

Autoregressive moving-average models for nonstationary signals and applications to speech waveforms

Doctoral Thesis

Author(s):

Dzung, Dacfe

Publication date:

1981

Permanent link:

<https://doi.org/10.3929/ethz-a-000243428>

Rights / license:

In Copyright - Non-Commercial Use Permitted

Diss. ETH Nr. 6847

AUTOREGRESSIVE MOVING-AVERAGE MODELS FOR
NONSTATIONARY SIGNALS AND APPLICATIONS TO
SPEECH WAVEFORMS

ABHANDLUNG

zur Erlangung des Titels eines
DOKTORS DER TECHNISCHEN WISSENSCHAFTEN

der

EIDGENOESSISCHEN TECHNISCHEN HOCHSCHULE ZUERICH

vorgelegt von

Dacfev Dzung

dipl. El.-Ing. ETH

geboren am 8. Juli 1952,

britischer Staatsangehöriger

Angenommen auf Antrag von

Prof. Dr. H. Melchior, Referent

Prof. Dr. W. Schaufelberger, Korreferent

1981

Abstract

The parametric description of random processes by autoregressive (AR-)models leads to some well-known methods for the compact representation of time series. By estimating the parameters from the data one obtains models which can be used for, e.g., classification, prediction, and efficient transmission of signals. Particular successful applications of these methods are found in speech processing. This work treats two extensions, both of which are particular relevant to speech signals.

First, moving-average (MA-)parameters are introduced in addition to the autoregressive parameters to produce a complete autoregressive moving-average (ARMA-)model. This model allows a better description of valleys in the spectrum of the signal. However, in general one has to solve a nonlinear optimization problem to find the parameters from the data, since one wants to determine the ARMA-model which is optimum (in the prediction error sense) to the data, even if the data itself has not been generated by an ARMA-system. The convergence of the iterative numerical optimization algorithms is studied in detail analytically and by simulations. Methods are given to choose suitable starting values for the iterations, to speed up convergence and to cope with ill-conditioned cases. Ill-conditioning, characterized by steep changes in the spectrum, particularly affects the otherwise fast convergent Newton-Raphson type algorithms. The simple steepest descent gradient algorithm is more robust, but slower.

The second extension treated in this work deals with nonstationarity, described by time-varying optimum parameters. An obvious ad-hoc method to obtain time-varying estimates of the parameters is simply to segment the data into blocks and to apply the parameter computation algorithm to each block independently. A refinement of this method is proposed, where one only computes the parameter for the new block, if a test has indicated that they have sufficiently changed from the previous block, thereby making use of stationary or only slowly time-varying parts in the signal to save computation. A different, more general approach to nonstationarity considers the time-variation of the parameters as a random walk and tracks their movement sample-wise using adaptive filters. Simple stochastic gradient and a new Kalman filter type algorithms are compared again analytically and by simulations. It is proved that the Kalman filter type algorithm is superior in most respects, but it requires considerably more computation.

The proposed algorithms are applied to speech signals in order to test them under realistic conditions. Typically, an ARMA-model with 8 autoregressive and 6 moving-average parameters represents a reasonable choice for speech signals, giving a better overall fit of the spectrum than purely autoregressive models with the same total number of parameters. - The good properties of these optimum estimates justifies the use of the somewhat more complex parameter

computation, in contrast to computationally simpler, but suboptimum schemes. A substantial amount of computation could further be saved by blockwise detecting stationarity, without any noticeable influence on the resynthesized signal. - The adaptive algorithms, particularly of the Kalman filter type, showed a tracking behaviour well suited for speech signals.

