Nichtlineare Dynamik und Zeitreihenanalyse

— 2. Übung —

1 Weißes Rauschen

Die Datenreihe *whitenoise.dat* wurde in das Plot– und Analyseprogramm Grace (Version 5.1.12 vom 19. April 2003, unter Debian GNU Linux) geladen (`xmgrace whitenoise.dat`). Es handelt sich dabei um unabhängige gaußverteilte Zufallszahlen mit einer Standardabweichung $\sigma = 1.0$. Es wurden die Abbildungen der ersten Iterierten

$$x(t+1) = f[x(t)]$$

und der zweiten Iterierten

$$x(t+2) = f[x(t)]$$

erstellt und geplottet (vgl. Abb. 1.2 und 1.3). Dazu wurden die Datenreihen zunächst dupliziert (*Edit/Data sets/Edit/Duplicate*). Anschließend wurde der erste Punkt der ersten und der letzte Punkt der zweiten Datenreihe gelöscht (*Data/Data set operations/Drop points*), für die zweite Iterierte entsprechend jeweils die ersten bzw. letzten zwei Datenpunkte. Zum Plot wurde die Funktion *Data/Transformations/Evaluate expression* verwendet, links die zweite Datenreihe markiert und als Formel '$x = sn.y$' eingegeben (wobei n für die Nummer der ersten Datenreihe steht).

Anschließend wurde der zugehörige Korrelationskoeffizient $C(x, x+n)$

$$C(x, x+1) = -0.01468402$$
$$C(x, x+2) = 0.004307217$$

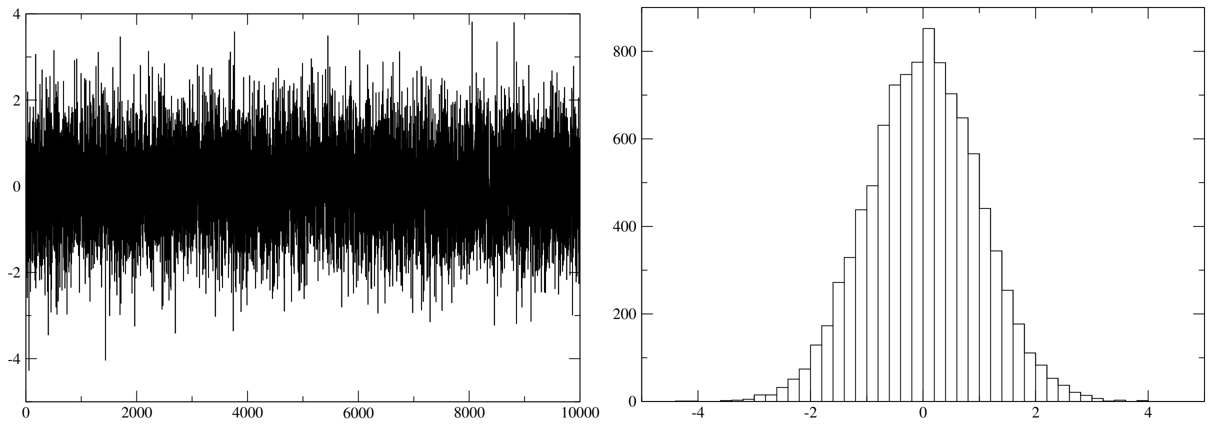

Abb. 1.1: Weißes Rauschen. Darstellung der Zeitreihe (x,t) (links) und Histogramm mit 50 bins (entsprechend einer "bin width" von 200 Samples). Aus dem Histogramm läßt sich gut die den Daten zugrundeliegende Verteilung ablesen: Es handelt sich um eine Gaußsche Normalverteilung.

