Methoden der angewandten Schwingungsphysik zur Analyse technischer Systeme

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von
Jörg Daniel Wichard
aus
Ziegenhain

Göttingen 2000

D7

Referent:	Prof. Dr. Werner Lauterborn
Korreferent:	Prof. Dr. Ingo Borchert
Tag der mündlichen Prüfung:	21.06.2000

Zusammenfassung

Die vorliegende Arbeit faßt die Ergebnisse meiner Tätigkeit als Doktorand in der Arbeitsgruppe „Nichtlineare Dynamik“ am III. Physikalischen Institut der Universität Göttingen zusammen. Die behandelten Problemfelder entstammen dem Verbundprojekt „Analyse von Maschinen- und Prozeßzuständen mit Methoden der Nichtlinearen Dynamik“ (AMP-NLD), welches vom Bundesministerium für Bildung und Forschung in Verbindung mit dem Verein Deutscher Ingenieure gefördert wurde. Unsere Projektpartner waren die Technische Hochschule Darmstadt und das Maschinenbauunternehmen Carl Schenck AG.

Die Zusammenarbeit hatte zum Ziel, die Methoden zur Untersuchung nichtlinearer dynamischer Systeme auf technische Prozesse und Produktionsabläufe auszudehnen und ihre praktische Einsatzfähigkeit zu testen. Unserem Industriepartner oblag dabei die Aufgabe, die möglichen Anwendungsfelder NLD-basierter Methoden zu spezifizieren und uns die benötigten Meßdaten zur Verfügung zu stellen. Dadurch erklärt sich der anwendungsorientierte Charakter der Fragestellungen, die zum Teil in den Maschinenbau und in die angewandte Mechanik hineinreichen. Dabei hat der Blick aus Richtung eines abstrakteren physikalischen Standpunktes viel zur Klärung der Probleme beigetragen.

Im ersten Teil dieser Arbeit wird ein Verfahren zur Schätzung der Unwucht eines Rotors im Hochlauf vorgestellt. An diesem Beispiel konnte die Tragfähigkeit von Konzepten der NLD im praktischen Einsatz unter Beweis gestellt werden. Zur Vorhersage der Unwucht benutzten wir ein lokales Modell und entwickelten auf dieser Grundlage ein Verfahren, das in der industriellen Fertigung zum Einsatz kommen wird.

Im zweiten Teil wird die Schwingung einer gleitgelagerten Welle untersucht. Aus mehreren Gründen ist es wünschenswert, die genaue Position einer rotierenden Welle im Lager zu jedem Zeitpunkt zu kennen, z.B. um Beschädigungen des Lagers durch Überlastung zu vermeiden oder um

entstehende Wellenrisse bereits im Frühstadium diagnostizieren zu können. Die Positionsbestimmung der Welle im Lager wurde bisher durch eine Meßanordnung mit zwei induktiven Abstandssensoren durchgeführt. Dabei wurde nicht berücksichtigt, daß die Inhomogenitäten der Wellenoberfläche das Sensorsignal stören und die Abstandsmessung verfälschen. Wir entwickelten ein Verfahren zur Signaltrennung, das zusammen mit einem erweiterten Meßaufbau eine Trennung der einzelnen Signalkomponenten ermöglicht.
Im letzten Teil der Arbeit steht die von Huang et al. entwickelte „Empirical Mode Decomposition" (EMD) im Vordergrund. Dieses relativ neue Verfahren zur Zeit-Frequenz-Analyse wird vorgestellt, und seine praktischen Einsatzmöglichkeiten werden diskutiert. Eine Grundlage der EMD bildet der Begriff des analytischen Signals, dessen heuristische Einführung wir durch ein mathematisches Argument stützen, indem wir die Definition des analytischen Signals mit einem Resultat der Funktionentheorie in Verbindung bringen.

Ergründe, wie die Dinge selber sind, indem du sie zerlegst in Stoff, Ursache, Zweck.

Marc Aurel, Selbstbetrachtungen, XII

Inhaltsverzeichnis

Kapitel 1

Einleitung

1.1 Datenanalyse in der industriellen und technischen Praxis

Zu Beginn des neuen Jahrtausends stehen wir am Anfang einer Epoche, die von vielen Menschen als das hereinbrechende „Informationszeitalter“ bezeichnet wird. Der Aufbau eines globalen Datennetzes, in dem man auf unübersehbare Mengen von Dokumenten zugreifen kann, läßt diese Bezeichnung gerechtfertigt erscheinen. Die rasante Entwicklung der letzten Jahre wird zu einer umfassenden Veränderung der Gesellschaft in fast allen Bereichen des täglichen Lebens führen, deren Auswirkungen kaum abzusehen sind.
Von diesen Umwälzungen bleibt auch die Arbeitswelt nicht unberührt. Die komplette interne Vernetzung der Betriebe ermöglicht eine produktionsbegleitende Datenerfassung meßbarer Größen und Parameter, von der Lagerhaltung der kleinsten Schraube bis hin zum Auslieferungstermin des fertigen Produkts. Dabei wird die manuelle Kontrolle und Steuerung vieler produktionstechnischer Anlagen durch den Menschen in zunehmendem Maße von rechnergesteuerten Überwachungsanlagen und Qualitätssicherungssystemen übernommen. Die stürmische Entwicklung der Rechnerleistung und der verfügbaren Menge an Massenspeicher ist dafür mitverantwortlich. Diese Ent-

wicklung wird nach wie vor durch das Mooresche Gesetz beschrieben[1]. So ist es heute üblich, Daten aus Produktions- und Fertigungsprozessen aufzunehmen und dauerhaft zu speichern, um die gesammelten Informationen einer nachgeschalteten Analyse zu unterziehen und die gewonnenen Erkenntnisse langfristig zur Verbesserung des Produktionsablaufs und zur Steigerung der Produktivität einzusetzen.
Ebenso ist es heute möglich, entscheidende Größen und Parameter eines Prozesses in Echtzeit zu überwachen und zu kontrollieren, um mögliche Schadensszenarien bereits im Vorfeld abfangen zu können. Aufgrund der enormen Ausfallkosten sind die Betreiber technischer Anlagen daran interessiert, entstehende Schäden und den Grad der Abnutzung wichtiger Anlagenkomponenten während des Betriebs der Anlagen zu erkennen, um sie bei der nächsten Routinewartung schnell und zielgerichtet beheben zu können.
Für viele dieser Anwendungen sind die bewährten Methoden aus der linearen System- und Signaltheorie von großem Nutzen. Diese konventionellen Analysemethoden gehören heute zum festen Bestandteil der Ingenieursausbildung und haben sich in der industriellen Praxis etabliert. Sie sind in der Regel mit geringem Aufwand zu implementieren und basieren auf einem weitgehend abgeschlossenen theoretischen Fundament. Zudem sind die Ansprüche an Rechenzeit- und Speicherbedarf nach heutigen Maßstäben eher bescheiden.
Die zunehmende Leistungsfähigkeit moderner Rechner ermöglicht den Einsatz wesentlich komplexerer Analysewerkzeuge, deren Erprobung und Entwicklung noch in vollem Gange ist. Zur Signalklassifikation und zur Erkennung spezifischer Grundmuster wurden z.B. verschiedene Clustering-Algorithmen eingesetzt [65, 64], die ihren Ursprung in der KI-Forschung haben. Auch der Einsatz neuronaler Netze zur Vorhersage und Klassifizierung von Signalen wurde bei unserem Industriepartner erfolgreich erprobt [66].
Die Liste solcher Beispiele läßt sich beliebig verlängern und zeigt, daß der Einsatz moderner Verfahren zur Datenanalyse in produktionstechnischen Syste-

[1] Von Gordon E. Moore 1965 formuliertes Gesetz, nach dem sich die Packungsdichte der Transistoren auf einem Mikroprozessor alle 18 Monate verdoppelt. Daraus ergibt sich für unsere Computer eine Vervierfachung der Speicherkapazität alle drei Jahre und eine Verzehnfachung der Geschwindigkeit etwa alle 3,5 Jahre.

men in vollem Gange ist und sich den angestammtem Methoden der linearen Systemtheorie erfolgreich zur Seite stellt.

1.2 Die Wahl der Methoden

Zu Beginn der sechziger Jahre wurde aufgrund der Ergebnisse numerischer Simulationen von E.M. Lorenz wiederentdeckt, daß Systeme, die einfachen nichtlinearen Gleichungen genügen, bereits äußerst komplexes Verhalten zeigen können [49]. Derartige Phänomene werden unter dem Begriff *Chaos* zusammengefaßt, um sie gegen stochastische Systeme mit einer Vielzahl von Freiheitsgraden abzugrenzen [47]. Aus dem Studium solcher Systeme entwickelte sich die Nichtlineare Dynamik (NLD), die sich als eigenständige Disziplin in der Mathematik [5, 22, 67, 78] und in der Physik [3, 4, 42, 48, 71] etabliert hat. Da sich die Erkenntnisse der NLD auf viele natürliche und technische Systeme übertragen lassen, ist aus dem ursprünglichen Spezialgebiet mit der Zeit ein breites Spektrum von Anwendungen erwachsen. Ausgehend von einfachen Modellen wurde ein begrifflicher und methodischer Apparat entwickelt, der detaillierte Einblicke in das komplexe Verhalten dynamischer Systeme erlaubt. Auf dieser Grundlage ist es in zunehmendem Maße möglich, ausgesuchte Systeme in technischen Anwendungen zu analysieren, zu optimieren und zu steuern, wie beispielweise die Leistung von Festkörperlasern [70, 69].

Um die Anwendungsfelder der NLD zu erweitern, wurden vom Bundesministerium für Bildung und Forschung Projekte zwischen Industrie und Hochschule mit dem Ziel gefördert, verfügbares Wissen von den Hochschulen in neue innovative Technologien zu überführen. Von diesem Leitgedanken wurde auch das Verbundprojekt „Analyse von Maschinen- und Prozeßzuständen mit Methoden der Nichtlinearen Dynamik“ getragen, in dessen Umfeld diese Arbeit entstanden ist. Im Vorfeld des Projektes hatte sich bereits gezeigt, daß sich auf der Basis nichtlinearer Modellbildung relevante Informationen aus technischen Systemen extrahieren lassen [83]. Gemeinsam mit unserem Industriepartner gingen wir daran, die Problemfelder einzugrenzen, in denen

sich NLD-basierte Analysemethoden einsetzen lassen.
Die Fragestellungen, denen wir im Rahmen dieses Projektes nachgegangen sind, und die Probleme, die wir erfolgreich lösen konnten, sind der Gegenstand dieser Arbeit.

1.3 Der Aufbau der Arbeit

Im zweiten Kapitel wird der Auswuchtvorgang eines Rotors untersucht. Im wesentlichen geht es darum, die Messung der Unwucht zu modellieren, um die Unwucht bereits in der Hochlaufphase schätzen zu können. Die Vorhersage durch lokale Modelle, die in der NLD zu den üblichen Vorhersageverfahren gehört, konnte hier erfolgreich eingesetzt werden. Die Patentanmeldung der daraus erwachsenen Erfindung wurde beim Deutschen Patent- und Markenamt eingereicht.
Die Untersuchung der Schwingungen einer gleitgelagerten Welle ist der Gegenstand des dritten Kapitels. Dabei stellen wir ein Verfahren zur Signaltrennung vor, das es erlaubt, die Bewegung der Welle im Lager genauer als bisher zu verfolgen. Dieses Verfahren wurde ebenfalls zur Patentreife gebracht.
Im vierten Kapitel stellen wir das Konzept des analytischen Signals vor, für dessen heuristische Einführung eine neue Begründung gegeben wird, die sich auf ein solides mathematisches Argument berufen kann. Wir erbringen den Beweis, daß die Definition des analytischen Signals über die Hilbert-Transformation zwingend wird, wenn das analytische Signal der Randwert einer holomorphen Funktion sein soll. Der Einsatz des analytischen Signals im Rahmen der Zeit-Frequenz-Analyse wird anhand der „Empirical Mode Decomposition“ im fünften Kapitel erläutert. Dieses relativ neue Konzept der nichtlinearen Zeitreihenanalyse wird vorgestellt und seine Leistungsfähigkeit an Datensätzen aus der Praxis getestet und diskutiert.
Eine abschließende Betrachtung der Ergebnisse findet sich am Ende der Arbeit.

Kapitel 2

Auswuchten im Hochlauf

In diesem Kapitel soll ein Verfahren vorgestellt werden, das im Rahmen des Verbundprojektes „Analyse von Maschinen- und Prozeßzuständen mit Methoden der Nichtlinearen Dynamik“ entstanden ist. Die Aufgabe bestand darin, die Unwucht eines Rotors während der Hochlaufphase zu schätzen. Es war bekannt, daß die Arbeiterinnen und Arbeiter an den Auswuchtmaschinen die Unwucht während der Hochlaufphase zu Beginn der Messung vorhersagen konnten, doch es war nicht gelungen diese menschliche Fähigkeit zu automatisieren. Wir konnten dieses Problem lösen, indem wir einen Ansatz aus der NLD verwendeten und an die praktischen Gegebenheiten anpaßten. Die daraus hervorgegangene Erfindung wurde als Patentantrag unter dem Titel „Unwuchtschätzung im Hochlauf mittels Nächster-Nachbarn-Suche“ beim Deutschen Patentamt in München eingereicht.

2.1 Begriffsbestimmung des Auswuchtens

Ein Rotor gilt als *ausgewuchtet*, wenn eine seiner zentralen Hauptträgheitsachsen mit der Verbindungslinie der Lagerstellen-Mittelpunkte, d.h. der von feststehenden Lagern erzwungenen Drehachse, zusammenfällt. Ein ausgewuchteter Rotor überträgt weder Fliehkräfte auf seine Lager noch veranlaßt er die Lager zu umlauffrequenten Schwingungen [25].

Bei der industriellen Produktion von Rotoren (z.B. gewickelte Kommutatoranker von Elektromotoren, Turbinenräder, Kurbelwellen) kommt es infolge von Materialinhomogenitäten oder fertigungsbedingten Toleranzen oft zu einer Verschiebung zwischen Drehachse und einer der Hauptträgheitsachsen. Ein solcher unausgewuchteter Rotor führt zu großen dynamischen Lagerkräften, die hohen Verschleiß und entsprechend kurze Lebensdauer von Welle und Lager zur Folge haben. Außerdem trägt ein schlecht ausgewuchteter Rotor auch zu lästiger Geräuschbildung durch Luft- und Körperschall bei. Um

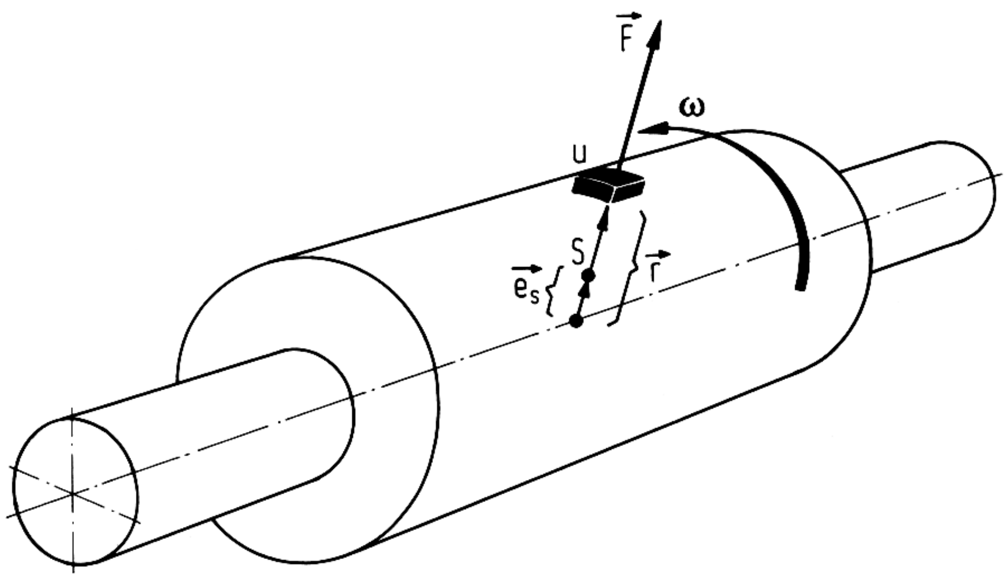


Abbildung 2.1: Zur Beschreibung der Unwucht denke man sich einen ausgewuchteten Rotor, an dem eine Überwuchtmasse u mit Radius $\vec{r}$ angebracht ist. Der Schwerpunkt S des Rotors, der vorher auf der Drehachse lag, verschiebt sich dadurch um $\vec{e}_s$. In Richtung von $\vec{e}_s$ wirkt die Fliehkraft $\vec{F}$.

solche Effekte weitgehend zu vermeiden, wird der Rotor ausgewuchtet. Dabei versucht man die Massenverteilung des Rotors so zu verändern, daß seine zentrale Hauptträgheitsachse mit der Drehachse zusammenfällt und dadurch alle am Rotor wirkenden Fliehkräfte miteinander im Gleichgewicht stehen.
Zur quantitativen Beschreibung der Unwucht denke man sich eine punktförmige Einzelmasse u bei $\vec{r}$ in der Radialebene eines ausgewuchteten Rotors der Masse m befestigt. Dadurch verlagert sich der Schwerpunkt

des Rotors um

$$\vec{e}_s = \frac{u}{m+u} \cdot \vec{r}$$

von der Drehachse (siehe Abbildung 2.1). Wenn der Rotor nun mit der Drehfrequenz ω umläuft, so wirkt auf ihn die Fliehkraft

$$\vec{F} = \omega^2 (m+u) \cdot \vec{e}_s = \omega^2 u \cdot \vec{r} \ ,$$

die als Lagerkraft in den starren Lagern des Rotors in Erscheinung tritt.
Die Einzelmasse u nennt man in der Auswuchttechnik auch *Überwuchtmasse* [72]. Das Produkt

$$\vec{U} \ := \ u \cdot \vec{r} \tag{2.1}$$

wird in der Literatur als *Unwucht* bezeichnet. Ziel des Auswuchtens ist es nun, diese Unwucht durch geeignete Gegenmaßnahmen (z. B. durch das Anbringen von Gegengewichten) zu beseitigen.
Für die technische Praxis ist hierbei der *Fundamentalsatz der Auswuchttechnik für starre Rotoren* besonders wichtig [25]:

> Sämtliche Unwuchten eines Rotors lassen sich stets zu zwei komplementären Unwuchten in zwei an sich beliebigen Radialebenen zusammenfassen, in Ausnahmefällen auch zu einer einzigen Unwuchtresultierenden.

Die Bedeutung dieses Satzes für die Auswuchttechnik wird klar, wenn man bedenkt, daß man die Ausgleichsgewichte in der Praxis nicht an beliebigen Stellen des Rotors anbringen kann, sondern diesbezüglich konstruktionsbedingten Einschränkungen unterworfen ist.
Aus diesem Grund legt man in der Regel zwei Auswuchtebenen fest, in denen der Auswuchtvorgang vorgenommen werden kann. In Abbildung 2.2 ist dies schematisch für einen Rotor dargestellt, dessen Auswuchtebenen auf die Stirnseiten gelegt worden sind. Die Unwucht des Rotors ist in die beiden Unwuchten U_1 und U_2 aufgeteilt, die in den entsprechenden Radialebenen liegen.

Um das Problem mathematisch zu beschreiben, kann man nun die einzelnen Unwuchten U_1 und U_2 als komplexwertige Größen des Rotors auffassen, indem man die jeweiligen Radialebenen, in denen die Einzelunwuchten liegen, mit der komplexen Zahlenebene identifiziert. Die Unwucht eines (idealen) starren Rotors ist unabhängig von seiner Drehzahl.
Zur Messung der Unwucht benutzt man in der technischen Praxis normalerweise dynamische Meßverfahren, die aus den Lagerkräften eines umlaufenden Rotors die Unwucht in den beiden Unwuchtebenen bestimmen. Gemessen werden in der Regel die am Lager auftretenden Schwingungen als Funktion der Frequenz des Rotors. Dies ist zum Beispiel beim *Wattmeter-Verfahren* der Fall, auf dem die von uns untersuchten Meßwerte beruhen[1]. Durch ein geeignetes Kalibrierungsverfahren ist es immer möglich, von den Schwingungen am Lager auf die verursachende Unwucht zurückzuschließen. Bei einer typischen Auswuchtmaschine, die das Wattmeter-Verfahren zur Bestimmung der Unwucht benutzt, gibt es zwei Anzeigeinstrumente (für jeweils eine der beiden Auswuchtebenen) in Form eines Polarzeigerdiagramms (siehe Abbildung 2.3). Während eines Unwuchtlaufs, also während der zu untersuchende Rotor auf eine bestimmte Drehzahl hochläuft, wandert ein Lichtzeiger über die Skalen der Anzeigeinstrumente. Hat der Rotor eine bestimmte Drehzahl erreicht, so ruht der Lichtzeiger auf der Skala und gibt dem Bediener des Gerätes Auskunft über die zum Auswuchten benötigten Ausgleichsgewichte, denn nach der Kalibrierung repräsentiert dieser Zeiger die komplexe Unwucht des Rotors. Dabei liefert die Länge des Vektors im Polarzeigerdiagramm die Information zur benötigten Auswuchtmasse, und der Polarwinkel gibt die radiale Richtung an, in welcher das Ausgleichsgewicht am Rotor angebracht werden muß.
Für die weiteren Betrachtungen sind die näheren Umstände der Datenaufnahme und des Meßverfahrens jedoch ohne Belang. Wesentlich ist nur, daß jede Unwuchtmessung an einem Rotor zwei Datensätze liefert, nämlich die frequenzabhängigen Schwingungen in den beiden komplexen Auswuchtebe-

[1] Eine ausführliche Beschreibung der verschiedenen Meßverfahren und eine theoretische Herleitung des Fundamentalsatzes der Auswuchttechnik finden sich in dem Buch von K. Federn [25], dem auch die beiden Abbildungen 2.1 und 2.2 entnommen wurden.

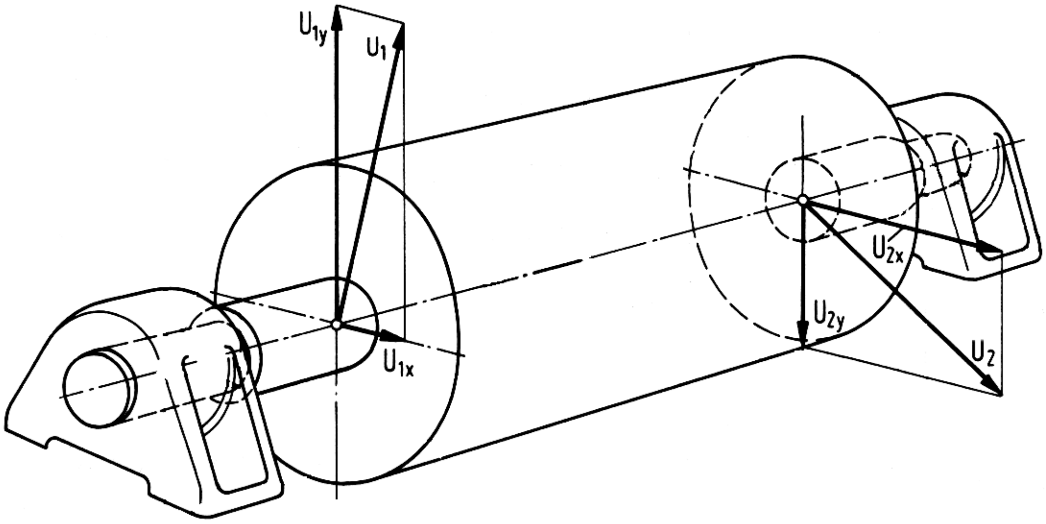


Abbildung 2.2: Die Abbildung veranschaulicht den Fundamentalsatz der Auswuchttechnik. Sämtliche Unwuchten eines Rotors wurden in den beiden komplementären Unwuchten U_1 und U_2 zusammengefaßt. Die Auswuchtebenen wurden auf die Stirnseiten des Rotors gelegt, um ein bequemes Anbringen der Ausgleichsgewichte zu ermöglichen.

nen, die der Unwucht des Rotors zugeordnet werden können. Den Hochlauf können wir also als Einschwingvorgang betrachten, wobei sich der Lichtzeiger auf den richtigen Wert bewegt und denselben erreicht, sobald sich der Rotor mit konstanter Drehzahl dreht. Die Bahn des Zeigers als Funktion der Drehzahl ω in einer Auswuchtebene wird dann beschrieben durch die Abbildung

$$\begin{aligned} U : \mathbf{R} &\rightarrow \mathbf{C} \\ \omega &\mapsto U(\omega) \,. \end{aligned} \tag{2.2}$$

Die Daten, die wir im folgenden untersuchen, wurden in äquidistanten Zeitschritten aufgenommen und einer numerischen Simulation des Wattmeter-Verfahrens unterzogen, um damit ein Bild zu erzeugen, das dem Polarzeigerdiagramm einer Auswuchtmaschine entspricht. In Abbildung 2.3 sind drei dieser Unwuchtläufe wiedergegeben.

Eine Unwuchtmessung mit n Meßwerten, die bei den Drehzahlen $\omega_1, \ldots, \omega_n$

aufgenommen wurden, läßt sich als ein Vektor im $\mathbf{C}^n$ auffassen

$$\vec{U} = \begin{pmatrix} U(\omega_1) \\ \vdots \\ U(\omega_n) \end{pmatrix} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} . \tag{2.3}$$

Aus den Einzeldarstellungen von Betrag und Phase in Abbildung 2.3 entnimmt man den typischen Verlauf einer Unwuchtmessung.
Für den weiteren Verlauf unserer Untersuchungen wird der Begriff des *Abstands zwischen zwei Unwuchtmessungen* von Bedeutung sein. Da wir einen Unwuchtlauf als einen Vektor im $\mathbf{C}^n$ auffassen können, liegt es nahe, den Abstand zwischen zwei Unwuchtmessungen als den euklidischen Abstand der beiden Vektoren im $\mathbf{C}^n$ aufzufassen. Dabei gehen wir von dem praxisrelevanten Fall aus, daß nur Unwuchtmessungen derselben Länge verglichen werden. Als Abstand $A(\vec{U}, \vec{V})$ zwischen zwei Unwuchtläufen $\vec{U}, \vec{V} \in \mathbf{C}^n$ wird folgender Ausdruck definiert:

$$\begin{aligned} A(\vec{U}, \vec{V}) &:= \|\vec{U} - \vec{V}\|^2 \\ &= \sum_{k=1}^{n} (u_k - v_k)\overline{(u_k - v_k)} \\ &= \sum_{k=1}^{n} \left(\Re[u_k - v_k]\right)^2 + \left(\Im[u_k - v_k]\right)^2 . \end{aligned} \tag{2.4}$$

Diese Abstandsdefinition werden wir in Abschnitt 2.4 noch erweitern, wenn wir die Drehinvarianz des Problems genauer untersuchen.