Zusammenfassung

Die parametrische Beschreibung von Zufallsprozessen durch autoregressive (AR-)Modelle führt zu verschiedenen bekannten Methoden zur kompakten Darstellung von Zeitreihen. Durch die Schätzung der Parameter aus den Daten erhält man Modelle, die z.B. für die Klassifikation, Prädiktion und effiziente Übertragung von Signalen benutzt werden können. Besonders in der Sprachsignalverarbeitung finden sich viele erfolgreiche Anwendungen dieser Methoden. Diese Arbeit behandelt zwei Erweiterungen:

In der ersten Erweiterung wird das autoregressive Modell durch Einführen von moving-average (MA-)Parametern zum allgemeinen autoregressiven moving-average (ARMA-)Modell erweitert. Dieses Modell erlaubt eine bessere Beschreibung von Tälern im Signalspektrum. Dabei muss jedoch im allgemeinen ein nichtlineares Optimierungsproblem gelöst werden, um die Parameter aus den Daten zu berechnen, da ein ARMA-Modell zu bestimmen ist, das die Daten optimal (nach dem Prädiktionsfehler-Kriterium) beschreibt, auch im Falle, wo diese Daten nicht exakt einem ARMA-Prozess entsprechen. Die Konvergenz iterativer numerischer Optimierungsalgorithmen wird eingehend analytisch und durch Simulationen untersucht und Methoden zur Wahl der Startwerte zur Iteration, zur Konvergenzbeschleunigung und zur Behandlung schlecht konditionierter Fälle gegeben. Schlechte Konditionierung, charakterisiert durch steile Änderungen im Spektrum, beeinflusst vor allem die Konvergenz der im allgemeinen schnellen Newton-Raphson Algorithmen. Der einfache Gradientenalgorithmus des steilsten Abstiegs ist robuster, aber langsamer.

Die zweite Erweiterung zum bekannten AR-Verfahren behandelt die Instationarität, beschrieben durch zeitlich veränderliche Parameter. In einer naheliegenden ad-hoc Methode zur Berechnung zeitvariabler Parameter werden die Daten in Blöcke segmentiert und die Parameterberechnung erfolgt für jeden einzelnen Block getrennt. Eine Verfeinerung dieser Methode wird vorgeschlagen, wo die Parameter nur dann neu berechnet werden, wenn ein Test darauf hinweist, dass diese Parameter stark von denjenigen des letzten Blocks abweichen. Dadurch nutzt man die stationären oder nur langsam veränderlichen Teile im Signal zur Reduktion des Rechenaufwandes aus. Ein anderer, allgemeinerer Ansatz für die Instationarität betrachtet die Zeitabhängigkeit der Parameter als eine Brown'sche Bewegung und verfolgt deren Änderungen fortlaufend mittels adaptiven Filtern. Einfache stochastische Gradientenalgorithmen und ein neuer Algorithmus vom Kalman Filter Typ werden wieder analytisch und durch Simulationen verglichen. Es wird insbesondere bewiesen, dass der Kalman Filter Typ Algorithmus in den meisten Fällen überlegen ist, er benötigt allerdings einen wesentlich höheren Rechenaufwand.

Die vorgeschlagenen Algorithmen wurden alle auf Sprachsignale angewandt, um sie unter realistischen Bedingungen zu prüfen. Gewöhnlich stellt ein ARMA-Modell mit 8 autoregressiven und 6 moving-average Parametern für

Sprachsignale eine gute Wahl dar, die im ganzen gesehen eine bessere Approximation liefert als rein autoregressive Modelle mit der gleichen Gesamtzahl von Parametern. - Die guten Eigenschaften der optimalen Modellparameterschätzwerte rechtfertigen die etwas aufwendigere Berechnung gegenüber einfacheren, aber suboptimalen Methoden. Ein wesentlicher Anteil an Berechnungen konnte ausserdem durch Detektion stationärer Signalausschnitte eingespart werden, dies ohne jeden bemerkbaren Effekt auf das resynthetisierte Sprachsignal. - Die adaptiven Algorithmen, insbesondere vom Kalman Filter Typ, zeigten ein den Sprachsignalen gut angepasstes Verfolgungsverhalten.