\pi} d\omega$ ebenfalls bekannt ist. Die Aufgabe, das Signal S^{t+m} aufgrund verrauschter Beobachtungen (zum Zeitpunkt $t-1$) linear zu schätzen, nennt man ein lineares Filter-Problem.

Es geht hier um orthogonale Projektion von S^{t+m} auf den unendlichdimensionalen Hilbertraum $\mathcal{H}_{t-1}^{-\infty}(\mathbf{X})$. Spätestens zum Zeitpunkt $(t-1)$, soll ein 'Schätzwert' $\hat{S}^t$ abgeliefert werden. Den Fall $m=0$ ist ähnlich zur oben betrachteten Einschnitt-Prädiktion. Im Unterschied zum früher betrachteten Szenario interessieren jetzt auch die Fälle $m \leq 0$. Mit spricht von einem Filtern mit der Zeitverzögerung $|m|$. Mit dem Schätzwert $\hat{S}^t \in \mathcal{H}_{t-1}^{-\infty}(X)$ versucht man das Signal zur Zeit $t-|m|$ im quadratischen Mittel optimal zu approximieren: $\text{var}(S^{t-|m|} - \hat{S}^t) = \min$. Die besonders genaue Wiederherstellung der des Signalprozesses kann man erwarten, wenn man eine beliebige Zeitverzögerung zulässt, $m \rightarrow -\infty$. Im Limes besteht das Filterproblem in der Aufgabe, das Signal S orthogonal auf den Raum $\mathcal{H}_{-\infty}^{-\infty}(\mathbf{X})$ zu projizieren. Darauf kommen wir unten.

Wie früher bei der Extrapolation ist es angebracht, zuerst für endliches m zu den Spektraldarstellungen überzugehen, bevor man zum Fall $m = -\infty$ übergeht.

$$\begin{aligned} \text{cov}(S^{t+m}, X^{t_k}) &= \text{cov}(S^{t+m}, S^{t_k}) = \int e^{i(t+m-t_k)\omega} d\mu_S(\omega), \\ \text{cov}(X^{t_j}, X^{t_k}) &= \text{cov}(S^{t_j}, S^{t_k}) + \text{cov}(N^{t_j}, N^{t_k}) = \int e^{i(t_j-t_k)\omega} (f_S(\omega) + f_N(\omega)) \frac{1}{2\pi} d\omega, \\ \text{cov}\left(S^{t+m} - \sum_j b_j X^{t_j}, X^{t_k}\right) &= \int e^{-it_k\omega} \left[e^{i(t+m)\omega} f_S - \sum_j b_j e^{it_j\omega} \cdot (f_S + f_N) \right] \frac{1}{2\pi} d\omega. \end{aligned}$$

Der beste lineare Schätzer aufgrund der Menge der Beobachtungen $\{X^{t_k} : t_k \in \tilde{T}\}$ ist dadurch gekennzeichnet, dass die eben berechneten Kovarianzen für alle t_k verschwinden. Betrachten wir $t=0$ und wie gewohnt $\tilde{T} = -\mathbb{N}$. Die Bedingung ist dann, dass die Fourier-Koeffizienten der Funktion in der Klammer [] für alle negativen Indizes verschwinden, was wir folgendermaßen 'symbolisch' notieren

$$\left[e^{im\omega} f_S - \sum_{j=1}^{\infty} b_j e^{-ij\omega} \cdot (f_S + f_N) \right] \sim c_0 + c_1 e^{i\omega} + c_2 e^{i2\omega} + \dots$$

Im Allgemeinen wissen wir nur, dass der Fußpunkt des Lots im $L^2(\mu)$ -Sinn durch endliche Linearkombinationen $\sum_1^\infty b_j^{(\varepsilon)} e^{-ij\omega}$ approximiert werden kann. Um die Notation nicht zu verkomplizieren, nehmen wir nun aber im folgenden an, dass Koeffizienten $b_j^{(opt)}$ existieren. Mit der Bezeichnung $B(e^{-i\omega}) = \sum_1^\infty b_j^{(opt)} e^{-ij\omega}$ gewinnt die Orthogonalitätsbedingung die Form

$$e^{im\omega} f_S - B(e^{-i\omega}) \cdot (f_S + f_N) \sim c_0 + c_1 e^{i\omega} + c_2 e^{i2\omega} + \dots$$

Interessant ist auch die Länge des Lots. Da für die optimalen b_j die Vektoren $\sum b_j^{(opt)} X^{-j}$ und $S^m - \sum b_j^{(opt)} X^{-j}$ zueinander senkrecht sind, gilt

$$\begin{aligned} \text{var} \left(S^m - \sum b_j^{(opt)} X^{-j} \right) &= \text{var}(S^m) - \text{var} \left(\sum b_j^{(opt)} X^{-j} \right) = \\ &= \int [f_S - |B(e^{-i\omega})|^2 (f_S + f_N)] \frac{1}{2\pi} d\omega. \end{aligned}$$