2.2 Problemstellung

Unsere Aufgabe bestand darin, den Wert der Unwucht aus dem Anfangsverlauf der Messung, also *im Hochlauf*, zu schätzen. Als Datenbasis standen mehrere Unwuchtmessungen zur Verfügung, die bei der Schenck RoTec GmbH von Herrn A. Buschbeck durchgeführt wurden. Der Versuch, die Unwucht im Hochlauf zu schätzen, wurde durch folgende Beobachtung motiviert: Die Arbeiterinnen und Arbeiter an den Auswuchtmaschinen waren in der Lage, Betrag und Phase der Unwucht aus dem Anfangsverlauf der Unwuchtmessung

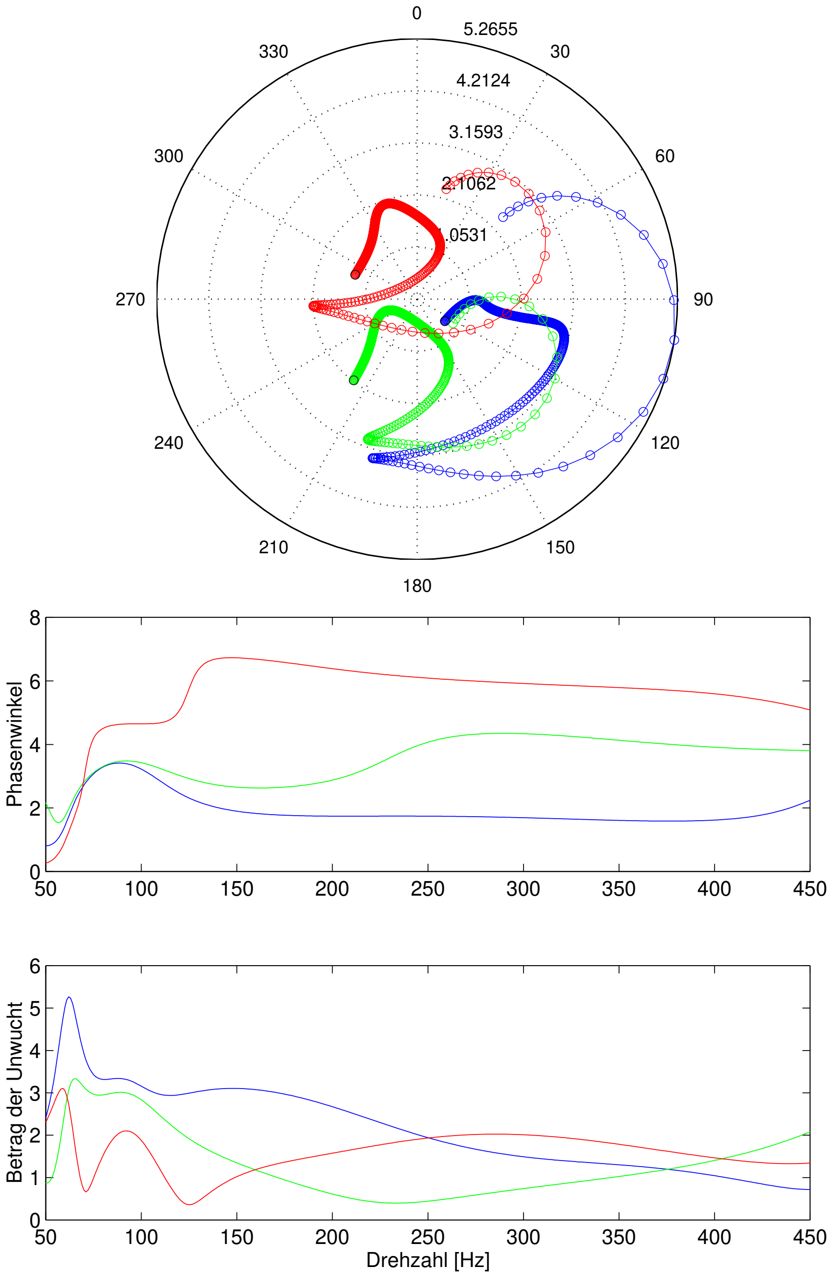


Abbildung 2.3: Die Meßdaten von drei verschiedenen Unwuchtläufen aus einer Serie von Elektroaggregaten. Im oberen Bild sieht man die Darstellung im Polarzeigerdiagramm, in den unteren beiden Abbildungen sind jeweils Betrag und Phase der Unwuchtläufe als Funktion der Drehzahl des Rotors dargestellt. Die Messung beginnt erst bei 50 Hz, da im unteren Drehzahlbereich die Resonanzfrequenzen der Auswuchtmaschine durchfahren werden. Für uns ist der Bereich um die maximale Drehzahl von 450 Hz interessant, für den wir Betrag und Phase der Unwucht aus dem Anfangsverlauf der Messung vorhersagen wollen.

abzuschätzen und konnten dadurch die Messung vor Erreichen der maximalen Drehzahl abbrechen, und somit den gesamten Ablauf des Auswuchtvorgangs beschleunigen. Diese Tatsache war seit geraumer Zeit bekannt, doch es mangelte an einer technischen Umsetzung dieser menschlichen Fähigkeit in einen automatisierten Prozeß. Da es möglich ist, den Endwert eines Unwuchtlaufs aus dem Anfangsverlauf vorherzusagen, liegt der Schluß nahe, daß die Information über den Endwert im Beginn des Laufs verborgen liegt. Die Aufgabe einer automatisierten Vorhersage liegt darin, diese Information zu extrahieren. Die ersten Versuche führten uns auf klassische Methoden wie *Principal Component Analysis* und *Archetypal Analysis* [17, 18]. Die Ergebnisse waren zwar erfolgversprechend, doch der Bedarf an Rechenzeit ließ die Anwendung im Echtzeitbetrieb als nahezu unmöglich erscheinen. Die Schätzung des Endwertes im Hochlauf soll sehr schnell geschehen, d.h. unmittelbar nach dem Abbruch der Unwuchtmessung im Anfangsverlauf. Auf der Suche nach einem echtzeitfähigen Verfahren führte ein Ansatz aus der nichtlinearen Datenanalyse zum Ziel.

2.3 Modellierung des Hochlaufs

Die Hochlaufphase eines unwuchtigen Rotors kann man durch *lokale Modelle* beschreiben und den Endwert des Unwuchtlaufs dadurch vorhersagen. Diese Methode der Vorhersage wird in der Nichtlinearen Dynamik häufig verwendet und soll im folgenden näher erläutert werden. Dazu müssen zuerst einige technische Begriffe eingeführt werden. Eine weitergehende Einführung in die NLD findet sich z.B. in [3, 5, 42]. Unter dem *Fluß* eines dynamischen Systems versteht man die stetig differenzierbare Abbildung

$$\begin{aligned} \varphi : \mathbf{R}^d \times \mathbf{R} &\to \mathbf{R}^d \\ (\vec{x}, t) &\mapsto \varphi(\vec{x}, t) \end{aligned} \tag{2.5}$$

mit den Eigenschaften

$$\begin{aligned} \varphi(t=0, \vec{x}) &= \vec{x}\ , \\ \varphi(s, \varphi(t, \vec{x})) &= \varphi(s+t, \vec{x}) \qquad \forall s, t \in \mathbf{R}\ ;\ \ \forall \vec{x} \in \mathbf{R}^d. \end{aligned} \tag{2.6}$$

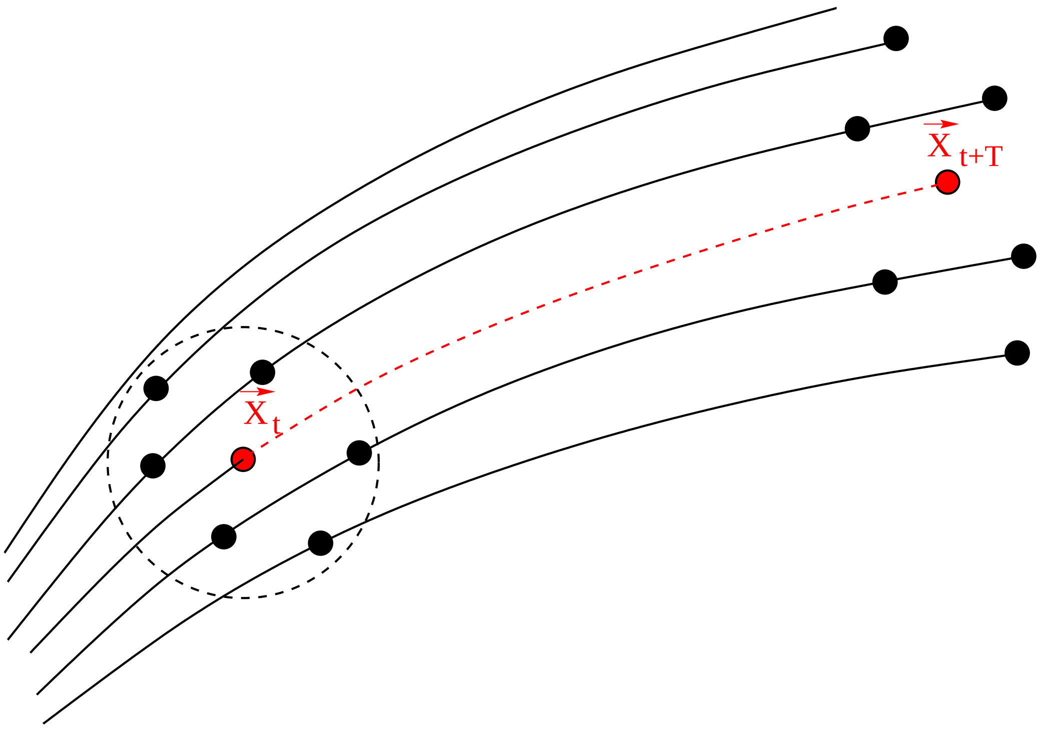


Abbildung 2.4: Diese Abbildung verdeutlicht die Vorhersage eines Systemzustands durch ein lokales Modell. Die zeitliche Entwicklung des Zustandes $\vec{x}_t$ soll vorhergesagt werden. Dazu sucht man sich die benachbarten Systemzustände (angedeutet durch die schwarzen Punkte im gestrichelten Kreis um $\vec{x}_t$) und versucht, eine Abbildung zu finden, welche die zeitliche Entwicklung dieser benachbarten Punkte beschreibt. Diese Abbildung verwendet man, um die zukünftige Entwicklung $\vec{x}_{(t+T)}$ vorherzusagen.

Jeder Zustand des Systems wird eindeutig und vollständig durch einen Punkt $\vec{x} \in \mathbf{R}^d$ in einem endlichdimensionalen reellen Vektorraum beschrieben. Ein gängiges Verfahren zur Vorhersage nichtlinearer dynamischer Systeme beruht auf der Verwendung *lokaler Modelle* [13, 24, 42, 53]. Von einem lokalen Modell spricht man, wenn man nur die Zustände in einem kleinen Gebiet des Phasenraumes oder des rekonstruierten Phasenraumes[2] zur Approximation des Flusses in dem entsprechenden Gebiet verwendet.
Grundlegend für diese Art der Modellierung ist die Stetigkeit des Flusses des

[2]Die Rekonstruktion des Phasenraumes kann dabei z.B. durch die Einführung von Zeitverzögerungskoordinaten geschehen [74, 68].

dynamischen Systems, aus der folgt, daß sich die Entwicklung eines Systemzustandes für einen kleinen Zeitraum nicht wesentlich von der Entwicklung eng benachbarter Zustände im Phasenraum (oder im Rekonstruktionsraum, der im allgemeinen ein diffeomorphes Bild der Dynamik im Phasenraum ist) unterscheidet. Dieses Verfahren ist in Abbildung 2.4 dargestellt. Für einen Zustand $\vec{x}_t$ des Systems möchte man die zukünftige Entwicklung zum Zeitpunkt $t + T$ wissen. Dazu sucht man die k nächsten Nachbarn des Zustands $\vec{x}_t$ und versucht, eine Abbildung zu finden, die diese nächsten Nachbarn auf die entsprechenden, um T weiterentwickelten Zustände abbildet. Diese Abbildung[3] benutzt man nun, um die zeitliche Entwicklung des Zustandes $\vec{x}_t$ vorauszusagen.
Diese Methode der Vorhersage mittels lokaler Modelle hat in der nichtlinearen Dynamik weite Verbreitung gefunden, da sie nicht die zum Teil sehr aufwendige Approximation der Modellgleichungen voraussetzt. Die einzige Schwierigkeit besteht in der effizienten Suche der nächsten Nachbarn (NN). Ein großes Problem der Nachbarsuche ist der Ressourcenbedarf, denn die Komplexität des Problems wächst mit der Anzahl der Datenpunkte, und bei großen Datensätzen oder der Suche in hochdimensionalen Räumen benötigen einfache Realisierungen der NN-Suche zuviel Zeit. Deshalb ist es notwendig, schnelle Algorithmen einzusetzen, um die geforderte Echtzeitfähigkeit zu gewährleisten. Ein solcher schneller NN-Algorithmus ist der von C. Merkwirth entwickelte ATRIA-Algorithmus (Advanced Triangle Inequality Algorithm) [55], der auch im Programmpaket TSTOOL zum Einsatz kommt, welches im Rahmen dieses Projektes entstand [56, 59, 60].
Bei diesem Algorithmus werden die Daten einer Vorverarbeitung unterzogen, in der ein Suchbaum aufgebaut wird, dessen Verwendung während der NN-Suche das Auffinden benachbarter Punkte extrem beschleunigt.
Der Aufbau des Suchbaumes geschieht durch rekursive Einteilung des Datensatzes in Cluster. Bei Erreichen der kritischen Clustergröße (ca. 60-200 Datenpunkte) bricht die Rekursion ab. Jeder Cluster wird durch sein Zentrum und seinen Durchmesser beschrieben. Diese Information wird später beim Durchsuchen des Baumes genutzt, um all jene Cluster von der Suche

[3] In der Literatur auch „Local Predictor“ genannt [13].

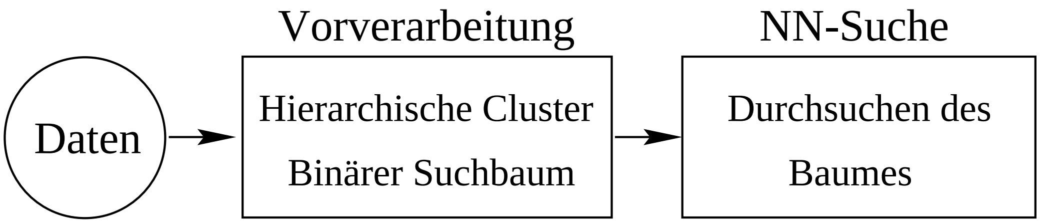


Abbildung 2.5: Schema der Verarbeitungsschritte

auszuschließen, die keine NN enthalten können. Die Verwendung eines in Vorverarbeitungs- und Suchphase aufgeteilten Algorithmus bringt einen großen Vorteil, denn das Erstellen des Suchbaumes auf einer großen Datenbasis kann bereits im Vorfeld geschehen und fällt damit zeitlich nicht ins Gewicht. Wenn man nun während des Hochlaufs den NN zu einer Unwuchtmessung sucht, so geht das Durchsuchen des Baumes sehr schnell, und der NN ist praktisch sofort bekannt.

Bei der Untersuchung des Unwuchtproblems lag es nun nahe, eine Vorhersage des Unwuchtlaufs durch ein lokales Modell zu versuchen. Natürlich liegt in diesem Fall nicht der stetige Fluß eines dynamischen Systems vor, der es erlauben würde, die lokalen Modelle zur Vorhersage eines Systemzustandes zu benutzen. Deshalb werden wir zuerst begründen müssen, warum der Einsatz lokaler Modelle trotzdem sinnvoll ist. Die Bahn des Lichtzeigers in der komplexen Auswuchtebene nehmen wir als eine stetige Funktion der Drehzahl des Rotors an. Die Bewegung des Lichtzeigers können wir, wie oben dargelegt, als Einschwingvorgang des Systems auf einen spezifischen Endwert der Unwucht auffassen. Bei der serienmäßigen Produktion von Rotoren in hoher Stückzahl kommt es bei den ansonsten völlig baugleichen Exemplaren gewöhnlich zu kleineren Unterschieden in der Massenverteilung aufgrund fertigungsbedingter Toleranzen. Unserem Ansatz liegt nun die Annahme zugrunde, daß Rotoren mit ähnlichen Masseverteilungen auch ähnliches Verhalten beim Hochlauf zeigen sollten, was sich *a posteriori* bestätigte.

Einen einzelnen Unwuchtlauf $\vec{U}$ mit k Meßwerten beschreiben wir nach Gleichung 2.3 als einen Vektor im $\mathbf{C}^k$. Der Endwert des Unwuchtlaufs, also die zu schätzende Unwucht in der komplexen Auswuchtebene, wird durch das k-te Element des Vektors $\vec{U}$ gegeben. Die Hochlaufphase wird durch die ersten n

Einträge $u_1, \ldots, u_n$ des Vektors $\vec{U}$ repräsentiert. Aus diesen Vektoren stellen wir die Datenbasis zusammen, in der die Suche nach den benachbarten Unwuchtläufen stattfinden soll. Der zugrundeliegende Vektorraum ist der $\mathbf{C}^n$.
Um die Unwucht zu einem vorgegebenen Unwuchtlauf $\vec{U}^{\text{ref}}$ im Hochlauf vorherzusagen, benutzt man den Vektor aus der Datenbasis $\vec{U}^{\text{NN}}$, der den geringsten Abstand zu $\vec{U}^{\text{ref}}$ besitzt. Dessen Endwert u_k^{NN} wird als Schätzung für u_k^{ref} benutzt.
Dieser Ansatz wurde zuerst an einem Datensatz getestet, der bei Unwuchtmessungen an Elektroaggregaten aufgenommen wurde. Er lieferte recht vielversprechende Ergebnisse, obwohl bei der Datenaufnahme die Bedingungen ungünstig waren, da die Motoren während des Auswuchtvorganges in ihren Gehäusen gelassen wurden. Auf diese Weise wurden den Schwingungssignalen der unwuchtigen Rotoren noch resonante Eigenschwingungen der Gehäuseteile überlagert. Ein Beispiel für die Vorhersage der Unwucht eines solchen Elektroaggregates ist in Abbildung 2.6 abgebildet. Die Schleife im Verlauf des Polarzeigerdiagramms wird als Durchfahren einer Resonanzfrequenz des Rotors oder des angegliederten schwingungsfähigen Systems interpretiert [25].
Nach diesem ersten Test der Unwuchtvorhersage wurden in einem zweiten Durchgang die Anker von Elektromotoren an einer Auswuchtmaschine ausgemessen, die in der industriellen Fertigung zum Einsatz kommt. Dabei wurden nur die Rotoren ohne Gehäuseteile ausgemessen, um die Störungen durch Resonanzeffekte gering zu halten. In Abbildung 2.7 sind fünf Unwuchtläufe und die entsprechenden Vorhersagen durch benachbarte Unwuchtläufe dargestellt. An diesem Beispiel sieht man die Ähnlichkeit der verschiedenen Unwuchtläufe. Obwohl sie sich in Betrag und Phase unterscheiden, scheint ihnen die gleiche Grundform zugrunde zu liegen.
Um einen Eindruck von der Leistungsfähigkeit der Unwuchtvorhersage zu bekommen, wurde das Verfahren an einem Datensatz aus 210 Unwuchtläufen in jeweils zwei Auswuchtebenen erprobt. Ein einzelner Unwuchtlauf wurde gemäß Gleichung 2.3 durch einen Vektor im $\mathbf{C}^{400}$ repräsentiert, die Hochlaufphase durch einen entsprechend verkürzten Vektor.

1. Ein Referenzlauf $\vec{U}^{\text{ref}}$ mit bekannter Unwucht u_{400}^{ref} wurde aus der Da-

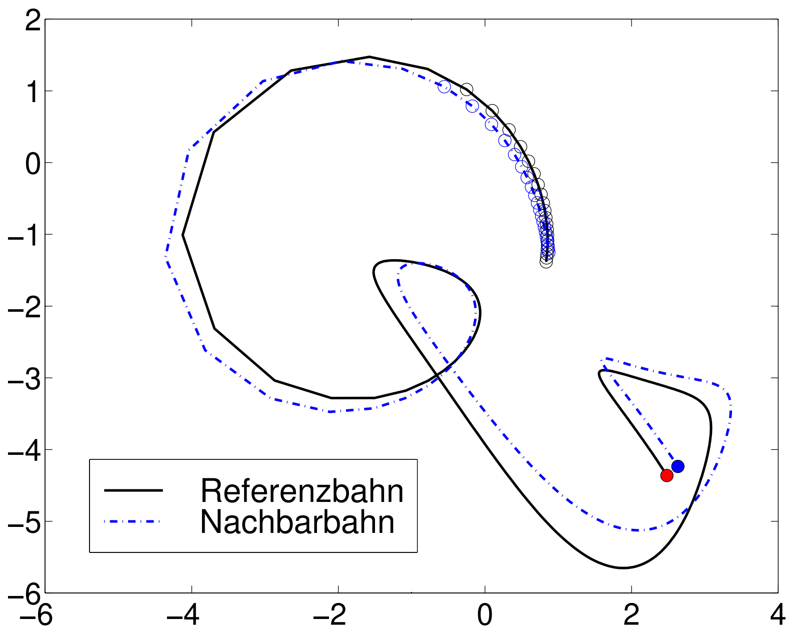


Abbildung 2.6: Diese Abbildung illustriert die NN-Vorhersage der Unwucht eines Elektroaggregats im Hochlauf. Die schwarze Linie ist die Bahn des Zeigers in der Auswuchtebene, der rote Punkt am Ende der Bahn ist die Unwucht bei konstanter Drehzahl, die geschätzt werden soll. Dazu werden die ersten 20 Werte des Unwuchtlaufs verwendet (durch schwarze Kreise am Anfang der Bahn gekennzeichnet) und mit allen anderen Unwuchtläufen der Datenbasis verglichen. Der Unwuchtlauf mit dem kleinsten euklidischen Abstand zum Referenzlauf wird zur Vorhersage der Unwucht benutzt (blauer Kreis am Ende der blau gekennzeichneten NN-Bahn). Die Schleife der Bahn ist ein Hinweis auf das Durchfahren einer Resonanzfrequenz des Systems.

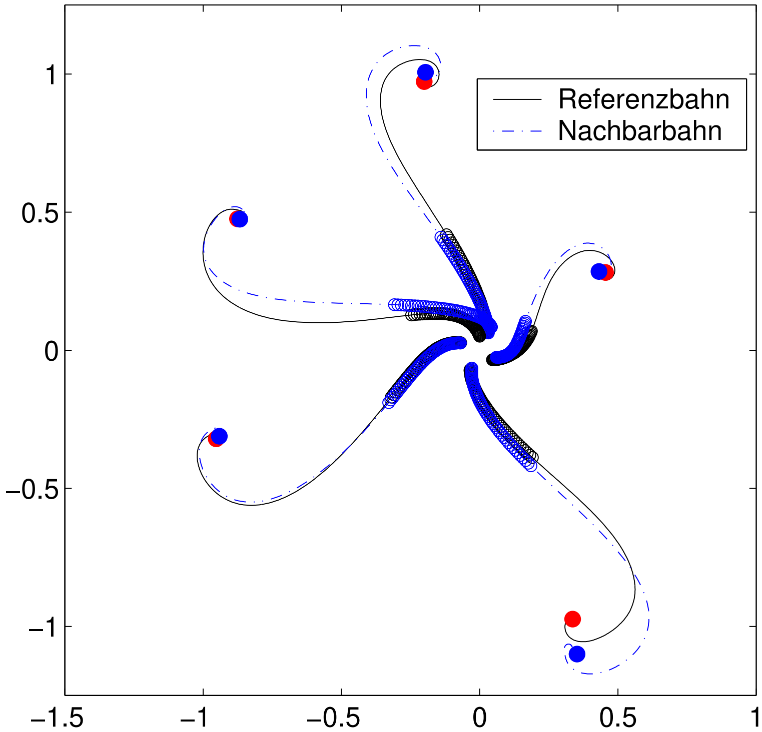


Abbildung 2.7: In dieser Abbildung ist die Unwuchtvorhersage für fünf verschiedene Rotoren (Anker von Elektromotoren) dargestellt. Die Unwucht der Referenzbahn ist durch einen roten Punkt markiert, die Vorhersage durch die benachbarte Bahn durch einen blauen Punkt. Die ersten 50 Meßwerte eines jeden Unwuchtlaufs, die zu Suche der nächsten Nachbarn und zur Vorhersage der Unwucht verwendet wurden, sind durch Kreise gekennzeichnet.

tenbasis ausgewählt.

2. Die Länge der Hochlaufphase wurde festgesetzt.

3. In der Datenbasis wurde unter allen Unwuchtläufen nach dem nächsten Nachbarn $\vec{U}^{\mathrm{NN}}$ des Referenzlaufs gesucht.

4. Die Unwucht u_{400}^{NN} wurde mit der Unwucht des Referenzlaufs verglichen und der Vorhersagefehler bestimmt.

Als Vorhersagefehler bezeichnen wir die Differenz zwischen Vorhersage der Unwucht u_{400}^{NN} und der Unwucht des Referenzlaufs u_{400}^{ref}, also die komplexe Zahl

$$u^{\mathrm{error}} \quad := \quad u_{400}^{\mathrm{NN}} - u_{400}^{\mathrm{ref}} \; .$$

Um die Fehler der Unwuchtvorhersage miteinander vergleichen zu können, müssen wir zu den Beträgen übergehen. Der absolute Fehler der Vorhersage wird dann durch den Betrag von u^{error} gegeben und der relative Fehler r_{rel} durch

$$r_{\mathrm{rel}} \quad := \quad \frac{\|u^{\mathrm{error}}\|}{\|u_{400}^{\mathrm{ref}}\|} \; .$$

In Abbildung 2.8 sind die Ergebnisse dargestellt. Der mittlere relative Fehler ist gegen die Länge der zur Vorhersage verwendeten Hochlaufphase aufgetragen. Dabei wurde über alle Vorhersagen gemittelt und die Standardabweichung σ_{rel} als Fehlerbalken eingezeichnet. Es zeigte sich, daß man ab einer Länge von 100 Meßpunkten des Hochlaufs bereits einen mittleren relativen Fehler von weniger als 10% erreicht. Dieser Fehler ist für die praktische Anwendung des Verfahrens akzeptabel und innerhalb der Fehlertoleranzen des Auswuchtvorgangs [12]. Für dieses Beispiel bedeutet das eine Verkürzung des Auswuchtvorgangs auf ein Viertel der Zeit, die ein kompletter Hochlauf erfordern würde.

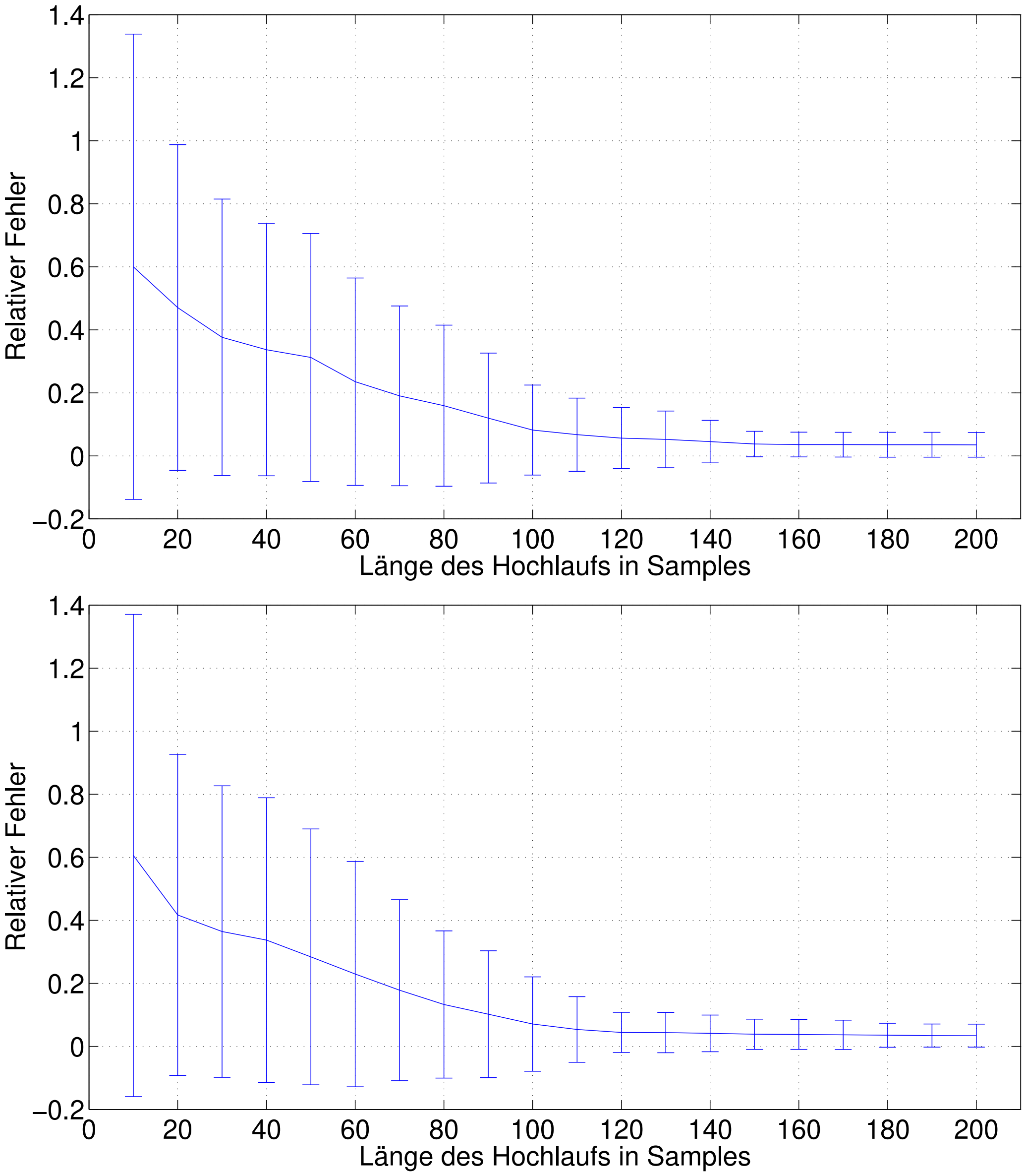


Abbildung 2.8: Der mittlere relative Fehler der Vorhersage als Funktion der Dauer der Hochlaufphase für die Auswuchtebene 1 (oben) und die Auswuchtebene 2 (unten). Ein einzelner Unwuchtlauf wurde durch 400 Meßwerte repräsentiert. Die Vorhersagen wurden über 208 Unwuchtläufe gemittelt und die Standardabweichung als Fehlerbalken eingezeichnet. Ab einer Länge der Hochlaufphase von 100 Samples, also rund einem Viertel des Unwuchtlaufs, liegt der mittlere relative Fehler unter 10%.