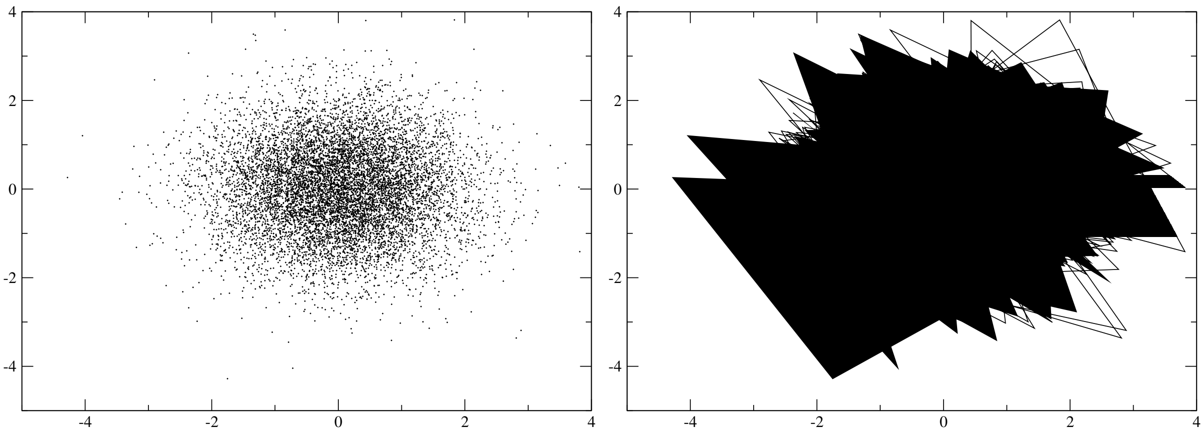

Abb. 1.2: Erste Iterierte des weißen Rauschens. Für eine Darstellung der ursprünglichen Zeitreihe vgl. Abb. 1.1. Links die reinen Datenpunkte, rechts die verbundenen Datenpunkte. Aus beiden Abbildungen geht deutlich hervor, daß aufeinanderfolgende Datenpunkte im Mitel nicht korreliert sind.

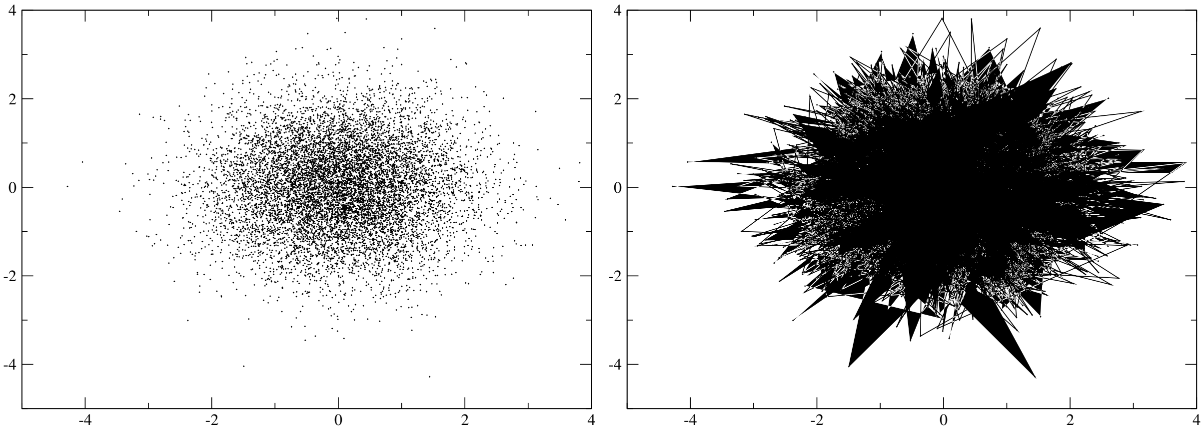

Abb. 1.3: Zweite Iterierte des weißen Rauschens. Für eine Darstellung der ursprünglichen Zeitreihe vgl. Abb. 1.1. Links die reinen Datenpunkte, rechts die verbundenen Datenpunkte. Aus beiden Abbildungen geht deutlich hervor, daß aufeinanderfolgende Datenpunkte im Mitel nicht korreliert sind.

über lineare Regression (*Data/Transformations/Regression*) bestimmt.

Zum Vergleich der so erhaltenen Korrelationskoeffizienten wurde die Autokorrelationsfunktion (AKF) $C(k)$ der Datenreihe berechnet (*Data/Transformations/Correlation,covariance*). Für *Maximum lag* wurde zunächst ein Wert von 1000 gewählt. Dieser Wert gibt an, bis zu welchem Wert k die AKF berechnet (numerisch geschätzt) wird. Die in der wissenschaftlichen Literatur empfohlenen Werte reichen dabei von $\frac{1}{4}$ bis zu $\frac{1}{10}$ der Länge der Datenreihe.

