Z. Wahrscheinlichkeitstheorie verw. Gebiete
68, 337–364 (1985)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Minimax-Robust Prediction of Discrete Time Series

Jürgen Franke

Fachbereich Mathematik der Universität Frankfurt, Robert-Mayer-Str. 10, D-6000 Frankfurt 1, Bundesrepublik Deutschland

Summary. We discuss a robust approach for predicting a weakly stationary discrete time series whose spectral density f is not exactly known. We assume that we know that $f \in \mathfrak{D}$, where $\mathfrak{D}$ is a convex set of spectral densities fulfilling some not too stringent conditions. We proof the existence of a "most indeterministic" density f^0 in $\mathfrak{D}$, and we show that the classical optimal linear predictor for a time series with spectral density f^0 is minimax-robust in the sense that it minimizes the maximal possible prediction error.

We investigate some special models $\mathfrak{D}$, and, in doing so, we illustrate a generally applicable method for determining the robust predictor. In particular, we discuss model sets $\mathfrak{D}$ which are defined by conditions on a finite part of the autocovariance sequence of the corresponding time series. These examples are of particular interest as the most indeterministic density is an autoregressive one, i.e. the robust predictor is finite. We discuss connections between this type of model set $\mathfrak{D}$ and maximum entropy and generalized maximum entropy spectral estimates.

1. Introduction

The classical theory of linear prediction of weakly stationary discrete time series, due mainly to Wiener and Kolmogorov, is based on complete information about the spectral measure of the time series, a situation rarely encountered in applications. Normally, one has only approximate knowledge about the spectral measure.

In Chap. 2 we develop a general approach to the problem of optimal linear prediction one time unit ahead in the case of incomplete information. For simplicity, we consider only time series with absolutely continuous spectral measure. We assume that our information consists of knowing that the spectral density is contained in some convex set $\mathfrak{D}$, which we call the spectral information set, and which is a subset of the space $L^1(\lambda)$ of functions on $(-\pi, \pi]$

integrable with respect to Lebesgue measure λ. Our approach is closely related to Huber's theory of robust estimation of location parameters. Martin [17] considers robustness with respect to certain types of innovation distributions. Our approach differs, in that we focus our attention on robustness with respect to the second moment structure, in the spirit of the classical theory.

We first investigate the notion of a "most indeterministic" spectral density f^0 in the set of densities $\mathfrak{D}$. This is the density in $\mathfrak{D}$ for which classical optimal prediction has maximal risk, and, in this sense, it represents the least favorable situation which is still compatible with the given spectral information. Then, after clarifying the appropriate notion of "predictor" in our situation, we show that, under quite general assumptions on $\mathfrak{D}$, the classical optimal predictor Π^0 for a time series with f^0 as spectral density is a good predictor uniformly with respect to all spectral densities in $\mathfrak{D}$. More precisely, we have

$$R(\Pi^0, f) \leq R(\Pi^0, f^0) \qquad \text{for all } f \in \mathfrak{D},$$

where $R(\Pi, f)$ is the error we have to expect when we predict one step ahead a time series with spectral density f by means of the predictor Π. Thus Π^0 is a minimax-robust predictor, in the sense that it minimizes the maximal possible prediction error. This property of Π^0 is an immediate consequence of the concavity and of the form of the directional derivative of the functional $\frac{1}{2\pi}\int \log\{f(\omega)\}\, d\omega$, the logarithm of the classical minimal prediction error.

The purpose of this paper is to describe a generally useful method for finding minimax linear filters in situations where the spectral properties of the time series concerned are only vaguely known. Robust one-step prediction has been chosen as illustrative example here, but our approach can be adapted to similar problems like q-step prediction or filtering of a signal in uncorrelated noise [32].

The problem of robust prediction has been solved by Hosoya [12] for the special additive noise model

$$\mathfrak{D}_{\mathrm{AN}} = \left\{ f \,\middle|\, f = (1-\varepsilon)\, g + \varepsilon f^*, \frac{1}{2\pi}\int f^*(\omega)\, d\omega = 1 \right\}$$

which is closely related to the ε-contamination model considered by Huber [13] in location parameter estimation. Hosoya's methods rely heavily on the special properties of this model. There is no indication how to prove similar results for other kinds of spectral information set, as Hosoya's crucial Lemmas 3 and 5 make essential use of rather special features of the model considered. By applying the theory of convex optimization in Banach spaces, we are able to prove the main result for arbitrary convex sets $\mathfrak{D}$. Simultaneously, we describe a method for explicitly deriving the robust predictor with respect to those sets.

In Chap. 3 we illustrate this method in two problems. These examples should find the interest of applied people since it turns out that the minimax-robust predictor is a finite predictor which can be evaluated. These results lead us to investigate which model sets $\mathfrak{D}$ have this desirable property. We furthermore tie up robust prediction problems with the theory of maximum entropy spectral estimates, an area of remarkable current interest (see Childers [6] for extensive references). In [31], the results of Chap. 2 are applied to the prediction of a time series which is the sum of a model process with known spectral density and a noise process with only vaguely known spectral density, which may be correlated in a not completely known manner.

Vastola and Poor [33] have independently studied the prediction of a time series with only vaguely known spectral density, and they have provided an interesting alternative to the approach of Chap. 2. The precise relationship between both methods is described in [32]. As discussed at the end of Chap. 2, our approach has the advantage that it allows for a straightforward generalization to non-convex spectral information sets.

2. A General Approach to Robust Prediction

$\{X_n, -\infty<n<\infty\}$ will denote a time series, i.e. a (weakly) stationary sequence of zero-mean complex random variables, with autocovariances

$$r_n=\mathscr{E} X_{n+k} \bar{X}_n, \qquad -\infty<n, k<\infty.$$

Let $\mathscr{X}$ be the closed, linear hull of the X_n, $-\infty<n<\infty$, with respect to mean-square convergence, and let $\mathscr{X}_-$ be the subspace of $\mathscr{X}$ generated by the X_n, $n<0$. We regard $\mathscr{X}$ as a subspace of the Hilbert space $\mathscr{H}$ of all complex-valued random variables Y which have mean zero and finite variance, where the scalar product of $Y, Z\in\mathscr{H}$ is given by $\mathscr{E} Y\bar{Z}$.

For sake of simplicity, we consider only situations where $\{X_n, -\infty<n<\infty\}$ has a spectral density f. Then, the autocovariances r_n are given by

$$r_n=\frac{1}{2\pi}\int e^{in\omega} f(\omega)\, d\omega, \qquad -\infty<n<\infty.$$

Here and in the following, the range of integration is always $(-\pi, \pi]$. Let $L^2(f)$ denote the Hilbert space of complex-valued functions on $(-\pi, \pi]$ which are square-integrable with respect to the measure with density f; $L^2_-(f)$ will denote the closed subspace generated by the set $\{e^{ik\omega}, k<0\}$.

We want to predict X_0 from the X_n, $n<0$. We consider only linear predictors, i.e. elements of $\mathscr{X}_-$. For $\Pi\in\mathscr{X}_-$, we measure the performance of the linear predictor Π by its mean-square error $\mathscr{E}|X_0-\Pi|^2$.

The classical prediction theory of Wiener and Kolmogorov is concerned with minimizing the mean-square prediction error, presupposing that the autocovariances r_n, $-\infty<n<\infty$, or, equivalently, the spectral density f of $\{X_n\}$ is known. $\mathscr{E}|X_0-\Pi|^2$ is minimized, if we choose Π as the orthogonal projection $P\{X_0|\mathscr{X}_-\}$ of X_0 onto the subspace $\mathscr{X}_-$. Therefore, the classical linear prediction problem is solved if we succeed in specifying a triangular scheme $\pi_{n,k}$, $k=1, \ldots, n$, $n>0$, of prediction coefficients such that

$$\sum_{k=1}^{n} \pi_{n,k} X_{-k} \to P\{X_0 | \mathscr{X}_-\} \tag{2.1}$$

in Hilbert space $\mathscr{X}$.

The spectral representation theorem [10 – Theorem II.2″] provides an isometry from the Hilbert space $L^2(f)$ onto the Hilbert space $\mathscr{X}$, where, in particular, $e^{in\omega}$ is mapped onto X_n, $-\infty < n < \infty$. The solution of the classical prediction problem is based on this isometry: one constructs a sequence

$$\pi_n(\omega) = \sum_{k=1}^{n} \pi_{n,k} e^{-ik\omega}$$

of trigonometric polynomials which converge in $L^2(f)$ to the orthogonal projection of the function 1 onto $L^2_-(f)$. Then, the coefficients $\pi_{n,k}$, $k=1, \ldots, n$, $n>0$, satisfy (2.1).

We now leave the classical framework by assuming that f is not known, but that we have some partial knowledge of f which can be summarized to $f \in \mathfrak{D}$. The spectral information set $\mathfrak{D}$ is a given subset of $L^1(\lambda)$. We are interested in linear predictors of X_0 based on the X_n, $n<0$, which result in a small mean-square prediction error if any f in $\mathfrak{D}$ is the spectral density of $\{X_n\}$.

Before proceeding further, we first have to clarify what we mean by a predictor of X_0. As stated above, in classical prediction theory, where f is known, there are two essentially equivalent notions of a linear predictor. In time domain, a linear predictor is any element of $\mathscr{X}_-$ or, equivalently, a linear predictor is given by a triangular scheme of prediction coefficients satisfying (2.1). In frequency domain, a linear predictor is given by a function in $L^2_-(f)$. However, if we do not know f then we do not know the space $L^2(f)$. Also, we do not know the subspace $\mathscr{X}$ of $\mathscr{H}$ as, given coefficients $\pi_{n,k}$, $k=1, \ldots, n$, $n>0$, we cannot always say if $\sum_{k=1}^{n} \pi_{n,k} X_{-k}$ converges in mean-square.

In order to overcome these problems, Hosoya [12] considers only functions in $\bigcap \{L^2_-(f);\ f \in \mathfrak{D}\}$ as representing linear predictors of X_0 under uncertainty about the spectral density f, which is described by the set $\mathfrak{D}$. To stress the procedural aspect of prediction, we prefer to consider linear predictors as specified by a triangular scheme of prediction coefficients for which mean-square convergence of the corresponding linear combinations of the X_n, $n<0$, holds for all time series $\{X_n\}$ with spectral density in $\mathfrak{D}$. We, therefore, introduce the following terminology:

Definition 1. Let $\mathfrak{D}$ be a subset of $L^1(\lambda)$.

a) A $\mathfrak{D}$-global linear predictor Π of X_0 is given by a sequence $\left\{\sum_{k=1}^{n} \pi_{n,k} X_{-k},\ n \geq 1\right\}$ of finite linear combinations of the observations X_n, $n<0$, which converges in mean-square if the spectral density of the time series $\{X_n\}$ is contained in $\mathfrak{D}$.

b) If the coefficients $\pi_{n,k}$ of a $\mathfrak{D}$-global linear predictor Π do not depend on n, then we say that Π is of infinite autoregressive type, and we write

$$\Pi = \sum_{k=1}^{\infty} \pi_k X_{-k}.$$

A sequence $\left\{\sum_{k=1}^{n} \pi_{n,k} X_{-k},\ n \geqq 1\right\}$ is a $\mathfrak{D}$-global linear predictor iff the trigonometric polynomials $\sum_{k=1}^{n} \pi_{n,k} e^{-ik\omega}$ converge in $L^2(f)$ for all $f \in \mathfrak{D}$. Therefore, a $\mathfrak{D}$-global linear predictor corresponds to a predictor in the sense of Hosoya. Without further assumptions on $\mathfrak{D}$, it is, however, not obvious if for every $\pi \in \bigcap \{L^2_-(f);\ f \in \mathfrak{D}\}$ there exists a fixed sequence $\left\{\sum_{k=1}^{n} \pi_{n,k} e^{-ik\omega},\ n \geqq 1\right\}$ of trigonometric polynomials converging to π in $L^2(f)$ for all $f \in \mathfrak{D}$. As the following results in the main do not depend on the particular notion of predictor which is adopted, if only it is general enough, we do not discuss this point further.