2.4 Die Drehinvarianz des Auswuchtproblems

Im folgenden werden wir unser Auswuchtproblem, also die Schätzung der dynamischen Unwucht nach Betrag und Phase, unter dem Gesichtspunkt der Symmetrie des Problems betrachten. Der Betrag und die Phase der Unwucht sind Größen, die fest mit dem Rotor verbunden sind, in dem Sinne, daß sie sich nicht ändern, wenn man den Rotor bei einem zweiten Unwuchtlauf in einer anderen Position in die Meßapparatur einlegt oder die Nullmarkierung verschiebt, auf die sich die Angabe der Phase der Unwucht bezieht. Diese Symmetrie bezüglich des Nullpunktes der Auswuchtebene werden wir nun ausnutzen. Ein Unwuchtlauf $\vec{U}$ mit n Meßwerten läßt sich nach Definition 2.3 darstellen als ein Vektor im $\mathbf{C}^n$. Im weiteren Verlauf werden wir die Polardarstellung dieses Vektors verwenden, also

$$\vec{U} \;=\; \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} r_1^u e^{i\theta_1^u} \\ \vdots \\ r_n^u e^{i\theta_n^u} \end{pmatrix} . \tag{2.7}$$

Eine Drehung der Bahn des Unwuchtlaufs in der Auswuchtebene um den Winkel φ wird beschrieben durch die Multiplikation mit einem Phasenfaktor:

$$D_\varphi(\vec{U}) \;:=\; e^{i\varphi}\vec{U} . \tag{2.8}$$

Besonders wichtig wird der Abstand zweier Unwuchtläufe $\vec{U}$ und $\vec{V}$ aus Definition 2.4, wenn man ihn als Funktion des Drehwinkels φ auffaßt:

$$A_\varphi(\vec{U},\vec{V}) \;:=\; \|\vec{U} - D_\varphi(\vec{V})\|^2 . \tag{2.9}$$

Die Bedeutung dieses Ausdrucks ergibt sich aus der Drehinvarianz des Auswuchtproblems. Für einen (bis auf die Überwuchtmasse) symmetrischen Rotor sollte die Auswahl des benachbarten Unwuchtlaufes nicht von dem Winkel φ abhängen. Um den benachbarten Unwuchtlauf zu einem gegebenen Referenzunwuchtlauf zu finden, werden alle Unwuchtläufe so gedreht, daß sie minimalen Abstand zum Referenzlauf haben. Unter allen dergestalt gedrehten Unwuchtläufen kann man nun den nächsten Nachbarn suchen, um den

Verlauf des Referenzdatensatzes vorherzusagen. In Abbildung 2.9 ist diese Situation veranschaulicht.
Wichtig für die Anwendung ist die Frage, ob die Funktion $A_\varphi(\vec{U},\vec{V})$ aus Definition 2.9 für zwei gegebene Unwuchtläufe $\vec{U}$ und $\vec{V}$ ein eindeutiges globales Minimum bezüglich des Drehwinkels besitzt und ob eventuell noch lokale Minima existieren, in die sich ein Suchalgorithmus verirren könnte. Antwort darauf gibt der folgende

Satz 2.1. *Für zwei Unwuchtläufe $\vec{U},\vec{V} \in \mathbf{C}^n$ hat die Abstandsfunktion $A_\varphi(\vec{U},\vec{V})$ die Darstellung*

$$A_\varphi(\vec{U},\vec{V}) = C + A\sin(B+\varphi)$$

mit den Konstanten $(C,A,B) \in \mathbf{R}_0^+ \times \mathbf{R}_0^+ \times [0,2\pi)$.

Beweis. Es seien $\vec{U},\vec{V} \in \mathbf{C}^n$ zwei Unwuchtläufe mit der Darstellung

$$\vec{U} = \begin{pmatrix} r_1^u e^{i\theta_1^u} \\ \vdots \\ r_n^u e^{i\theta_n^u} \end{pmatrix} \quad \text{und} \quad \vec{V} = \begin{pmatrix} r_1^v e^{i\theta_1^v} \\ \vdots \\ r_n^v e^{i\theta_n^v} \end{pmatrix}$$

und der Drehwinkel $\varphi \in [0,2\pi)$. Die Abstandsfunktion $A_\varphi(\vec{U},\vec{V})$ wird dann zu:

$$\begin{aligned}
A_\varphi(\vec{U},\vec{V}) &= \|\vec{U} - e^{i\varphi}\vec{V}\|^2 \\
&= \sum_{k=1}^n \left[\left(\Re\left[r_k^u e^{i\theta_k^u} - r_k^v e^{i(\theta_k^v+\varphi)}\right]\right)^2 + \left(\Im\left[r_k^u e^{i\theta_k^u} - r_k^v e^{i(\theta_k^v+\varphi)}\right]\right)^2\right] \\
&= \sum_{k=1}^n \left[(r_k^u\cos(\theta_k^u) - r_k^v\cos(\theta_k^v+\varphi))^2 + (r_k^u\sin(\theta_k^u) - r_k^v\sin(\theta_k^v+\varphi))^2\right] \\
&= \sum_{k=1}^n \left[(r_k^u)^2 + (r_k^v)^2 - 2r_k^u r_k^v\left(\cos(\theta_k^u)\cos(\theta_k^v+\varphi) + \sin(\theta_k^u)\sin(\theta_k^v+\varphi)\right)\right] \\
&= \sum_{k=1}^n \left[(r_k^u)^2 + (r_k^v)^2 - 2r_k^u r_k^v\cos(\theta_k^u - \theta_k^v - \varphi)\right] .
\end{aligned}$$

Dieser Ausdruck läßt sich zunächst einmal zusammenfassen zu

$$A_\varphi(\vec{U},\vec{V}) = C + \sum_{k=1}^n a_k\sin(b_k+\varphi)$$

mit den Konstanten

$$\begin{aligned} C &:= \sum_{k=1}^{n} \left[(r_k^u)^2 + (r_k^v)^2 \right] , \qquad (2.10) \\ a_k &:= -2 r_k^u r_k^v , \\ b_k &:= \frac{\pi}{2} - (\theta_k^u - \theta_k^v) . \end{aligned}$$

Es muß noch gezeigt werden, daß folgendes gilt:

$$\sum_{k=1}^{n} a_k \sin(b_k + \varphi) = A \sin(B + \varphi) . \qquad (2.11)$$

Die linke Seite in Gleichung 2.11 läßt sich umschreiben zu

$$\sum_{k=1}^{n} a_k \sin(b_k + \varphi) = \cos(\varphi) \sum_{k=1}^{n} a_k \sin(b_k) + \sin(\varphi) \sum_{k=1}^{n} a_k \cos(b_k) .$$

Wir setzen

$$A := \sqrt{ \left(\sum_{k=1}^{n} a_k \sin(b_k) \right)^2 + \left(\sum_{k=1}^{n} a_k \cos(b_k) \right)^2 } \qquad (2.12)$$

und müssen nun zwei Fälle unterscheiden:
1. Fall: $A \neq 0$
In diesem Fall existiert eine eindeutig[4] bestimmte Konstante $B \in [0, 2\pi)$ mit:

$$\begin{aligned} \sin(B) &= \frac{1}{A} \sum_{k=1}^{n} a_k \sin(b_k) , \qquad (2.13) \\ \cos(B) &= \frac{1}{A} \sum_{k=1}^{n} a_k \cos(b_k) . \end{aligned}$$

2. Fall: $A = 0$
Im Fall $A = 0$ läßt sich B beliebig wählen:

$$\begin{aligned} A \sin(B) &= 0 , \\ A \cos(B) &= 0 . \end{aligned}$$

Für diesen Fall setzen wir $B = 0$.
Damit ist Satz 2.1 bewiesen. □

[4]Dabei benutzen wir einen Satz aus der Analysis [35]: Gilt für zwei Zahlen $x, y \in \mathbf{R}$ die Beziehung $x^2 + y^2 = 1$, so existiert genau ein $t \in [0, 2\pi)$ mit $x = \sin(t)$ und $y = \cos(t)$.

Aus diesem Satz folgt für $A \neq 0$ die Existenz eines eindeutigen Minimums für den Abstand zweier Unwuchtläufe. Allerdings bleibt noch der mathematisch unbefriedigende Fall $A = 0$, bei dem die Konstante $B = 0$ gesetzt wird. In diesem Fall existiert kein eindeutiges Minimum der Abstandsfunktion. Aber dieser Fall besitzt keine praktische Relevanz. Bei der Auswertung der Experimente kam dieser Fall auch niemals vor. Die Erklärung dafür erhält man, wenn man versucht, den Fall $A = 0$ zu konstruieren, d.h. man muß zwei Unwuchtläufe finden, deren Abstand sich bei der oben definierten Drehung nicht ändert.

Dazu muß nach Gleichung 2.12 gelten:

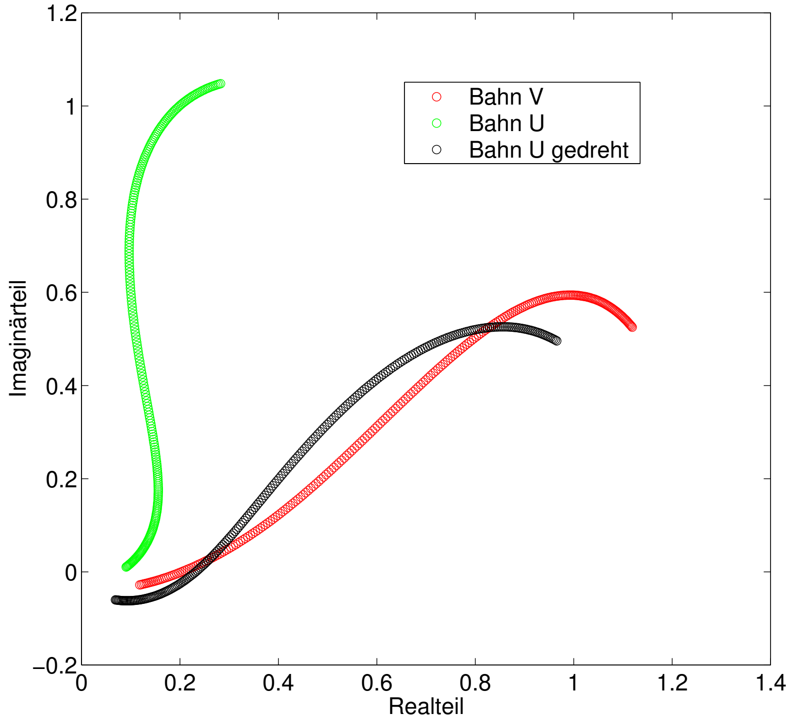


Abbildung 2.9: Diese Figur zeigt die Referenzbahn $\vec{V}$ (rot) und eine weitere Bahn $\vec{U}$ (grün), die um den Winkel $\varphi = 5.45$ RAD gedreht wurde (schwarz), um den Abstand $A_\varphi(\vec{U}, \vec{V})$ aus Definition 2.9 zu minimieren.

$$\sum_{k=1}^{n} a_k \sin(b_k) = 0 \quad \text{und} \quad \sum_{k=1}^{n} a_k \cos(b_k) = 0\,. \tag{2.14}$$

Dies ist sicher der Fall, wenn alle a_k verschwinden, also $\vec{U} = \vec{0}$ oder $\vec{V} = \vec{0}$.

Abgesehen davon, daß dieser Fall ohnehin äußerst selten auftreten wird, kann man ihn auch einfach abfangen, indem man eine untere Grenze für den Betrag der Unwucht angibt, ab der nicht mehr ausgewuchtet wird und zusätzlich alle Unwuchtläufe mit $\vec{U} = \vec{0}$ aus der Datenbasis entfernt.
Wesentlich interessanter ist der Fall $A = 0$ für $\vec{U} \neq 0$ und $\vec{V} \neq 0$, bei dem sich die Anteile der einzelnen Komponenten kompensieren. Solche Unwuchtläufe kann man erzeugen, indem man die beiden n-dimensionalen Vektoren $\vec{U}$ und $\vec{V}$ jeweils um einen Eintrag u_{n+1} bzw. v_{n+1} erweitert, so daß gilt:

$$A_\varphi(\vec{U}, \vec{V}) + (u_{n+1} - v_{n+1})\overline{(u_{n+1} - v_{n+1})} = konst.$$

Diese Argumentation läßt sich nun umkehren. Sollte einmal der Fall $A = 0$ auftreten, dann kann man durch Verlängern (bzw. Verkürzen) der Hochlaufphase um einen Meßwert bereits wieder den Fall $A \neq 0$ erhalten. Dies ist eine unmittelbare Folgerung aus Satz 2.1, wenn wir annehmen, daß die hinzugenommenen (bzw. weggenommenen) Meßwerte von Null verschieden sind.
Da nach Satz 2.1 für $A \neq 0$ immer ein globales Minimum für den Abstand zweier Unwuchtläufe als Funktion des Drehwinkels existiert, ergibt sich die Frage, wie dieses Minimum möglichst schnell gefunden werden kann. Aus der expliziten Formel für die Konstanten A, B und C in dem Ausdruck

$$A_\varphi(\vec{U}, \vec{V}) = C + A\sin(B + \varphi)$$

können wir den Winkel φ_{min}, der den kleinsten Abstand liefert, direkt berechnen. Die Konstante C aus Gleichung 2.10 und die Amplitude A aus Gleichung 2.12 sind nur bei der Berechnung des Abstands nötig, die Lage des Minimums hingegen wird nur durch B bestimmt. Aus Gleichung 2.13 entnimmt man die Beziehung

$$B = \arctan\left(\frac{\sum_{k=1}^{n} a_k \sin(b_k)}{\sum_{k=1}^{n} a_k \cos(b_k)}\right) . \tag{2.15}$$

Die Vorhersage der Unwucht im Hochlauf unter Berücksichtigung der Drehinvarianz läuft im Grunde analog zum oben beschriebenen Fall, allerdings werden die Unwuchtläufe der Datenbasis erst gedreht bis sie den geringsten Abstand zum Referenzlauf haben. Auf den gedrehten Unwuchtläufen wird dann eine Suche des nächsten Nachbarn durchgeführt, und mit diesem wird die Unwucht im Hochlauf vorhergesagt.

2.5 Vergleich der beiden Verfahren

Nun stehen zwei Verfahren zur Verfügung, um die Unwucht im Hochlauf zu schätzen. Man kann einfach die Datenbasis nach dem nächsten Nachbarn durchsuchen (statisches Verfahren) oder man kann die Unwuchtläufe vor der Suche gegeneinander drehen. Es erhebt sich die Frage, welches Verfahren in der Praxis zum Einsatz kommen sollte. Dabei sind folgende Kriterien ausschlaggebend:

- Die Qualität der Vorhersage, also der mittlere Vorhersagefehler und die zugehörige Standardabweichung.
- Die Geschwindigkeit der Vorhersage.
- Die Größe der Datenbasis, dabei gilt: je kleiner, desto besser. Eine kleine Datenbasis läßt sich relativ schnell erstellen und durchsuchen.

Diese Kriterien haben nicht alle das gleiche Gewicht. Unser Industriepartner zeigte sich vorallem darüber besorgt, daß der Aufwand zum Erstellen einer geeigneten Datenbasis den potentiellen Kunden verschrecken könnte. Aus diesem Grund hatte die zweite Methode, welche die Drehinvarianz des Problems ausnutzt, besonderen Reiz. Obwohl der Rechenaufwand größer ist, da jede Bahn erst einmal gedreht werden muß, ist die benötigte Datenbasis erheblich kleiner. In unserem Beispiel ergibt sich aus der achtzähligen Symmetrie des untersuchten Rotors eine Verkleinerung der Datenbasis um den Faktor acht. Dieses Argument war für unseren Industriepartner Grund genug, dieser Methode den Vorzug zu geben.
In Abbildung 2.10 wird der relative Vorhersagefehler der beiden Verfahren verglichen. Dabei zeigt sich, daß der relative Fehler der statischen Methode für kurze Hochlaufphasen größer ist als der relative Fehler der gedrehten Bahnen. Ab einer Länge von 90 Punkten (Ebene 1) bzw. 80 Punkten (Ebene 2) sinkt der mittlere relative Vorhersagefehler der statischen Methode unter den Fehler der gedrehten Bahnen. Für dieses Verhalten fand sich bisher keine Erklärung. Die Standardabweichung des mittleren Vorhersagefehlers

der gedrehten Bahnen ist auf jeden Fall immer kleiner als bei der statischen Methode.
Bei der Vorhersage in Abbildung 2.11 haben wir die ersten 50 Werte des Hochlaufs nicht bei der Erstellung der Datenbasis berücksichtigt. Nur die Punkte zwischen dem 50. und dem 150. Meßwert wurden als Referenzwerte herangezogen. In diesem Bereich des Unwuchtlaufs steckt besonders viel Dynamik, deshalb lag der Schluß nahe, dort besonders viel Information über den Endwert zu finden. Die Ergebnisse bestätigen diese Vermutung. Hier reichen 40 Meßwerte, um den mittleren relativen Fehler in die Nähe der Zehn-Prozent-Marke zu bringen. Die folgende Tabelle faßt den Vergleich noch einmal zusammen.

Bewertungsgröße		Vorhersage	
		statisch	gedreht
mittlerer rel. Fehler	(lange Hochlaufphase)	+	−
Standardabweichung	(lange Hochlaufphase)	−	+
mittlerer rel. Fehler	(kurze Hochlaufphase)	−	+
Standardabweichung	(kurze Hochlaufphase)	−	+
Rechenzeitbedarf		+	−
Größe der Datenbasis		−	+

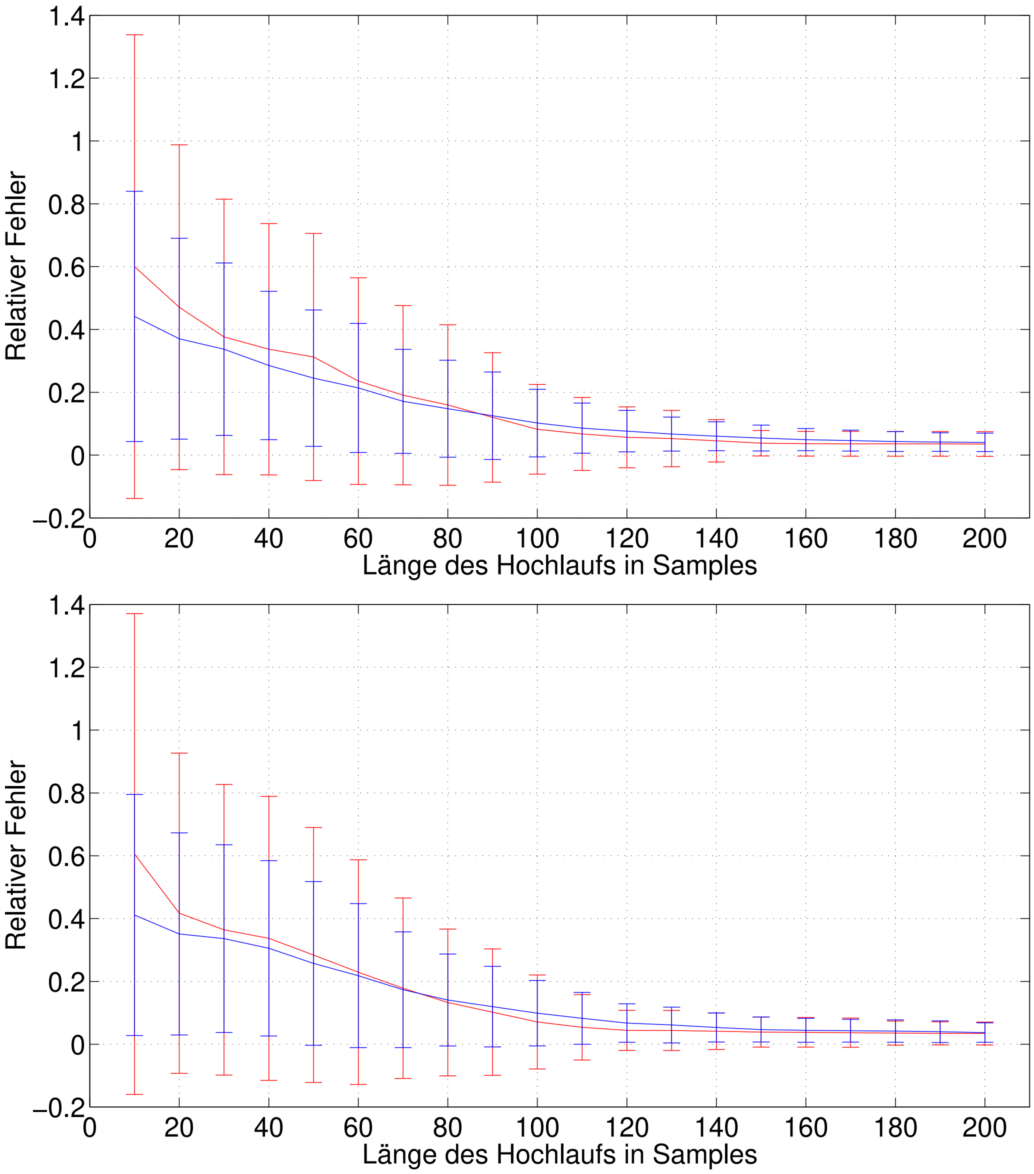
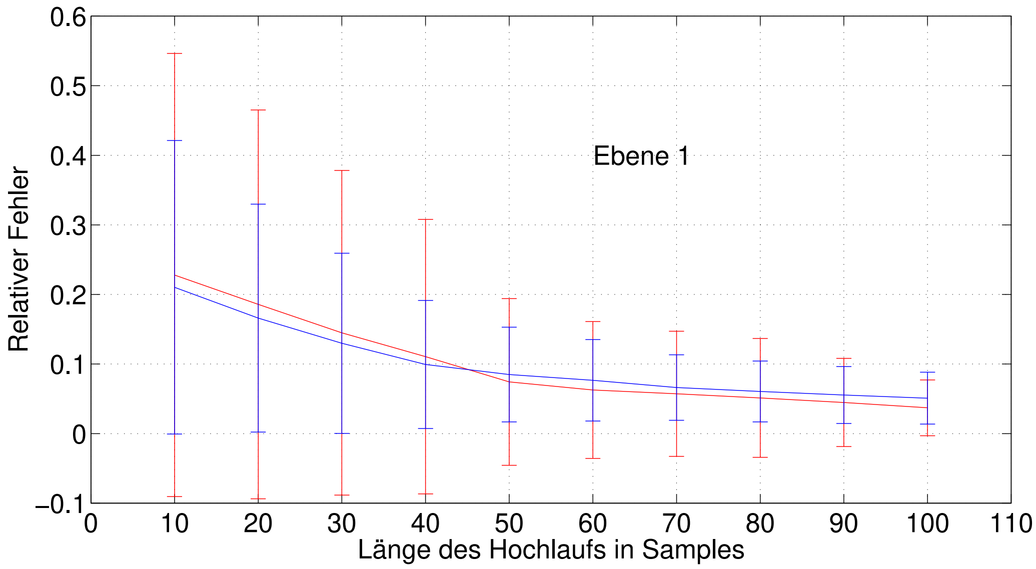


Abbildung 2.10: Diese Abbildung zeigt den Vergleich zwischen den beiden Methoden der Vorhersage für die beiden Auswuchtebenen. Die rote Kurve ist der Vorhersagefehler aus Abbildung 2.8. Die blaue Kurve repräsentiert den mittleren relativen Vorhersagefehler mit der Methode der gedrehten Bahnen. Die Standardabweichung (Fehlerbalken) ist kleiner als bei der statischen Methode, d.h. die Vorhersage mit der Methode der gedrehten Bahnen ist die zuverlässigere.



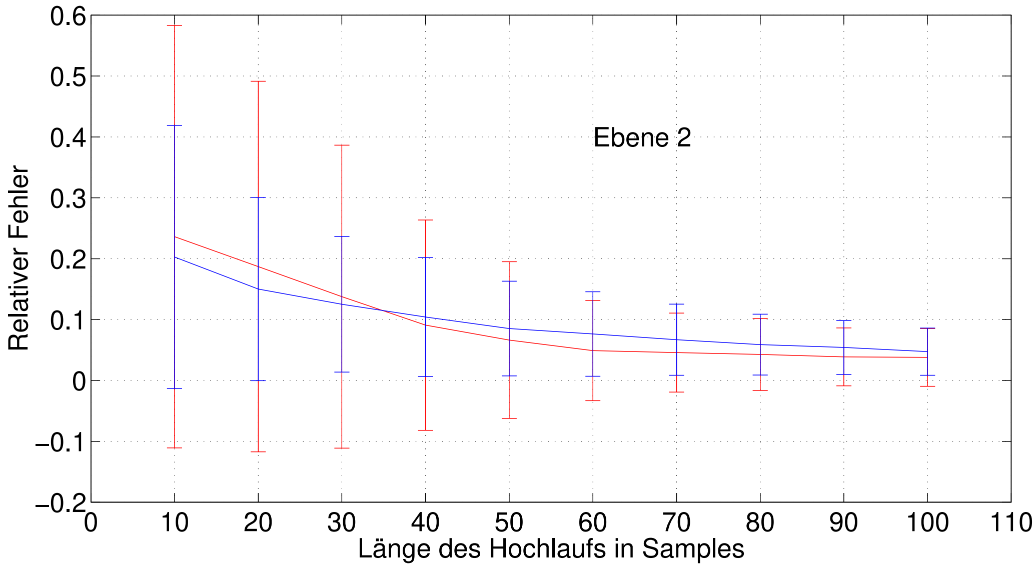


Abbildung 2.11: Diese Abbildung zeigt den Vergleich zwischen den beiden Methoden der Vorhersage, allerdings wurden die ersten 50 Meßwerte des Hochlaufs ausgespart. Der für die Vorhersage relevante Abschnitt des Hochlaufs liegt hier zwischen dem 50. und dem 150. Meßwert. Die rote Kurve stellt den Vorhersagefehler der statischen Methode dar. Als blaue Kurve ist der mittlere relative Vorhersagefehler der Methode der gedrehten Bahnen dargestellt.