Zur Darstellung der AKF vergleiche Abb. 1.4. Einmal wurde die gesamte AKF im Überblick dargestellt, einmal ein Ausschnitt der ersten zwanzig Werte, um die Werte für kleine k deutlich ablesen zu können. Die Werte für $k = 1$ und $k = 2$

$$k_1 = -0,014628454$$
$$k_2 = 0,0043875906$$

wurden über die Funktion *Edit/Data sets/Edit/Edit data/In spreadsheed* ermittelt. Ein Vergleich mit den aus der (linearen) Regression der ersten und zweiten Iterierten ermittelten Korrelationskoeffizienten zeigt, daß sich die Werte nicht signifikant voneinander unterschei-

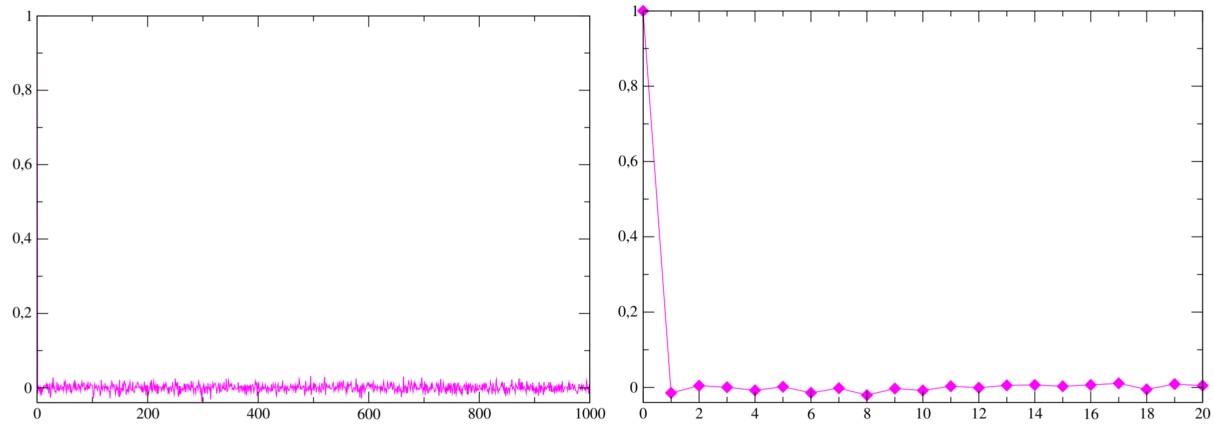

Abb. 1.4: Autokorrelationsfunktion (AKF) des weißen Rauschens, berechnet mit einem *maximum lag* von 1000 (entsprechend 10% der Länge der Datenreihe). Für eine Darstellung der ursprünglichen Zeitreihe vgl. Abb. 1.1. Links die gesamte AKF im Überblick, rechts ein Ausschnitt der ersten zwanzig Werte.