It is convenient to extend the notion of a $\mathfrak{D}$-global linear predictor of X_0 to that of a generalized predictor of X_0. Then, we can talk about candidates for $\mathfrak{D}$-global linear predictors of X_0 without having to worry about convergence immediately.

Definition 2. a) A generalized linear predictor Π of X_0 is given by a sequence $\left\{\sum_{k=1}^{n} \pi_{n,k} X_{-k},\ n \geqq 1\right\}$ of finite linear combinations of the observations X_n, $n<0$, where $\pi_{n,k}$, $k=1, \ldots, n$, $n>0$, are arbitrary complex numbers.

b) Let f be a spectral density. We call a generalized linear predictor $\Pi = \left\{\sum_{k=1}^{n} \pi_{n,k} X_{-k},\ n \geqq 1\right\}$ a Wiener predictor with respect to f if (2.1) holds for time series $\{X_n\}$ with spectral density f.

c) Let $\Pi = \left\{\sum_{k=1}^{n} \pi_{n,k} X_{-k},\ n \geqq 1\right\}$ be a generalized linear predictor and f be a spectral density. If the sequence $\pi(\omega) = \sum_{k=1}^{n} \pi_{n,k} e^{-ik\omega}$ converges in $L^2(f)$ then we define the prediction error $R(\Pi, f)$ by

$$R(\Pi, f) = \lim_{n \to \infty} \mathscr{E} \left| X_0 - \sum_{k=1}^{n} \pi_{n,k} X_{-k} \right|^2$$
$$= \lim_{n \to \infty} \frac{1}{2\pi} \int |1 - \pi_n(\omega)|^2 f(\omega)\, d\omega$$

Otherwise, we put $R(\Pi, f) = \infty$.

The last line of Definition 2 corresponds to common usage of convex optimization theory, where convex functionals, which are not defined everywhere, are set to ∞ outside their domain of definition.

In searching for a good predictor of X_0 under spectral uncertainty, described by $\mathfrak{D}$, we adopt a minimax approach. Therefore, we introduce the following terminology:

Definition 3. a) A $\mathfrak{D}$-global linear predictor Π^r is called a minimax-robust linear predictor with respect to $\mathfrak{D}$ if it satisfies

$$\max_{f \in \mathfrak{D}} R(\Pi^r, f) = \min_{\Pi} \max_{f \in \mathfrak{D}} R(\Pi, f),$$

where the minimum is taken over all generalized linear predictors of X_0.

b) A spectral density $f^0 \in \mathfrak{D}$ is called most indeterministic in $\mathfrak{D}$ if

$$\min_{\Pi} R(\Pi, f^0) = \max_{f \in \mathfrak{D}} \min_{\Pi} R(\Pi, f),$$

i.e. f^0 is least favorable in $\mathfrak{D}$ for one-step prediction.

Under appropriate assumptions we shall show that a Wiener predictor with respect to a spectral density, which is most indeterministic in $\mathfrak{D}$, is minimax-robust with respect to $\mathfrak{D}$.

To begin with let us remark that by the theorem of Szegö, Krein and Kolmogorov [10 – Theorem III.3]

$$\min_{\Pi} R(\Pi, f) = \exp\left(\frac{1}{2\pi} \int \log\{f(\omega)\}\, d\omega\right) \tag{2.2}$$

It is more convenient to work with the exponent $I(f) = \frac{1}{2\pi} \int \log\{f(\omega)\}\, d\omega$, which we call the entropy of the spectral density f. (The relation of the functional I to Shannon's notion of entropy per degree of freedom for Gaussian time series will be made explicit in Chap. 3).

By (2.2) and by the isometry of $L^2(f)$ to the Hilbert space $\mathscr{X}$ generated by a time series with spectral density f we have

$$\exp I(f) = \inf \frac{1}{2\pi} \int \left|1 - \sum_{k=1}^{n} c_k e^{-ik\omega}\right|^2 f(\omega)\, d\omega$$

where the infimum is taken over all $n \geqq 1$ and complex $c_1, \ldots, c_n$. As infimum of functionals which are continuous with respect to the w^*-topology of $L^1(\lambda)$, $\exp\{I(f)\}$ is upper semicontinuous itself.

By (2.2), a spectral density f^0 is most indeterministic in $\mathfrak{D}$ iff the entropy functional I assumes its maximum in $\mathfrak{D}$ at f^0. The strict concavity of I on the set $\{f;\ I(f) > -\infty\}$ and the upper semicontinuity of I in the w^*-topology of $L^1(\lambda)$ immediately imply the following theorem.

Theorem 1. *Let $\mathfrak{D}$ be a set of spectral densities on which I is not identically $-\infty$.*

a) *If $\mathfrak{D}$ is convex then there exists at most one most indeterministic spectral density in $\mathfrak{D}$.*

b) *If $\mathfrak{D}$ is w^*-compact in $L^1(\lambda)$ then there exists a most indeterministic spectral density in $\mathfrak{D}$.*

By a result of Ioffe and Tihomirov [14 – Proposition 9.1.2.2], b) can be replaced by the equivalent statement: if $\mathfrak{D}$ is uniformly summable in the w^*-topology of $L^1(\lambda)$, then there exists a most indeterministic spectral density in the w^*-closure of $\mathfrak{D}$.

The second part of the theorem is not so useful for verifying the existence of a most indeterministic spectral density in $\mathfrak{D}$, as most of the particular spectral information sets $\mathfrak{D}$ discussed in the literature are not w^*-compact. However, frequently the natural generalizations of these spectral information sets to subsets of the set of all spectral measures on $(\pi, \pi]$ are compact in the w^*-topology. It is not difficult to generalize Theorem 1b) to sets $\mathfrak{D}$ of arbitrary spectral measures (Theorem 1′ below), and, in this more general framework, it is useful for infering the existence of a most indeterministic spectral measure.

As we are primarily interested in spectral information sets $\mathfrak{D} \subseteq L^1(\lambda)$ which are not w^*-compact, it is convenient to assume convexity of $\mathfrak{D}$. Then, convex optimization methods provide not only necessary, but also sufficient conditions on a most indeterministic spectral density, as given in the following proposition. We shall discuss the convexity assumption in more detail at the end of this chapter.

Proposition 1. a) *Let $\mathfrak{D}$ be a convex set of spectral densities such that I is not identically $-\infty$ on $\mathfrak{D}$. f^0 is most indeterministic in $\mathfrak{D}$ iff*

$$\frac{1}{2\pi}\int \{f(\omega)/f^0(\omega)\}\, d\omega \leqq 1 \qquad \textit{for all } f \in \mathfrak{D}.$$

b) *Let $\mathfrak{D}$ be an arbitrary convex set in $L^1(\lambda)$ such that the set of functions in $\mathfrak{D}$ which are essentially bounded from below is dense in $\mathfrak{D}$. Let f^0 be a spectral density in $\mathfrak{D}$ for which $1/f^0$ is essentially bounded. Then, f^0 is a most indeterministic spectral density in $\mathfrak{D}$ iff*

$$\frac{1}{2\pi}\int \{f(\omega)/f^0(\omega)\}\, d\omega \leqq 1 \qquad \textit{for all } f \in \mathfrak{D}.$$

Proof. By Lemma 1 of the appendix f^0 is most indeterministic in a convex set $\mathfrak{D} \subset L^1(\lambda)$ iff $I_0'(f) \leqq I_0'(f^0)$ for all $f \in \mathfrak{D}$, where $I_0'(f)$ denotes the directional derivative of I at f^0 in direction f.

By Lemma 2 (i) of the appendix

$$I_0'(f) = \frac{1}{2\pi}\int \{f(\omega)/f^0(\omega)\}\, d\omega$$

for all nonnegative $f \in L^1(\lambda)$ and all f^0 with $I(f^0) > -\infty$, from which a) follows.

If f^0 is nonnegative and $1/f^0$ is essentially bounded then by Lemma 2 (i) (2.1) is true for all $f \in L^1(\lambda)$ which are essentially bounded from below. As these are dense in $\mathfrak{D}$ we conclude b) from Lemma 2 (ii) and (iii). □

The motivation for stating part b) of Proposition 1 is our intention to use later on convex optimization methods for determining most indeterministic spectral densities. If we work with sets $\mathfrak{D}$ of spectral densities then we always have to deal with the additional constraint of nonnegativity. For calculating a most indeterministic spectral density, it is more convenient to forget about this constraint and work with general subsets of $L^1(\lambda)$. The boundedness assumption on $1/f^0$ is satisfied for most sets $\mathbf{D}$, as spectral densities which are not bounded away from 0 have comparatively small entropy.

The following proposition replaces Hosoya's theorem 1 [12] for our context of general model sets $\mathfrak{D}$.

Proposition 2. *Let $\mathfrak{D}$ be a convex set of spectral densities such that I is not identically $-\infty$ on $\mathfrak{D}$, and let f^0 be the most indeterministic density in $\mathfrak{D}$. Let $\Pi^0=\left\{\sum_{k=1}^{n}\pi_{n,k}^0 X_{-k},\ n\geqq 1\right\}$ be a Wiener predictor with respect to f^0, and let $\pi^0(\omega)$ denote the $L^2(f^0)$-limit of the trigonometric polynomials*

$$\pi_n^0(\omega)=\sum_{k=1}^{n}\pi_{n,k}^0 e^{-ik\omega}.$$

Then

(i) $\pi^0(\omega)\in L_-^2(f)$ *for all* $f\in\mathfrak{D}$

(ii) $\frac{1}{2\pi}\int|1-\pi^0(\omega)|^2 f(\omega)\,d\omega\leqq\frac{1}{2\pi}\int|1-\pi^0(\omega)|^2 f^0(\omega)\,d\omega$ *for all* $f\in\mathfrak{D}$.