Beispiel 3.4.1. Der AR(1)-Prozess $\mathbf{X}_0 = (I - \phi_1 B)^{-1} \mathbf{Z}_0$ (mit $\text{var} Z^t = \sigma^2$) wurde von einem weissen Rauschen $n \cdot \mathbf{N}$ mit der Intensität n^2 überlagert. Wir beobachten $\mathbf{X}_n = \mathbf{X}_0 + n \cdot \mathbf{N}$. Der beste Schätzer für S^t aufgrund der Beobachtungen in $\mathcal{H}_{t-1}^{-\infty}(X_h)$ ist auch der beste Prädiktor für X^t , weil sich S^t und X^t um eine Zufallsgröße unterscheiden, die mit allem Beobachteten unkorreliert ist. Die beste lineare Schätzung ist jetzt aber nicht wie im Fall ohne Rauschen $\varphi_1 \cdot X_n^{t-1}$. Der verrauschte Prozess ist kein AR(1)-Prozess, kein 'Markov-Prozess'. Die letzte Beobachtung ist nicht der beste Prädiktor; wir müssen, intuitiv gesprochen, über die früheren Beobachtungswerte glätten, um die eventuelle Irreführung durch das Rauschen möglichst abzumildern. Was da genau zu geschehen hat, wollen wir nun ausführlich und genau studieren.

Der verrauschte Prozess $\mathbf{X}_n$ hat die Spektraldichte

$$\begin{aligned} f_n(\omega) &= \frac{\sigma^2}{|1 - \phi_1 e^{-i\omega}|^2} + n^2 = \frac{\sigma^2}{|1 - \phi_1 e^{-i\omega}|^2} \left(1 + \left(\frac{n}{\sigma}\right)^2 |1 - \phi_1 e^{-i\omega}|^2 \right) = \\ &= \frac{\sigma^2}{|1 - \phi_1 e^{-i\omega}|^2} \cdot |A|^2 \cdot |1 - \tilde{\phi} e^{-i\omega}|^2 \end{aligned}$$

mit Koeffizienten $|A|^2$ und $-\tilde{\phi} = \frac{B}{A}$, die wir im Abschnitt über ARMA-Konstruktionen als Beispiel berechnet haben. In unserem speziellen Fall ergibt sich

$$\begin{aligned} -\tilde{\phi} &= \frac{\bar{A}B}{|A|^2} = \frac{1}{|A|^2} \left(\frac{n}{\sigma}\right)^2 (-\phi_1) \\ |A|^2 &= (|A|^2 + |B|^2) \left(1 + |\tilde{\phi}|^2\right)^{-1} = \frac{1}{1 + |\tilde{\phi}|^2} \left(1 + \left(\frac{n}{\sigma}\right)^2 (1 + |\phi_1|^2)\right). \end{aligned}$$

$\mathbf{X}_n$ hat die Kovarianzstruktur eines ARMA(1, 1)-Prozesses. Es existiert ein weisses Rauschen $\mathbf{Z}_n$, sodass

$$(I - \phi_1 B)X_n = (I - \tilde{\phi} B)Z_n.$$

Die Varianz des Prädiktionsfehlers bei optimaler Einschrittprädiktion ist $\sigma^2|A|^2$; und daraus ergibt sich auch die Varianz des Schätzfehlers bei der Schätzung des Signals

$$\text{var}(S^t - \hat{X}_n^t) = \text{var}(X^t - \hat{X}_n^t) - n^2$$

Wir wissen, wie man in einem ARMA-Prozess X_n^t optimal vorhersagt, wenn man die Werte $\dots X_n^{t-2}, X_n^{t-1}$ und die Innovationen $\dots Z_n^{t-2}, Z_n^{t-1}$ registriert hat. Man benötigt nur (X_n^{t-1}, Z_n^{t-1}) . Der optimale Prädiktor ist

$$\hat{X}_n^t = \phi_1 X_n^{t-1} - \tilde{\phi} Z_n^{t-1}$$

Das kann folgendermaßen verstanden werden: Bei einer unerwartet großen Abweichung des Beobachtungswerts zur Zeit $t - 1$ von dem geschätzten Wert $\hat{X}_n^{t-1}$ darf man nicht (wie bei einem AR(1)-Prozess) $\phi_1 X_n^{t-1}$ als Prädiktor für die nächste Vorhersage wählen; man muss entsprechend korrigieren. Der Korrekturbedarf ist allerdings gering, wenn die Intensität n^2 des Rauschens klein ist; denn dann ist $|\tilde{\phi}|$ klein.