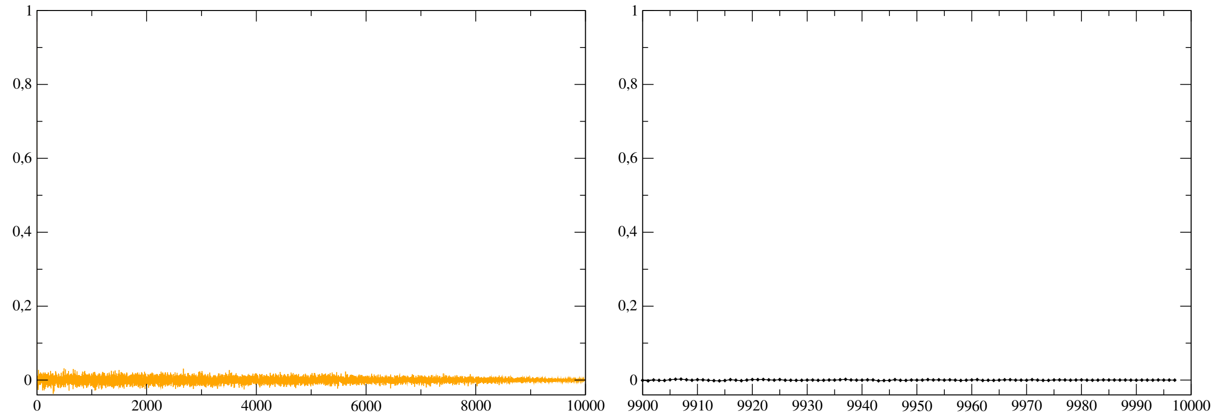

Abb. 1.5: Autokorrelationsfunktion (AKF) des weißen Rauschens, berechnet mit einem *maximum lag* von 9998 (entsprechend 99.98% der Länge der Datenreihe). Für eine Darstellung der ursprünglichen Zeitreihe vgl. Abb. 1.1. Links die gesamte AKF im Überblick, rechts ein Ausschnitt der letzten hundert Werte. Deutlich sichtbar ist, daß mit abnehmender Zahl der Werte, über die gemittelt wird, die Schwankungsbreite der Werte für k abnimmt.

den, was zu erwarten war.

In einem zweiten Schritt wurde ein *maximum lag* von 9998 — und damit nicht mehr von 10%, sondern von 99.98% der Länge der Datenreihe — zur Berechnung der AFK eingesetzt. Die resultierende Funktion ist in Abb. 1.5 dargestellt. Deutlich zu sehen ist hier, daß mit abnehmender Zahl der Werte, über die gemittelt wird, die Schwankungsbreite der Werte für k abnimmt. Das ist ein Artefakt der Berechnung der AKF. Hieran wird deutlich, warum die AKF nur mit einem Wert für *maximum lag* deutlich kleiner als die Länge der Datenreihe berechnet werden sollte.

2 Periodische Signale mit Rauschen

Die Datenreihe *sn1000.10.dat* wurde in Grace geladen (vgl. Abb. 2.1). Es handelt sich hierbei um eine Sinuszeitreihe mit einer Frequenz von 1000 Hz und einer Abtastfrequenz von 44100

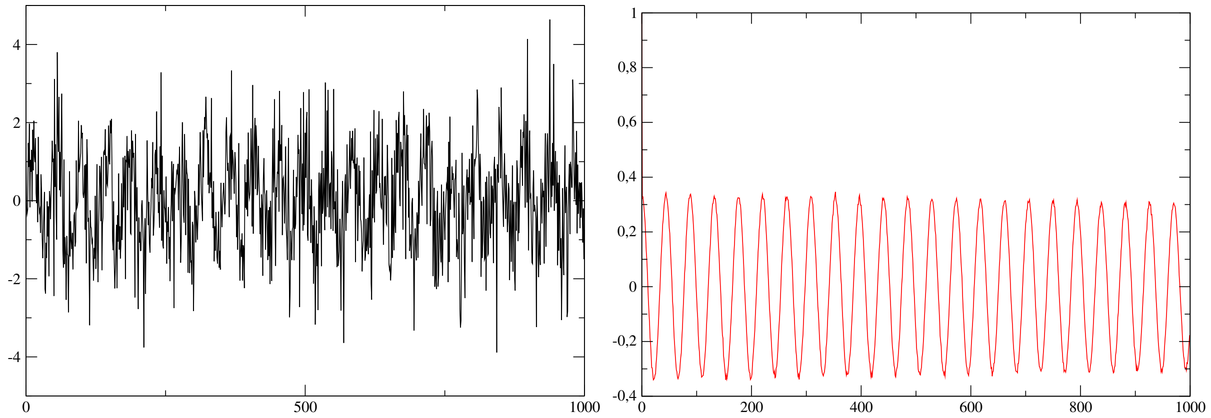

Abb. 2.1: Periodische Signale mit Rauschen. Links ist eine Sinuszeitreihe mit einer Frequenz von 1000 Hz und einer Abtastfrequenz von 44100 Hz mit starkem Rauschanteil abgebildet. Die Standardabweichung der gaußverteilten Zufallszahlen des Rauschanteils betrug $\sigma_{\mathrm{R}} = 1$. Gezeigt ist aus Gründen der Auflösung nur das erste Zehntel der Zeitreihe. Rechts abgebildet ist die Autokorrelationsfunktion (AKF) $C(k)$ der Zeitreihe mit einem *maximum lag* von 1000. Da die AKF einer periodischen Zeitreihe ihrerseits periodisch ist und die identische Periodenlänge aufweist, kann sie zur Bestimmung der Periode stark verrauschter Signale verwendet werden, wie hier demonstriert ist.