Kapitel 3

Rekonstruktion von Wellenbahn und Wellenprofil

In diesem Kapitel soll ein Verfahren vorgestellt werden, das ebenfalls im Rahmen des Verbundprojektes „Analyse von Maschinen- und Prozeßzuständen mit Methoden der Nichtlinearen Dynamik“ entstanden ist. Es handelt sich dabei um ein Verfahren zur Signaltrennung an umlaufenden Rotoren. Die Anregung zur Untersuchung dieses Problems erhielten wir von der C. Schenck AG. Es stellte sich heraus, daß wir durch eine geeignete Erweiterung der bestehenden Meßapparatur und eine nachgeschaltete Datenanalyse die geforderte Signaltrennung erreichen konnten. Da von Seiten unseres Industriepartners Interesse daran bestand, das Verfahren zur Patentreife zu bringen, wurde das Problem weiter von uns verfolgt und zum Abschluß gebracht. Der Patentantrag wurde beim Deutschen Patent- und Markenamt unter dem Titel „Verfahren zur berührungslosen Schwingungsmessung an rotierenden Maschinen“ eingereicht.

3.1 Problemstellung

Bei der Überwachung und Steuerung rotierender Maschinen ist die Kenntnis der Bewegung, insbesondere der maximalen Auslenkung einer Welle im Lager, von großer Bedeutung. In Gleitlagern befindet sich zwischen der Oberfläche der Welle und der Innenwand des Lagers ein dünner Flüssigkeitsfilm, der für die Schmierung der Welle im Lager sorgt. Bei zu großen Auslenkungen der rotierenden Welle im Lager wird dieser Flüssigkeitsfilm überbrückt, und die Welle kommt in direkten Kontakt mit der Lagerwand. Im Maschinenbau nennt man diesen Vorgang *Anschleifen* [31]. Dabei kann es zu Beschädigungen von Lager und Welle kommen, was in der Regel kostenintensive Reparaturarbeiten nach sich zieht[1].

Ferner kann eine Analyse der Wellenbahn bei der Diagnose von Maschinenschäden hilfreich sein, so lassen sich zum Beispiel auf diese Weise Risse im Wellenquerschnitt detektieren [32, 79]. Um eventuelle Lagerschäden bereits in der Frühphase erkennen zu können, muß man die Position der Welle im Lager zu jedem Zeitpunkt kennen. Diese Aufgabe wurde bei unserem Industriepartner, der C. Schenck AG in Darmstadt, bisher durch eine Meßapparatur gelöst, wie sie in Abbildung 3.1 skizziert ist. Sie besteht aus zwei Abstandssensoren, die im Winkel von 90° gegeneinander montiert sind. Bei der genaueren Untersuchung der durch diese Methode rekonstruierten Wellenbahnen zeigte sich allerdings, daß bei einem Meßaufbau mit zwei Sensoren ein systematischer Fehler auftritt, der von den Inhomogenitäten der Welle herrührt. Im folgenden werden wir diesen Meßaufbau beschreiben und eine Erweiterung vorschlagen, die durch eine nachgeschaltete Auswertung der Meßdaten den systematischen Fehler weitgehend herausrechnet. Der Preis dafür ist ein zusätzlicher Meßaufnehmer, der in den alten, Zwei-Sensoren-Aufbau eingegliedert werden muß.

3.2 Meßaufbau (zwei Sensoren)

In Abbildung 3.1 ist der Meßaufbau mit zwei Sensoren skizziert. Die Meßaufnehmer arbeiten induktiv und liefern ein Wechselspannungssignal, dessen genaue Interpretation in Abschnitt 3.3 gegeben wird. Dabei ist wichtig, daß sich diese Wechselspannung in eine Größe umrechnen läßt, die im wesent-

[1] In unserem Fall untersuchten wir die Lager von Kraftwerksgeneratoren, wo jeder reparaturbedingte Ausfall mit erheblichen Kosten verbunden ist.

lichen der Auslenkung der Welle aus der Ruhelage entspricht. Das Signal enthält allerdings noch zusätzliche Komponenten, die im folgenden als die *Inhomogenitäten der Welle* bezeichnet werden und die man in zwei Kategorien unterteilen kann:

1. **Mechanisch bedingte Meßfehler im Schwingungssignal**
   Diese entstehen, wenn eine Abstandsänderung gemessen wird, die nicht von einer Verlagerung oder Bewegung der Welle herrührt. Die Ursachen hierfür sind Kratzer, Risse, Rost oder andere Ablagerungen, eine unrunde Welle, abgeflachte Wellenteile oder Gravierungen der Wellenoberfläche.

2. **Elektrisch bedingte Meßfehler im Schwingungssignal**
   Elektrisch bedingte Meßfehler entstehen durch Inhomogenitäten der elektrischen und magnetischen Eigenschaften der Wellenoberfläche, wenn also die elektrische Leitfähigkeit oder die Permeabilität des Materials entlang der Meßspur schwankt oder wenn lokale Magnetfelder auf der Wellenoberfläche vorhanden sind.

Die Sensoren sind im Winkel von 90° zueinander montiert und tasten die Welle in äquidistanten Zeitschritten ab, die mit der (als konstant angenommenen) Drehzahl synchronisiert sind. Bei den Datensätzen, die wir von der Firma Schenck erhalten haben, wurden von jedem Sensor 64 Meßwerte pro Umdrehung aufgenommen. Insgesamt wurden 16 Umdrehungen aufgezeichnet. In Abbildung 3.2 ist ein solcher Datensatz visualisiert. Die bisherige Methode zur Rekonstruktion der Wellenbahn, also der Bewegung des Wellenmittelpunktes, bestand darin, die beiden Sensorsignale gegeneinander aufzutragen. Durch die Inhomogenitäten der Welle, die dem „reinen“ Signal der Bahn überlagert sind, ergibt sich ein systematischer Fehler. Entdeckt wurde dieser Fehler, als man einmal eine zweite Meßanordnung um 45° versetzt zum primären Meßaufbau anbrachte und die rekonstruierten Wellenbahnen miteinander verglich und dabei eine Abweichung feststellte.

Um die Fehler bei der Rekonstruktion der Wellenbahn zu vermeiden, entschloß man sich zu einem zweistufigen Meßverfahren. Im ersten Schritt wurde

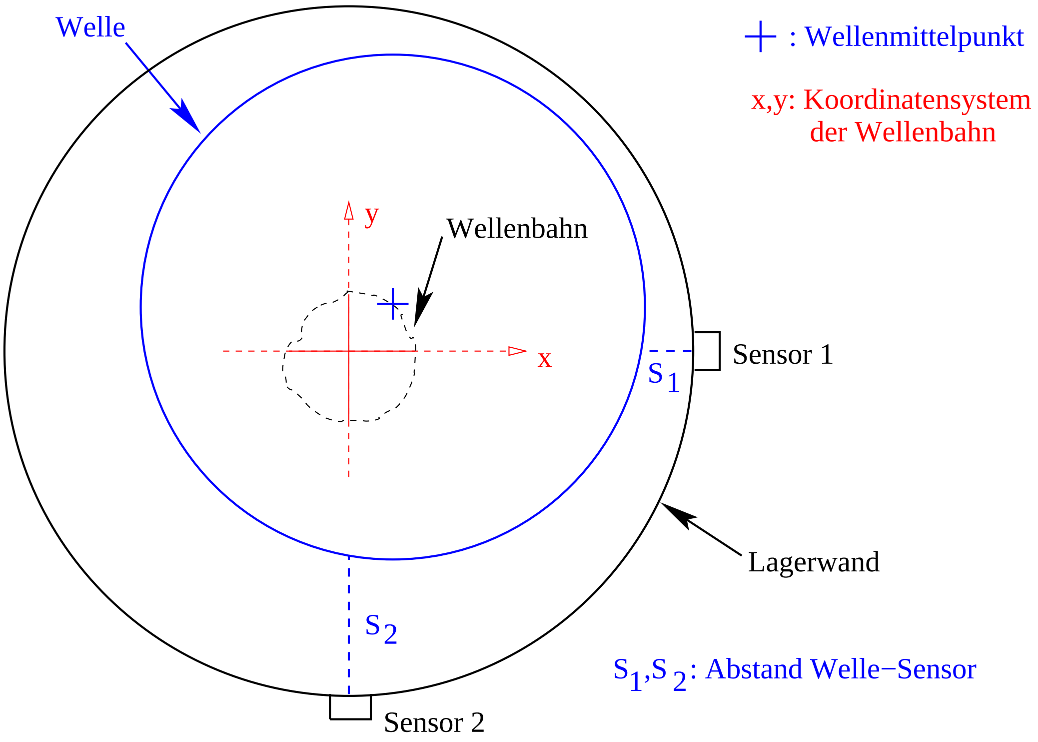


Abbildung 3.1: Die Abbildung zeigt den bisher verwendeten Aufbau mit zwei Abstandssensoren zur Bestimmung der Wellenbahn im Lager. Der Abstand der Welle zur Lagerwand ist zur besseren Anschauung nicht maßstabsgetreu dargestellt. Der Mittelpunkt der Welle ist durch das blaue Kreuz markiert, er bewegt sich im Lager auf der *Wellenbahn*, die als gestrichelte Linie angedeutet ist. Die Wellenbahn wird im lagerfesten Koordinatensystem beschrieben. Der Ursprung dieses Koordinatensystems liegt in der Ruheposition der Welle. Die beiden Sensoren sind in einem Winkel von 90° zueinander montiert.

die Welle sehr langsam, sozusagen im „Leerlauf“, vermessen, um die Welleninhomogenitäten zu bestimmen. Dabei ging man davon aus, daß die Auslenkungen der Wellenbahn erst im Normalbetrieb der Maschine einsetzen, also bei hohen Drehzahlen und Belastungen des Rotors[2]. Im zweiten Schritt wurden diese Daten dann dazu benutzt, die Welleninhomogenitäten aus dem Sensorsignal zu eliminieren. Dieses Verfahren hat jedoch mehrere Nachteile. Zum einen kann man nicht mit Sicherheit davon ausgehen, daß die Welle im Lager keine Mittelpunktsbewegung mehr ausführt, wenn sie langsam und ohne Belastung umläuft. Zum anderen können sich die Materialeigenschaften der Welle während des Betriebs verändern. So werden z.B. die Messungen der Welleninhomogenitäten bei Kraftwerksgeneratoren vor der Inbetriebnahme durchgeführt, also im „kalten“ Zustand. Nach der Inbetriebnahme erwärmt sich die Welle und der Wellendurchmesser verändert sich durch die thermische Ausdehnung des Materials.
Wie wir später sehen werden, kann man diese Probleme durch Hinzunahme eines dritten Sensors umgehen. Unser modifizierter Meßaufbau versetzt uns sogar in die Lage, Veränderungen der Welle während des Betriebs zu diagnostizieren.

3.3 Die Interpretation des mittelwertfreien Sensorsignals

Zunächst werden wir eine Interpretation des mittelwertfreien Sensorsignals geben. Dabei beziehen sich alle Bezeichnungen auf Abbildung 3.3.
Bei der Überwachung der Wellenbewegung im Lager erfolgt die Abstandsmessung berührungslos über induktive Meßaufnehmer. Diese liefern ein Spannungssignal, das sich in zwei Komponenten aufteilen läßt. Zum einen in den Gleichanteil, mit dem diese Sensoren betrieben werden und zum anderen in den Wechselspannungsanteil, der durch die Bewegung der Welle

[2]In unserem Fall ging es in erster Linie um die Verbesserung schwingungsdiagnostischer Überwachungssysteme für Turbogeneratoren in Kraftwerken. Die Arbeitsdrehzahl eines üblichen Gas-Dampf-Generators beträgt 3000 U/min, die Drehzahl einer Dampfturbine liegt bei ca. 6000 U/min.

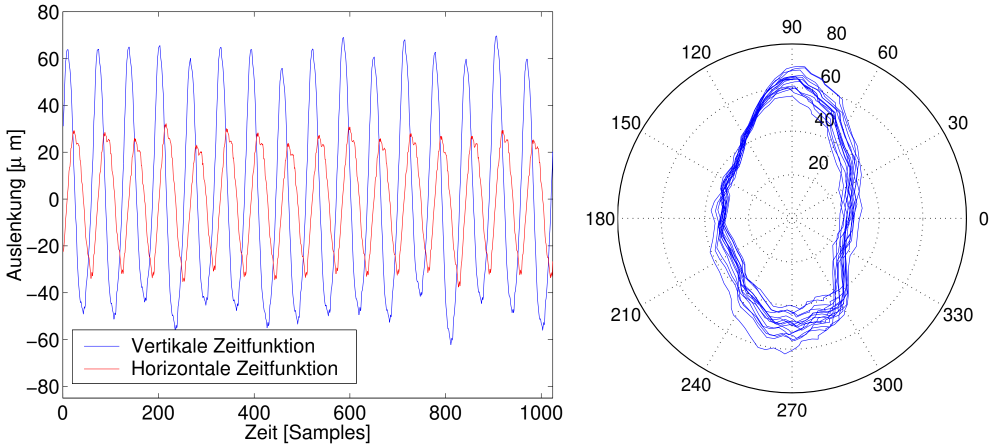


Abbildung 3.2: In der linken Abbildung sind die beiden Sensorsignale gegen die Zeit aufgetragen, also die Auslenkung der Welle aus der Ruhelage und die überlagerten Inhomogenitäten der Wellenoberfläche für sechzehn Umdrehungen der Welle. Das Standardverfahren zur Rekonstruktion der Wellenbahn besteht nun darin, die vertikale Zeitfunktion gegen die horizontale Zeitfunktion aufzutragen. Dabei macht man einen systematischen Fehler, denn man erhält nicht das Bild der Bahn, sondern die Bahn wird durch die überlagerten Inhomogenitäten der Wellenoberfläche verzerrt wiedergegeben.

und die Welleninhomogenitäten hervorgerufen wird. Aus meßtechnischen Gründen wird der Gleichanteil durch einen Hochpaßfilter unterdrückt. Die Information über die Auslenkung der Welle ist im Wechselspannungsanteil enthalten.

In Abbildung 3.3 ist die Situation zwischen Sensor und Wellenoberfläche skizziert. In der Mitte des Lagers befindet sich das rot gekennzeichnete Koordinatensystem zur Beschreibung der Wellenbahn. Bezüglich dieses lagerfesten Systems hat der Mittelpunkt der Welle die Koordinaten (x, y). Der Sensor „sieht" in diesem Fall aber nur die Bahnkomponente in Richtung der x-Achse des Koordinatensystems. Die Idealgestalt der Welle ist durch das Kreisprofil mit dem Radius R gegeben und als gestrichelter blauer Kreis dargestellt. Das reale Wellenprofil zeigt nun Abweichungen von dieser Idealgestalt, die mit ΔR bezeichnet werden und die es neben den Bahnkoordinaten zu bestimmen gilt. In der oberen Figur von Abbildung 3.3 ist die Nullstellung der Welle im Lager als gestrichelter roter Halbkreis angedeutet. Wir gehen davon aus, daß sich in der Ruhestellung der Wellenmittelpunkt ebenfalls im Zentrum des Lagers befindet, ansonsten muß das Koordinatensystem zur Beschreibung der Wellenbahn vom Lagerzentrum in diese Ruhestellung verlegt werden. Bezüglich dieser Stellung mißt der Sensor die Position der Welle, also gerade die relative Schwankung des Abstands von Wellenprofil und Nullstellung, denn die Nullstellung ist ja gerade der Gleichanteil (bzw. der Mittelwert) des Signals, der durch die Hochpaßfilterung eliminiert worden ist. In der unteren Figur von Abbildung 3.3 ist die Situation noch einmal vergrößert dargestellt. Das Sensorsignal setzt sich aus der Bahnkomponente X und der Abweichung des Wellenprofils ΔR zusammen. Der absolute Abstand von Wellenoberfläche und Sensor Sx geht ebensowenig in die Betrachtung ein, wie der Wellendurchmesser R oder der Innendurchmesser des Lagers.

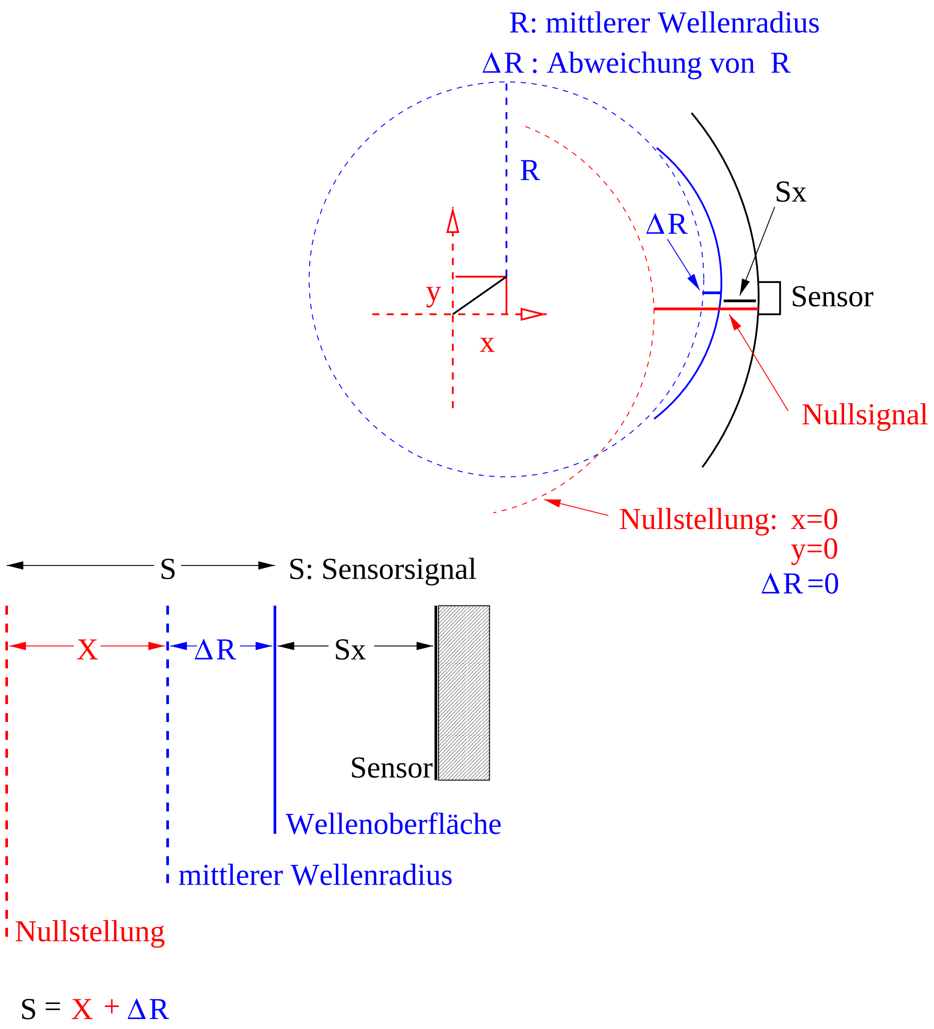


Abbildung 3.3: Zur Interpretation des mittelwertfreien Sensorsignals: In der oberen Figur sieht man das Koordinatensystem zur Beschreibung der Wellenbahn (rot). Der rote gestrichelte Halbkreis symbolisiert die „Nullstellung“ einer idealen kreisrunden Welle, der blaue Kreis ist die ideale Welle mit den Bahnkoordinaten (x, y). Die „reale“, von der Idealgestalt um ΔR abweichende Wellenoberfläche ist als durchgezogene blaue Linie in der Nähe des Sensors angedeutet. Die untere Figur gibt die Verhältnisse am Sensor wieder. Das mittelwertfreie Sensorsignal setzt sich aus der Bahnkomponente und der Abweichung der Wellenoberfläche vom idealen Kreisprofil zusammen.

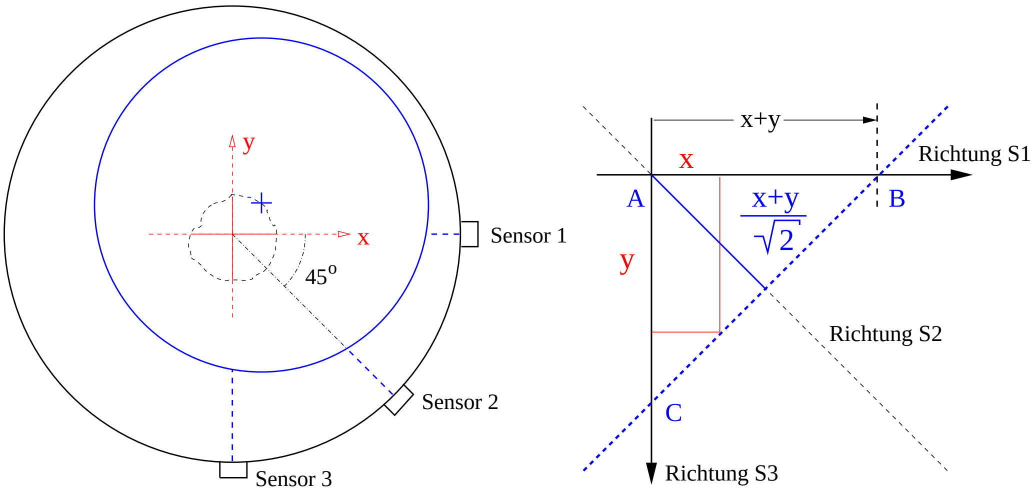


Abbildung 3.4: In der linken Abbildung sieht man den modifizierten Meßaufbau zur Bestimmung von Wellenbahn und Wellenprofil. Die drei Sensoren sind um jeweils 45° versetzt angeordnet. In der rechten Abbildung ist die Zusammensetzung des Signals am zweiten Sensor skizziert.

3.4 Modifizierter Meßaufbau (drei Sensoren)

Im folgenden wird der modifizierte Meßaufbau mit drei Sensoren beschrieben, der es ermöglicht, die gewünschte Signaltrennung zwischen der Wellenbahn und den Inhomogenitäten der Welle zu erreichen. Zunächst werden folgende Voraussetzungen gemacht:

- Der maximale Radius der Wellenbahn ist klein gegenüber dem Radius der Welle.
- Die Drehfrequenz der Welle ist bekannt und während der Messung konstant.
- Bei jedem Umlauf wird die Welle von jedem Sensor jeweils 64 mal abgetastet.
- Es werden 16 Umläufe der Welle aufgezeichnet.
- Während der Messung ändert sich das Wellenprofil nicht.

- Die Sensoren messen die relativen Schwankungen des Abstands zur Welle in Richtung des Lagermittelpunktes, d.h. das Signal ist mittelwertfrei bzw. hochpaßgefiltert.

Diese Voraussetzungen entsprechen der experimentellen Situation während des Betriebes.
In Abbildung 3.4 ist der modifizierte Meßaufbau skizziert. Dabei wurde der dritte Sensor zwischen die beiden vorhandenen Sensoren montiert, also unter einem Winkel von 45° zu Sensor 1 und Sensor 3. Es sind auch andere Positionen möglich, doch das ändert nichts an den grundsätzlichen Überlegungen. Da der Sensor 2 nicht in Richtung der Achsen des lagerfesten Koordinatensystems montiert ist, ist er in der Lage, Anteile von beiden Bahnkomponenten zu detektieren. Im rechten Teil von Abbildung 3.4 ist die Geometrie des Problems veranschaulicht. Die Basis des rechtwinkligen Dreiecks ABC hat die Länge $\sqrt{2(x+y)^2}$. Somit ergibt sich für die Komponente B_{xy} der Bahn in Richtung von Sensor 2:

$$B_{xy} = \frac{x+y}{\sqrt{2}}.$$

Die Idee des Lösungsansatzes besteht darin, die Rekonstruktion von Wellenbahn und Welleninhomogenitäten auf die Lösung eines linearen Gleichungssystems zu reduzieren, indem man die Geometrie des Systems ausnutzt. Dabei ist es wichtig, daß die Welle mit konstanter Drehzahl läuft und die Datenaufnahme in äquidistanten Zeitschritten erfolgt. Die Anzahl der Meßwerte pro Umdrehung muß ein Vielfaches von acht sein. Dadurch wird gewährleistet, daß jeder Sensor pro Umdrehung einmal denselben Bereich der Welleoberfläche „sieht", d.h. dieselbe Inhomogenität der Wellenoberfläche mißt, der jeweils noch das Bahnsignal überlagert ist.
Um das zugehörige Gleichungssystem aufzustellen, wird der Wellenquerschnitt gleichmäßig in jeweils acht korrespondierende Radien $R_1, \dots, R_8$ unterteilt. Jeder Radius repräsentiert einen Bereich auf der Wellenoberfläche. Die Inhomogenitäten der Welle in diesen Bereichen werden mit $\Delta R_1, \dots, \Delta R_8$ bezeichnet. Das Koordinatensystem zur Beschreibung der Wellenbahn legen wir in die Mitte des Lagers (Nullstellung der Welle). In Abbildung 3.5 ist dies

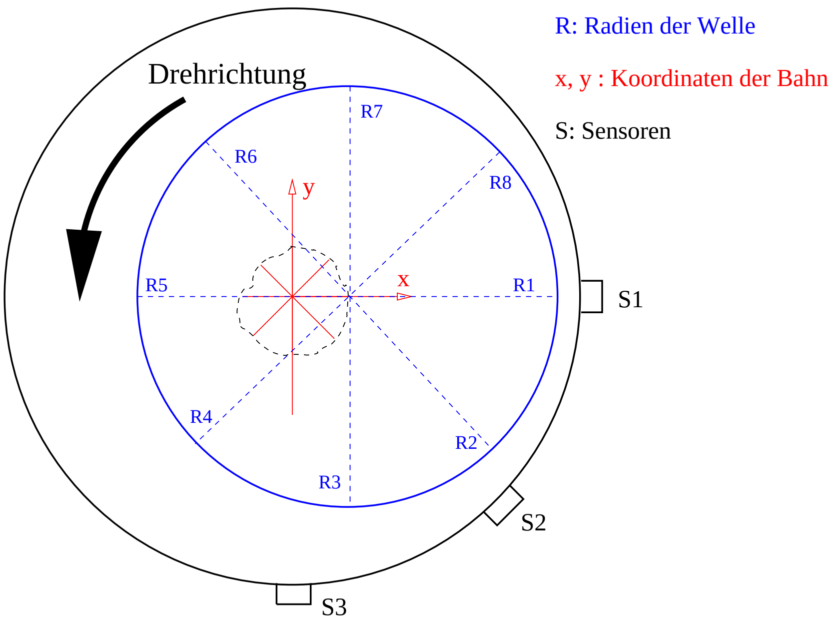


Abbildung 3.5: Die Abbildung zeigt die gleichmäßige Unterteilung der Welle in acht korrespondierende Radien. Jeder Radius markiert einen Bereich auf der Oberfläche der Welle, der während einer Umdrehung von jedem der drei Sensoren einmal abgetastet wird.

noch einmal dargestellt. Während einer Umdrehung der Welle läuft jeder dieser acht Radien an allen drei Sensoren vorbei, und zwar zu fest definierten Zeitpunkten bzw. definierten Drehwinkeln φ_i der Welle. In Abbildung 3.6 ist dies für zwei verschiedene Drehwinkel veranschaulicht. Im letzten Abschnitt wurde gezeigt, daß sich das Sensorsignal aus der Bahnkomponente und der Inhomogenität der Welle zusammensetzt. Dies benutzen wir nun zum Aufstellen des linearen Gleichungssystems. Betrachten wir zunächst den ersten Umlauf der Welle, also jene Sensorsignale, die bei den Drehwinkeln

$$\varphi_k \quad = \quad (k-1) \cdot \frac{\pi}{4} \ \text{, für } k = 1, \ldots, 8 \ ,$$

vorliegen. Für φ_1 bis φ_8 liefern die drei Sensoren S_1, S_2, S_3 folgendes Gleichungssystem:

$$\begin{aligned}
\varphi_1 = 0\text{:} \qquad S_1(\varphi_1) &= x_1 + \Delta R_1 \\
S_2(\varphi_1) &= \frac{x_1 - y_1}{\sqrt{2}} + \Delta R_2 \\
S_3(\varphi_1) &= -y_1 + \Delta R_3
\end{aligned}$$