Wir können den Prädiktor natürlich auch allein mit den X schreiben

$$\begin{aligned} \hat{X}_n^t &= \phi_1 X_n^{t-1} - \tilde{\phi} Z_n^{t-1} = \phi_1 X_n^{t-1} - \tilde{\phi}(X_n^{t-1} - \hat{X}_n^{t-1}) = \\ &= (\phi_1 - \tilde{\phi})(X_n^{t-1} + \tilde{\phi} X_n^{t-1} + \tilde{\phi}^2 X_n^{t-2} + \dots) \end{aligned}$$

In dieser Darstellung sieht man, dass die länger zurückliegenden Beobachtungswerte nur mit geometrisch abnehmenden Gewichten zu berücksichtigen sind.

Unendliche Zeitverzögerung Im Fall von unendlicher Zeitverzögerung ergibt sich das Lot von S^m in trivialer Weise aus der Formel für das Lot von S^0 . Die Bedingung an die 'Filter-Koeffizienten' lautet hier

$$\begin{aligned} \pi_{\mathcal{H}_\infty^-(\mathbf{x})} S^m &= \sum_{-\infty}^{\infty} \phi_j^{(\infty)} X^{m-j} \quad \text{mit} \\ 0 &= f_S - \left(\sum \phi_j^{(\infty)} e^{-ij\omega} \right) (f_S + f_N) = f_S - \Phi^{(\infty)}(\omega) \cdot (f_S + f_N). \end{aligned}$$

Die Abweichung des gesendeten Signals $\mathbf{S}$ von dem aus den empfangenen X^t linear rekonstruierten Prozess hat die Varianz

$$\begin{aligned} \text{var} \left(S^0 - \sum \phi_j^{(\infty)} X^{-j} \right) &= \int f_S(\omega) \frac{1}{2\pi} d\omega - \int |\Phi^{(\infty)}(\omega)|^2 [f_S + f_N](\omega) \frac{1}{2\pi} d\omega = \\ &= \int \left[1 - \frac{f_S}{f_S + f_N} \right] d\mu_S(\omega). \end{aligned}$$

Der für die Wiederherstellung des Eingangssignals optimale Filter (mit beliebiger Zeitverzögerung) hat die Transfer-Funktion $\Phi^{(\infty)}(\omega)$. Die Ingenieure nennen den Quotienten $\|\mu_S\|/\|\mu_N\|$ den Signal-Noise-Ratio SNR des gestörten Kanals. Der SNR ist ein Maß für die technische Qualität eines Nutzsignals, welches von einem Rauschen überlagert wird.

Wir wollen das alles an einem Spezialfall konkretisieren, den Norbert Wiener und A. N. Kolmogorov in den Jahren 1940-1941 (unabhängig voneinander) behandelt haben. Die Ingenieure nennen das Resultat häufig den Dekonvolutionsfilter von N. Wiener.

Beispiel 3.4.2. Ein Signalprozess $\mathbf{S}$ wurde durch einen TLF T_a geschickt. ($a(\cdot)$ ist die Impulsantwort-Funktion, $A(e^{-i\omega}) = \sum a(h)e^{-ih\omega}$ ist die Transferfunktion). Wir empfangen durch Rauschen überlagerte Beobachtungswerte

$$Y^t = (T_a S)^t + N^t = \int a(h)S(t-h) dh + N^t.$$

Gesucht ist eine Linearkombination der Beobachtungswerte sodass $S^t - \sum b(h) Y^{t-h}$ im Mittel klein ist. Die Spektraldichte des Rauschens sei bekannt. Wenn man nun auch etwas über die statistischen Verhältnisse bei den gesendeten Signalen weiss, dann kann man daran gehen, aus den 'mit $a(\cdot)$ gefalteten' Signale und dann noch gestört empfangenen Werten die ursprünglichen Signale zu rekonstruieren. Wir nehmen an, dass $\mathbf{S}$ ein stationärer Prozess ist mit dem Spektralmaß $d\mu_S(\omega)$ ist. Wir haben dann

$$\begin{aligned} \text{cov}(S^u, Y^s) &= \text{cov}\left(S^u, \sum a(h)S^{s-h}\right) = \\ &= \sum \overline{a(h)} \cdot \int e^{i(u-s+h)\omega} d\mu_S(\omega) \\ &= \int \overline{A(e^{-i\omega})} \cdot e^{i(u-s)\omega} d\mu_S(\omega), \\ \text{cov}(Y^t, Y^u) &= \text{cov}((T_a Y)^t, (T_a Y)^u) + \text{cov}(N^t, N^u) = \\ &= \int e^{i(t-u)\omega} |A(e^{-i\omega})|^2 d\mu_S(\omega) + \int e^{i(t-u)\omega} d\mu_N(\omega) = \\ &= \int e^{i(t-u)\omega} \left[|A(e^{-i\omega})|^2 + \frac{f_N}{f_S}(\omega) \right] d\mu_S(\omega) \end{aligned}$$

und mit der Abkürzung $g(\omega) = |A(e^{-i\omega})|^2 + \frac{f_N}{f_S}(\omega)$

$$\text{cov}\left(S^0 - \sum b(h) Y^{-h}, Y^s\right) = \int \left[\overline{A(e^{-i\omega})} e^{is\omega} - \sum b(h) e^{-ih\omega} e^{-is\omega} g(\omega) \right] d\mu_S(\omega).$$