Proof. The proof of (i) is broken into three parts.

a) If $I(f)>-\infty$, then f is the spectral density of a purely nondeterministic time series, and it admits the canonical factorization

$$f(\omega)=|h(\omega)|^2$$

where

$$h(\omega)=\sum_{k=0}^{\infty}\bar{a}_k e^{ik\omega},$$

$$|a_0|^2=\exp I(f),\qquad \sum_{k=0}^{\infty}|a_k|^2<\infty,$$

$$\pi(\omega)=1-a_0/\overline{h(\omega)}\in L_-^2(f),$$

and $\{e^{ik\omega}/\overline{h(\omega)},\ -\infty<k<\infty\}$ is an orthogonal basis of $L^2(f)$ such that $\{e^{ik\omega}/\overline{h(\omega)},\ k<0\}$ generates the same subspace as $\{e^{ik\omega},\ k<0\}$ (see e.g. [10]). $\pi(\omega)$ is the function in $L^2(f)$ which corresponds to the classically optimal predictor, i.e. to a Wiener predictor with respect to f.

b) Let $f^0(\omega)=|h_0(\omega)|^2$ be the factorization of f^0, and $|a_0^0|^2=\exp\{I(f^0)\}$. From Proposition 1 we know that $f(\omega)/f^0(\omega)$ is integrable for all $f\in\mathfrak{D}$. Therefore,

$$\pi^0(\omega)=1-a_0^0/\overline{h_0(\omega)}\in L^2(f)\qquad \text{for all } f\in\mathfrak{D},$$

and so we have particularly $\pi^0(\omega)\in L_-^2(f)$ in the deterministic case $I(f)=-\infty$, where $L_-^2(f)$ coincides with $L^2(f)$.

c) Consider a sequence of functions $\psi_n(\omega)$ which converges to 0 in $L^2(f^0)$. Let $f\in\mathfrak{D}$ be the density of a purely nondeterministic time series and let h give its factorization $f(\omega)=|h(\omega)|^2$. By the Cauchy-Schwartz inequality and Proposition 1 a):

$$|\int \{f(\omega)\,\psi_n(\omega)\,e^{-ik\omega}/h(\omega)\}\,d\omega|^2 = |\int \{\psi_n(\omega)\,e^{-ik\omega}\,h_0(\omega)\,\overline{h(\omega)/h_0(\omega)}\}\,d\omega|^2$$
$$\leqq \int \{|\psi_n(\omega)|^2 f^0(\omega)\}\,d\omega \int \{f(\omega)/f^0(\omega)\}\,d\omega$$
$$\leqq 2\pi \int \{|\psi_n(\omega)|^2 f^0(\omega)\}\,d\omega$$
$$\text{for all } -\infty < k < \infty,\ n \geqq 1.$$

We conclude that $\psi_n(\omega)$ converges weakly to 0 in the Hilbert space $L^2(f)$. Applying this argument to $\psi_n(\omega) = \pi^0(\omega) - \pi_n^0(\omega)$, we get

$$\pi_n^0(\omega) \to \pi^0(\omega) \qquad \text{weakly in } L^2(f). \tag{2.3}$$

As one-sided trigonometric polynomials, the $\pi_n^0(\omega)$ are contained in $L^2_-(f)$, and, by b), $\pi^0(\omega) \in L^2(f)$. As a closed subspace of the Hilbert space $L^2(f)$, $L^2_-(f)$ is weakly closed (e.g. [29], p. 82), and we conclude from (2.3) that $\pi^0(\omega) \in L^2_-(f)$.

d) (ii) follows from Proposition 1 and

$$|1 - \pi^0(\omega)|^2 f^0(\omega) = \exp\{I(f^0)\} \text{ a.s.} \quad \square$$

By Proposition 2, we have $\pi^0 \in L^2_-(f)$ for all $f \in \mathfrak{D}$. If f^0, the most indeterministic spectral density in $\mathfrak{D}$, satisfies certain conditions we can even specify a sequence of one-sided trigonometric polynomials which converges to $\pi^0(\omega)$ in $L^2(f)$-norm for all $f \in \mathfrak{D}$. This sequence of polynomials then determines the desired minimax-robust linear predictor with respect to $\mathfrak{D}$.

Theorem 2. *Let $\mathfrak{D}$ be a convex set of spectral densities such that I is not identically $-\infty$ on $\mathfrak{D}$. Let f^0 be the most indeterministic density in $\mathfrak{D}$, and let $1/f^0$ be essentially bounded. Let $\pi^0(\omega)$ be the function in $L^2(f^0)$ which corresponds to the classically optimal predictor of X_0, provided $\{X_n\}$ is a time series with spectral density f^0.*

(i) *$\pi^0(\omega) \in L^2_-(\lambda)$, i.e. there exist π_k^0, $k \geqq 1$, such that*

$$\pi^0(\omega) = \sum_{k=1}^{\infty} \pi_k^0\, e^{-ik\omega}, \qquad \text{and} \quad \sum_{k=1}^{\infty} |\pi_k^0|^2 < \infty.$$

The Cesaro means

$$\pi_n^0(\omega) = \sum_{k=1}^{n} \left(1 - \frac{|k|}{n}\right) \pi_k^0\, e^{-ik\omega}, \qquad n \geqq 1,$$

converge to $\pi^0(\omega)$ in $L^2(f)$-norm for every spectral density f.

(ii) *The Wiener predictor $\Pi^0 = \left\{\sum_{k=1}^{n} \left(1 - \frac{|k|}{n}\right) \pi_k^0 X_{-k},\ n \geqq 1\right\}$ with respect to f^0 is $\mathfrak{D}$-global and minimax-robust with respect to $\mathfrak{D}$.*

(iii) *If $1/f^0$ is even a function of bounded variation over $(-\pi, \pi]$, then the series $\sum_{k=1}^{\infty} \pi_k^0$ of prediction coefficients converges absolutely, and the predictor of infinite autoregressive type $\Pi_a^0 = \prod_{k=1}^{\infty} \pi_k^0 X_{-k}$ is minimax-robust with respect to $\mathfrak{D}$.*

Proof. (i) By a theorem of Masani [18 – Theorem 2.8, Lemma 2.7] $1/f^0 \in L^1(\lambda)$ implies the existence of coefficients $\pi_1^0, \pi_2^0, \ldots$ such that for $m \to \infty$

$$\sum_{k=1}^{m} \pi_k^0 e^{-ik\omega} \to 1 - a_0^0/\overline{h_0(\omega)} = \pi^0(\omega) \quad \text{in } L^2(\lambda),$$

where, as in the proof of Proposition 2, $h_0(\omega)$ denotes the canonical factor of f^0. As $1/f^0(\omega) = 1/|h_0(\omega)|^2$ is essentially bounded, $\pi^0(\omega)$ is essentially bounded, too. By theorems of Zygmund [28 – Theorems 3.4, 3.9] the Cesaro-means $\pi_n^0(\omega)$ converge almost everywhere to $\pi^0(\omega)$, and the sequence $\pi_n^0(\omega)$ is essentially bounded uniformly in n. By Lebesgue's theorem we get for every spectral density f:

$$\pi_n^0(\omega) \to \pi^0(\omega) \quad \text{in } L^2(f).$$

(ii) From (i) we conclude for every spectral density f

$$R(\Pi^0, f) = \frac{1}{2\pi} \int |1 - \pi^0(\omega)|^2 f(\omega)\, d\omega.$$

As f^0 is most indeterministic in $\mathfrak{D}$, the last relation and Proposition 2(ii) imply

$$\max_{f \in \mathfrak{D}} R(\Pi^0, f) = R(\Pi^0, f^0) = \max_{f \in \mathfrak{D}} \min_{\Pi} R(\Pi, f).$$

This is the desired minimax property of Π^0.

(iii) If $1/f^0$ is of bounded variation then $\pi^0(\omega) = 1 - a_0^0/\overline{h_0(\omega)}$ is of bounded variation, too. Moreover, since its Fourier-coefficients vanish for $k \geqq 0$ it is in the Hardy-space H^1. By a theorem of Hoffman [11 – p. 71] $\pi^0(\omega)$ is absolutely continuous and its Fourier-series $\sum_{k=1}^{\infty} \pi_k^0 e^{-ik\omega}$ converges absolutely. As in part (i) we conclude the minimax property of the infinite autoregressive predictor $\Pi_a^0 = \sum_{k=1}^{\infty} \pi_k^0 X_{-k}$. □

Boundedness of $1/f^0$ is a rather weak condition, as we have discussed below Proposition 1. In most practically interesting examples of $\mathfrak{D}$ the even stronger assumption of Theorem 2(iii) is fulfilled, and in these situations the minimax robust predictor is of infinite autoregressive type.

Due to Masani [18 – Theorem 5.2], in classical prediction theory there is a criterion for this event to happen: f^0 should be essentially bounded, and $1/f^0$ should be integrable. We impose an analogous, in some sense stronger condition: $1/f^0$ should be of bounded variation. From this we get that the minimax-robust predictor is of infinite autoregressive type and, additionally, convergence properties of this predictor which are uniform for all spectral densities.

For concrete model sets $\mathfrak{D}$ we can explicitly determine the most indeterministic density by means of Proposition 1. We illustrate this calculation with an example.

Parzen [25] considered $\frac{1}{2\pi}\int\{f(\omega)-g(\omega)\}^2\,d\omega$ as a measure of closeness between spectral densities f and g. For fixed spectral density g and $\varepsilon>0$ we choose

$$\mathfrak{D}_2=\left\{f\in L^1(\lambda)\left|\frac{1}{2\pi}\int\{f(\omega)-g(\omega)\}^2\,d\omega\leqq\varepsilon\right.\right\}$$

as the subset of $L^1(\lambda)$ representing our knowledge of the spectral density of the time series which we want to predict.

As discussed below Proposition 1, it is convenient from a technical viewpoint to consider $\mathfrak{D}_2$ instead of the subset of all spectral densities in $\mathfrak{D}_2$.

Proposition 3. *Let g be a spectral density essentially bounded from above. Then*

$$f^0(\omega)=g(\omega)/2+\{g^2(\omega)/4+c\}^{1/2}$$

is the most indeterministic density in $\mathfrak{D}_2$, where $c>0$ is uniquely determined by $\frac{1}{2\pi}\int\{f^0(\omega)-g(\omega)\}^2\,d\omega=\varepsilon$.

Proof. Consider $f^0\in\mathfrak{D}_2\cap L^\infty(\lambda)$ for which $1/f^0$ is bounded. By Proposition 1b) f^0 is most indeterministic in $\mathfrak{D}_2$ iff

$$\Phi(f)=\frac{1}{2\pi}\int\{f(\omega)/f^0(\omega)\}\,d\omega\leqq 1\qquad\text{for all } f\in\mathfrak{D}_2. \tag{2.4}$$

(2.4) says that Φ is a support functional of the convex set $\mathfrak{D}_2$ in f^0.