Hz mit einer Standardabweichung σ_{R} der gaußverteilten Zufallszahlen des Rauschanteils von $\sigma_{\mathrm{R}} = 1$. Die Amplitude des Sinus betrug ebenfalls 1.

Zur Bestimmung der Periode des stark verrauschten Signals kann die Autokorrelationsfunktion (AKF) herangezogen werden. Dabei macht man sich die Tatsache zunutze, daß die AKF eines periodischen Signals ihrerseits wieder periodisch ist und dieselbe Periodenlänge aufweist wie das Ausgangssignal.

Zur Abschätzung der Periodenlänge wurde die Länge (in Samples) für je zehn und zwanzig Perioden der AKF in Grace am Graphen abgelesen und durch die Sampling–Frequenz $f_s = 44\,100$ Hz geteilt, um die Periodendauer in Sekunden zu erhalten. Für zehn Perioden ergab sich eine Abschätzung von 44.04 Samples pro Periode, für 20 Perioden von 44.11 pro Periode, was umgerechnet sehr genau einer Periodendauer von $\frac{1}{1000}$ Sekunde entspricht. Dieser Wert war zu erwarten, da die Datenreihe laut Angabe eine Sinus–Zeitreihe mit der Frequenz $f = 1000$ Hz ist, die mit einer Sampling–Frequenz von $f_s = 44\,100$ Hz aufgenommen wurde.

3 Filtern

Wurde bei der ersten Übung der Mittelwertfilter (*running average filter*) hinsichtlich seiner Fähigkeit, sowohl das Rauschen als auch die Periodizität eines Signales zu reduzieren bzw. zu eliminieren, charakterisiert, soll er hier mit dem Dreiecksfilter

$$y_i = \frac{1}{4}\left(x_{i-1} + 2x_i + x_{i+1}\right)$$

mit den Gewichten (*weights*)

$$\beta_{-1} = \frac{1}{4} \qquad \beta_0 = \frac{1}{2} \qquad \beta_1 = \frac{1}{4}$$

verglichen werden. Dafür wurde in Grace ein Datensatz (*data set*) mit den Gewichten angelegt (*Edit/Data sets/Edit/Create new/In spreadsheed*), wobei die x–Werte als Index dienten