Im 'nichtkausalen' Fall, (d. h. im Fall $m = -\infty$) ist zu fordern, dass diese Kovarianz für alle $s \in \mathbb{Z}$ verschwindet. Dies fordert das $\overline{A(e^{-i\omega})} - B(e^{-i\omega}) \cdot g(\omega) = 0$. Die Transferfunktion der optimalen 'Dekonvolution' ist also

$$B(e^{-i\omega}) = \frac{\overline{A(e^{-i\omega})}}{g(\omega)} = \frac{1}{A(e^{-i\omega})} \cdot \frac{|A(e^{-i\omega})|^2}{|A(e^{-i\omega})|^2 + \frac{f_N}{f_S}(\omega)}.$$

Diese Formel beschreibt in übersichtlicher Weise, wie die optimale Dekonvolution sich verhält, wenn die Störung durch das Rauschen klein ist. T_b ist im Grenzfall der zu T_a reziproke Filter. Wenn $A(\cdot)$ und die Spektraldichten f_S und f_N rationale Funktionen der Variablen $e^{-i\omega}$ sind dann kann man die Wienerdekonvolution offenbar mit einem ARMA-Filter realisieren.

3.5 Maße mit orthogonalen Werten

Wir wollen die Idee noch etwas weiter verfolgen, dass ein schwachstationärer Prozess (in gewissen Hinsichten) als eine Kurve in einem Hilbertraum $\mathcal{H}$ verstanden werden kann. Es sollte aber auch deutlich werden, dass bei dieser Sichtweise viele für den Stochastiker interessante Fragen aus dem Blick verschwinden. Andererseits sollte klar werden, dass nicht alle Konstruktionen, die aus der Sicht der Hilbertraumtheorie (speziell der Spektraltheorie) naheliegen, hineinführen in die Assoziationszusammenhänge der stochastischen Prozesse.

Maße mit orthogonalen Werten

In der elementaren Maßtheorie betrachtet man bekanntlich nicht nur Maße mit nichtnegativen Werten, sondern auch signierte und komplexwertige Maße auf einem messbaren Raum $(S, \mathfrak{G})$. Jeder $\mathfrak{G}$ -messbaren Menge S wird eine komplexe Zahl $\sigma(S)$ zugeordnet, sodass gilt

1. $\sigma(\emptyset) = 0$, $\sup\{|\sigma(S)| : S \in \mathfrak{G}\} < \infty$,
2. $\sigma(S_1 + S_2) = \sigma(S_1) + \sigma(S_2)$ für disjunkte Mengen
3. $S = \sum_1^\infty S_j \implies \left| \sigma(S) - \sum_1^N \sigma(S_j) \right| \rightarrow 0$.

(Das Zeichen $+$ bedeutet hier die disjunkte Vereinigung)

Als Varianten dieses Begriff betrachten wir im Folgenden 'Maße mit orthogonalen Werten in einem Hilbertraum' und dann (in stochastischer Einfärbung) sog 'unkorrelierte stochastischen Maße' auf dem Grundraum $(S, \mathfrak{G})$, welcher bei uns hier der Raum $S = \mathbb{R}/2\pi$ sein wird.

Definition. Es sei $(S, \mathfrak{G})$ ein messbarer Raum und $\mathcal{H}$ ein Hilbertraum. Ein 'Maß mit orthogonalen Werten' $M(\cdot)$ ordnet jedem $S \in \mathfrak{G}$ einen Vektor $M(S) \in \mathcal{H}$ zu, sodass gilt

1. $M(\emptyset) = 0$, $\sup\{\|M(S)\|^2 : S \in \mathfrak{G}\} < \infty$,
2. $S_1 \cap S_2 = \emptyset \implies M(S_1) \perp M(S_2)$, $M(S_1 + S_2) = M(S_1) + M(S_2)$,
3. $S = \sum_1^\infty S_j \implies \|M(S) - \sum_1^n M(S_j)\|^2 \rightarrow 0$.