From this and Lemma 5 of the appendix we get that f^0 is most indeterministic in $\mathfrak{D}_2$ iff $\frac{1}{2\pi}\int\{f^0(\omega)-g(\omega)\}^2\,d\omega=\varepsilon$ and there exists $c>0$ such that

$$\int\{f(\omega)/f^0(\omega)\}\,d\omega=\frac{1}{c}\int\{f^0(\omega)-g(\omega)\}\,f(\omega)\,d\omega\qquad\text{for all } f\in L^1(\lambda). \tag{2.5}$$

As f^0 is nonnegative, (2.5) is equivalent to

$$f(\omega)=g(\omega)/2+\{g^2(\omega)/4+c\}^{1/2}\quad\text{a.s.}$$

The right-hand side is indeed in $L^\infty(\lambda)$ and essentially bounded away from 0. Moreover, there exists a unique $c>0$ such that

$$\begin{aligned}\varepsilon&=\frac{1}{2\pi}\int\{f^0(\omega)-g(\omega)\}^2\,d\omega\\&=\frac{1}{2\pi}\int\{g^2(\omega)/4+c\}^{1/2}-\{g(\omega)/2\}^2\,d\omega\end{aligned}$$

as the right-hand side is continuous and increasing in c, and it vanishes for $c=0$. $\square$

Once f^0 is known, the minimax-robust predictor can be derived from the classical best linear predictor for a time series with spectral density f^0. As $f^0(\omega) \geqq c^{1/2} > 0$ a.s., $1/f^0$ is of bounded variation if g is of bounded variation. Then, the minimax-robust predictor with respect to $\mathfrak{D}_2$ is of infinite autoregressive type.

In essentially the same manner the most indeterministic density in Hosoya's additive noise model

$$\mathfrak{D}_{\mathrm{AN}} = \left\{ f \in L^1(\lambda) \,\middle|\, f = (1-\varepsilon)\, g + \varepsilon f^*, \frac{1}{2\pi} \int f^*(\omega)\, d\omega = 1 \right\}$$

can be determined (see [8 – Chap. 5] for the details). The support functionals of $\mathfrak{D}_{\mathrm{AN}}$ in f^0 are of the form

$$\Phi(f) = \frac{1}{2\pi} \int \{\varphi(\omega) + 1/c\}\, f(\omega)\, d\omega,$$

where $c > 0$ and $\varphi \in L^\infty(\lambda)$ satisfying

$$\begin{aligned} &\varphi(\omega) \leqq 0 \quad \text{a.s.} \quad && \text{and} \\ &\varphi(\omega) = 0 \quad \text{a.s.} \quad && \text{on } \{\omega \mid f^0(\omega) > (1-\varepsilon)\, g(\omega)\}. \end{aligned}$$

From this and Proposition 1b) we get

$$f^0(\omega) = \max\{c, (1-\varepsilon)\, g(\omega)\} \quad \text{a.s.},$$

where $c > 0$ is uniquely determined by

$$\frac{1}{2\pi} \int \{f^0(\omega) - (1-\varepsilon)\, g(\omega)\}\, d\omega = \varepsilon.$$

By Theorem 2, f^0 determines the minimax-robust predictor of $\mathfrak{D}_{\mathrm{AN}}$. This is the result which Hosoya [12] derived using a different kind of argument.

It is of theoretical interest that most results of this chapter extend easily to robust prediction of time series with arbitrary spectral measures. The reason is that the minimal prediction error for a time series with spectral measure μ depends only on the absolutely continuous part of μ. To be precise, consider the following extension of Definition 2c):

Definition 2'. Let $\Pi = \left\{ \sum_{k=1}^{n} \pi_{n,k} X_{-k},\ n \geqq 1 \right\}$ be a generalized linear predictor and μ be a spectral measure. If the sequence $\pi_n(\omega) = \sum_{k=1}^{n} \pi_{n,k}\, e^{-ik\omega}$ converges in mean-square with respect to μ then we define the prediction error $R(\Pi, \mu)$ by

$$R(\Pi, \mu) = \lim_{n \to \infty} \frac{1}{2\pi} \int |1 - \pi_n(\omega)|^2\, d\mu(\omega).$$

Otherwise, we put $R(\Pi, \mu) = \infty$.

By the theorem of Szegö, Krein and Kolmogorov,

$$\min_{\Pi} R(\Pi, \mu) = \exp\left(\frac{1}{2\pi} \int \log\{f_\mu(\omega)\}\, d\omega\right),$$

where f_μ is the density of the absolutely continuous part of the measure μ. By the same argument, which we have used for proving the upper semicontinuity of $\min_{\Pi} R(\Pi, f)$ as a functional of f, we get that $\min_{\Pi} R(\Pi, \mu)$ is upper semicontinuous as a functional of μ in the w^*-topology. This implies the following generalization of Theorem 1 b), which we state for later reference.

Theorem 1'. *Let $\mathfrak{M}$ be a w^*-compact set of spectral measures. Then, there exists a most indeterministic spectral measure μ^0 in $\mathfrak{M}$, i.e. a measure μ^0 satisfying*

$$\min_{\Pi} R(\Pi, \mu) \leqq \min_{\Pi} R(\Pi, \mu^0) \qquad \textit{for all } \mu \in \mathfrak{M}.$$

We have shown [8 – Theorem 2] that Theorem 2(iii) remains true for an arbitrary convex set $\mathfrak{M}$ of spectral measures if the most indeterministic measure is absolutely continuous and

$$\frac{1}{2\pi} \int \{1/f^0(\omega)\}\, d\mu(\omega) \leqq 1 \qquad \text{for all } \mu \in \mathfrak{M}.$$

(Remark that by Proposition 1a) the inequality holds for all absolutely continuous measures anyway.) The proof of this result is quite similar to the proof of Theorem 1 (ii); therefore, we do not give it here. The additional assumptions are easily checked for, e.g., those sets $\mathfrak{M}$, which are straightforward generalization of Hosoya's spectral information set $\mathfrak{D}_{AN}$, described above, and of the spectral information sets discussed in the following Chap. 3 [8].

In this chapter, we have considered only convex spectral information sets $\mathfrak{D}$. We have needed the convexity of $\mathfrak{D}$ for proving that the condition

$$\frac{1}{2\pi} \int \{f(\omega)/f^0(\omega)\}\, d\omega \leqq 1 \qquad \text{for all } f \in \mathfrak{D} \tag{2.6}$$

of Proposition 1 is sufficient for f^0 to be most indeterministic in $\mathfrak{D}$. The condition is still necessary if $\mathfrak{D}$ is not convex, but looks locally around f^0 like a convex set. A precise definition of this rather weak property of $\mathfrak{D}$, which requires too much preparations to repeat it here, has been given by Girsanov [30 – Chap. 6]; he calls those constraints regular at f^0. In proving Proposition 2 and Theorem 2, we have used only the necessity of the condition of Theorem 2. Therefore, Theorem 2 continues to hold for spectral information sets $\mathfrak{D}$ given by constraints on f which are regular at f^0 in the sense of Girsanov.

The difference between convex sets $\mathfrak{D}$ and sets $\mathfrak{D}$ which are only regular at f^0 is best seen by looking at the robust prediction procedure suggested by the results of this chapter. As a first step, we have to find a spectral density f^0 in $\mathfrak{D}$ which satisfies (2.6). This is a more or less straightforward task if we use the tools of convex optimization theory in $L^1(\lambda)$. If $\mathfrak{D}$ is convex then, by sufficiency of (2.6) and by Theorem 1a), we know that such a f^0 is the unique most indeterministic spectral density in $\mathfrak{D}$. If $\mathfrak{D}$ is not convex, but only regular at f^0 in the sense of Girsanov, then we have to verify by some ad hoc means that f^0, satisfying (2.6), really is most indeterministic in $\mathfrak{D}$. Finally, provided f^0 satisfies the assumptions of Theorem 2, we predict X_0 by means of the classically optimal linear predictor pretending that we are confronted with the least favorable situation, i.e. that $\{X_n\}$ is a time series with spectral density f^0.

3. Finite Robust Predictors and Maximum Entropy Spectral Densities

In this chapter we consider models $\mathfrak{D}$ where we can expect the minimax-robust predictor to have finite memory. This feature would be most desirable from the practical point of view.

In a first approach we assume exact information about some autovariances r_k of the time series $\{X_k, -\infty < k < \infty\}$. We restrict our attention to real time series for simplicity. We are given a positive definite sequence $\{c_k, 0 \leqq k \leqq M\}$ of real numbers. Our information about the process to be predicted is described by the set of spectral densities

$$\mathfrak{D}_c = \left\{ f \,\middle|\, \frac{1}{2\pi} \int \cos(k\omega) f(\omega)\, d\omega = c_k,\ 0 \leqq k \leqq M \right\}. \tag{3.1}$$

We want to predict X_0 robustly with respect to the set $\mathfrak{D}_c$. As it is convex, by Theorem 1 there exists at most one most indeterministic spectral density f^0 in $\mathfrak{D}_c$. To calculate f^0, by Proposition 1 and Theorem 2 we only have to determine the support functionals of $\mathfrak{D}_c$. This is done in Lemma 3 of the appendix. Before formulating the result, let us point out some connections between this robust prediction model and some concepts in spectral estimation theory.

Lacoss [15] discussed spectral estimation problems where the time span during which data have been obtained is of the order of periods of interest. He pointed out that common techniques of estimating the spectral density suffer considerably from inadequate resolution of neighboring peaks of spectral density and possibly from undesirable shifts in the frequency of such peaks. To circumvent both effects Burg [4, 5] proposed estimating the spectral density by means of the maximum entropy (ME) method. His criticism of the common spectral estimation methods was based on the reasoning that those make rather unrealistic assumptions about the extension of the data outside the known interval of observations; e.g. the autocorrelation approach of Blackman and Tukey [2] assumes a zero extension. Burg proposed to use the observations for extrapolating the unknown data in such a way that the resulting spectral density estimate f_E exhibits maximum entropy under all spectral densities which agree with the available data. Here, a spectral density is said to agree with the data iff its autocovariances coincide with some autocovariance estimates $C_0, \ldots, C_M$ based on the data.

If we pass a strictly stationary, e.g. Gaussian, time series through a linear filter with frequency response function h then, by a theorem of Shannon [26, 1] the resulting change in entropy is $\frac{1}{2\pi}\int \log|h(\omega)|^2\,d\omega$. In particular, let $\{Y_n, -\infty<n<\infty\}$ be a purely nondeterministic Gaussian time series with spectral density f. This process can be generated by passing white Gaussian noise $\{W_n, -\infty<n<\infty\}$ through a suitable linear filter:

$$Y_n=\sum_{k=0}^{\infty} a_k W_{n-k} \quad \text{for all } n$$

where

$$\left|\sum_{k=0}^{\infty} a_k e^{-ik\omega}\right|^2 = f(\omega) \quad \text{a.s.}$$

From Shannon's theorem and his definition of "entropy per degree of freedom" for a strictly stationary process we conclude that the entropy of the time series $\{Y_n, -\infty<n<\infty\}$ is the sum of the entropy of the one-dimensional distribution of the innovations W_n and of $\frac{1}{2\pi}\int \log\{f(\omega)\}\,d\omega$.

Due to the above reasoning, the ME-spectral density f_E with respect to the autocovariance estimates $C_0, \ldots, C_M$ is defined as the solution of the extremum problem

$$\frac{1}{2\pi}\int \log\{f(\omega)\}\,d\omega=\max!$$

under the constraints

$$\frac{1}{2\pi}\int \cos(k\omega)\,f(\omega)\,d\omega=C_k, \qquad 0\leqq k\leqq M.$$

In this formulation the connection between ME-spectral estimation and robust prediction is obvious. The ME-spectral estimate f_E is the most indeterministic density in $\mathfrak{D}_c$ for the choice $c_k=C_k$, $0\leqq k\leqq M$.