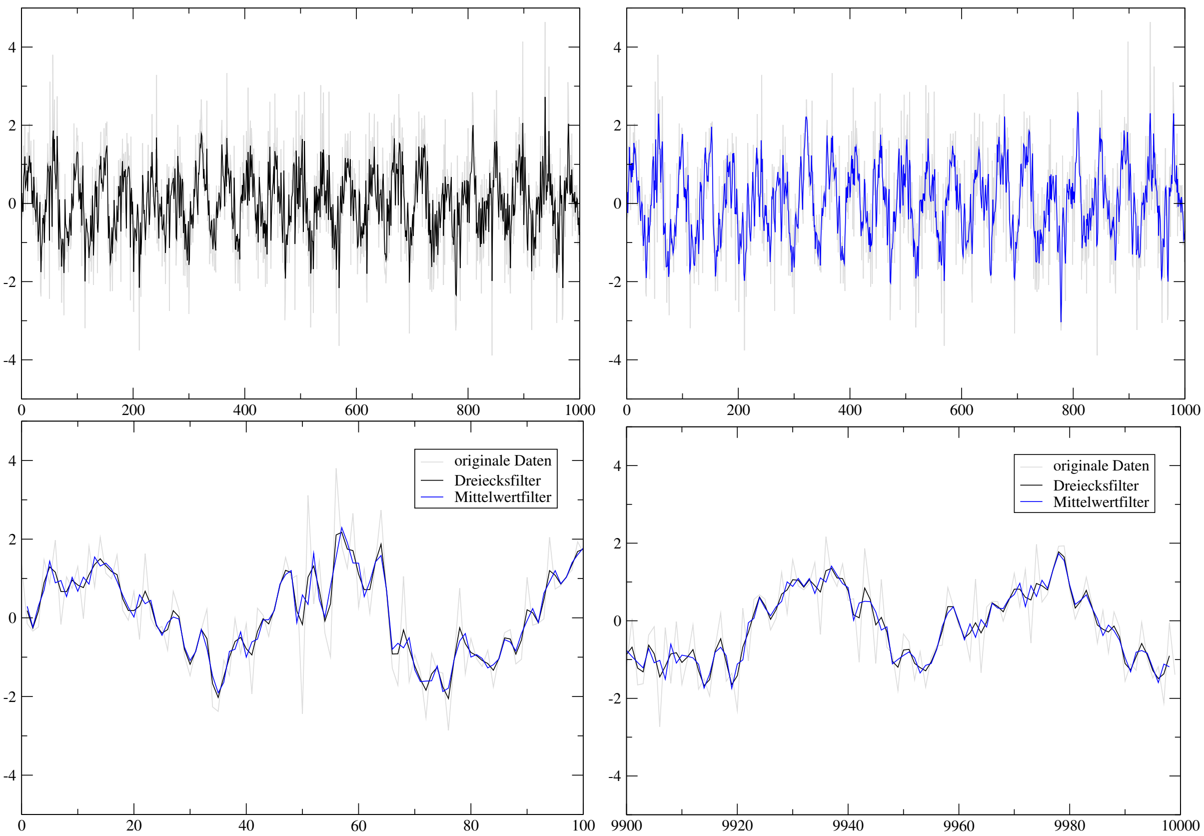

Abb. 3.1: Glättung des verrauschten periodischen Signals mittels Dreiecks– und Mittelwertfilter (*running average filter*). Verwendet wurde dieselbe Zeitreihe wie in Abb. 2.1 mit einer Standardabweichung des gaußschen Rauschens von 1.0. Obere Zeile: Im linken Graphen ist die Glättung mittels Dreiecksfilter, im rechten mittels *Running average*–Filter der Fensterbreite 3 zu sehen. Im Hintergrund bei beiden Abbildungen das ursprüngliche Signal mit Rauschanteilen (große Amplitude). Gezeigt ist immer nur ein Ausschnitt aus der Zeitreihe. Untere Zeile: Vergleich der Wirkung der beiden Filter für jeweils einen kleinen Ausschnitt aus der Zeitreihe. Der Dreiecksfilter folgt den Schwankungen des Ausgangssignales besser und reduziert deshalb die Varianz schwächer als der Mittelwertfilter.

und für die y–Werte die Datenwerte (Gewichte) eingetragen wurden. Dieser Filter wurde dann auf die schon in der vorangegangenen Aufgabe genutzte Datenreihe *sn1000.10.dat* (verrauschtes Sinus–Signal mit $f_s = 44\,100$ Hz und $f = 1000$ Hz, Standardabweichung des Rauschanteils $\sigma_\mathrm{R} = 1$) angewandt (*Data/Transformations/Digital filter*).

Anschließend wurde dieselbe Datenreihe mit einem Mittelwertfilter (*running average filter*) der Fensterbreite 3 gefiltert (*Data/Transformations/Running averages*). Ein Vergleich beider Filterarten ist in der Abb. 3.1 zu sehen. Zur Charakterisierung der Filter und ihrer Unterschiede wurde die Reduktion der Varianz bzw. Standardabweichung der Datenreihe betrachtet:

$$\begin{aligned}\sigma_x &= 1.21823\\ \sigma_\triangle &= 0.92622\\ \sigma_\mathrm{RA} &= 0.90363\end{aligned}$$

Es bedeuten: σ_x: Standardabweichung der ursprünglichen Datenreihe; $\sigma_\triangle$: Standardabweichung nach Filterung mit dem Dreiecksfilter; σ_RA: Standardabweichung nach Filterung mit dem Mittelwertfilter (*running average filter*, RA). Aus den Daten geht eindeutig hervor, daß der Mittelwertfilter die Standardardabweichung und damit die Varianz stärker reduziert.