Viele der aus der elementaren Maßtheorie vertrauten Konstruktionen kann man leicht auf die Maße mit orthogonalen Werten übertragen. Man beweist beispielsweise mit den bekannten Argumenten den

Satz 3.5.1. *Wenn $\tilde{\mathfrak{S}}$ eine Mengenalgebra ist, die $\mathfrak{G}$ erzeugt und $\tilde{M}(\cdot)$ eine Funktion auf $\tilde{\mathfrak{S}}$ mit den Eigenschaften 1), 2), 3), dann existiert genau eine Fortsetzung zu einem Maß mit orthogonalen Werten.*

Satz 3.5.2. Zu einem Maß mit orthogonalen Werten $dM(\cdot)$ auf $(S, \mathfrak{G})$ (mit Werten in $\mathcal{H}$) erhält man ein klassisches reellwertiges Maß $d\mu(\cdot)$ auf $(S, \mathfrak{G})$, wenn man definiert

$$\mu(S) = \|M(S)\|^2 \quad \text{für } S \in \mathfrak{G}.$$

Man nennt $\mu(\cdot)$ manchmal das Intensitätsmaß zu $M(\cdot)$.

Satz 3.5.3. Zum Maß mit orthogonalen Werten $dM(\cdot)$ gehört auch ein 'Integral'

$$I(\cdot) : L^2(S, \mathfrak{G}, d\mu) \ni f \longmapsto I(f) = \int f(s) dM(s) \in \mathcal{H}.$$

Dieses 'Integral' $I(\cdot)$ ist eine lineare Abbildung in den Hilbertraum $\mathcal{H}$. Für Treppenfunktionen $f = \sum \alpha_j 1_{S_j}$ mit paarweise disjunkten S_j ergibt sich das M -Integral als eine orthogonale Summe

$$I(f) = \int \sum \alpha_j 1_{S_j} dM = \sum \alpha_j M(S_j)$$

mit dem Normquadrat $\|I(f)\|^2 = \sum |\alpha_j|^2 \mu(S_j) = \int |f(s)|^2 d\mu(s)$.

Allgemein gilt $\|I(f)\|^2 = \int |f(s)|^2 d\mu(s)$, insbesondere haben wir

$$\int |f^{(n)}(s) - f(s)|^2 d\mu(s) \rightarrow 0 \implies \|I(f^{(n)}) - I(f)\|^2 \rightarrow 0.$$

Das 'Integral' $I(\cdot)$ zum Maß mit orthogonalen Werten $M(\cdot)$ ist eine Isometrie des Hilbertraums $L^2(S, \mathfrak{G}, d\mu)$ auf den Teil-Hilbertraum von $\mathfrak{H}$, der von den $M(S)$ mit $S \in \mathfrak{G}$ aufgespannt wird.

$$\langle I(f), I(g) \rangle_{\mathcal{H}} = \langle f, g \rangle_{\mu} = \int f(s) \cdot \overline{g(s)} d\mu(s) \quad \text{für } f, g \in L^2(S, \mathfrak{G}, d\mu).$$

Beispiel 3.5.1 (Atomares Leistungsmaß). Es sei $(S, \mathfrak{G})$ ein messbarer Raum, $\{s_j : j \in J\}$ eine abzählbare Familie von Punkten in S , und $\{v(s_j) : j \in J\}$ eine Familie paarweise orthogonaler Elemente eines Hilbertraums $\mathcal{H}$ mit $\sum_j \|v(s_j)\|^2 < \infty$. Für $S \in \mathfrak{G}$ sei $M(S) = \sum_{\{j:s_j \in S\}} v(s_j)$. $M(\cdot)$ ist dann ein Maß mit orthogonalen Werten. Das Leistungsmaß dazu ist das Maß $\mu(\cdot)$ mit $\mu(S) = \sum_{\{j:s_j \in S\}} \|v(s_j)\|^2$ für $s \in \mathfrak{G}$.

Beispiel 3.5.2 (Prozesse mit orthogonalen Zuwächsen über der Parametermenge $T = \mathbb{R}$). Es sei $\mathbf{Z} = \{Z^s : s \in \mathbb{R}\}$ eine Familie zentrierter komplexwertiger quadratintegrabler Zufallsgrößen mit

1. $\sup\{\text{var}(Z^{s_2} - Z^{s_1})\} < \infty$
2. $s_1 < s_2 < s_3 < s_4 \implies \text{cov}(Z^{s_4} - Z^{s_3}, Z^{s_2} - Z^{s_1}) = 0$.
3. $\forall s \quad s^n \downarrow s \quad \text{var}(Z^{s^n} - Z^s) \rightarrow 0$.

Die Familie $\mathbf{Z}$ nennt man dann einen zentrierten rechtsstetigen Prozess mit orthogonalen Zuwächsen.