Edward and Fitelson [7] and van den Bos [3] derived explicit expressions for f_E. Makhoul [16] and Morf et al. [19, 20] discussed computationally well-behaved algorithms for determining f_E, which are based on the Levinson-Durbin algorithm for estimating autoregression coefficients. We derive the form of f_E once more as illustration for the general method of determining most indeterministic densities for special models $\mathfrak{D}$.

Proposition 4. *For $M\geqq 0$ and a positive definite sequence $c_0, \ldots, c_M$ of real numbers, let $\mathbf{C}$ denote the Toeplitz matrix $(c_{j-k})_{0\leqq j,k\leqq M-1}$, and $\mathbf{c}=(c_1, \ldots, c_M)^T$. Let $\mathbf{a}=(a_1, \ldots, a_M)^T$ be the solution of the Yule-Walker equations* [24 – Chap. 3.5.4]

$$\mathbf{C}\mathbf{a}=-\mathbf{c}, \tag{3.2a}$$

and

$$\sigma^2=\sum_{k=1}^{M} c_k a_k + c_0. \tag{3.2b}$$

The unique most indeterministic f^0 *in the spectral information set* $\mathfrak{D}_c$, *given by* (3.1), *is*

$$f^0(\omega)=\sigma^2 \Big/ \left|1+\sum_{k=1}^{M} a_k e^{-ik\omega}\right|^2. \tag{3.3}$$

The finite-memory predictor

$$\Pi^0=-\sum_{k=1}^{M} a_k X_{-k}$$

is minimax-robust with respect to $\mathfrak{D}_c$.

Proof. By Proposition 1 and by Lemma 3 of the appendix, a spectral density f^0, for which $1/f^0$ is essentially bounded, is most indeterministic in $\mathfrak{D}_c$ iff there exist Lagrange multipliers $\lambda_0, \ldots, \lambda_M$ such that

$$\frac{1}{2\pi}\int f(\omega)/f^0(\omega)\,d\omega=\sum_{k=0}^{M}\frac{1}{2\pi}\int \lambda_k \cos(k\omega) f(\omega)\,dw$$

$$\text{for all } f\in L^1(\lambda).$$

This is only possible iff

$$f^0(\omega)=1 \Big/ \left\{\sum_{k=0}^{M}\lambda_k \cos(k\omega)\right\} \quad \text{a.s.} \tag{3.4}$$

As $f^0>0$ a.s. and $f^0 \in L^1(\lambda)$, by a theorem of Hannan [10 – Theorem II.10] (3.4) implies the existence of real numbers $\gamma_0>0, \gamma_1, \ldots, \gamma_M$ such that the zeroes of the complex polynomial $\sum_{k=0}^{M} \bar{\gamma}_k z^k$ lie outside of the unit circle and

$$f^0(\omega)=1 \Big/ \left|\sum_{k=0}^{M} \gamma_k e^{-ik\omega}\right|^2.$$

Therefore, f^0 is the spectral density of an autoregressive time series of order M. By results of Pagano [23], there exists exactly one such spectral density in $\mathfrak{D}_c$, and it is given by (3.3) with $\mathbf{a}$, σ^2 satisfying (3.2). As Π^0 is the classically optimal predictor for time series with spectral density f^0, it is minimax-robust with respect to $\mathfrak{D}_c$, by Theorem 2. □

Along the same lines as Proposition 3 one can handle the case where only part of the autocovariances up to lag M are known. The appropriate model set is

$$\mathfrak{D}_{c,\mathbf{K}}=\left\{f \,\middle|\, \frac{1}{2\pi}\int \cos(k\omega) f(\omega)\,d\omega=c_k,\ k\in\mathbf{K}\right\},$$

where $\mathbf{K}$ is a subset of $\{0, \ldots, M\}$ with $0\in\mathbf{K}$.

If c_k, $k\in\mathbf{K}$, are the values of autocovariance estimates then the most indeterministic spectral density in $\mathfrak{D}_{c,\mathbf{K}}$ coincides with the *ME*-type spectral estimate derived by Newman [22].

From the practical viewpoint it is rather unrealistic to assume exact knowledge of some autocovariances and no information at all about the rest. We now consider the case that only an approximate knowledge of some autocovariances is given.

Let $\mathfrak{K}$ be a convex set in $\mathbb{R}^{M+1}$ for some $M \geqq 0$. We want to predict a (for simplicity real) time series from which we know that its covariance vector $\mathbf{r} = (r_0, \ldots, r_M)^T$ is in $\mathfrak{K}$.

Theorem 3. *Let $\mathfrak{K}$ be a compact, convex set in $\mathbb{R}^{M+1}$ with nonempty interior and containing at least one $\mathbf{r}$ for which the sequence $r_0, \ldots, r_M$ is positive definite. Let*

$$\mathfrak{D}_{\mathfrak{K}} = \left\{ f \,|\, (r_0, \ldots, r_M)^T \in \mathfrak{K} \text{ for } r_k = \frac{1}{2\pi} \int \cos(k\omega) f(\omega)\, d\omega,\ 0 \leqq k \leqq M \right\}$$

There exists a unique most indeterministic spectral density f^0 in $\mathfrak{D}_{\mathfrak{K}}$. Let $r_k^0 = \frac{1}{2\pi} \int \cos(k\omega) f^0(\omega)\, d\omega$, $0 \leqq k \leqq M$. Furthermore, we have

(i) *f^0 is the unique spectral density in $\mathfrak{D}_{\mathfrak{K}}$ for which there exist $\lambda_0, \ldots, \lambda_M \in \mathbb{R}$ such that*

$$f^0(\omega) = 1 \Bigg/ \left\{ \sum_{k=0}^{M} \lambda_k \cos(k\omega) \right\} \quad \text{a.s.} \tag{3.5}$$

and

$$\sum_{k=0}^{M} \lambda_k r_k \leqq \sum_{k=0}^{M} \lambda_k r_k^0 = 1 \quad \textit{for all } \mathbf{r} \in \mathfrak{K}.$$

In particular, f^0 is the spectral density of an autoregression of order at most M.

(ii) *Let $\gamma_0 > 0$, $\gamma_1, \ldots, \gamma_M \in \mathbb{R}$ be given by*

$$\sum_{k=0}^{M} \lambda_k \cos(k\omega) = \left| \sum_{k=0}^{M} \gamma_k e^{-ik\omega} \right|^2$$

and

$$\Pi^0 = -\sum_{k=1}^{M} \gamma_k X_{-k} / \gamma_0$$

i.e. Π^0 is the classically optimal predictor for a time series with spectral density given by (3.5). *Then, Π^0 is minimax-robust with respect to $\mathfrak{D}_{\mathfrak{K}}$.*

Proof. a) $\mathfrak{D}_{\mathfrak{K}}$ satisfies the assumptions of Theorem 4 below. In particular, (3.8) follows from the compactness of $\mathfrak{K}$. By Theorem 4, there exists a unique most indeterministic density f^0 in $\mathfrak{D}_{\mathfrak{K}}$, and it has to be the spectral density of an autoregression such that, particularly, $1/f^0(\omega)$ is essentially bounded.

b) By Proposition 1 and Lemma 4 of the appendix $f^0 \in \mathfrak{D}_{\mathfrak{K}}$ is most indeterministic in $\mathfrak{D}_{\mathfrak{K}}$ iff there exist real Lagrange multipliers $\lambda_0, \ldots, \lambda_M$ such that

$$\frac{1}{2\pi} \int \{f(\omega)/f^0(\omega)\}\, d\omega = \sum_{k=0}^{M} \frac{1}{2\pi} \int \lambda_k \cos(k\omega) f(\omega)\, d\omega \quad \text{for all } f \in L^1(\lambda),$$

and

$$\sum_{k=0}^{M} \lambda_k r_k \leqq \sum_{k=0}^{M} \lambda_k r_k^0 \quad \text{for all } \mathbf{r} \in \mathfrak{K}.$$

The rest follows completely analogous to the proof of Proposition 4. □

Theorem 3 reduces the problem of determining the most indeterministic spectral density in $\mathfrak{D}_{\mathfrak{K}}$, i.e. of solving an extremum problem in $L^1(\lambda)$, to the finite-dimensional problem of constructing a hyperplane $\left\{\mathbf{r} \in \mathbb{R}^{M+1} \,\middle|\, \sum_{k=0}^{M} \lambda_k r_k = 1\right\}$ in Euclidean space $\mathbb{R}^{M+1}$ which is tangential to the convex set $\mathfrak{K}$ in $\mathbf{r}^0$, where

$$r_k^0 = \frac{1}{2\pi} \int \cos(k\omega) \Big/ \left\{\sum_{j=0}^{M} \lambda_j \cos(j\omega)\right\} d\omega \qquad 0 \leqq k \leqq M.$$

Corollary 1. *Let T be a convex, continuous functional on $\mathbb{R}^{M+1}$, which is Gateaux-differentiable in all positive definite $\mathbf{r} = (r_0, \ldots, r_M)^T$. Let $\alpha > \inf\{T(\mathbf{r}), \mathbf{r}$ positive definite$\}$ and $\mathfrak{K} = \{\mathbf{r} \mid T(\mathbf{r}) \leqq \alpha\}$. Let f^0 be the most indeterministic spectral density in $\mathfrak{D}_{\mathfrak{K}}$ and $\mathbf{r}^0 = (r_0^0, \ldots, r_M^0)^T$ its covariance vector. Let*

$$T_0'(\mathbf{r}) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \{T(\mathbf{r}^0 + \varepsilon \mathbf{r}) - T(\mathbf{r}^0)\} = \sum_{k=0}^{M} t_k r_k, \qquad \mathbf{r} \in \mathbb{R}^{M+1}$$

be the Gateaux-derivative of T in $\mathbf{r}^0$.

There exists a Lagrange multiplier $\beta > 0$ such that

$$f^0(\omega) = 1 \Big/ \left\{\beta \sum_{k=0}^{M} t_k \cos(k\omega)\right\} \qquad \text{a.s.}$$

and β is uniquely determined by the condition $T(\mathbf{r}^0) = \alpha$.

Proof. By Theorem 3 it suffices to specify the set of support functionals $\sum_{k=0}^{M} \lambda_k r_k$ of $\mathfrak{K}$ in $\mathbf{r}^0$. By results of Ioffe and Tihomirov [14 – §4.2 and Proposition 4.3.2] these are of the form $\beta T_0'$ with $\beta > 0$, i.e. $\lambda_k = \beta t_k$, $0 \leqq k \leqq M$. □

Apart from the robust prediction setting, Theorem 3 characterizes the class of generalized maximum entropy estimates which take into account the variability of the covariance estimate appearing in the original *ME*-method. Newman [21] has discussed a special generalized *ME*-spectral estimate. He considered the neighborhood

$$\left\{f \,\middle|\, \sum_{k=-M}^{M} w_k \left|\frac{1}{2\pi} \int e^{-ik\omega} f(\omega)\, d\omega - C_k\right|^2 \leqq \sigma^2\right\} \tag{3.6}$$

where $\sigma^2 > 0$, $C_k = C_{-k}$, $w_k = w_{-k}$, $0 \leqq k \leqq M$, and $w_0, \ldots, w_M$ are some positive weights accounting for the different degree of confidence placed in each estimated autocovariance value.