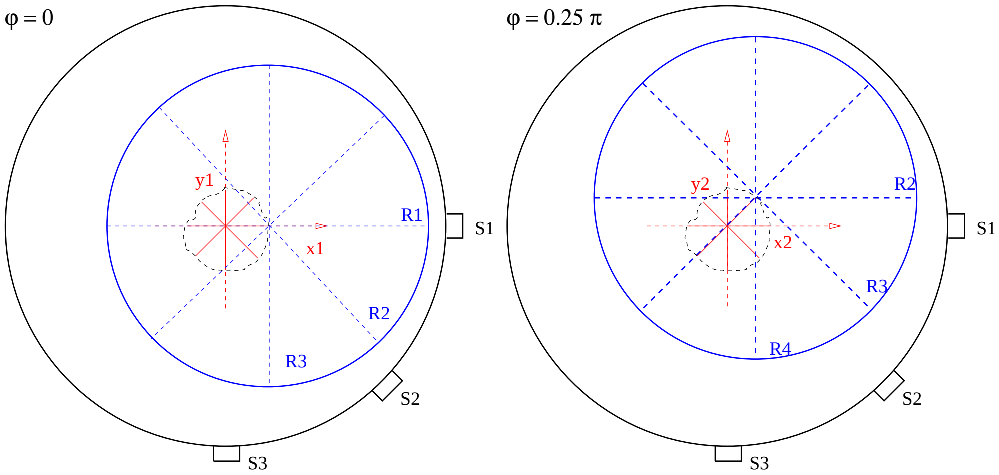


Abbildung 3.6: Die Position der Welle auf ihrer Bahn für die beiden Drehwinkel $\varphi_0 = 0$ und $\varphi_1 = \frac{\pi}{4}$.

$$\varphi_2 = \pi/4: \qquad \begin{aligned} S_1(\varphi_2) &= x_2 + \Delta R_2 \\ S_2(\varphi_2) &= \frac{x_2 - y_2}{\sqrt{2}} + \Delta R_3 \\ S_3(\varphi_2) &= -y_2 + \Delta R_4 \end{aligned}$$

$$\varphi_3 = \pi/2: \qquad \begin{aligned} S_1(\varphi_3) &= x_3 + \Delta R_3 \\ S_2(\varphi_3) &= \frac{x_3 - y_3}{\sqrt{2}} + \Delta R_4 \\ S_3(\varphi_3) &= -y_3 + \Delta R_5 \end{aligned}$$

$$\varphi_4 = 3\pi/4: \qquad \begin{aligned} S_1(\varphi_4) &= x_4 + \Delta R_4 \\ S_2(\varphi_4) &= \frac{x_4 - y_4}{\sqrt{2}} + \Delta R_5 \\ S_3(\varphi_4) &= -y_4 + \Delta R_6 \end{aligned}$$

$$\varphi_5 = \pi: \qquad \begin{aligned} S_1(\varphi_5) &= x_5 + \Delta R_5 \\ S_2(\varphi_5) &= \frac{x_5 - y_5}{\sqrt{2}} + \Delta R_6 \\ S_3(\varphi_5) &= -y_5 + \Delta R_7 \end{aligned}$$

$$\varphi_6 = 5\pi/4: \qquad \begin{aligned} S_1(\varphi_6) &= x_6 + \Delta R_6 \\ S_2(\varphi_6) &= \frac{x_6 - y_6}{\sqrt{2}} + \Delta R_7 \\ S_3(\varphi_6) &= -y_6 + \Delta R_8 \end{aligned}$$

$$\varphi_7 = 3\pi/2: \qquad \begin{aligned} S_1(\varphi_7) &= x_7 + \Delta R_7 \\ S_2(\varphi_7) &= \frac{x_7 - y_7}{\sqrt{2}} + \Delta R_8 \\ S_3(\varphi_7) &= -y_7 + \Delta R_1 \end{aligned}$$

$$\varphi_8 = 7\pi/4: \qquad \begin{aligned} S_1(\varphi_8) &= x_8 + \Delta R_8 \\ S_2(\varphi_8) &= \frac{x_8 - y_8}{\sqrt{2}} + \Delta R_1 \\ S_3(\varphi_8) &= -y_8 + \Delta R_2 \end{aligned} \tag{3.1}$$

Nach einem Umlauf erhalten wir auf diese Weise ein lineares Gleichungssytem mit 24 Gleichungen (8 Drehwinkel mit jeweils drei Sensorsignalen) und 24 Unbekannten (8 Radiuskomponenten und 2×8 Bahnkomponenten). Nach einem vollen Umlauf befindet sich die Welle wieder in ihrer ursprünglichen Stellung, allerdings kann sich der Wellenmittelpunkt unter Umständen auf einer neuen Position befinden, denn die Wellenbahn muß nicht in sich geschlossen sein. Aus den Daten eines Umlaufs kann man bereits die Wellenbahn und die Welleninhomogenitäten gewinnen, indem man versucht, das entsprechende Gleichungssystem zu lösen. Durch Umsortieren kann man das Gleichungssystem 3.1 in die bequemere Matrizenschreibweise überführen zu

$$\vec{S} = A \cdot \vec{r}\,, \tag{3.2}$$

mit dem Vektor $\vec{S}$ der Sensorsignale, dem Vektor $\vec{r}$ der Bahnkomponenten und Welleninhomogenitäten und einer dünnbesetzten 24×24-Matrix A mit den Matrixelementen $a_{ij} \in \{-1, 0, \frac{1}{\sqrt{2}}, 1\}$. Zuerst erhebt sich die Frage nach der Auflösbarkeit des Gleichungssystems 3.2. Die Matrix A ist singulär und hat den Rang 22, es gilt also

$$\dim(\ker(A)) = 2.$$

Das Gleichungssystem besitzt demnach keine eindeutige Lösung, sondern die Lösungen bilden einen 2-dimensionalen affinen Unterraum des $\mathbf{R}^{24}$. In diesem Lösungsraum ist die Lösung $\vec{r}_0$ besonders ausgezeichnet, die das Residuum minimiert:

$$\min_{\vec{r}\in\mathbf{R}^{24}} \|A\cdot\vec{r} \;-\; \vec{S}\|^2 \;=\; \|A\cdot\vec{r}_0 \;-\; \vec{S}\|^2\ . \tag{3.3}$$

Diese Lösung hat zwei Eigenschaften, die sie für unser Problem interessant machen:

1. Die Lösung $\vec{r}_0$ ist eindeutig, und es existiert ein numerisch stabiler Algorithmus, um $\vec{r}_0$ zu berechnen.
2. Die Lösung $\vec{r}_0$ ist die *physikalisch sinnvolle Lösung* des Problems.

Zu 1:
Für eine reelle $(m \times n)$-Matrix, also $A \in \mathrm{Mat}(m, n, \mathbf{R})$, existiert genau eine Moore-Penrose-Inverse $A^\natural \in \mathrm{Mat}(n, m, \mathbf{R})$. Diese Matrix liefert diejenige Lösung, die das Residuum aus Gleichung 3.3 minimiert. Es gilt also

$$\vec{r}_0 \;=\; A^\natural\cdot\vec{S}. \tag{3.4}$$

Der Beweis hierzu findet sich bei Stoer [73].
Zur Berechnung der Moore-Penrose-Inversen kann man auf die Singulärwertzerlegung (SVD) zurückgreifen, die zur numerischen Behandlung des Problems in verschiedenen bewährten Programmbibliotheken zur Verfügung steht, z.B. in den *Numerical Recipes in C* [61]. Bei der SVD wird die $(m \times n)$-Matrix A zerlegt in das Produkt einer zeilenweise orthogonalen $(m\times n)$-Matrix U, einer diagonalen $(n\times n)$-Matrix W und einer orthogonalen $(n \times n)$-Matrix $V^\dagger$, also

$$A \;=\; U\,W\,V^\dagger \tag{3.5}$$

mit

$$U^\dagger\,U = E_n, \qquad V^\dagger\,V = E_n$$

und

$$W = \begin{pmatrix} w_1 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & w_n \end{pmatrix} .$$

Die Diagonalelemente der Matrix W nennt man die *singulären Werte* von A. Die Anzahl der von Null verschiedenen singulären Werte liefert die Dimension des Bildes der zur Matrix A gehörenden linearen Abbildung. Die MOORE-PENROSE-Inverse $A^\natural$ ergibt sich dann als

$$A^\natural = V\, W^{-1}\, U^\dagger$$

mit der „Inversen“

$$W^{-1} = \begin{pmatrix} \tilde{w}_1 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \tilde{w}_n \end{pmatrix} ,$$

wobei die Diagonalelemente wie folgt definiert sind:

$$\tilde{w}_i := \begin{cases} 1/w_i & \text{für } w_i \neq 0 \\ 0 & \text{für } w_i = 0 \end{cases} .$$

Zu 2:

Um die Lösung $\vec{r}_0$ aus Gleichung 3.4 physikalisch zu interpretieren, muß man die Struktur des Lösungsraumes genauer untersuchen. Der Kern von A, also die Lösungen des homogenen Gleichungssystems

$$A \cdot \vec{r}_0 = 0 , \tag{3.6}$$

ist ein zweidimensionaler Untervektorraum des $\mathbf{R}^{24}$. Alle Lösungen des Gleichungssystems 3.2 erhält man, indem man eine spezielle Lösung des Gleichungssystems mit den Elementen des Kerns kombiniert. Ist also $\vec{r}_0$ die spezielle Lösung der MOORE-PENROSE-Inversen, so erhält man alle Lösungen von $\vec{S} = A \cdot \vec{r}$ in der Form

$$\vec{r} = \vec{r}_0 + \vec{y} \;\text{ mit }\; \vec{y} \in \ker(A) .$$

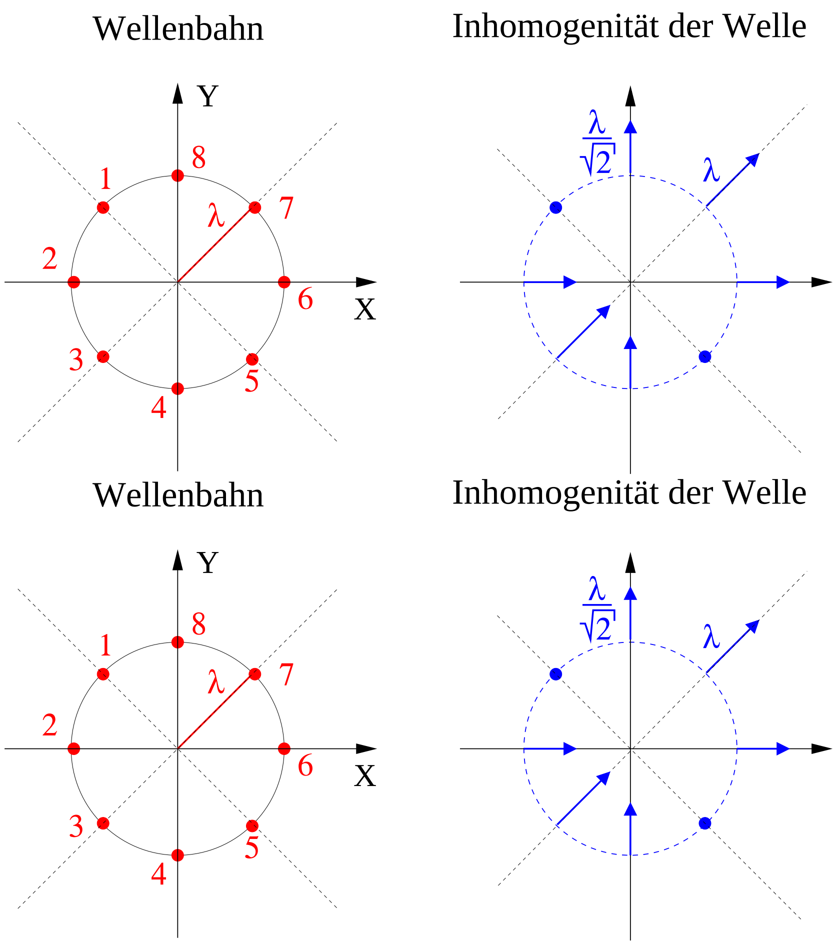


Abbildung 3.7: Hier sind zwei Vektoren visualisiert, die zusammen eine Orthogonalbasis des Kerns der Abbildung 3.6 bilden. Jeder dieser Vektoren aus dem $\mathbf{R}^{24}$ läßt sich übersetzen in acht Positionen der Welle mit den zugehörigen Welleninhomogenitäten. Das Diagramm oben links zeigt die acht Positionen der Wellenbahn auf einem Kreis mit dem Radius λ (links). Zu diesen acht Punkten der Wellenbahn gehört das Wellenprofil auf der rechten Seite, das für den Drehwinkel $\varphi = 0$ dargestellt ist. Der gestrichelte blaue Kreis gibt das ideale Kreisprofil wieder, und die blauen Pfeile geben die Abweichung der Wellenoberfläche vom idealen Kreisprofil in radialer Richtung an (der Durchmesser ist nicht maßstabsgetreu eingezeichnet). Wenn sich die Welle zum Drehwinkel $\varphi = 0$ auf dem Punkt 1 ihrer Bahn befindet, dann liefern die drei Sensoren, die sich in Richtung der Punkte 4, 5 und 6 befinden, gerade das Nullsignal, denn die Auslenkung der Welle auf ihrer Bahn wird genau durch die Inhomogenitäten der Welle kompensiert. Das gilt auch für die Punkte 2 bis 8 der Wellenbahn, wenn sich die Welle (und damit ihr Profil) um jeweils $\pi/4$ weitergedreht hat. Die gleiche Betrachtung gilt für die beiden unteren Diagramme des zweiten Basisvektors.

Um die Struktur des Kerns aufzuklären, kann man bei der Singulärwertzerlegung der Matrix A sowohl die Orthogonalbasis des Bildes von A als auch die Orthogonalbasis des Kerns berechnen [21, 61]. Die beiden orthogonalen Basisvektoren des Kerns lassen sich wieder in Bahnkomponenten und Welleninhomogenitäten zurückübersetzen. In Figur 3.7 ist dies veranschaulicht. Zum Verständnis dieser Figur muß man sich klarmachen, daß diese beiden Basisvektoren den Lösungsraum des homogenen Gleichungssystems 3.6 aufspannen, also den Fall, in dem alle drei Sensoren in den acht Positionen der Welle (siehe Figur 3.6) das Nullsignal liefern. Dieses kann zwei Ursachen haben. Entweder befindet sich der Mittelpunkt der Welle im Ursprung des lagerfesten Koordinatensystems und die Wellenoberfläche zeigt keine Inhomogenitäten, oder die Auslenkung der Welle aus der Ruhelage wird in allen acht Positionen genau durch die Inhomogenität der Wellenoberfläche kompensiert. Dieser letzte Fall ist aber in der Praxis äußerst unwahrscheinlich, da die Inhomogenitäten der Wellenoberfläche mindestens um eine Größenordnung kleiner sind als die Auslenkungen der Wellenbahn [80].

Bei allen möglichen Lösungen des Gleichungssystems 3.2 kann man also davon ausgehen, daß sich die Bewegung der Wellenbahn und die Inhomogenitäten der Welle nicht kompensieren. Dadurch schränkt man den Lösungsraum auf eine einzige Lösung ein, nämlich auf die Lösung, die durch die Moore-Penrose-Inverse gegeben ist.

Im allgemeinen Fall des inhomogenen Gleichungssystems 3.2 gelten die gleichen Überlegungen wie im homogenen Fall. Der Vektor $\vec{r}_0$, der das Residuum aus Gleichung 3.3 minimiert, steht orthogonal auf dem Kern der Abbildung und verschiebt die Moore-Penrose-Lösung des homogenen Gleichungssystems auf den affinen zweidimensionalen Unterraum, in dem die Lösungen des inhomogenen Gleichungssystems liegen. Die Lösung mit der kleinsten euklidischen Norm ist auch in diesem Fall die Lösung, die nicht davon ausgeht, daß sich die Bewegung des Wellenmittelpunktes und die Inhomogenitäten der Welle genau kompensieren. Aus diesem Grund ist die Lösung $\vec{r}_0$ die physikalisch sinnvolle Lösung des Problems.

Die bisherigen Betrachtungen bezogen sich auf acht verschiedene Drehwinkel der Welle und lieferten für diese die Position des Wellenmittelpunktes und

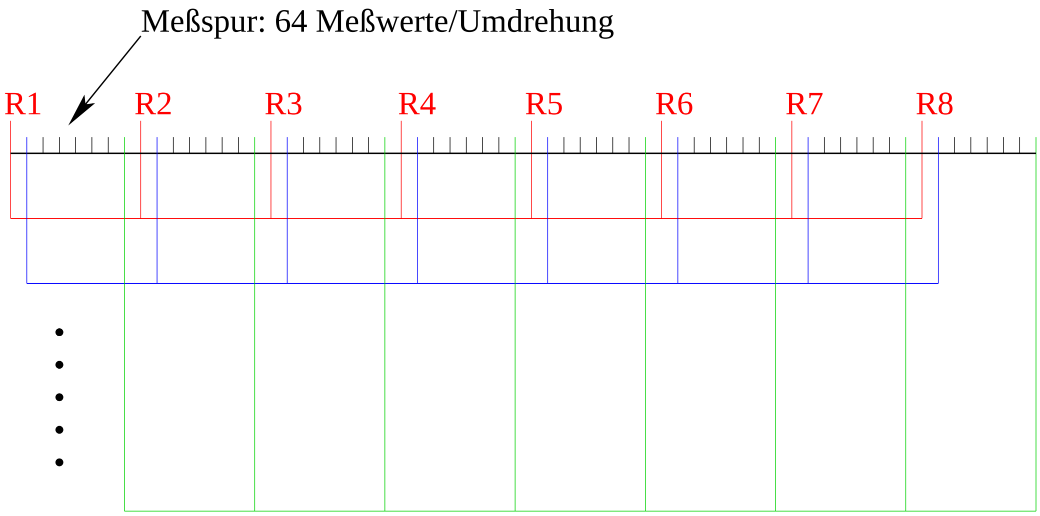


Abbildung 3.8: Die Abbildung zeigt die 64 Punkte der Wellenoberfläche, die von den Sensoren während eines Umlaufs ausgemessen werden, einmal abgewickelt. Aus diesen 64 Meßwerten wählt man jeweils acht korrespondierende Radien R1 bis R8 aus, um mit ihnen das Gleichungssystem 3.1 aufzustellen. Auf diese Weise verfährt man mit allen 64 Meßwerten eines Umlaufs, um die Bewegung der Welle zu rekonstruieren

die Inhomogenitäten der Wellenoberfläche. Nun wird die Welle während einer Umdrehung 64 mal abgetastet, so daß man das beschriebene Verfahren auch für die verbleibenden Meßwerte in gleicher Weise anwenden kann. Dazu wählt man aus den 64 Meßwerten eines Umlaufs jeweils acht korrespondierende Radien gemäß Abbildung 3.5 aus und stellt das zugehörige Gleichungssystem in der Art von Gleichung 3.1 auf. In Abbildung 3.8 ist diese Vorgehensweise veranschaulicht. Wenn man alle 64 Meßwerte eines Umlaufs verwendet hat und die zugehörigen Gleichungssysteme gelöst hat, kann man die Ergebnisse in geeigneter Weise zusammenfügen und erhält ein Bild der Wellenbahn und der Welleninhomogenitäten.

Wie bereits erwähnt, werden bei diesem Verfahren gleich 16 Umläufe hintereinander aufgezeichnet. Man kann das bisher beschriebene Verfahren für einen Umlauf der Welle problemlos auf die allgemeine Situation übertragen. Dabei wird angenommen, daß sich die Inhomogenitäten der Welle während der Messung nicht ändern. Das ist eine realistische Annahme, da die Mes-

sung von 16 Umläufen bei 3000 U/min nur 0,32 Sekunden dauert, also eine Zeitspanne, in der z.B. die thermische Ausdehnung der Welle nicht von Bedeutung ist.
Um die 16 Umläufe in einem Gleichungssystem zusammenzufassen, müssen wir zuerst eine Funktion zur Indizierung einführen. Für $i \in \mathbf{N}$ soll gelten:

$$\mathrm{mod}_8(i) \quad := \quad \begin{cases} i \mod 8 & \text{für } i \not\equiv 0 \mod 8 \\ 8 & \text{für } i \equiv 0 \mod 8 \end{cases} .$$

Das Gleichungssystem für 16 Umläufe hat dann folgende Gestalt:

Für $i = 1, \ldots, 128$:

$$\begin{aligned} S_1(\varphi_i) &= x_i + \Delta R_{\mathrm{mod}_8(i)} \\ S_2(\varphi_i) &= \frac{x_i - y_i}{\sqrt{2}} + \Delta R_{\mathrm{mod}_8(i+1)} \\ S_3(\varphi_i) &= -y_i + \Delta R_{\mathrm{mod}_8(i+2)} . \end{aligned}$$

Dies führt insgesamt auf $3 \times 8 \times 16 = 384$ lineare Gleichungen mit 8 Unbekannten für die Welle und $2 \times 8 \times 16$ Unbekannten der Bahn. Dieses Problem läßt sich wieder in der Matrizenschreibweise formulieren als

$$\vec{S} = A \cdot \vec{r} \,, \tag{3.7}$$

mit dem Vektor $\vec{S}$ der Sensorsignale, dem Vektor $\vec{r}$ der Bahnkomponenten und Welleninhomogenitäten und der 384×256-Matrix A. Für dieses Gleichungssystem läßt sich wieder mit Hilfe einer Singulärwertzerlegung eine Moore-Penrose-Inverse finden, welche die Lösung des Gleichungssystems 3.7 mit der kleinsten euklidischen Norm liefert. Die Matrix hat den Rang 254, für den Kern von A gilt wieder

$$\dim(\ker(A)) \quad = \quad 2 \,,$$

wobei die Betrachtungen über die Struktur des Kerns analog zu dem oben diskutierten Fall der 24×24-Matrix für eine Umdrehung der Welle verlaufen. Die graphische Darstellung der Basis des Kerns hat die gleiche Form wie in Abbildung 3.7, mit dem Unterschied, daß nun 16 Umdrehungen der Welle betrachtet werden müssen. Dadurch erhält man 16 identische Diagramme

für die Bahnkomponenten der 16 Umdrehungen in Form des Diagramms in Abbildung 3.7.
Die Lösung des Gleichungssystems 3.7 durch die MOORE-PENROSE-Pseudoinverse ist ebenfalls wieder die physikalisch sinnvolle Lösung, da sie die Fälle ausschließt, in denen die Bewegung der Welle genau durch die Inhomogenität der Wellenoberfläche ausgeglichen wird.

3.5 Betrachtung systematischer Fehlerquellen

In diesem Abschnitt soll auf die systematischen Fehlerquellen eingegangen werden, die bei dem vorgestellten Verfahren von Bedeutung sind. Zunächst gibt es einen systematischen Fehler, der in der Geometrie des Aufbaus begründet ist. In Abbildung 3.9 sind die Verhältnisse veranschaulicht, die zwischen Wellenoberfläche und Sensor bestehen. Der Sensor 3 vermißt nicht den Bereich der Wellenoberfläche um den Radius R_6, sondern den benachbarten Bereich $\tilde{R}_6$, bedingt durch die Auslenkung der Welle. Dadurch ändert sich der Abstand zwischen Welle und Sensor S_3 um den Betrag

$$\Delta_3 = \frac{x^2}{R_6} .$$

Nimmt man für Radius und Auslenkung der Welle realistische Werte an, so ergeben sich für den Fehler Δ_3 Werte, die eine Größenordnung unter den Inhomogenitäten der Welle und zwei Größenordnungen unter der Auslenkung der Welle liegen. Diesen Fehler kann man in der Praxis vernachlässigen.
Eine weitere Fehlerquelle liegt in der Richtcharakteristik der induktiven Abstandssensoren. Es ist klar, daß ein solcher induktiver Meßaufnehmer niemals punktgenau messen kann, also nicht den exakten Abstand zwischen einem definierten Punkt auf der Wellenoberfläche und einem definierten Punkt auf der Oberfläche des Sensors liefert. Vielmehr mißt ein solcher Meßaufnehmer einen ganzen Bereich der Wellenoberfläche aus und liefert dementsprechend einen mittleren Abstand als Ergebnis. Dies ist aber ein generelles Problem der induktiven Abstandsmessung.

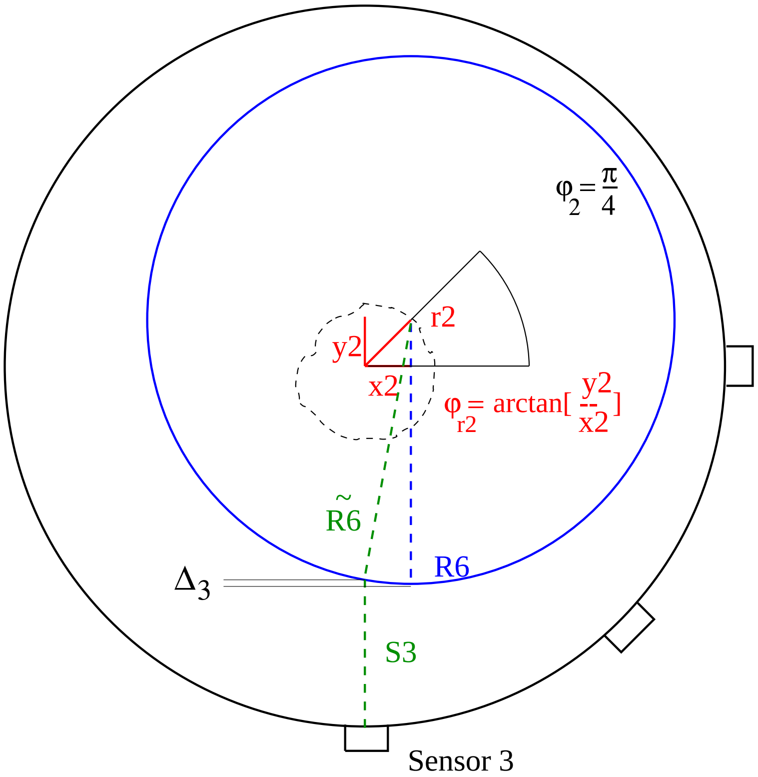


Abbildung 3.9: Darstellung des systematischen Fehlers der Rekonstruktion für den Drehwinkel der Welle $\varphi_2 = \frac{\pi}{4}$.

Außerdem besitzen beide Verfahren (2- und 3-Sensoren) noch eine systematische Fehlerquelle, die in der Hochpaßfilterung der Signale ihre Ursache hat. Da die Sensoren auf diese Weise nur die Schwankungen der Abstände um eine angenommene Ruhestellung angeben (siehe Abschnitt 3.3), ist die wirkliche Lage dieser Ruhestellung nicht bekannt. In unserem Fall haben wir diese besagte Ruhestellung in die Mitte des Lagers gelegt, so wie es auch in der Praxis bei der 2-Sensoren-Methode gemacht wird.

3.6 Numerische Simulation

In diesem Abschnitt sollen einige Ergebnisse vorgestellt werden, die aus einer numerischen Simulation der beiden Verfahren gewonnen wurden. In einem separaten Programm wurden Datensätze erzeugt, welche die Sensorsignale in einem Meßaufbau gemäß Abbildung 3.4 simulieren. Diese Datensätze wurden verwendet, um das Wellenprofil und die Wellenbahn zu rekonstruieren und mit den wahren Werten zu vergleichen. Alle Sensorsignale wurden von ihrem Mittelwert befreit. Es wurden die typischen Abmaße eines realen Meßaufbaus zugrundegelegt [80]:

- Der Innendurchmesser des Lagers beträgt 30,0 cm.
- Der mittlere Radius der Welle beträgt 14,9 cm.
- Der mittlere Radius der Bahn beträgt 0,05 cm.
- Die Schwankungen im Wellenprofil liegen eine Größenordnung unter dem Bahnradius (0,006 cm).