Es existiert genau ein Maß mit orthogonalen Maßen $M(\cdot)$ auf dem borel'schen Parameterraum $(S, \mathfrak{S}) = (\mathbb{R}, \mathfrak{B})$ mit

$$M((s', s'']) = Z^{s''} - Z^{s'} \quad \text{für alle halboffenen Intervalle } (s', s''].$$

Das dazugehörige Leistungsmaß $\mu(ds)$ ist das Maß zur Verteilungsfunktion $F(\cdot)$:

$$F(s) = \lim_{s' \rightarrow -\infty} \text{var}(Z^s - Z^{s'}).$$

Das Integral bzgl. $M(\cdot)$ bezeichnet man auch mit

$$I(f) = \int f dM = \int f(s) dZ^s \quad \text{für } f \in L^2(\mathbb{R}, \mathfrak{B}, dF(\cdot)).$$

Man beachte aber, dass es sich nicht um ein stochastisches Integral im Sinne der modernen Theorie der stochastischen Integration handelt. Vom Integrator dZ^s ist nicht gefordert, dass es sich um ein Semimartingal handelt.

bf Konstruktion: Es sei $\{Z(\lambda) : \lambda \in (0, 2\pi]\}$ eine rechtsstetige Schar von Elementen eines Hilbertraums $\mathcal{H}$ mit orthogonalen Zuwächsen und $\|Z(0)\|^2 = 0$, $\|Z(2\pi)\|^2 = \sigma^2 < \infty$. Es existiert genau ein Maß mit orthogonalen Werten auf der Borel algebra über $\mathbb{R}/2\pi$ mit $M((\lambda', \lambda'']) = Z(\lambda'') - Z(\lambda')$ für $0 < \lambda' < \lambda'' \leq 2\pi$. Das Intensitätsmaß $\mu(\cdot)$ ist das borelsche Maß mit der Verteilungsfunktion $F(\lambda) = \|Z(\lambda)\|^2$.

Für $n \in \mathbb{Z}$ sei $X^n = \int e^{in\lambda} dM(\lambda)$. Für $h \in \mathbb{Z}$ gilt dann

$$\langle X^{n+h}, X^n \rangle_{\mathcal{H}} = \left\langle \int e^{i(n+h)\lambda} dM(\lambda), \int e^{in\lambda} dM(\lambda) \right\rangle = \int f(s) \cdot \overline{g(s)} d\mu(s).$$

Wir gelangen mit der Konstruktion noch näher an die Welt der schwachstationären Zeitreihen, wenn wir annehmen, dass der Hilbertraum $\mathcal{H}$ ein Raum zentrierter quadratintegrierbarer Funktionen (auf irgendeinem Wahrscheinlichkeitsraum) ist. Diese Funktionen interpretieren wir als Zufallsgrößen mit dem Erwartungswert = 0. Für die Zufallsgrößen $Y = \int f(\lambda) dM(\lambda)$, $W = \int g(\lambda) dM(\lambda)$. Die X^n bilden eine zentrierte schwachstationäre Zeitreihe zum Spektralmaß μ .

$$\text{var} X^n = \sigma^2, \quad \text{cov}(X^{n+h}, X^n) = \gamma(h) = \int_0^{2\pi} f(s) \cdot \overline{g(s)} d\mu(s)$$

Der folgende Satz kann als Umkehrung dieser Konstruktion verstanden werden.

Satz 3.5.4. *Es sei $\{X^n : n \in \mathbb{Z}\}$ eine zentrierte schwachstationäre Zeitreihe mit der Kovarianzfolge $\gamma(h) = \int_0^{2\pi} e^{ih\omega} d\mu(\omega) = \int_0^{2\pi} e^{ih\omega} dF(\omega)$. ($F(\cdot)$ rechtsstetig auf $(0, 2\pi]$.) Es existiert dann ein wohlbestimmtes Maß mit zentrierten unkorrelierten Werten auf der Borel algebra über $(0, 2\pi]$, sodass*

$$X^n = \int e^{in\omega} dM(\omega) = \int e^{in\omega} dZ(\omega), \quad \text{wobei } Z(\lambda) = \int_0^\lambda dM(\omega).$$

Beweis. Den Raum aller Linearkombinationen der Zufallsgrößen X^n machen wir zu einem Prä-Hilbertraum mit der Norm

$$\left\| \sum a_j X^j \right\|^2 = \sum_{jk} a_j \bar{a}_k \gamma(j-k) = \int \left| \sum_j a_j e^{ij\omega} \right|^2 d\mu(\omega).$$

Wir erhalten eine isometrische Abbildung dieses Prä-Hilbertraums auf den Raum der trigonometrischen Polynome im Raum $L^2(\mathbb{R}/2\pi, \mathfrak{B}, \mu)$ (wobei $\mathfrak{B}$ die Borelalgebra ist), wenn wir abbilden

$$\sum a_j X^j \longmapsto \sum a_j e^{ij\omega}$$

Die Fortsetzung auf die Vervollständigung ist ein Hilbertraumisomorphismus des von den X^n erzeugten Hilbertraums $\mathcal{H}$ auf den Raum $L^2(\mathbb{R}/2\pi, \mathfrak{B}, \mu)$. Jedes bzgl. μ quadratintegrierbare g ist das Bild genau einer Zufallsgrößen Y aus $\mathcal{H}$; und es gilt $\text{var}Y = \int |g(\omega)|^2 d\mu(\omega)$. Das Urbild der Indikatorfunktion des Intervalls $(0, \lambda]$ sei mit $Z(\lambda)$ bezeichnet. Die Schar $\{Z(\lambda) : 0 < \lambda \leq 2\pi\}$ ist der gesuchte Prozess mit unkorrelierten Zuwächsen. Für $0 < \lambda' < \lambda'' \leq 2\pi$ gilt in der Tat

$$\text{cov}(Z(\lambda'), Z(\lambda'')) = \int 1_{(0, \lambda']} 1_{(0, \lambda'']} d\mu = \mu((0, \lambda']) = F(\lambda').$$