(3.6) coincides with our model set $\mathfrak{D}_{\mathfrak{K}}$ for the special choice

$$\mathfrak{K}=\left\{\mathbf{r}\,\middle|\,\sum_{k=0}^{M} w_k\,|r_k-C_k|^2\leqq\sigma^2\right\}$$

The form of the ME-spectral density in (3.6) which Newman derived, can easily be shown to follow directly from Corollary 1.

In our opinion (3.6) has the drawback that the C_k are treated as non-related estimates of varying reliability. It should be preferable to take into account that $C_0, \ldots, C_M$ are highly correlated components of a covariance estimator and to consider appropriate neighborhoods $\mathfrak{K}$.

It is a somewhat surprising and satisfactory result that by Theorem 3 the minimax-robust predictor is finite even if our information about some auto-covariances of the time series to be predicted is of a rather general nature. This fact can be understood by a simple and intuitive argument.

Consider a time series $\{X_n,\ -\infty<n<\infty\}$ with spectral measure μ and positive definite covariance sequence $\{c_k,\ -\infty<k<\infty\}$. For the moment we allow for spectral measures with non-vanishing singular part.

Let $M\geqq 0$. As we have assumed that $c_0, \ldots, c_M$ is positive definite, by results of Pagano [23] there exists a unique solution $\mathbf{a}=(a_1, \ldots, a_M)^T$ of the Yule-Walker equations (3.2a), and there exists an autoregression $\{Y_n,\ -\infty<n<\infty\}$ of order at most M with autoregression coefficients $a_1, \ldots, a_M$ and innovation variance σ^2, given by (3.2b), i.e. we have

$$Y_n+\sum_{k=1}^{M} a_k Y_{n-k}=\sigma V_n \qquad \text{for all } n,$$

where $V_n,\ -\infty<n<\infty$, are uncorrelated random variables with mean zero and unit variance.

Let f_M denote the spectral density of $\{Y_n,\ -\infty<n<\infty\}$. f_M is uniquely determined by the fact that it is the spectral density of an autoregression of order at most M and by the requirement

$$\frac{1}{2\pi}\int e^{-ik\omega} f_M(\omega)\,d\omega=c_k, \qquad 0\leqq k\leqq M.$$

By (3.2) $a_1, \ldots, a_M$ are also the solution of

$$\mathscr{E}\left(X_0+\sum_{k=1}^{M} a_k X_{-k}\right)^2=\min!,$$

i.e. $-a_1, \ldots, -a_M$ are the coefficients of the best finite-memory linear predictor Π_M of X_0 on the basis of $X_{-1}, \ldots, X_{-M}$. The minimal finite-memory prediction error for the time series $\{X_n, -\infty<n<\infty\}$ is always greater or equal to the minimal prediction error. For the autoregression $\{Y_n,\ -\infty<n<\infty\}$, the best linear predictor coincides with the best finite-memory linear predictor Π_M of the length M. We conclude

$$\min_{\Pi} R(\Pi,\mu)\leqq R(\Pi_M,\mu)=R(\Pi_M,f_M)=\min_{\Pi} R(\Pi,f_M). \tag{3.7}$$

From these considerations and Theorems 1 and 2 follows

Theorem 4. *Let $M \geqq 0$. Let $\mathfrak{D}$ be a convex set of spectral densities on which I is not identically $-\infty$. Let $\mathfrak{M}$ be the set of absolutely continuous spectral measures with density in $\mathfrak{D}$, and let $\bar{\mathfrak{M}}$ be the w^*-closure of $\mathfrak{M}$. Assume that $\bar{\mathfrak{M}}$ is w^*-compact and*

if $\mu \in \bar{\mathfrak{M}}$ and f_M is the spectral density of an autoregression of order at most M with

$$\frac{1}{2\pi} \int e^{-ik\omega} f_M(\omega)\, d\omega = \frac{1}{2\pi} \int e^{-ik\omega}\, d\mu(\omega), \qquad |k| \leqq M, \tag{3.8}$$

then $f_M \in \mathfrak{D}$.

There exists a unique most indeterministic f^0 in $\mathfrak{D}$, and it is the spectral density of an autoregression of order at most M. In particular, the minimax-robust predictor for $\mathfrak{D}$ has finite memory of length at most M.

Proof. Let f_μ denote the density of the absolutely continuous part of the measure μ. By Theorem 1′, $\min_\Pi R(\Pi, \mu)$ assumes its maximum on the w^*-compact set $\bar{\mathfrak{M}}$, i.e. there exists a spectral measure μ^0 in $\bar{\mathfrak{M}}$ with

$$\min_\Pi R(\Pi, \mu^0) \geqq \min_\Pi R(\Pi, \mu) \qquad \text{for all } \mu \in \bar{\mathfrak{M}}.$$

Let

$$r_k^0 = \frac{1}{2\pi} \int e^{-ik\omega}\, d\mu^0(\omega) \qquad k = 0, \ldots, M.$$

As there exists $f \in \mathfrak{D}$ with $I(f) > -\infty$ we have

$$\min_\Pi R(\Pi, \mu^0) \geqq \min_\Pi R(\Pi, f) > 0.$$

Therefore, μ^0 is not singular, and $r_0^0, \ldots, r_M^0$ is a positive definite sequence. By the remarks which we have made in advance of the statement of the theorem, there exists a spectral density f_M^0 of an autoregression of order at most M with

$$r_k^0 = \frac{1}{2\pi} \int e^{-ik\omega} f_M^0(\omega)\, d\omega \qquad k = 0, \ldots, M.$$

By (3.7)

$$\min_\Pi R(\Pi, f_M^0) \geqq \min_\Pi R(\Pi, \mu^0) \geqq \min_\Pi R(\Pi, f) \qquad \text{for all } f \in \mathfrak{D}.$$

By (3.8) $f_M^0 \in \mathfrak{D}$. Together with Theorem 1 a), we conclude that f_M^0 is the unique most indeterministic spectral density in $\mathfrak{D}$.

As f_M^0 is an autoregressive spectral density, $1/f_M^0$ is essentially bounded. The rest follows from Theorem 2. □

The assumptions of Theorem 4 are fulfilled for the sets $\mathfrak{D}_\mathfrak{K}$ of Theorem 3, and the finite memory of the minimax-robust predictor for these models follows. Theorem 4, however, gives only an existence result, whereas Theorem 3 gives a unique characterization of f^0 in terms of the set $\mathfrak{K}$.

Loosely speaking, Theorem 4 says that the minimax-robust predictor for X_0 depends only on the last M observations $X_{-1}, \ldots, X_{-M}$ if our knowledge of the time series $\{X_k, \ -\infty<k<\infty\}$ is restricted to information about some finite part of its time domain structure.

Appropriately formulated, analogous results are true for e.g. interpolation or prediction over more than one time lag into the future, as can be shown by considerations completely analogous to Theorem 4. However, the dominating role played by autoregressive processes in this chapter depends on the problem of predicting robustly exactly one time unit ahead. E.g., if we want to predict further into the future then autoregressions as the processes which correspond to most indeterministic spectral densities are replaced with certain mixed autoregressive-moving average processes [31].

Appendix

Some Results from Convex Optimization Theory

In this appendix we consider an optimization problem of the form

$$L(p)=\min! \qquad \text{under the constraint } p\in\mathbb{Q}. \tag{A1}$$

Here, L is a convex functional from some Banach space $\mathbb{E}$ into the extended real line $(-\infty, \infty]$. $\mathbb{Q}$ is some convex subset of $\mathbb{E}$.

Firstly, we introduce some notation. $\chi_{\mathbb{Q}}$ denotes the convex indicator of $\mathbb{Q}$, i.e.

$$\chi_{\mathbb{Q}}(p)=0 \text{ for } p\in\mathbb{Q}, \qquad =\infty \text{ for } p\notin\mathbb{Q}.$$

$\partial L(p^0)$ denotes the subdifferential of L at the point p^0, i.e. the set of all continuous linear functionals Φ on $\mathbb{E}$ with

$$\Phi(p-p^0)\leqq L(p)-L(p^0) \qquad \text{for all } p\in\mathbb{E}.$$

For $p^0\in\mathbb{Q}$, the subdifferential $\partial\chi_{\mathbb{Q}}(p^0)$ coincides with the set of support functionals of the convex set $\mathbb{Q}$ in p^0, i.e. the set of continuous linear functionals Φ on $\mathbb{E}$ with

$$\Phi(p-p^0)\leqq 0 \qquad \text{for all } p\in\mathbb{Q}.$$

L'_0 denotes the directional derivative of L in the direction p at the point p^0, i.e. for $L(p^0)<\infty$ we set

$$L'_0(p)=\infty \qquad \text{if } L(p^0+\varepsilon p)=\infty \text{ for all } \varepsilon>0$$

$$L'_0(p)=\lim_{\varepsilon\to 0+} \frac{1}{\varepsilon}\{L(p^0+\varepsilon p)-L(p^0)\} \qquad \text{else.}$$

We are interested in necessary and sufficient conditions that for given $p^0\in\mathbb{Q}$ we have $L(p)\geqq L(p^0)$ for all $p\in\mathbb{Q}$. The main theorem of convex optimization theory (see e.g. [27]) says:

Let L be finite everywhere. Let $\operatorname{int}(\mathbb{Q}) \neq \emptyset$ or let L be continuous in some point $p \in \mathbb{Q}$. Then, p^0 is a solution of (A1) iff

$$0 \in \partial L(p^0) + \partial \chi_{\mathbb{Q}}(p^0)$$

and the latter condition is equivalent to

$$L'_0(p - p^0) \geqq 0 \quad \text{for all } p \in \mathbb{Q}.$$

We have to modify this result, as we want to consider a nowhere continuous, non-finite functional L and sets $\mathbb{Q}$ with empty interior.

Lemma 1. *Let L be a convex functional from $\mathbb{E}$ into $(-\infty, \infty)$ and $\mathbb{Q} \subset \mathbb{E}$ convex.*

(i) *Every point p^0 of local minimum for L on $\mathbb{Q}$ with $L(p^0) < \infty$ is also a point of global minimum.*

(ii) *Let $L(p^0) < \infty$ and $p^0 \in \mathbb{Q}$. p^0 is a solution of* (A1) *iff*

$$L'_0(p - p^0) \geqq 0 \quad \textit{for all } p \in \mathbb{Q}.$$

Proof. The first part of the lemma is well known for everywhere finite convex functionals (e.g. [9] – Theorem 15.1), and the usual proof applies to points of finite local minimum for extended real-valued convex functionals, too.