Ferner wurden die Sensordaten mit künstlichem Meßrauschen versehen, und zwar mit einer $\mathcal{N}(0, 1)$-normalverteilten Pseudozufallszahlenfolge, die mit einem vorgegebenen Signal-Rausch-Verhältnis (SNR) zu den reinen Sensordaten hinzuaddiert wurde.
Die Ergebnisse der Simulation sind in den folgenden Abbildungen dargestellt. Das vorgestellte Verfahren wurde zusätzlich mit der bisher von der Fa. Schenck verwendeten Methode verglichen, die darin besteht, die Signale von Sensor 1 und Sensor 3 gegeneinander aufzutragen (2-Sensoren-Methode). Dabei bleibt die Inhomogenität des Wellenprofils unberücksichtigt, was zu Fehlern in der Rekonstruktion der Wellenbahn führt.
Bei der Auswertung der Simulationen zeigte sich, daß unser modifiziertes Verfahren in der Lage ist, die Wellenbahn und die Schwankungen des Wellenprofils zu detektieren. Auch bei verrauschten Daten liefert es bessere Ergebnisse als die bisherige 2-Sensoren-Methode (siehe Abbildung 3.11 und 3.12).
Bei der Simulation der Wellenbahn in Abbildung 3.13 wurde auf einen Datensatz zurückgegriffen, der uns von Herrn Dr. Weigel zur Verfügung gestellt worden ist. Dabei wurde eine Wellenbahn aus einer realen Messung modifiziert und mit einer künstlichen Welleninhomogenität überlagert. Auch in

diesem Fall liefert unser Verfahren bessere Ergebnisse als die herkömmliche Methode.

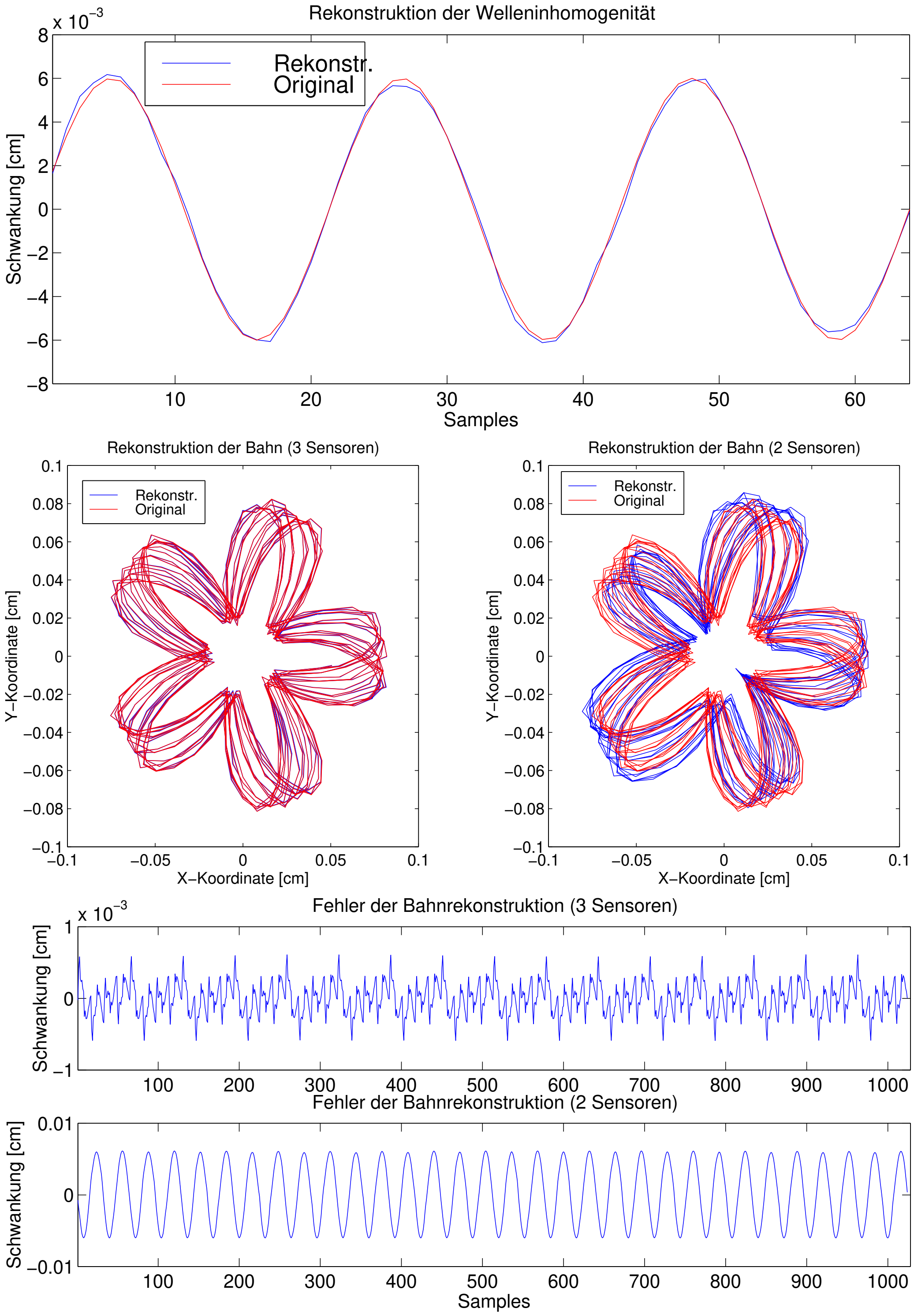


Abbildung 3.10: Rekonstruktion von Wellenbahn und Wellenprofil für einen Datensatz ohne Meßrauschen.

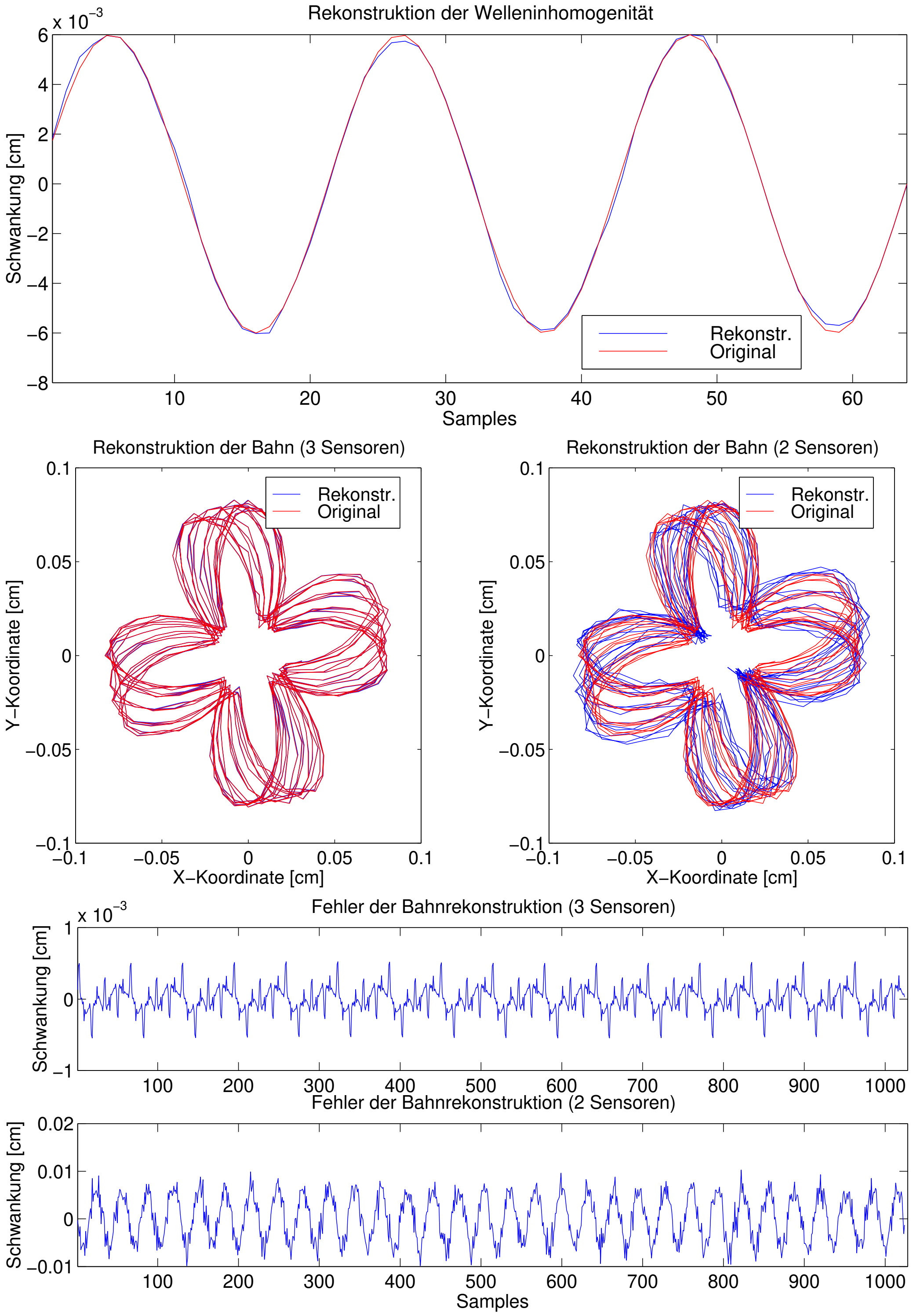


Abbildung 3.11: Rekonstruktion von Wellenbahn und Wellenprofil für einen Datensatz mit künstlichem Meßrauschen (SNR: 50 dB).

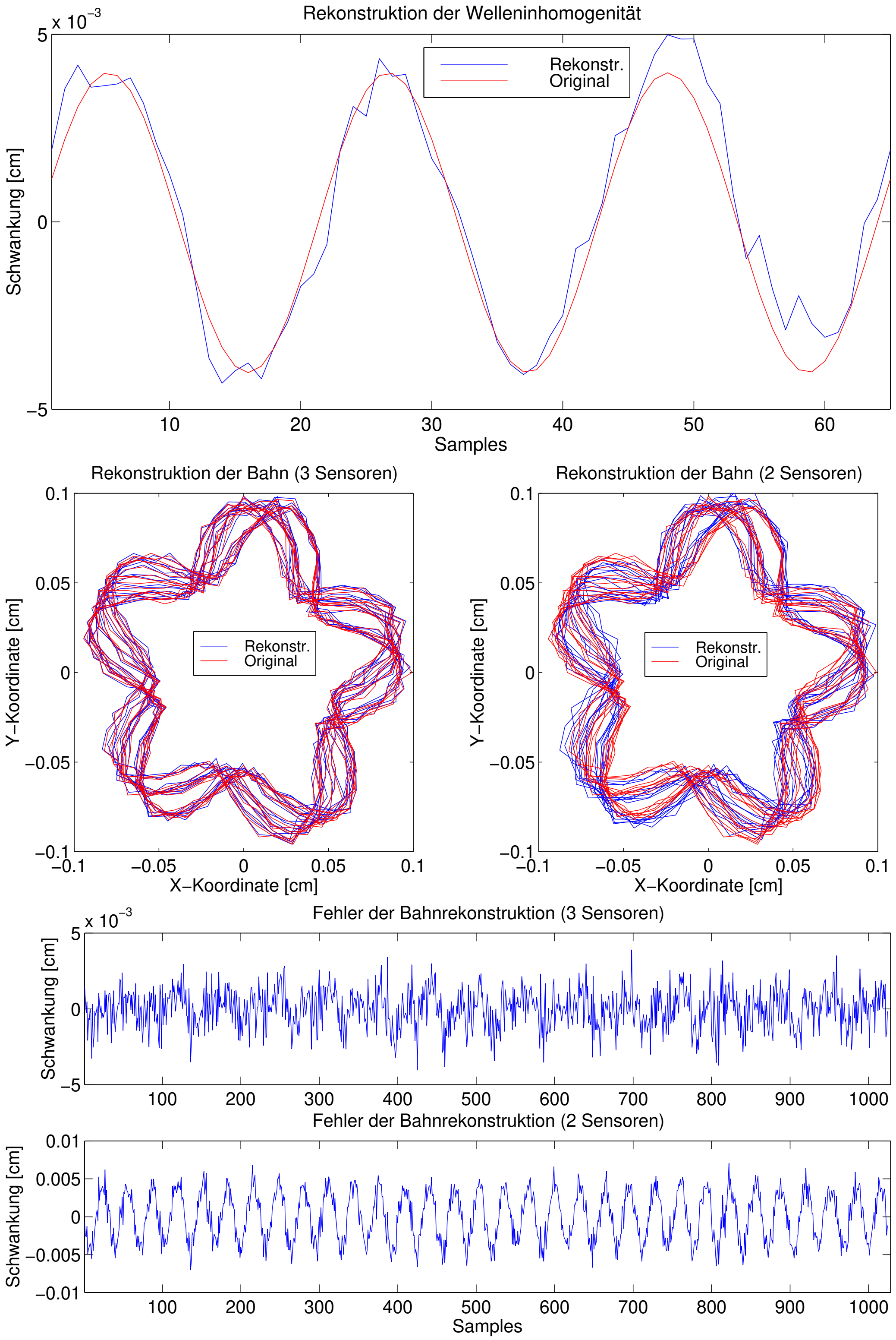


Abbildung 3.12: Rekonstruktion von Wellenbahn und Wellenprofil für einen Datensatz mit künstlichem Meßrauschen (SNR: 40 dB).

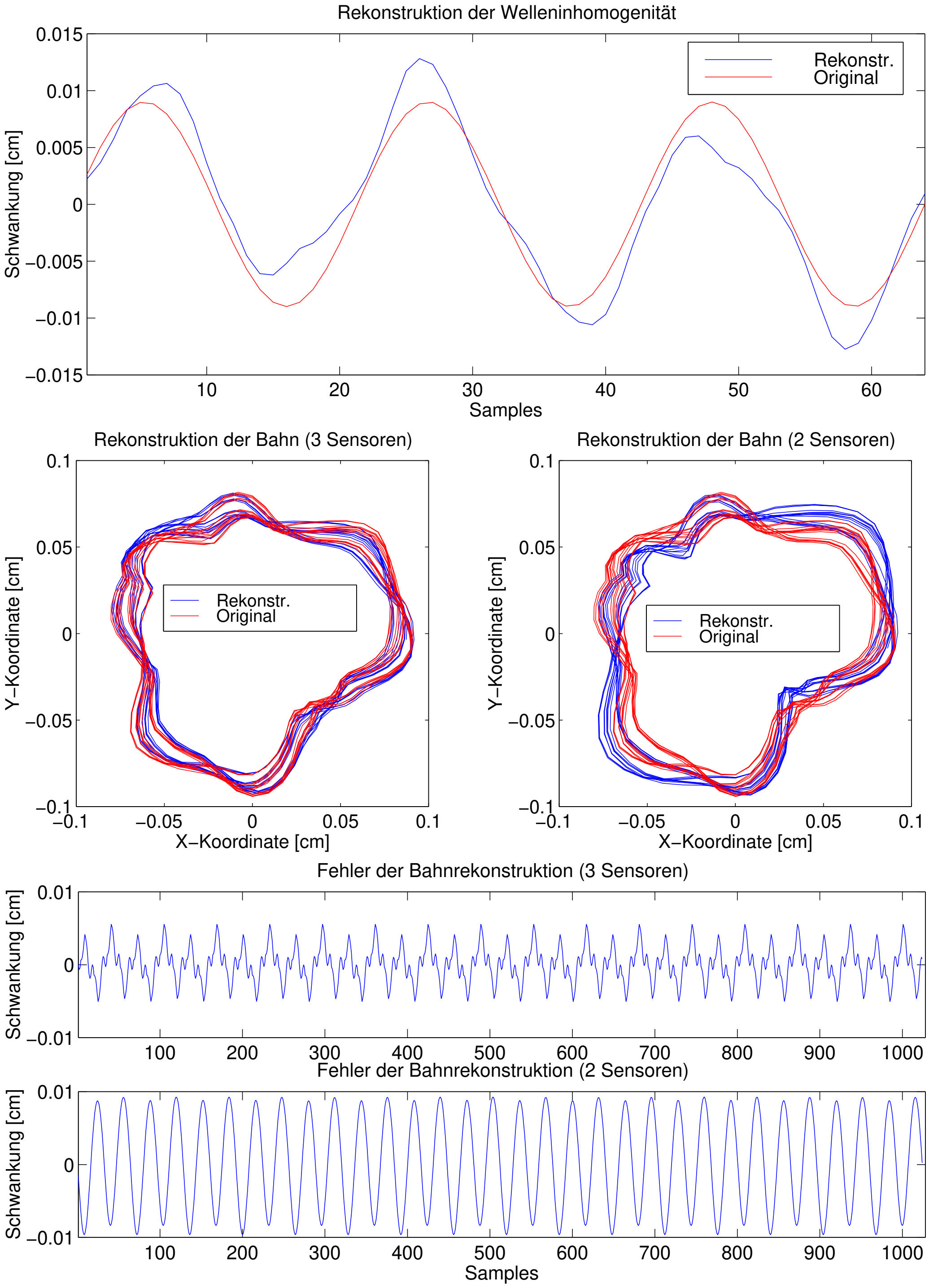


Abbildung 3.13: Rekonstruktion von Wellenbahn und Wellenprofil für einen Datensatz ohne künstliches Meßrauschen. Zur Simulation der Wellenbahn wurde einer der „echten" Datensätze aus einem 2-Sensoren-Meßaufbau herangezogen und mit einem künstlichen Wellenprofil überlagert.

Kapitel 4

Die Hilbert-Transformation und das analytische Signal

In diesem Kapitel werden wir das Konzept des *analytischen Signals* vorstellen und seine Herleitung näher begründen. Die Bedeutung dieses Konzeptes wird in Kapitel 5 klar, wenn wir die instantane Frequenz einführen und deren Anwendung in der Zeit-Frequenz-Analyse anhand der EMD erläutern.

4.1 Die Hilbert-Transformation

Das analytische Signal wird über die Hilbert-Transformation definiert, aus diesem Grund sollen im folgenden einige ihrer Eigenschaften erläutert werden. Für ein Signal $s(t)$ wird die Hilbert-Transformation $\mathcal{H}\left[s(t)\right]$ gegeben durch

$$\begin{aligned} \mathcal{H}\left[s(t)\right] &:= \lim_{\epsilon \searrow 0} \left[\int_{-\infty}^{0-\epsilon} \frac{s(u)}{\pi(t-u)} \mathrm{d}u + \int_{0+\epsilon}^{\infty} \frac{s(u)}{\pi(t-u)} \mathrm{d}u \right] \\ &= \text{p.v.} \int_{-\infty}^{\infty} \frac{s(u)}{\pi(t-u)} \mathrm{d}u \;. \end{aligned} \tag{4.1}$$

Für Signale $s(t)$ mit

$$\int_{-\infty}^{\infty} (s(t))^2 \, \mathrm{d}t \quad < \quad \infty$$

existiert das Integral in Gleichung 4.1 fast überall im uneigentlichen Sinne und wird der Cauchysche Hauptwert (p.v.) des Integrals genannt [75]. Alle nachfolgenden Integrale der Hilbert-Transformation sind immer im Sinne des Cauchyschen Hauptwertes zu verstehen.
Die Hilbert-Transformation ist eine Integraltransformation mit folgenden Eigenschaften [9]:

$$\begin{array}{llrcl}
\text{Linearität} & : & \mathcal{H}\left[\alpha x(t) + \beta y(t)\right] & = & \alpha \mathcal{H}\left[x(t)\right] + \beta \mathcal{H}\left[y(t)\right] \ , \\
\text{Antisymmetrie} & : & \mathcal{H}\left[\mathcal{H}\left[s(t)\right]\right] & = & -s(t) \ , \\
\text{Energieerhaltung} & : & \int\limits_{-\infty}^{\infty} (s(t))^2 \, \mathrm{d}t & = & \int\limits_{-\infty}^{\infty} (\mathcal{H}\left[s(t)\right])^2 \, \mathrm{d}t \ , \\
\text{Orthogonalität} & : & \int\limits_{-\infty}^{\infty} s(t) \mathcal{H}\left[s(t)\right] \mathrm{d}t & = & 0 \ .
\end{array}$$

4.2 Das analytische Signal

Das Konzept des analytischen Signals wurde 1946 von D. Gabor [30] in die Signalverarbeitung eingeführt und hat seitdem weite Verbreitung gefunden. Die Idee besteht in der „Fortsetzung“ eines reellen Zeitsignals $s(t)$ in die komplexe Ebene. Dabei wird das ursprüngliche Signal als Realteil eines komplexwertigen Signals aufgefaßt. Der zugehörige Imaginärteil wird aus der Hilbert-Transformierten $\mathcal{H}\left[s(t)\right]$ des Realteils gewonnen. Das analytische Signal ist dann gegeben durch

$$\tilde{s}(t) \quad = \quad s(t) + i\mathcal{H}\left[s(t)\right] \ . \tag{4.2}$$

Gabor gibt für dieses Vorgehen ein heuristisches Argument an, das hier kurz erläutert werden soll. Angenommen, das Signal habe die Gestalt

$$s(t) \quad = \quad \alpha \cos(\omega t) + \beta \sin(\omega t) \ ,$$

dann läßt sich $s(t)$ als Realteil von

$$\tilde{s}(t) \quad = \quad (\alpha - i\beta) e^{i\omega t}$$

schreiben. Der Imaginärteil von $\tilde{s}(t)$ ist dann gerade die Hilbert-Transformierte des ursprünglichen Signals

$$\begin{aligned} \Im\left[\tilde{s}(t)\right] &= \mathcal{H}\left[s(t)\right] \\ &= \alpha \sin(\omega t) - \beta \cos(\omega t) \ . \end{aligned}$$

Bei komplizierteren Signalen wird dieses Verfahren auf die Fourierdarstellung angewandt.

Gabor läßt sich bei der Einführung des analytischen Signals von dem Gedanken leiten, den mathematischen Formalismus der Quantenmechanik für die Signalverarbeitung zu nutzen. In seiner Arbeit *Theory of Communication* [30] schreibt er dazu:

> In order to apply the simple and elegant formalism of quantum mechanics, it will be convenient first to express the signals amplitude $s(t)$ in somewhat different form.

Das analytische Signal benutzt er, um die die Zeit-Frequenz-Unschärferelation $\Delta t \Delta f \geq 1/2$ für Signale abzuleiten, analog zur Heisenbergschen Unbestimmtheitsrelation der Wellenmechanik[1]. Die Zeit-Frequenz-Unschärferelation spielt bei der Bestimmung des spektralen Ausflösungsvermögens in der Zeit-Frequenz Analyse eine entscheidente Rolle [15, 16].

[1]Bei der Ableitung der Zeit-Frequenz-Unschärferelation folgt er einer Idee W. Paulis und H. Weyls [81], die den heuristischen Ansatz Heisenbergs [33] verallgemeinerten und einen strengen Beweis der Unschärferelation gaben, der auf der Schwarzschen Ungleichung basiert.

Über die mathematische und physikalische Begründung des analytischen Signals ist seit der richtungsweisenden Arbeit Gabors viel geschrieben worden. Einige Autoren vertreten die Ansicht, das analytische Signal besitze nur „physikalisch relevante Frequenzen", denn für ein analytisches Signal gemäß Gleichung 4.2 enthält das Fourierspektrum nur positive Frequenzen [8]. Dieser Standpunkt ist allerdings nicht haltbar, worauf z.B. Cohen hingewiesen hat [16].

4.3 Die Hilbert-Transformation und die Fortsetzung des Signals

Im folgenden wird eine Begründung für die Einführung des analytischen Signals gegeben, die sich auf ein Argument aus der Funktionentheorie stützt. Dabei wird sich zeigen, daß die Definition des analytischen Signals über die Hilbert-Transformation unter bestimmten Voraussetzungen zwingend wird, nämlich genau dann, wenn man das Signal $\tilde{s}(t)$ als den Randwert einer analytischen Funktion darstellen will.
Die physikalische Motivation zur Untersuchung solcher Randwertprobleme stammt aus der Potentialtheorie, wo man vor der Aufgabe steht, eine im Inneren eines Gebietes harmonische[2] Funktion zu finden, die auf dem Rand des Gebietes vorgegebene Werte annimmt. Problemstellungen dieser Art wurden bereits von Gauß und Dirichlet untersucht (Dirichlet-Problem).
Real- und Imaginärteile holomorpher lassen sich als harmonische Funktionen auffassen, wenn man die komplexe Ebene mit dem $\mathbf{R}^2$ identifiziert [29]. Überlegungen dieser Art veranlaßten B. Riemann, sich eingehender mit dem Studium harmonischer Funktionen zu befassen, und in seiner Dissertationsschrift von 1851 benutzte er ein Theorem im Beweis seines Abbildungssatzes, welches erst zu Beginn des 20. Jahrhunderts von D. Hilbert im Zusammenhang mit der Theorie der linearen Integralgleichungen endgültig bewiesen wurde [37] und seitdem als „Riemann-Hilbertsches Problem" bezeichnet wird.

[2]Eine Funktion $f(x, y)$ im $\mathbf{R}^2$ heißt *harmonisch*, wenn sie die Laplace- oder Potentialgleichung erfüllt: $\triangle f = 0$.

Dabei geht es darum, eine in einem gewissen Gebiet analytische Funktion zu finden, die eine gegebene Beziehung zwischen den Randwerten ihres Real- und Imaginärteils erfüllt. Eine eingehende Beschreibung der historischen Entwicklung des Randwertproblems und die Auswirkungen auf den Fortgang der Funktionentheorie findet sich in Felix Kleins Werk „Entwicklung der Mathematik im 19. Jahrhundert“ [45].
In gewissem Sinne lassen sich die nachfolgenden Überlegungen auch als Spezialfall des Riemann-Hilbertschen-Problems für unbeschränkte Gebiete formulieren; allerdings bezahlt man diesen abstrakten Standpunkt mit einem erheblichen begrifflichen Aufwand, und aus diesem Grund werden wir uns dem Hauptergebnis dieses Abschnitts auf elementarem Wege nähern. An das Signal $s(t)$ werden folgende Bedingungen gestellt:

- $s(t)$ ist ein energiebeschränktes Signal:

$$\int_{-\infty}^{\infty} |s(t)|^2 \, \mathrm{d}t \quad < \quad \infty \,. \tag{4.3}$$

- $s(t)$ ist hölderstetig, d.h. es gibt zwei Konstanten $K > 0$ und $0 < \alpha \leq 1$, so daß für alle $t_1, t_2 \in \mathbf{R}$ gilt

$$|s(t_1) - s(t_2)| \leq K \cdot |t_1 - t_2|^\alpha \,. \tag{4.4}$$

- $s(t)$ verschwindet im Unendlichen:

$$\lim_{t \to \pm\infty} s(t) \quad = \quad 0 \,. \tag{4.5}$$

Für die weiteren Überlegungen benötigen wir einen Satz, der in der Funktionentheorie unter dem Namen *Plemelj-Sochozki-Formeln* bekannt ist und der für unsere Zwecke in einer modifizierten und weniger allgemeinen Fassung Anwendung findet [40, 28]. Die nachfolgende Version dieses Satzes ist dem Buch von E. Meister [54] entnommen, in welchem sich auch der Beweis findet (S.108 ff.).

Satz 4.1 (Plemelj-Sochozki-Formeln). *Es sei f eine auf $\mathbf{R} \cup \{\infty\}$ hölderstetige reellwertige Funktion mit dem Grenzwert $\lim\limits_{|x|\to\infty} f(x) = 0$, dann definiert*

$$F(z) \quad = \quad \frac{1}{2\pi i} \int\limits_{-\infty}^{\infty} \frac{f(x)}{x-z} dx$$

eine in $\mathbf{C} \setminus \mathbf{R} = H^+ \cup H^-$ stückweise holomorphe Funktion, die eine hölderstetige Randfunktion

$$F^o(x) := \lim_{y\to 0} F(x+iy) = \begin{cases} \frac{f(x)}{2} + \frac{1}{2\pi i} \int\limits_{-\infty}^{\infty} \frac{f(t)}{t-x} dt & : \quad y > 0 \\ -\frac{f(x)}{2} + \frac{1}{2\pi i} \int\limits_{-\infty}^{\infty} \frac{f(t)}{t-x} dt & : \quad y < 0 \end{cases}$$

besitzt und im Unendlichen die Randwerte

$$F^o(\infty) := \lim_{z\to\infty} F(z) = 0$$

annimmt. Hierbei bezeichnet H^+ die obere und H^- die untere Halbebene, das Integral ist wieder im Sinne des Cauchyschen Hauptwerts zu verstehen.

