BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 16, Number 1, January 1987

MAXIMUM ENTROPY AND THE MOMENT PROBLEM

H. J. LANDAU

Introduction. The trigonometric moment problem stands at the source of several major streams in analysis. From it flow developments in function theory, in spectral representation of operators, in probability, in approximation, and in the study of inverse problems. Here we connect it also with a group of questions centering on entropy and prediction. In turn, this will suggest a simple approach, by way of orthogonal decomposition, to the moment problem itself.

In statistical estimation, one often wants to guess an unknown probability distribution, given certain observations based on it. There are generally infinitely many distributions consistent with the data, and the question of which of these to select is an important one. The notion of entropy has been proposed here as the basis of a principle of salience which has received considerable attention. We will show that, in the context of spectral analysis, this idea is linked to a certain question of prediction by the trigonometric moment problem, and that all three strongly illuminate one another. The phenomena we describe are known, but our object is to unify them conceptually and to reduce the analytic intricacy of the arguments. To this end, we give a completely elementary discussion, virtually free of calculation, which shows that all the facts, including those concerning the moment problem, can be understood as direct consequences of orthogonal decomposition in a finite-dimensional space. We then describe how, in its continuous version, this leads to a view of second-order Sturm-Liouville differential equations, and conclude with some questions concerning the connection between combinatorial ideas and orthogonality in this problem.

Entropy and statistical inference. Suppose that we are interested in the distribution of some quantity, but know only the values of certain averages defined by that distribution, which are insufficient to specify it uniquely. For example, we might have tossed a six-sided die fifty times, wishing to find how often each face appeared, but were able to observe only the average value of these faces. What should we select as an appropriate distribution, on the strength of the available information? Various criteria have been proposed to

Received by the editors May 21, 1986.

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 42A70; Secondary 42A05, 62M15, 94A17, 60G25.