Let p^0 be a solution of (A1). By convexity of $\mathbb{Q}$

$$p^\varepsilon = p^0 + \varepsilon(p - p^0) \in \mathbb{Q} \quad \text{for all } 0 \leqq \varepsilon \leqq 1,\ p \in \mathbb{Q}.$$

By minimality of $L(p^0)$

$$L'_0(p - p^0) = \lim_{\varepsilon \to 0+} \frac{1}{\varepsilon}\{L(p^\varepsilon) - L(p^0)\} \geqq 0 \quad \text{for all } p \in \mathbb{Q}$$

The other direction follows from the convexity of L, as

$$L'_0(p - p^0) \leqq \frac{1}{\varepsilon}\{L(p^\varepsilon) - L(p^0)\} \leqq L(p) - L(p^0) \quad \text{for all } 0 \leqq \varepsilon \leqq 1,\ p \in \mathbb{Q}. \quad \square$$

In the following, we consider the special case

$$\mathbb{E} = L^1(\lambda), \qquad L(f) = \frac{-1}{2\pi} \int \log\{f^+(\omega)\}\, d\omega$$

L is convex, as for $0 < \gamma < 1$ and $f, g \in L^1(\lambda)$:

$$L(\{1-\gamma\} f + \gamma g) \leqq \frac{-(1-\gamma)}{2\pi} \int \log\{f^+(\omega)\}\, d\omega - \frac{\gamma}{2\pi} \int \log\{g^+(\omega)\}\, d\omega$$

The right-hand side is $+\infty$ unless $f(\omega), g(\omega) > 0$ a.s.

L has rather good directional differentiability properties. In particular, if $\operatorname{essinf}_\omega f^0(\omega) > 0$ then L'_0 is nearly identical to a continuous linear functional on $L^1(\lambda)$. Therefore, it is possible to modify the proof of the main theorem of convex optimization theory to cover the problem

$$L(f)=\min! \qquad \text{under the constraint } f\in\mathbb{Q} \tag{A2}$$

without assuming $\operatorname{int}(\mathbb{Q})\neq\emptyset$.

Lemma 2. (i) *If* $L(f^0)<\infty$ *then the directional derivative of* L *in direction* f *at* f^0 *is*

$$L'_0(f)=\infty \qquad \text{if } L(f^0+\varepsilon f)=\infty \text{ for all } \varepsilon>0$$

$$L'_0(f)=\frac{-1}{2\pi}\int\{f(\omega)/f^0(\omega)\}\,d\omega \qquad \text{else.}$$

(ii) *Let* $\operatorname{essinf}_\omega f^0(\omega)>0$. *Then,*

$$\Phi_0(f)=\frac{-1}{2\pi}\int\{f(\omega)/f^0(\omega)\}\,d\omega$$

is a continuous linear functional, and

$$\partial L(f^0)=\{\Phi_0\}.$$

(iii) *Let* $\mathbb{Q}\neq\emptyset$ *be a convex set in* $L^1(\lambda)$ *with the property: for all* $g\in\mathbb{Q}$ *there exist* $f_n\in\mathbb{Q}$, $n\geqq 1$, *such that*

$$\operatorname{essinf}_\omega f_n(\omega)>-\infty,\ n\geqq 1, \text{ and } f_n\to g \qquad \text{in } L^1(\lambda).$$

Let $f^0\in\mathbb{Q}$ *and* $\operatorname{essinf}_\omega f^0(\omega)>0$. *Then,* f^0 *is a solution of the optimization problem* (A2) *iff*

$$0\in\partial L(f^0)+\partial\chi_{\mathbb{Q}}(f^0).$$

Proof. a) Let $L(f^0+\varepsilon f)<\infty$ for some $\varepsilon>0$. Then $f^0(\omega)+\varepsilon f(\omega)>0$ a.s. By concavity of the logarithm

$$\frac{1}{\varepsilon}\{\log(f^0(\omega)+\varepsilon f(\omega))-\log f^0(\omega)\}$$
$$\to f(\omega)/f^0(\omega) \qquad \text{monotonically a.s.}$$

From this follows

$$\frac{1}{\varepsilon}\{L(f^0+\varepsilon f)-L(f^0)\}=\frac{-1}{2\pi\varepsilon}\int\{\log(f^0(\omega)+\varepsilon f(\omega))-\log f^0(\omega)\}\,d\omega$$
$$\to\frac{-1}{2\pi}\int\{f(\omega)/f^0(\omega)\} \qquad \text{monotonically for } \varepsilon\to 0+.$$

b) We show for a continuous linear functional Φ:

$$\Phi\in\partial L(f^0) \qquad \text{iff } \Phi(f-f^0)\leqq L'_0(f-f^0) \qquad \text{for all } f\in L^1(\lambda). \tag{A3}$$

If $\Phi\in\partial L(f^0)$, then by definition

$$\Phi(f)-\Phi(f^0)\leqq L(f)-L(f^0) \qquad \text{for all } f\in L^1(\lambda).$$

Therefore, for $\varepsilon\in(0,1)$ and $f\in L^1(\lambda)$

$$\begin{aligned}\Phi(f-f^0)&=\frac{1}{\varepsilon}\{\Phi(f^0+\varepsilon(f-f^0))-\Phi(f^0)\}\\&\leqq\frac{1}{\varepsilon}\{L(f^0+\varepsilon(f-f^0))-L(f^0)\}.\end{aligned}$$

The right-hand side decreases for $\varepsilon\to 0+$ monotonically to $L'_0(f-f^0)$.

Due to the same reason

$$L'_0(f-f^0)\leqq L(f)-L(f^0)\qquad\text{for all } f\in L^1(\lambda).$$

From this follows the other direction of (A3).

c) Let $\operatorname{essinf}_\omega f^0(\omega)>0$. Then Φ_0 is a continuous linear functional, and it coincides with L'_0 on the set $\{f\,|\,L'_0(f)<\infty\}$. By (A3), $\Phi_0\in\partial L(f^0)$.

Let $\Phi\in\partial L(f^0)$, given by

$$\Phi(f)=\frac{1}{2\pi}\int\varphi(\omega)\,f(\omega)\,d\omega,\qquad\varphi\in L^\infty(\lambda).$$

By (A3), we have in particular for all $f\in L^\infty(\lambda)$

$$\begin{aligned}0&\leqq L'_0(f-f^0)-\Phi(f-f^0)\\&=\frac{-1}{2\pi}\int\{\varphi(\omega)+1/f^0(\omega)\}\{f(\omega)-f^0(\omega)\}\,d\omega.\end{aligned}$$

We conclude $\varphi(\omega)=-1/f^0(\omega)$ a.s., i.e. $\Phi=\Phi_0$.

d) As usual in convex optimization theory, we set $\bar{L}(f)=L(f)+\chi_{\mathbb{Q}}(f)$, such that (A2) is equivalent to the unconstrained optimization problem:

$$\bar{L}(f)=\min!$$

Then, by definition of the subdifferential:

$$\begin{aligned}&f^0 \text{ solves (A2) iff}\quad 0\in\partial\bar{L}(f^0)\\&\partial\bar{L}(f^0)\supseteq\partial L(f^0)+\partial\chi_{\mathbb{Q}}(f^0).\end{aligned}$$

The assumptions of the Moreau-Rockafellar theorem [14–0.3.3], which provides the reverse inclusion, are not fulfilled in our situation. To complete the proof of (iii) we have to show

$$\partial\bar{L}(f^0)\subseteq\partial L(f^0)+\partial\chi_{\mathbb{Q}}(f^0)$$

directly.

e) Let $\operatorname{essinf}_\omega f^0(\omega)>0$ and $\mathbb{Q}$ fulfil the assumptions of (iii). By convexity of $\mathbb{Q}$

$$\bar{L}'_0(f-f^0)=L'_0(f-f^0)\qquad\text{for all } f\in\mathbb{Q}\tag{A4}$$

Analogously to the proof of (A3) we can show for continuous, linear functionals Φ:

$$\Phi\in\partial\bar{L}(f^0)\quad\text{iff } \Phi(f-f^0)\leqq\bar{L}'_0(f-f^0) \text{ for all } f\in\mathbb{Q}.\tag{A5}$$

Let $\Phi \in \partial \bar{L}(f^0)$ and $\Psi = \Phi - \Phi_0$. Let $f \in \mathbb{Q}$ and $\operatorname{essinf}_\omega f(\omega) > -\infty$. Then $\Phi_0(f - f^0) = L'_0(f - f^0)$, and by (A4) and (A5):

$$\Psi(f - f^0) = \Phi(f - f^0) - L'_0(f - f^0) \leqq 0. \tag{A6}$$

For $g \in \mathbb{Q}$ we can choose a sequence $f_n \in \mathbb{Q}$ such that $\operatorname{essinf}_\omega f_n(\omega) > -\infty$, $n \geqq 1$, and $f_n \to g$. By (A6) and continuity of Ψ:

$$\Psi(g - f^0) = \lim_{n \to \infty} \Psi(f_n - f^0) \leqq 0 \quad \text{for all } g \in \mathbb{Q}.$$

Thus, $\Phi = \Phi_0 + \Psi$ where $\Phi_0 \in \partial L(f^0)$ by c) and $\Psi \in \partial \chi_{\mathbb{Q}}(f^0)$. □

In the following, we derive explicit representations of the subdifferentials $\partial \chi_{\mathbb{Q}}(f^0)$ for some special convex sets $\mathbb{Q}$.

Lemma 3. *Let Ψ_j, $j = 1, \ldots, n$, be continuous linear functionals on $L^1(\lambda)$, and $c_1, \ldots, c_n \in \mathbb{R}$. Let*

$$\mathbb{Q} = \{f \in L^1(\lambda) \mid \Psi_j(f) = c_j, j = 1, \ldots, n\},$$

For arbitrary $f^0 \in \mathbb{Q}$

$$\partial \chi_{\mathbb{Q}}(f^0) = \left\{ \Phi \,\middle|\, \Phi = \sum_{j=1}^n \lambda_j \Psi_j;\ \lambda_1, \ldots, \lambda_n \in \mathbb{R} \right\}.$$

Proof. Let $\mathbb{Q}_{\pm j} = \{f \mid \pm \Psi_j(f) \leqq \pm c_j\}$, $j = 1, \ldots, n$. $\mathbb{Q}$ is the intersection of $\mathbb{Q}_1, \ldots, \mathbb{Q}_n$, $\mathbb{Q}_{-1}, \ldots, \mathbb{Q}_{-n}$ all of which are convex sets with nonempty interior. By Propositions 4.3.1 and 4.3.2 of Ioffe and Tihomirov [14]

$$\begin{aligned} \partial \chi_{\mathbb{Q}}(f^0) &= \sum_{j=1}^n \{\partial \chi_{\mathbb{Q}_j}(f^0) + \partial \chi_{\mathbb{Q}_{-j}}(f^0)\} \\ &= \left\{ \sum_{j=1}^n \lambda_j \Psi_j;\ \lambda_1, \ldots, \lambda_n \in \mathbb{R} \right\}. \end{aligned}$$

Here, we have used that the Gateaux derivative of a continuous linear functional Ψ coincides with Ψ and, therefore, the support functionals of the level set $\{f \mid \Psi(f) \leqq c\}$ are of the form $\beta \Psi$, $\beta \geqq 0$, by Proposition 4.3.2 of Ioffe and Tihomorov [14]. □