Und für alle $n \in \mathbb{Z}$ gilt $\int e^{in\omega} dZ(\omega) = X^n$.

Beachte: Wenn man einen schwachstationären Prozess, der nicht gaussisch ist, mit Hilfe des eben konstruierten Maßes mit orthogonalen Werten darstellt,

$$X^t = \int e^{it\omega} dZ(\omega),$$

dann darf man nicht erwarten, dass die Anteile in disjunkten Frequenzbereichen irgendwie stochastisch unabhängig sind. Sie sind nur unkorreliert. Wenn man sich mit den Pfaden eines (strikt) stationären Prozesses befassen will, dann bleibt es ziemlich rätselhaft, wie die verschiedenen Frequenzbereiche zusammenwirken. Die Reichweite der mit der Spektralzerlegung stochastischer Prozesse verbundenen Einsichten wird (meines Erachtens) häufig überschätzt. Sie ist nicht zu vergleichen mit der Bedeutung der formal verwandten Spektraldarstellung einparametrischer Gruppen unitärer Operatoren in der Quantentheorie.

Prozesse mit fastperiodischen Pfaden; Schlussbemerkungen. Die folgenden Konstruktionen von Prozessen zu einem vorgegebenen Spektralmaß sollten die Lücke zwischen der linearen Theorie und den Fragen um das Pfadverhalten beleuchten. Sie sind nach Meinung vieler Lehrbücher lehrreich; sie liefern aber Prozesse, deren Pfadverhalten aus der Sicht der Stochastik (wie wohl auch aus der Sicht der Anwendungen) uninteressant erscheint. Ähnlich wie die L^2 -Theorie passen sie nur auf einige wenige Aspekte der stationären Zeitreihen. Es sei W eine komplexe standardnormalverteilte Zufallsgröße,

und Λ eine davon unabhängige Zufallsgröße mit der Verteilung ν auf $\mathbb{R}/2\pi$. Die Folge $X^n = \sigma W e^{in\Lambda}$ ist dann eine zentrierte gaussische Zeitreihe $\mathbf{X}$ mit dem Spektralmaß $d\mu(\omega) = \sigma^2 d\nu(\omega)$. In der Tat gilt $\mathfrak{E}X^n = 0$ und

$$\text{cov}(X^{n+h}, X^n) = \mathfrak{E}X^{n+h} \cdot \bar{X}^n = |\sigma|^2 \mathfrak{E}|W|^2 \cdot \mathfrak{E}e^{ih\Lambda} = |\sigma|^2 \cdot \int e^{ih\omega} d\nu(\omega).$$

Die Kovarianzfolge ist die Folge der charakteristischen Koeffizienten des Maßes $d\mu(\omega) = |\sigma|^2 d\nu(\omega)$. Die Pfade sind (fast)periodisch mit einer zufälligen Kreisfrequenz und einer zufälligen komplexen Amplitude. Wenn der Zufall die komplexe Amplitude σW und die zufällige Kreisfrequenz spezifiziert hat, dann liegt der gesamte Pfad fest. Wenn man einen Pfad in irgendeiner unendlichen Vergangenheit beobachtet hat, dann kann man offenbar den weiteren Verlauf exakt vorhersagen. Die Bestimmung der Kreisfrequenz kann allerdings nicht durch eine lineare Funktion der X^n bewerkstelligt werden. Wenn das Spektralmaß eine logarithmisch integrable Dicht besitzt, dann ist $\mathbf{X}$ im Sinne der Wold-Zerlegung regulär. Das Lot von X^0 auf den Hilbertraum $\text{span}\{X^{-1}, X^{-2}, \dots\}$ hat eine positive Länge. In diesem Falle existiert ein weisses Rauschen $\mathbf{Z}$ mit einer gewissen Intensität σ_∞^2 und eine quadratsummable Folge, sodass für alle $n \in \mathbb{Z}$ gilt $X^n = Z^n + \sum \chi_j Z^{n-j}$.