Betrachten wir nun eine auf H^+ holomorphe Funktion

$$F(z) = U(x,y) + iV(x,y) \ , \quad y > 0$$

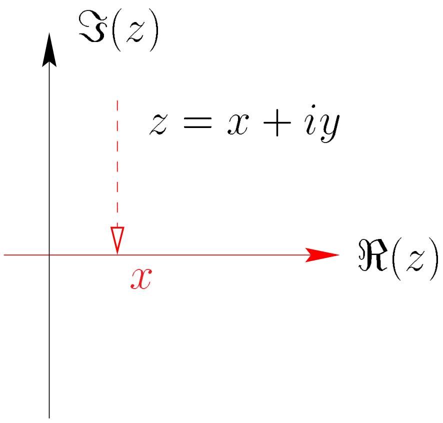
mit hölderstetigen Randwerten $F^o(x) = U(x,0) + iV(x,0)$, die der Bedingung

$$\lim_{|x|\to\infty} F^o(x) = 0$$

genügen sollen. Die Funktion $F(z)$ läßt sich in H^+ über das Cauchyintegral durch die Randwerte darstellen

$$F(z) \quad = \quad \frac{1}{2\pi i} \int\limits_{-\infty}^{\infty} \frac{F^o(x)}{x-z} \mathrm{d}x \ .$$

Diese Situation ist in Abbildung 4.1 veranschaulicht. Die Plemelj-Sochozki-



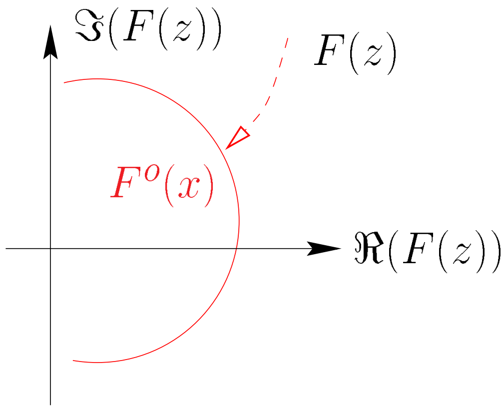


Abbildung 4.1: Das obere Bild zeigt die reelle Achse als Rand der oberen Halbebene H^+. In der unteren Abbildung sieht man die in H^+ analytische Funktion F bei der Annäherung an ihre hölderstetige Randfunktion F^o.

Formeln liefern uns die Beziehungen zwischen Real- und Imaginärteil von $F^o(x)$:

$$\begin{aligned}
\lim_{y \searrow 0} F(z) &= \frac{1}{2}\Big(U(x,0) + iV(x,0)\Big) + \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{U(t,0) + iV(t,0)}{t - x}\mathrm{d}t \\
&= \frac{1}{2}\left(U(x,0) - \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{V(t,0)}{x - t}\mathrm{d}t\right) + \frac{i}{2}\left(V(x,0) + \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{U(t,0)}{x - t}\mathrm{d}t\right) \\
&= \frac{1}{2}\Big(U(x,0) - \mathcal{H}\left[V(x,0)\right]\Big) + \frac{i}{2}\Big(V(x,0) + \mathcal{H}\left[U(x,0)\right]\Big)\,.
\end{aligned}$$

Soll also die Funktion $F^o(x)$ Randwert einer analytischen Funktion sein, so müssen Real- und Imaginärteil über die Hilbert-Transformation zusammenhängen:

$$\begin{aligned} U(x,0) &= -\mathcal{H}\left[V(x,0)\right] \ , \\ V(x,0) &= \mathcal{H}\left[U(x,0)\right] \ . \end{aligned} \tag{4.6}$$

Diese Gleichungen sind in der Elektrodynamik auch als *Dispersionsrelationen* oder *Kramers-Kronig-Relationen* bekannt und werden dort für die verallgemeinerte Suszeptibilität aufgestellt [46].
Für die Signalverarbeitung liefert Gleichung 4.6 eine mathematische Begründung für die Definition des *analytischen Signals*, nämlich als Randwert einer *analytischen Funktion*. Wenn ein reelles Zeitsignal $s(t)$ die Bedingungen 4.3 bis 4.5 erfüllt (und Signale aus „echten" physikalischen Systemen sind im allgemeinen energiebeschränkt, von endlicher Dauer und lassen sich durch hölderstetige Funktionen approximieren), dann gibt es nur eine Möglichkeit, einen Imaginärteil zu $s(t)$ hinzuzufügen, so daß ein „analytisches" Signal entsteht. Der Imaginärteil muß über die Hilbert-Transformation eingeführt werden, denn Gleichung 4.6 liefert eine notwendige und hinreichende Bedingung dafür, Randwert einer analytischen Funktion zu sein. Diese Aussagen lassen sich noch verallgemeinern, wenn man die Voraussetzung 4.4 (Hölderstetigkeit) fallen läßt und sich auf beliebige L_2-Funktionen bezieht, allerdings gelten die Randwerteigenschaften dann nur noch *fast überall* auf $\mathbf{R}$.
In diesem Kapitel haben wir ein mathematisches Argument für die Einführung des analytischen Signals über die Hilbert-Transformation geliefert. Ein weiterer interessanter Ansatz findet sich bei Vakemann [76, 77]. Auf den Zusammenhang zwischen analytischem Signal und dem Kausalitätsprinzip wird bei Papoulis hingewiesen [58].

Kapitel 5

Die Empirische Modenzerlegung (EMD)

„Wenn Du scharf sehen kannst, dann sieh hin“ (so hat jemand gesagt) „und urteile so weise wie möglich.“

Marc Aurel, Selbstbetrachtungen, VIII

Im folgenden werden wir ein Verfahren zur Zeit-Frequenz-Analyse vorstellen, das von Huang et al. [39] als „Empirical Mode Decomposition“ (EMD) in die Datenanalyse eingeführt wurde. Der Grundgedanke besteht in der Einführung von fundamentalen Moden[1], aus denen man sich das zu analysierende Signal zusammengesetzt denkt. Diese Zerlegung wird als „empirisch“ bezeichnet, weil die fundamentalen Moden nicht *a priori* gegeben sind, sondern in einem iterativen Verfahren aus dem Signal extrahiert werden.

Für jede dieser fundamentalen Moden kann man mit Hilfe des zugehörigen analytischen Signals einen momentanen Wert der Frequenz angeben, aus dem sich ein spezielles Spektrogramm aufbauen läßt. Durch die Beschränkung auf fundamentale Moden verliert das Verfahren an Allgemeingültigkeit gegenüber den klassischen Fourier-Methoden der Zeit-Frequenz-Analyse, aber man gewinnt dadurch eine adaptive Basis zur Beschreibung des Signals.

[1]In der Originalarbeit von Huang et al. [39] wird eine solche fundamentale Mode als „intrinsic mode function“ (IMF) bezeichnet.

Bei diesem Ansatz geht man davon aus, daß sich ein schwingungsfähiges System aus verschiedenen Subsystemen zusammensetzt, die auf charakteristischen Energie- und Zeitskalen schwingen. Bisher wurde die EMD in erster Linie in der Geophysik eingesetzt, z.B. zur Analyse seismischer Daten [50], zur Untersuchung atmosphärischer Schwankungen [82] und zum Studium des nichtlinearen Verhaltens von Wasserwellen [38].

5.1 Zeit-Frequenz-Analyse

Die Zeit-Frequenz-Analyse ist ein bevorzugtes Werkzeug zur Untersuchung nichtstationärer Prozesse. Sie ist ein Beispiel dafür, wie sich eine Idee (in diesem Fall die harmonische Analyse periodischer Signale) weiterentwickelt und sich immer größere Anwendungsfelder erschließt. Wenn man ein schwingungsfähiges System betrachtet, dessen Eigenschaften sich mit der Zeit ändern, dann wird das Spektrum allein wenig über diese Änderungen aussagen. Hingegen werden viele Kurzzeitspektren, deren zeitliche Entwicklung man verfolgen kann, die Dynamik des Prozesses gut erfassen.
Das Konzept der gefensterten Fourierspektren wurde im Laufe der Zeit erweitert. So zum Beispiel um das Skalogramm, welches auf der Theorie der Wavelets aufbaut [14, 41] oder die Wigner-Ville-Verteilung. Diese beiden Darstellungen lassen sich in der verallgemeinerten Cohen-Klasse der Zeit-Frequenz-Kerne zusammenfassen [16, 15].
Der große Vorteil dieser Methoden ist die klare theoretische Formulierung des Konzeptes. So hat man in der Fourieranalyse ein orthogonales Funktionensystem zur Verfügung, das als Basis für eine große Klasse von Signalen geeignet ist. Aber die Funktionen (in diesem Beispiel die harmonischen Funktionen), die als Funktionenbasis Verwendung finden, sind *a priori* gegeben. Dadurch unterliegen sie starken Beschränkungen bezüglich der erreichbaren Auflösung im Spektrum, die durch die Zeit-Frequenz-Unschärfe gegeben ist. Eine Erweiterung des Fourierschen Konzeptes stellen die evolutionären Spektren dar, die von Priestley in die Zeit-Frequenz-Analyse eingeführt wurden [62, 63]. Dabei wird den Koeffizienten im Fourierspektrum eine schwache Zei-

tabhängigkeit zugestanden. Allerdings ist die Schätzung evolutionärer Spektren schwierig und noch Gegenstand aktueller Forschung [23, 43, 19].
Die Verallgemeinerung der evolutionären Spektren durch die Einführung einer zeitabhängigen Phase wurde von Detka et al. vorgeschlagen [20].
Ein weiterer Ansatz stammt von Mallat et al. [52]. Darin wird das Signal in eine Summe von Wellenformen zerlegt, die aus einer redundanten Familie von Basisfunktionen ausgewählt werden. Dabei stand der adaptive Charakter dieser Zerlegung im Vordergrund, den man z.B. in der digitalen Bildverarbeitung zur effektiven Datenkompression nutzen kann [51].
Die erreichbare Auflösung im Spektrogramm ist eine Frage der verwendeten Basisfunktionen, deren Auswahl sich nach dem Vorwissen über das zu analysierende Signal richtet. Deshalb bedeutet die Zeit-Frequenz-Unschärfe in der Signalverarbeitung keine prinzipielle Schranke, obwohl die formale Analogie zur Quantenmechanik diesen Schluß nahelegt.

5.2 Das Konzept der instantanen Frequenz

Im folgenden soll das Konzept der *instantanen Frequenz* vorgestellt werden, welches eng verwandt ist mit dem Konzept des analytischen Signals. Der Begriff der instantanen Frequenz ist eine zwiespältige Angelegenheit. Auf der einen Seite ist er intuitiv einleuchtend und erscheint völlig natürlich. Betrachtet man zum Beispiel einen „Chirp“, also ein Sinussignal mit linear ansteigender Frequenz, dann ist die Bedeutung der instantanen Frequenz naheliegend. Der Hörer eines solchen Chirps würde sofort die Tonhöhe mit der instantanen Frequenz in Verbindung bringen.
Auf der anderen Seite wird die Interpretation der instantanen Frequenz schwierig, wenn im Signal mehr als eine Frequenzkomponente enthalten ist. Dies ist einer der Gründe, warum eine allgemeingültige Definition der instantanen Frequenz noch immer aussteht [10, 11].
Zuerst werden wir die Definition der instantanen Frequenz vorstellen, die über das analytische Signal gegeben wird. Es sei $s(t)$ ein reelles Zeitsignal

und $\tilde{s}(t)$ das zugehörige analytische Signal in der Form

$$\begin{aligned} \tilde{s}(t) &= s(t) + i\mathcal{H}[s(t)] \\ &= A(t)e^{i\varphi(t)} . \end{aligned} \tag{5.1}$$

Dabei ist $A(t)$ die zeitabhängige Amplitude und $\varphi(t)$ die zeitabhängige Phase. In der Radiotechnik interpretiert man die zeitabhängige Amplitude als die Einhüllende, bzw. als den AM-Anteil, und die Phase als den FM-Anteil des Signals. Als instantane Frequenz $\omega(t)$ definiert man nun die Zeitableitung der Phase [16]

$$\omega(t) = \frac{\mathrm{d}\varphi(t)}{\mathrm{d}t} . \tag{5.2}$$

Versucht man nun, dieses Konzept auf Signale mit verschiedenen Frequenzkomponenten anzuwenden, so gelangt man schnell zu Widersprüchen zur Anschauung. Betrachtet man zum Beispiel das Signal

$$s(t) = A_1 \cos(\omega_1 t) + A_2 \cos(\omega_2 t) \tag{5.3}$$

mit dem analytischen Signal

$$\begin{aligned} \tilde{s}(t) &= A_1 e^{i\omega_1 t} + A_2 e^{i\omega_2 t} \\ &= A(t)e^{\varphi(t)} . \end{aligned}$$

Für Amplitude und Phase des analytischen Signals erhalten wir

$$\begin{aligned} A(t) &= \sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos((\omega_2 - \omega_1)t)} , \\ \varphi(t) &= \arctan\left(\frac{A_1\sin(\omega_1 t) + A_2\sin(\omega_2 t)}{A_1\cos(\omega_1 t) + A_2\cos(\omega_2 t)}\right) . \end{aligned}$$

Die instantane Frequenz erhalten wir aus der Ableitung der Phase

$$\varphi'(t) = \frac{1}{2}(\omega_1 + \omega_2) + \frac{1}{2}(\omega_2 - \omega_1)\frac{A_2^2 - A_1^2}{A^2(t)} .$$

In Abbildung 5.1 ist die instantane Frequenz für unterschiedliche Amplituden A_1, A_2 der verwendeten harmonischen Signale dargestellt. In einem der

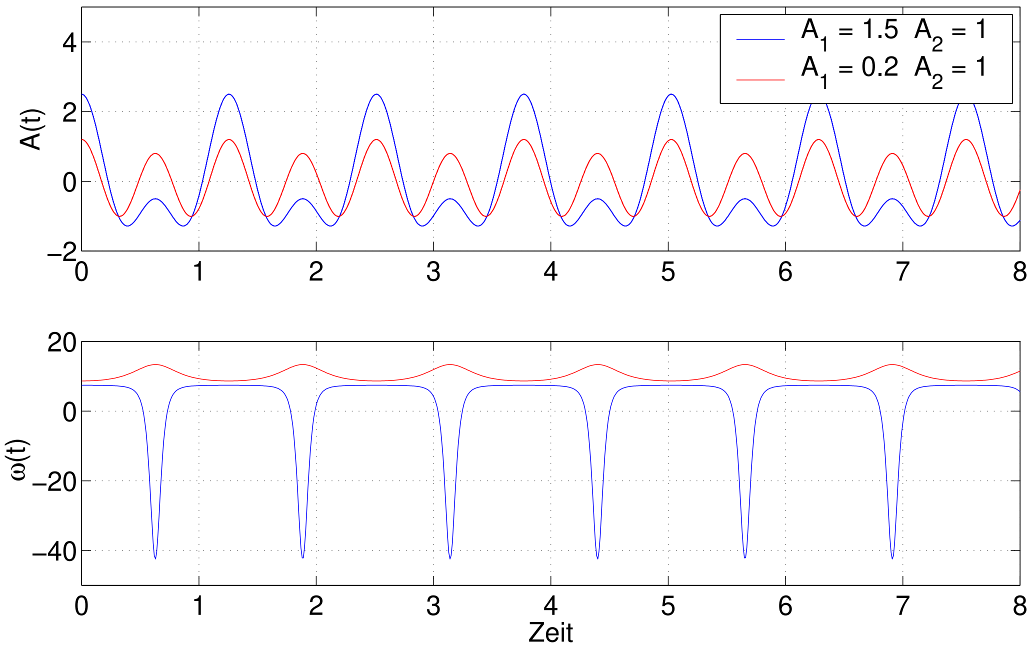


Abbildung 5.1: Das Schwingungssignal aus Gleichung 5.3 für $\omega_1 = 5$ und $\omega_2 = 10$ sowie die zugehörigen instantanen Frequenzen als Funktion der Zeit.

beiden Fälle beobachtet man das Auftreten von negativen instantanen Frequenzen, die der Anschauung zuwiderlaufen, da diese Frequenzen im Spektrum des Signals keine Repräsentation besitzen. Diese negativen instantanen Frequenzen kommen durch die negativen lokalen Maxima des Signals zustande.

Das Konzept der instantanen Frequenz und des analytischen Signals wurde bereits erfolgreich zur Analyse technischer Systeme eingesetzt [27, 26], allerdings beschränkte sich die Anwendung auf mittelwertfreie Signale mit wenigen Frequenzkomponenten [38]. Um multikomponente Signale zu untersuchen, muß man mehrere, auf die einzelnen Komponenten bezogene instantane Frequenzen einführen [6, 10, 11, 15, 16]. Die Zerlegung des Signals in einzelne Komponenten mit definierter instantaner Frequenz ist das Ziel der EMD.

5.3 Die fundamentalen Moden eines Signals

Die fundamentalen Moden des Signals bilden den Kern der EMD. Diese Moden werden aus dem Signal in einem iterativen Verfahren extrahiert. Jede Mode repräsentiert einen Oszillator, der mit zeitlich veränderlicher Amplitude und Frequenz schwingen kann. Eine fundamentale Mode der EMD entspricht formal einer harmonischen Funktion in der Fourieranalyse.

An eine fundamentale Mode werden zwei Bedingungen gestellt:

1. Die Anzahl der lokalen Extrema und die Anzahl der Nulldurchgänge muß entweder gleich sein oder darf sich nur um Eins unterscheiden.

2. Die Einhüllende der lokalen Maxima und die Einhüllende der lokalen Minima addieren sich in jedem Punkt zu Null.

Die erste Bedingung stellt sicher, daß die instantane Frequenz einer fundamentalen Mode zu jedem Zeitpunkt einen wohldefinierten positiven Wert besitzt. Diese Bedingung sorgt nämlich gerade dafür, daß die lokalen Maxima der fundamentalen Mode positiv und die lokalen Minima immer kleiner als Null sind. In der zweiten Bedingung wird die Symmetrie der Einhüllenden einer fundamentalen Mode gefordert. Dadurch lassen sich unerwünschte Fluktuationen der instantanen Frequenz unterdrücken. In Abbildung 5.2 ist eine fundamentale Mode dargestellt. Die Einhüllende der Extrema wurde durch einen kubischen Spline approximiert.

Die fundamentalen Moden werden in einem iterativen Verfahren berechnet, das in der Tabelle 5.1 beschrieben ist. Dabei bezeichnet $s(t)$ das reelle Signal, das in die k fundamentalen Moden $c_1(t), \ldots, c_k(t)$ und das Residuum $r(t)$ zerlegt wird. Das Signal $s(t)$ besitzt nach der EMD die Darstellung

$$s(t) \;=\; \sum_{i=1}^{k} c_i(t) \;+\; r(t)\,. \tag{5.4}$$

Jede fundamentale Mode wurde so konstruiert, daß sie sich durch ihre instantane Amplitude und ihre instantane Phase darstellen läßt [39, 38]. Es sei

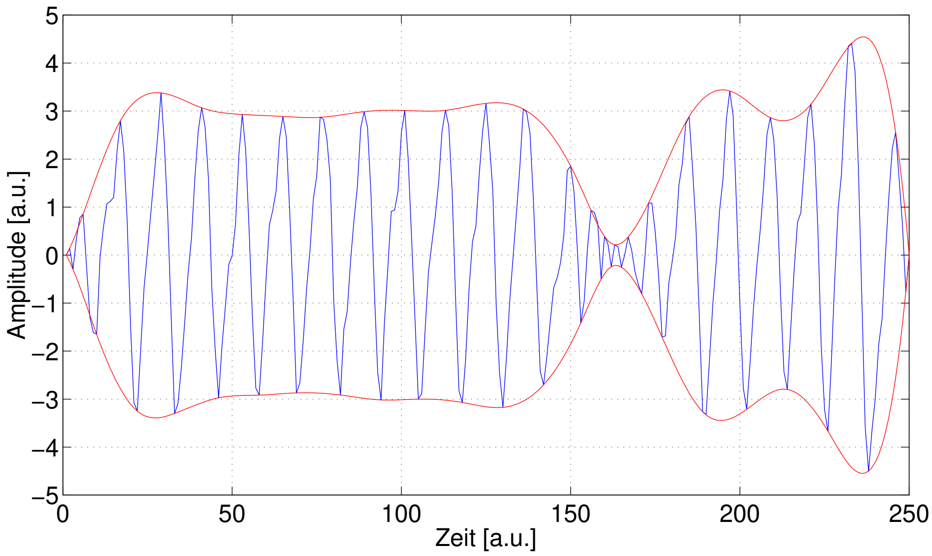


Abbildung 5.2: Diese Abbildung zeigt exemplarisch eine fundamentale Mode eines Signals (blaue Kurve). Die Einhüllende der lokalen Maxima und die Einhüllende der lokalen Minima ist jeweils rot dargestellt. Nach jedem Extremum folgt ein Nulldurchgang der fundamentalen Mode, und die Einhüllenden der Extrema sind symmetrisch bezüglich der Zeitachse.

$c_j(t)$ die j-te fundamentale Mode des Signals. Diese Mode können wir als reelles Signal auffassen und durch Anwendung der Hilbert-Transformation in das analytische Signal $\tilde{c}_j(t)$ verwandeln

$$\begin{aligned} \tilde{c}_j(t) &= c_j(t) + i\mathcal{H}\left[c_j(t)\right] \\ &= A_j(t)e^{i\varphi_j(t)} , \end{aligned}$$

wobei $A_j(t)$ die zeitabhängige Amplitude und $\varphi_j(t)$ die zeitabhängige Phase bezeichnet. Das Signal $s(t)$ hat dann die Gestalt

$$\begin{aligned} s(t) &= \sum_{j=1}^{k} c_j(t) \ + \ r(t) \\ &= \Re\left(\sum_{j=1}^{k} A_j(t)e^{i\varphi_j(t)}\right) \ + \ r(t) . \end{aligned}$$

1. Zur Initialisierung werden $r_0(t) = s(t)$ und $i = 1$ gesetzt.
2. Zur Berechnung der i-ten fundamentalen Mode sind folgende Schritte erforderlich:
   (a) Zur Initialisierung werden $h_0(t) = r_i(t)$ und $j = 1$ gesetzt.
   (b) Die lokalen Extrema von $h_{j-1}(t)$ werden bestimmt.
   (c) Durch die lokalen Maxima und durch die lokalen Minima von $h_{j-1}(t)$ wird jeweils ein kubischer Spline gelegt, um die obere und untere Einhüllende von $h_{j-1}(t)$ zu approximieren.
   (d) Der Mittelwert $m_{j-1}(t)$ aus der oberen und der unteren Einhüllenden wird gebildet.
   (e) Zwischen $h_{j-1}(t)$ und $m_{j-1}(t)$ wird die Differenz gebildet: $h_j(t) = h_{j-1}(t) - m_{j-1}(t)$.
   (f) Falls $h_j(t)$ das Abbruchkriterium erfüllt, hat man die i-te fundamentale Mode gefunden: $c_i(t) := h_j(t)$. Anderenfalls setzt man $j = j + 1$ und geht wieder zu 2b. Die Berechnung wird abgebrochen, wenn sich die Anzahl der Extrema und die Anzahl der Nulldurchgänge von $h_j(t)$ höchstens um Eins unterscheiden.
3. Die fundamentale Mode $c_i(t)$ wird von $r_{i-1}(t)$ abgezogen: $r_i(t) := r_{i-1}(t) - c_i(t)$.
4. Wenn das Residuum $r_i(t)$ noch mehr als zwei Extrema besitzt, setzt man $i = i + 1$ und geht wieder zu 2. Anderenfalls ist die EMD beendet und es verbleibt das Residuum $r(t) := r_i(t)$. Das Signal $s(t)$ besitzt dann die Darstellung $s(t) = \sum_{k=1}^{i-1} c_k(t) \; + \; r(t)$.

Tabelle 5.1: Algorithmus zur empirischen Modenzerlegung des Signals $s(t)$

Die instantane Frequenz der k-ten fundamentalen Mode erhält man gemäß Definition 5.2 durch die Ableitung der Phase:

$$\omega_k(t) \quad = \quad \frac{\mathrm{d}\varphi_k(t)}{\mathrm{d}t} \; .$$

Auf dieser Grundlage kann man nun ein modifiziertes Spektrogramm aufbauen, indem man in der Zeit-Frequenz-Ebene die instantane Frequenz der einzelnen Moden gegen die Zeit aufträgt und die zugehörigen instantanen Amplituden farbkodiert. Dieses Spektrogramm wurde von Huang et al. als *Hilbert-Spektrum* bezeichnet [39].

5.4 Anwendungen

Die EMD ist zur Analyse oszillatorischer Signale gedacht, also zur Untersuchung von Systemen, die aus schwingungsfähigen Subsystemen zusammengesetzt sind. Die Schwingungen dieser Subsysteme versucht man durch die Extraktion der fundamentalen Moden zu detektieren. Es macht offensichtlich keinen Sinn, unbekannte stochastische Prozesse mit der EMD untersuchen zu wollen. Wie bereits erwähnt, wurde die EMD bereits erfolgreich zur Untersuchung geophysikalischer Daten eingesetzt. In den untersuchten Systemen liegen verschiedene Subsysteme vor, die auf ihren eigenen Energie- und Zeitskalen schwingen und deren Verhalten sich in den fundamentalen Moden trennen läßt.

Als Beispiel für die breiten Anwendungsmöglichkeiten der EMD soll ein Datensatz aus der Geophysik untersucht werden. Es handelt sich hier zwar nicht um ein technisches System, aber es illustriert die Bedeutung der fundamentalen Moden eindrucksvoll. Am Mauna Loa Observatorium auf Hawaii wird seit Ende der fünfziger Jahre der CO_2-Gehalt der Erdatmosphäre gemessen. Eine ausführliche Beschreibung des Forschungsprogramms[2] und der meßtechnischen Details findet sich in der Arbeit von Pales und Keeling [57]. Die von uns untersuchten Daten dokumentieren den monatlichen Mittelwert der CO_2-Konzentration der Atmosphäre vom Januar 1959 bis zum Dezember 1998.

[2]Die zugehörigen Daten sind im Internet frei verfügbar unter der URL: `http://ingrid.ldgo.columbia.edu/SOURCES/.KEELING/.MAUNA_LOA.cdf/` .