Lemma 4. *Let $a_1, \ldots, a_n \in L^\infty(\lambda)$ and $\mathbf{A}$ be the continuous linear operator from $L^1(\lambda)$ into $\mathbb{R}^n$ given by*

$$(\mathbf{A}f)_j = \int a_j(\omega) f(\omega) \, d\omega \qquad j = 1, \ldots, n,\ f \in L^1(\lambda)$$

Let $\mathfrak{K}$ be a convex set in $\mathbb{R}^n$ with nonempty interior, and

$$\mathbb{Q} = \{f \in L^1(\lambda) \mid \mathbf{A}f \in \mathfrak{K}\}.$$

For arbitrary $f^0 \in \mathbb{Q}$

$$\partial \chi_{\mathbb{Q}}(f^0) = \left\{ \Phi \mid \Phi(f) = \sum_{j=1}^n \lambda_j \int a_j(\omega) f(\omega) \, d\omega;\ \lambda_1, \ldots, \lambda_n \in \mathbb{R} \right.$$
$$\left. \text{such that } \sum_{j=1}^n \lambda_j \{c_j - (\mathbf{A}f^0)_j\} \leqq 0 \text{ for all } \mathbf{c} \in \mathfrak{K} \right\}$$

Proof. Let $\mathbf{A}^*$ be the dual operator of $\mathbf{A}$. By Theorem 4.2.2 of Ioffe and Tihomirov [14]

$$\partial\chi_{\mathbb{Q}}(f^0)=\mathbf{A}^*\,\partial\chi_{\mathfrak{K}}(\mathbf{A}f^0).$$

Let $\Lambda(\mathbf{c})=\sum_{j=1}^{n}\lambda_j c_j$, $\mathbf{c}\in\mathbb{R}^n$. The image of Λ under $\mathbf{A}^*$ is given by

$$(\mathbf{A}^*\Lambda)(f)=\sum_{j=1}^{n}\lambda_j\int a_j(\omega)\,f(\omega)\,d\omega,\qquad f\in L^1(\lambda).$$

By definition, $\partial\chi_{\mathfrak{K}}(\mathbf{A}f^0)$ consists of the continuous linear functionals Λ on $\mathbb{R}^n$ with

$$\Lambda(\mathbf{c}-\mathbf{A}f^0)\leqq 0\qquad\text{for all }\mathbf{c}\in\mathfrak{K}.$$

This finishes the proof. □

Lemma 5. *Let g be a nonnegative function in $L^\infty(\lambda)$, $\varepsilon>0$ and*

$$\mathbb{Q}=\left\{f\in L^1(\lambda)\,\middle|\,\frac{1}{2\pi}\int\{f(\omega)-g(\omega)\}^2\,d\omega\leqq\varepsilon\right\}$$

Let $f^0\in\mathbb{Q}\cap L^\infty(\lambda)$ and

$$\Psi_0(f)=\frac{1}{2\pi}\int\{f^0(\omega)-g(\omega)\}\,f(\omega)\,d\omega.$$

Then

$$\partial\chi_{\mathbb{Q}}(f^0)\begin{cases}=\{0\}\\ =\{\gamma\Psi_0,\gamma\geqq 0\}\end{cases}\qquad\text{if }\frac{1}{2\pi}\int\{f^0(\omega)-g(\omega)\}^2\,d\omega\begin{cases}<\varepsilon\\ =\varepsilon\end{cases}$$

Proof. a) If $\frac{1}{2\pi}\int\{f^0(\omega)-g(\omega)\}^2\,d\omega<\varepsilon$ then for every $h\in L^\infty(\lambda)$ there exists $\delta>0$ such that $f^0\pm\delta h\in\mathbb{Q}$. From this we conclude

$$\Phi(h)=0\qquad\text{for all }h\in L^\infty(\lambda)\text{ and all }\Phi\in\partial\chi_{\mathbb{Q}}(f^0).$$

As $L^\infty(\lambda)$ is dense in $L^1(\lambda)$ we get $\partial\chi_{\mathbb{Q}}(f^0)=\{0\}$.

b) In the following we assume $\frac{1}{2\pi}\int\{f^0(\omega)-g(\omega)\}^2\,d\omega=\varepsilon$. Firstly, we consider $\mathbb{Q}$ as subset of $L^2(\lambda)$:

$$\mathbb{Q}=\{f\in L^2(\lambda)\,|\,G(f)\leqq\varepsilon\},\qquad G(f)=\frac{1}{2\pi}\int\{f(\omega)-g(\omega)\}^2\,d\omega.$$

G is a finite, convex functional on $L^2(\lambda)$. It is Gateaux-differentiable in f^0 with derivative

$$\begin{aligned}G_0'(f)&=\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\{G(f^0+\varepsilon f)-G(f^0)\}\\&=\frac{1}{\pi}\int\{f^0(\omega)-g(\omega)\}\,f(\omega)\,d\omega\\&=2\Psi_0(f).\end{aligned}$$

Let L_i^* denote the dual space of $L^i(\lambda)$, $i=1,2$. By proposition 4.3.2 and example 4.2.1.1 of Ioffe and Tihomirov [14] we have

$$\{\Phi \in L_2^* \mid \Phi(f) \leqq \Phi(f^0) \text{ for all } f \in \mathbb{Q}\} = \{\gamma \Psi_0, \gamma \geqq 0\}. \tag{A7}$$

c) As $L_1^* \subseteq L_2^*$

$$\begin{aligned}\partial \chi_{\mathbb{Q}}(f^0) &= \{\Phi \in L_1^* \mid \Phi(f) \leqq \Phi(f^0) \text{ for all } f \in \mathbb{Q}\}\\ &\subseteq \{\gamma \Psi_0, \gamma \geqq 0\}.\end{aligned}$$

As f^0, $g \in L^\infty$ we have $\Psi_0 \in L_1^*$. Together with (A7) this implies the other inclusion. □

Acknowledgement. I wish to thank Prof. H. Dinges for encouraging discussions and Prof. H. Rieder for bringing Hosoya's work to my attention. I would also like to thank and unknown referee for his helpful detailed comments. This work has been supported by the Deutsche Forschungsgemeinschaft in connection with Sonderforschungsbereich 123 "Stochastische Mathematische Modelle".

Literatur

1. Bartlett, M.S.: An introduction to stochastic processes. Cambridge: University Press 1956
2. Blackman, R.B., Tukey, J.W.: The measurement of power spectra from the point of view of communications engineering. New York: Dover 1959
3. van den Bos, A.: Alternative interpretation of maximum entropy spectral analysis. IEEE Trans. Information Theory **17**, 493–494 (1971)
4. Burg, J.P.: Maximum entropy spectral analysis. In: Modern Spectrum Analysis. D.G. Childers (ed.). New York: Wiley 1978
5. Burg, J.P.: A new analysis technique for time series data. In: Modern Spectrum Analysis. D.G. Childers (ed.). New York: Wiley 1978
6. Childers, D.G.: Modern Spectrum Analysis. IEEE Press Selected Reprint Series. New York: Wiley 1978
7. Edward, J.A., Fitelson, M.M.: Notes on maximum entropy processing. IEEE Trans. Information Theory **19**, 232–234 (1973)
8. Franke, J.: A general approach to the robust prediction of discrete time series. Preprint No. 74. Heidelberg: Sonderforschungsbereich 123 "Stochastische Mathematische Modelle", 1980
9. Girsanov, I.V.: Lectures on mathematical theory of extremum problems. Berlin-Heidelberg-New York: Springer 1972
10. Hannan, E.J.: Multiple time series. New York-London-Sidney-Toronto: Wiley 1970
11. Hoffman, K.: Banach spaces of analytic functions. Englewood Cliffs, N.J.: Prentice Hall 1962
12. Hosoya, Y.: Robust linear extrapolations of second-order stationary processes. Ann. Probability **6**, 574–584 (1978)
13. Huber, P.: Robust estimation of a location parameter. Ann. Math. Statist. 35, 73–101 (1964)
14. Ioffe, A.D., Tihomirov, V.M.: Theory of extremal problems. Amsterdam-New York-Oxford: North Holland 1979
15. Lacoss, R.T.: Data adaptive spectral analysis methods. Geophysics **36**, 61–675 (1971)
16. Makhoul, J.: Stable and efficient lattice methods for linear prediction. IEEE Trans. Acoust. Speech Signal Process. **25**, 423–428 (1977)
17. Martin, R.D.: Robust estimation for time series autoregressions. In Robustness in Statistics. R.L. Launer, G.N. Wilkinson (eds.). New York-San Francisco-London: Academic Press 1979
18. Masani, P.R.: The prediction theory of multivariate stochastic processes III. Acta Math. **104**, 141–162 (1960)
19. Morf, M., Dickinson, B., Kailath, T., Vieira, A.: Efficient solution of covariance equations for linear prediction. IEEE Trans. Acoust. Speech Signal Process. **25**, 429–433 (1977)

20. Morf, M., Lee, D.T., Nickolls, J.R., Vieira, A.: A classification of algorithms of ARMA models and ladder realizations. IEEE Internat. Conf. Record. Acoust. Speech Signal Process. 13–19 (1977)
21. Newman, W.I.: Extension to the maximum entropy method. IEEE Trans. Information Theory **23**, 89–93 (1977)
22. Newman, W.I.: Extension to the maximum entropy method II. IEEE Trans. Information Theory **25**, 705–708 (1979)
23. Pagano, M.: When is an autoregressive scheme stationary? Comm. Statist. **1**, 533–544 (1973)
24. Priestley, M.B.: Spectral Analysis and Time Series. London: Academic Press 1981
25. Parzen, E.: Some recent advances in time series modeling. IEEE Trans. Automatic Control **19**, 723–730 (1974)
26. Shannon, C.E., Weaver, W.: The mathematical theory of communication. Urbana: The University of Illinois Press 1949
27. Zeidler, E.: Vorlesungen über nichtlineare Funktionalanalysis III: Variationsmethoden und Optimierung. Leipzig: Teubner 1977
28. Zygmund, A.: Trigonometric series vol. I. Cambridge: University Press 1959
29. Weidmann, J.: Lineare Operatoren im Hilbertraum. Stuttgart: Teubner 1976
30. Girsanov, I.V.: Lectures on Mathematical Theory of Extremum Problems. Lecture Notes in Economics and Mathematical Systems No. 67. Berlin-Heidelberg-New York: Springer 1972
31. Franke, J.: On the Robust Prediction and Interpolation of a Discrete Time Series in the Presence of Correlated Noise. J. Time Series Analysis 5 (To appear)
32. Franke, J., Poor, V.H.: Minimax-robust Filtering and Finite-length Robust Predictors. In: Robust and Nonlinear Time Series Analysis. J. Franke, W. Härdle, R.D. Martin (eds.). Lecture Notes in Statistics 26. Berlin-Heidelberg-New York: Springer (To appear)
33. Vastola, K.S., Poor, V.H.: Robust Wiener-Kolmogorov Theory. IEEE Trans. Information Theory 30 (To appear)

Received April 14, 1981