Die weiteren Konstruktionen werden übersichtlicher, wenn man Prozesse in kontinuierlicher Zeit betrachtet, $\mathbf{X} = (X^t)_{t \in \mathbb{R}}$. Das kontinuierliche Analogon zu der obigen Konstruktion benützt σW wie oben und eine Zufallsgröße Λ mit einer Verteilung $d\nu(\omega)$ auf $\mathbb{R}$. Die Pfade sind periodisch. Die Kovarianzfunktion ist die charakteristische Funktion des Maßes $\sigma^2 d\nu(\omega)$. Man kann zeigen: Die Bedingung an die Spektraldichte, welche die Regularität im kontinuierlichen Fall garantiert, lautet

$$\int_{-\infty}^{\infty} \ln f(\omega) \frac{d\omega}{1 + \omega^2} > -\infty.$$

Im regulären Fall existiert ein quadratintegrable Funktion $\chi(s)$ und ein kontinuierliches 'weisses Rauschen' dZ^s , sodass $X^t = \int_0^\infty \chi(s) dZ^s$. Unter einem 'weissen Rauschen' (über der Zeitachse $\mathbb{R}$) ist hier ein Maß mit orthogonalen Werten auf $(\mathbb{R}, \mathfrak{B})$ zu verstehen, dessen Intensitätsmaß die Gleichverteilung ist. Das 'weisse Rauschen' zu unserem Prozess mit periodischen Pfaden erscheint als ein rätselhaftes Konstrukt.

Verallgemeinerte Konstruktion:

Es seien $W_1, W_2, \dots$ unabhängig standardnormalverteilt und $\sigma_1, \sigma_2, \dots$ Zahlen mit $\sum_m |\sigma_m|^2 = \sigma^2$. $\Lambda_1, \Lambda_2, \dots$ sei eine Folge von Zufallsgrößen mit den Verteilungen $\nu_1, \nu_2, \dots$, die von den W^m unabhängig ist. Die Summe $X^t = \sum_m \sigma_m W_m \cdot e^{it\Lambda_m}$ liefert einen stationären Prozess mit dem Spektralmaß $\mu(d\omega) = \sum |\sigma_m|^2 \nu_m(d\omega)$. In der Tat gilt

$$\begin{aligned} \text{cov}(X^{t+h}, X^t) &= \sum_{mn} \sigma_m \bar{\sigma}_n \mathfrak{E}(W_m \cdot \bar{W}_n) \mathfrak{E}(e^{i(t+h)\Lambda_m} \cdot e^{-it\Lambda_n}) = \\ &= \sum_m |\sigma_m|^2 \mathfrak{E}e^{ih\Lambda_m} = \sum_m |\sigma_m|^2 \int e^{ih\omega} d\nu_m(\omega). \end{aligned}$$

Wir können hier an den Fall denken, wo die $d\nu_m(\cdot)$ paarweise trägerfremd sind. Wir gewinnen die $|\sigma_m|^2 d\nu_m(\omega)$ aus dem Spektralmaß $\mu(d\omega)$ durch eine disjunkte Zerlegung der Menge der Frequenzen: $\mathbb{R} = \sum I_m$ (oder im zeitdiskreten Fall $\mathbb{R}/2\pi = \sum I_m$). Mit $|\sigma_m|^2 = \mu(I_m)$ und $d\nu_m(\omega) = |\sigma_m|^{-2} 1_{I_m}(\omega) d\mu(\omega)$, erhalten wir $d\mu$ als positive Linearkombination von Wahrscheinlichkeitsmaßen mit disjunkten Trägern.

Es ist eine gewagte Vorstellung, wenn man in Prozessen dieser Art Verwandte erblicken will zu gaussischen Prozessen der folgenden Art. Es sei $\sigma(\omega)$ eine quadratintegrale Funktion auf $\mathbb{R}/2\pi$ (bzw. auf $\mathbb{R}$ im zeitkontinuierlichen Fall) und $d\mu(\omega) = |\sigma(\omega)|^2 d\omega$. ($W(\omega) : \omega \in \mathbb{R}/2\pi$) sei eine Brown'sche Bewegung. Daraus konstruieren wir zu jedem t das stochastische Integral $X^t = \int e^{it\omega} \sigma(\omega) dW(\omega)$. Die Schar $(X^t)_t$ ist dann ein schwachstationärer Prozess mit dem Spektralmaß $|\sigma(\omega)|^2 d\omega$, über dessen Pfadverhalten man durchaus einiges sagen kann. Man denke an die Konstruktion des Ornstein-Uhlenbeck-Prozesses.

Die Pfade der oben zu einer Partition $\sum I_m$ konstruierten Prozesse sind fastperiodisch. Bei der letzten Konstruktion als veritables stochastisches Integral ist das ganz anders. Man muss tiefer in die Theorie der stochastischen Integrale eindringen, wenn man bei der zuletzt vorgestellten spektral zerlegten Prozess das Pfadverhalten diskutieren will. Die oben diskutierte Begriffsbildung eines Maßes mit orthogonalen Werten liefert keinen passenden Ansatz.