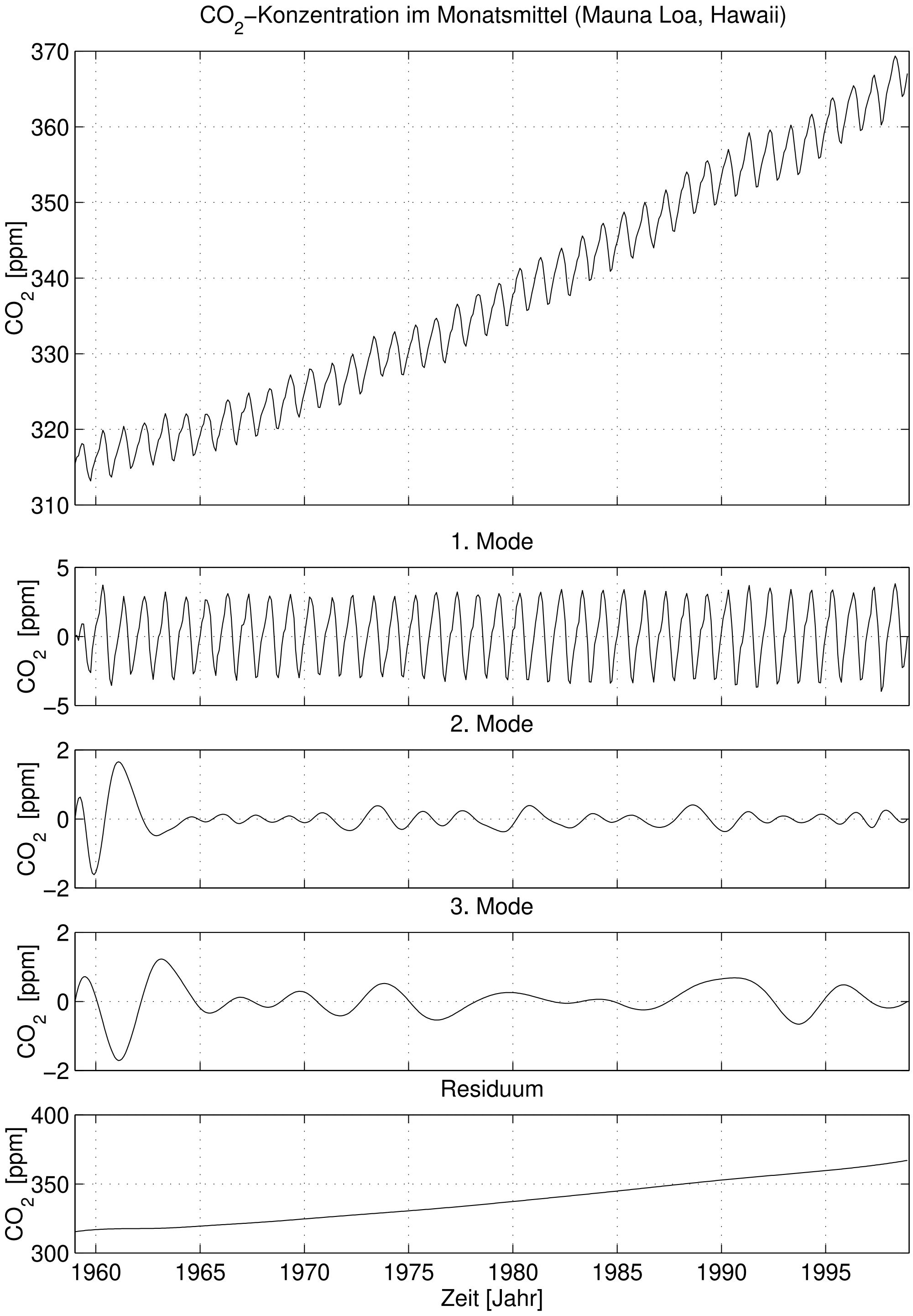


Abbildung 5.3: Das obere Diagramm zeigt die CO_2-Konzentration im monatlichen Mittel von Jan. 1959 bis Dez. 1998, gemessen am Mauna Loa Observatorium auf Hawaii. Darunter sind in absteigender Folge die fundamentalen Moden des Signals und das Residuum dargestellt.

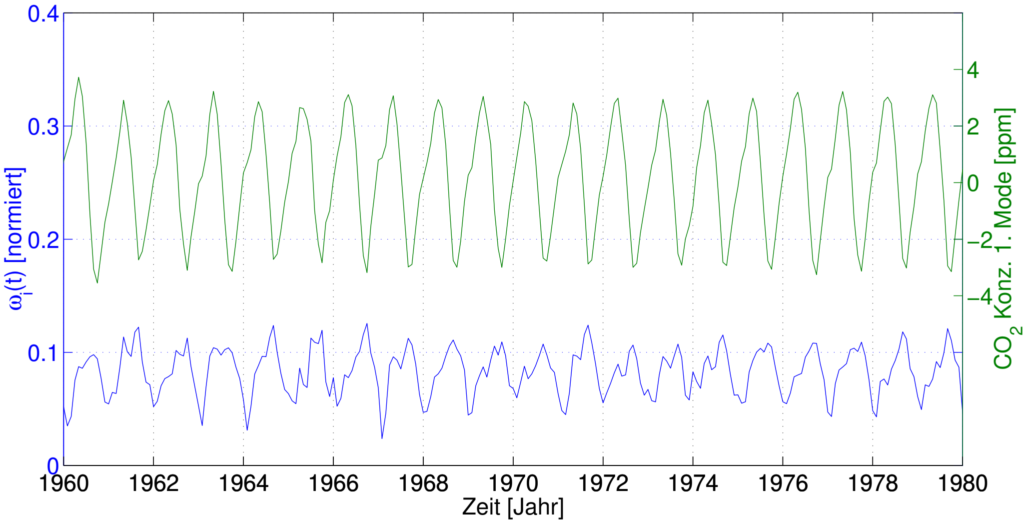


Abbildung 5.4: Die Abbildung zeigt die erste fundamentale Mode des Mauna Loa Datensatzes und die zugehörige instantane Frequenz im Zeitraum von 1965 bis 1990. Die Einbrüche der instantanen Frequenz zeigen den Beginn des Südsommers an.

In Abbildung 5.3 sind die Daten dieser Messung visualisiert. Man erkennt deutlich die jahreszeitlichen Schwankungen der CO_2-Konzentration, die von einem Langzeittrend überlagert werden. Die EMD des Signals liefert 3 fundamentale Moden und das Residuum. Die Ergebnisse der EMD sind ebenfalls in Abbildung 5.3 zu sehen. Die 1. Mode zeigt die jährlichen Schwankungen und das Residuum den kontinuierlichen Anstieg der CO_2-Konzentration im beobachteten Zeitraum. Die starken Schwankungen am Anfang der 2. und 3. fundamentalen Mode sind mit großer Wahrscheinlichkeit numerische Artefakte. Solche Randeffekte treten häufig auch bei Spektrogrammen und Skalogrammen auf und sind nur schwer zu unterdrücken [7]. Sehr interessant ist das Verhalten der 2. und 3. Mode nach 1974. Die Schwankungen dieser Moden deuten einen kurzzeitigen Rückgang der CO_2-Konzentration an, der von Keeling et al. mit der Ölkrise in Zusammenhang gebracht wurde [44].
Einen Ausschnitt der 1. fundamentalen Mode und der zugehörigen instantanen Frequenz sieht man in Abbildung 5.4. Betrachtet man zunächst die 1. fundamentale Mode der CO_2-Konzentration, dann sieht man die dominierende jährliche Schwankung. Durch die ungleichmäßige Land-Wasser-Verteilung

zwischen Nord- und Südhalbkugel und durch die unterschiedliche Vegetation spielt der Sommer auf der Nordhalbkugel bei dieser Oszillation die Hauptrolle. Mit dem Beginn der Vegetationsperiode im Frühling fällt die CO_2-Konzentration drastisch ab, weil CO_2 durch Photosynthese in den Pflanzen gebunden wird – die instantane Frequenz erreicht ihren höchsten Wert. Mit dem Ende der Vegetationsperiode auf der Nordhalbkugel steigt der CO_2-Gehalt der Atmosphäre wieder an. Dieser Anstieg wird durch den Südsommer etwas gebremst, in den Jahren 1964 und 1967 sieht man dies deutlich als „Knick“ in der 1. Mode und besonders an der extrem niedrigen instantanen Frequenz. Am Absinken der instantanen Frequenz sieht man deutlich den retardierenden Einfluß, den der Südsommer auf den Anstieg der CO_2-Konzentration ausübt.

5.5 Anmerkungen zur EMD

Die empirische Modenzerlegung ist kein unumstrittenes Verfahren zur Datenanalyse. Obwohl die EMD durchaus erfolgreich eingesetzt werden konnte, besitzt sie doch einige Schwachpunkte, die im folgenden diskutiert werden sollen.

Ein wesentlicher Kritikpunkt ist der lückenhafte theoretische Hintergrund des Verfahrens. Huang et al. behaupten, die EMD sei nach Konstruktion vollständig, und Gleichung 5.4 sei eine Identität [39]. Bisher wurde allerdings die Konvergenz des Verfahrens noch nicht bewiesen. Dieser Beweis ist aber notwendig, um die Vollständigkeit des Verfahrens „nach Konstruktion“ behaupten zu können.

Ein Nachteil des iterativen Verfahrens besteht darin, daß es bisher noch nicht möglich ist, die fundamentalen Moden des Signals in geschlossener Form anzugeben. Dadurch ist ein analytischer Zugang zur EMD verwehrt, und es ist im Grunde nicht möglich, saubere Beweise bezüglich der Vollständigkeit und der Orthogonalität der Basis zu führen. Wenn sich die EMD als Standardverfahren der Zeit-Frequenz-Analyse etablieren soll, so muß diese theoretische

Lücke geschlossen werden[3].

Ein weiterer Kritikpunkt betrifft die Wahl des Interpolationsverfahrens zur Berechnung der Einhüllenden. Neben den kubischen Splines, die von Huang et al. vorgeschlagen wurden, verwendeten wir die Akima-Interpolation, die an die kubischen Splines angelehnt ist, allerdings auf die Stetigkeit der zweiten Ableitung verzichtet [36, 2]. Als wir die Ergebnisse verglichen, zeigten sich erhebliche Abweichungen zwischen den fundamentalen Moden, die wir durch die beiden Methoden erhielten. Da es keine zwingende Notwendigkeit für die Verwendung kubischer Splines gibt, ist das Ergebnis der EMD in diesem Fall von der Wahl der Interpolationsmethode abhängig.

Bei der Anwendung der EMD auf Datensätze aus der technischen Praxis zeigten sich noch einige Mängel. Das Verfahren in seiner ursprünglichen Form [39] zeigte sich anfällig gegen den Einfluß von Rauschen. Dabei hatten wir den Eindruck, daß sich die fundamentalen Moden bei verrauschten Daten nicht sauber trennen lassen. Dieses Phänomen wird von Huang et al. in einer neueren Arbeit beschrieben und als *mode mixing* bezeichnet [38]. Um diesem störenden Einfluß zu begegnen, schlagen Huang et al. vor, die fundamentalen Moden jeweils auf einen bestimmten Frequenzbereich einzuschränken.

Zusammenfassend kann man sagen, daß sich die EMD trotz theoretischer Unklarheiten bei der Untersuchung von Systemen bewährt hat, die zum einen aus oszillierenden Subsystemen aufgebaut sind und zum anderen eine im wesentlichen unverrauschte Messung gestatten.

[3] Es ist bekannt, daß die wesentlichen Konvergenzsätze für Fourierreihen nicht zeitgleich mit ihrer heuristischen Einführung durch J.B.J. Fourier bewiesen wurden, sondern erst im Laufe der nachfolgenden Dekaden, als klar wurde, welche Tragweite die Fourieranalyse besitzt [34].

Kapitel 6

Schlußbetrachtung

Ein Ziel dieser Arbeit lag in der Entwicklung und Erprobung neuer Analyseverfahren zur Untersuchung technischer Systeme. Dieses Ziel konnte im Falle des Auswuchtproblems durch Auswahl und Modifikation bekannter Methoden der Nichtlinearen Dynamik erreicht werden. Die Vorhersage eines Systemzustandes durch benachbarte Zustände ist ein gängiges Verfahren, das seit langem erfolgreich in der NLD eingesetzt wird. Für dieses Verfahren haben wir ein neues Anwendungsfeld erschlossen, indem wir es an die technischen Besonderheiten des Auswuchtvorgangs anpaßten und seine Leistungsfähigkeit unter Berücksichtigung der Drehinvarianz entscheidend verbesserten.

Bei der Untersuchung der Bewegung einer umlaufenden Welle stand die Signaltrennung im Vordergrund. Es war bekannt, daß die Wellenbahnen rotierender Maschinen ein breites Spektrum nichtlinearer Effekte aufweisen, wie beispielsweise subharmonische, quasiperiodische und chaotische Schwingungen [1]. Außerdem waren bei unserem Industriepartner bereits nichtlineare Modelle zur Rekonstruktion der Wellenbahn erprobt worden [83]. Aus diesem Grund schien der Einsatz NLD-basierter Methoden erfolgversprechend. Es zeigte sich jedoch beim Fortgang unserer Bemühungen, daß sich die gewünschte Signaltrennung durch eine Erweiterung der bestehenden Meßanordnung auf die Lösung eines linearen Gleichungssystems zurückführen läßt. Durch eine eingehende Untersuchung des Lösungsraums konnten wir zeigen, daß unsere Methode die physikalisch relevante Lösung des Problems liefert.

Auf der Grundlage dieser Untersuchung konnte das Verfahren der Rekonstruktion von Wellenbahn und Wellenprofil zur Anwendungsreife gebracht werden.

Der Schwerpunkt des zweiten Teils dieser Arbeit lag in der Untersuchung des analytischen Signals und dessen Anwendung in der Empirical Mode Decomposition. Es wurde gezeigt, daß der Imaginärteil des analytischen Signals über die Hilbert-Transformation eingeführt werden muß, wenn das analytische Signal der Randwert einer holomorphen Funktion sein soll. Durch diese Überlegungen konnten wir der heuristischen Einführung des analytischen Signals ein mathematisches Argument zur Seite stellen. Das Konzept des analytischen Signals ist ein wesentlicher Bestandteil der EMD. Dieses Verfahren zur Zeit-Frequenz-Analyse wurde vorgestellt und an einem Datensatz erprobt, wo sich seine Leistungsfähigkeit unter Beweis stellte. Allerdings bereitete die Anwendung des Verfahrens auf verrauschte Daten aus der technischen Praxis noch einige Schwierigkeiten, so daß sich hier noch Raum für weitere Untersuchungen ergibt.

Literaturverzeichnis

[1] M.L. Adams und I.A. Abu-Mahfouz. Exploratory research on chaos concepts as diagnostic tools for assessing rotating machinery vibration signatures. In *Proceedings of the Fourth International Conference in Rotor Dynamics*, 1994.

[2] H. Akima. A new method of interpolation and smooth curve fitting based on local procedures. *J. Assoc. Comput. Mach.*, 17:589–602, 1970.

[3] J. Argyris, G. Faust, und M. Haase. *Die Erforschung des Chaos.* Vieweg, 1994.

[4] V.I. Arnold und A. Avez. *Problèmes Ergodiques de la Mécanique Classique.* Gauthier-Villars, 1967.

[5] D.K. Arrowsmith und C.M. Place. *Dynamische Systeme.* Spektrum Akademischer Verlag, Heidelberg, 1990.

[6] F. Auger und P. Flandrin. Improving the Readability of Time-Frequency and Time-Scale Representation by the Reassignment Method. *IEEE Trans. on Signal Processing*, 43(5):1068–1089, 1995.

[7] F. Auger, P. Flandrin, P. Gonçalvès, und O. Lemoine. *Time-Frequncy Toolbox for use with MATLAB, Tutorial.* CNRS (France) & Rice University (USA), 1996.

[8] A.V. Balakrishnan. Representation of signals and systems. In A.V. Balakrishnan, Herausgeber, *Communication Theory*, Band 6 von *Inter-*

University Electronics Series, Kapitel 1, Seiten 1–26. McGraw-Hill Book Company, New York, 1968.

[9] J.S. Bendat und A.G. Piersol. *Random Data.* John Wiley & Sons, New York, zweite Auflage, 1986.

[10] B. Boashash. Estimating and Interpreting the Instantaneous Frequency of a Signal – Part 1: Fundamentals. *Proc. IEEE*, 80(4):519–538, 1992.

[11] B. Boashash. Estimating and Interpreting the Instantaneous Frequency of a Signal – Part 2: Algorithms and Applications. *Proc. IEEE*, 80(4):539–568, 1992.

[12] A. Buschbeck. Private Mitteilung. Schenck RoTec GmbH, Darmstadt, 2000.

[13] M. Casdagli. Nonlinear prediction of chaotic time series. *Physica D*, 35:335–356, 1989.

[14] Ch.K. Chui. *An Introduction to Wavelets.* Academic Press, Boston, 1992.

[15] L. Cohen. Time-Frequency Distributions – A Review. *Proceedings of the IEEE*, 77(7):941–981, 1989.

[16] L. Cohen. *Time-Frequency Analysis.* Signal Processing Series. Prentice Hall, 1995.

[17] A. Cutler und L. Breiman. Archetypal Analysis. *Technometrics*, 36(4):338–347, 1994.

[18] A. Cutler und E. Stone. Moving Archetypes. *Physica D*, 107:1–16, 1997.

[19] C. Detka und A. El-Jaroudi. The Generalized Evolutionary Spectrum. *IEEE Trans. on Signal Processing*, 44(11):2877–2881, 1996.

[20] C.S. Detka und A. El-Jaroudi. Variable Frequency Evolutionary Spectrum. In *IEEE Int. Conference on Acoustics, Speech and Signal Processing*, Seiten 1553–1556, 1995.

[21] J.J. Dongarra, J.R. Bunch, C.B. Moler, und G.W. Stewart. *LINPACK Users' Guide*. SIAM, 1979.

[22] J.-P. Eckmann und D. Ruelle. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.*, 57(3):617–656, 1985.

[23] M. Emresoy und A. El-Jaroudi. Evolutionary Spectrum Estimation using Positivity Constrained Deconvolution. *IEEE Trans. on Signal Processing*, 47(3):889–893, 1999.

[24] J.D. Farmer und J.J. Sidorowich. Predicting chaotic time series. *Phys. Rev. Lett.*, 59(8):845–848, 1987.

[25] K. Federn. *Auswuchttechnik*, Band 1. Springer-Verlag, Berlin, 1977.

[26] M. Feldman. Non-Linear Free Vibration Identification via the Hilbert Transform. *Journal of Vibration and Control*, 208(3):475–489, 1997.

[27] M. Feldman und S. Seibold. Damage Diagnosis of Rotors: Application of Hilbert-Transform and Multi-Hypothesis Testing. *Journal of Vibration and Control*, (4):421–445, 1999.

[28] W. Fischer und I. Lieb. *Ausgewählte Kapitel aus der Funktionentheorie*. Vieweg, 1988.

[29] W. Fischer und I. Lieb. *Funktionentheorie*. Vieweg, 6. Auflage, 1992.

[30] D. Gabor. Theory of Communication. *Journal of the IEE*, 93:429–457, 1946.

[31] H. Gasch, R. und Pfützner. *Rotordynamik*. Springer-Verlag, Berlin, 1975.

[32] R. Gasch. Kleiner Beitrag zur Behandlung des dynamischen Verhaltens einer rotierenden Welle mit angerissenem Querschnitt. ILR-Bericht 8, TU-Berlin, ISBN 3 7983 0551 X, 1975.

[33] W. Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *ZS. f. Physik*, 43:172 ff., 1927.

[34] H. Heuser. *Lehrbuch der Analysis 2*, Kapitel XXIX, Seite 689 ff. Teubner, Stuttgart, vierte Auflage, 1988.

[35] H. Heuser. *Lehrbuch der Analysis 1.* Teubner, Stuttgart, 1990.

[36] D. Hilberg. Akima-interpolation. *Magazin für Computertechnik*, (6):206–214, 1989.

[37] D. Hilbert. *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen.* Teubner, Berlin, erste Auflage, 1912.

[38] N.E. Huang, Z. Shen, und S.R. Long. A New View of Nonlinear Water Waves: The Hilbert Spectrum. *Annu. Rev. Fluid Mech.*, 31:417–457, 1999.

[39] N.E. Huang, Z. Shen, S.R. Long, M.L. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung, und H.H. Liu. The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. *Proc. R. Soc. London*, A(454):903–995, 1998.

[40] A. Hurwitz und R. Courant. *Vorlesungen über allgemeine Funktionentheorie und elliptische Funktionen.* Springer-Verlag, Berlin, 1925.

[41] G. Kaiser. *A Friendly Guide to Wavelets.* Birkhäuser, Basel, 1995.

[42] H. Kantz und T. Schreiber. *Nonlinear time series analysis.* Cambridge University Press, Cambridge UK, 1997.

[43] S.A. Kayhan, A. El-Jaroudi, und L.F. Chaparro. Data-Adaptive Evolutionary Spectral Estimation. *IEEE Trans. on Signal Processing*, 43(1):204–213, 1995.

[44] C.D. Keeling, J.F.S. Chine, und T. P. Whorf. Increased activity of northern vegetation inferred from atmospheric CO_2 measurements. *Nature*, 382:146–149, 1996.

[45] F. Klein. *Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert*, Band I und II. Chelsea Publishing Company, New York, 1956.

[46] L.D. Landau und E.M. Lifschitz. *Statistische Physik Teil I*, Band 5 von *Lehrbuch der theoretischen Physik*. Akademie-Verlag, Berlin, 1987.

[47] W. Lauterborn. Methoden der Chaosphysik. *Kleinheubacher Berichte*, 32:585–598, 1989.

[48] R.W. Leven, B.P. Koch, und B. Pompe. *Chaos in dissipativen Systemen*. Akademie Verlag, Berlin, 1994.

[49] E.M. Lorenz. Deterministic Nonperiodic Flow. *J. Atmos. Sci.*, 20:130–141, 1963.

[50] I. Magrin-Chagnolleau und R.G. Baraniuk. Empirical Mode Decomposition based Frequency Attributes. In *Proceedings of the 69th SEG Meeting*, Seiten 1949–1952, Houston, Texas, 1999.

[51] S. Mallat. Applied Mathematics Meets Signal Processing. In *International Congress of Mathematicians*. http://www.mathematik.uni-bielefeld.de/ICM98/proceedings/ps/ps.html, 1998.

[52] S. Mallat und Z. Zhang. Matching Pursuits with Time-Frequency Dictionaries. *IEEE Trans. on Signal Processing*, 41:3397–3415, 1993.

[53] J. McNames. A nearest trajectory strategy for time series prediction. In *Proceedings of the International Workshop on Advanced Black-Box Techniques for Nonlinear Modeling*. K. U. Leuven Belgium, 1998.

[54] E. Meister. *Randwertaufgaben der Funktionentheorie*. Teubner, Stuttgart, 1983.

[55] C. Merkwirth. Fast Exact and Approximate Nearest Neighbor Searching for Nonlinear Signal Processing. In *Proceedings of the ECCTD*. Stresa, Italy, 1999.

[56] C. Merkwirth, U. Parlitz, und W. Lauterborn. Tools for Nonlinear Data Analysis. In *Proceedings of the NOLTA'98*, Band 3, Seiten 1253–1255, 1998.

[57] J. C. Pales und C.D Keeling. The concentration of atmospheric carbon dioxide in Hawaii. *J. Geophys. Res.*, 70:6053–6076, 1965.

[58] A. Papoulis. *Signal Analysis*. McGraw-Hill Book Comp., New York, 1977.

[59] U. Parlitz und W. Lauterborn. TSTOOL – Datenanalyse mit Methoden der Nichtlinearen Dynamik. VDI-Faltblatt INFO PHYS TECH, Nr. 23, Mai 1999.

[60] U. Parlitz, C. Merkwirth, J. Wichard, M. Wiesenfeldt, und W. Lauterborn. Analysewerkzeuge für Maschinen- und Prozeßzustände. In *Technische Anwendungen von Erkenntnissen der Nichlinearen Dynamik*, Seiten 53–57, Düsseldorf, 1999. VDI-Technologiezentrum Physikalische Technologien.

[61] W.H. Press, S.A. Teukolsky, W.T. Vetterling, und B.P. Flannery. *Numerical Recipes in C : The Art of Scientific Computing*. Cambridge University Press, Cambridge, zweite Auflage, 1992.

[62] M.B. Priestley. *Spectral Analysis and Time Series*, Band 2. Academic Press, London, 1981.

[63] M.B. Priestley. *Non-linear and Non-stationary Time Series Analysis*. Academic Press, London, 1988.

[64] F. Reine. Classification of Faults in Combustion Engines. In *Proceedings of the IDA'95*, 1995.

[65] F. Reine. Signal Level Clustering for Data Analysis. In *Proceedings of the IDA'95*, 1995.

[66] F. Reine und R. Zöller. Characterization of Time Series by Dynamical Clustering. In *Proceedings of the EUFIT '96*, Band 3, Seiten 1519–1524, 1996.

[67] D. Ruelle. *Chaotic evolution and strange attractors*. Cambridge University Press, 1989.

[68] T. Sauer, J.A. Yorke, und M. Casdagli. Embedology. *J.Stat.Phys.*, 65, 1991.

[69] A. Schenck zu Schweinsberg. *Stabilisierung eines frequenzverdoppelten Festkörperlasers.* Doktorarbeit, Georg-August-Universität Göttingen, 2000.

[70] H. G. Schuster, Herausgeber. *Handbook of Chaos Control.* Wiley-VCH Verlag GmbH, Weinheim, 1999.

[71] H.G. Schuster. *Deterministic Chaos.* VCH Verlagsgesellschaft mbH, Weinheim, 1988.

[72] A. Stodola. *Dampf- und Gasturbinen mit einem Anhang über die Aussichten der Wärmekraftmaschinen.* Springer-Verlag, Berlin, 6. Auflage, 1924.

[73] J. Stoer. *Numerische Mathematik 1.* Springer-Verlag, Berlin, fünfte Auflage, 1989.

[74] F. Takens. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence*, Lect. Notes Math. Springer-Verlag, Berlin, 1981.

[75] E.C. Titchmarsh. *Theory of Fourier Integrals.* Oxford Univ. Press, New York, 1948.

[76] D. Vakeman. On the Analytic Signal, the Teager-Kaiser Energy Algorithm, and Other Methods for Defining Amplitude and Frequency. *IEEE Trans. on Signal Processing*, 44(4):791–797, 1996.

[77] D. Vakeman. Analytic waves. *International Journal of Theoretical Physics*, 36(1):227–247, 1997.

[78] P. Walters. *An introduction to ergodic theory.* Springer-Verlag, 1982.

[79] M. Weigel. Dokumentation des Maschinenzustandes mittels schwingungsdiagnostischer Überwachungssysteme. In *Tagungsband zur SIRM'97, Kassel 4.-6. März*, 1997.

[80] M. Weigel. Private Mitteilung. Brüel & Kjaer SCHENCK, Condition Monitoring System, Darmstadt, 1999.

[81] H. Weyl. *Gruppentheorie und Quantenmechanik.* Hirzel, Leipzig, 1928.

[82] X. Zhu, Z. Shen, S.D. Eckermann, M. Bittner, I. Hirota, und J.H. Yee. Gravity Wave Characteristics in the Middle Atmosphere Derived from the Empirical Mode Decomposition Method. *J.Geophys.Res.*, 102(D14):16545–16561, 1997.

[83] R. Zöller. *Methoden der Nichtlinearen Dynamik zur Zustandsanalyse technischer Systeme.* Doktorarbeit, Technische Hochschule Darmstadt, 1997.

Danksagung

Herzlich bedanken möchte ich mich bei Prof. Dr. W. Lauterborn, der diese Arbeit ermöglicht hat. An die freundliche und offene Atmosphäre in der Arbeitsgruppe „Nichtlineare Dynamik“ werde ich mich gern erinnern. Besonderen Dank schulde ich PD Dr. Ulrich Parlitz für seinen Rat in vielen Fragen und die zahlreichen Anregungen, die ich von ihm erhalten habe.

Allen Mitgliedern der Arbeitsgruppe „Nichtlineare Dynamik“ bin ich sehr zu Dank verpflichtet für viele interessante Diskussionen, die zur Fertigstellung dieser Arbeit beigetragen haben.

Ich bedanke mich bei den Mitarbeitern der Werkstätten des Hauses, den Hard- und Softwarespezialisten und den allzeit gutgelaunten Mitarbeiterinnen im Büro, die mir beim Papierkrieg stets mit Rat und Tat zur Seite standen.

Bei meinen Eltern, meiner Familie und meinen Freunden möchte ich mich ganz besonders bedanken für ihre Unterstützung.

Bei Fabian, Gisa, Irena, Jochen, CD, Martin, Thomas, Olgert, Robert und Tobias bedanke ich mich für die Abwechslung bei den gemeinsamen Proben und Auftritten mit der Institutsband.

Lebenslauf

Jörg Daniel Wichard
geboren am 01.08.1969 in Ziegenhain

1976 - 1980	Besuch der Kellerwaldschule in Jesberg
1980 - 1989	Besuch der Jugenddorf-Christophorusschule Oberurff, Abschluß mit Abitur
1989 - 1990	Wehrdienst im Fernmeldebataillon 320 in Frankenberg
1990 - 1997	Studium der Physik und Mathematik an der Georg-August-Universität Göttingen
1997	Diplom in Physik
1997 - 2000	Wissenschaftlicher Mitarbeiter am III. Physikalischen Institut der Universität Göttingen