Title
LINEAR COMPUTATION OF FUNCTION APPROXIMATION WITH APPLICATION TO FITTING AND EDITING TWO-DIMENSIONAL DATA

Author
Young, Jonathan D.

Publication Date
1963-03-26

UCRL-10734

University of California

Ernest O. Lawrence Radiation Laboratory

LINEAR COMPUTATION OF FUNCTION APPROXIMATION WITH APPLICATION TO FITTING AND EDITING TWO-DIMENSIONAL DATA

Berkeley, California

UCRL-10734
UC-32 Math. and Comp.
TID-4500 (19th Ed.)

UNIVERSITY OF CALIFORNIA

Lawrence Radiation Laboratory
Berkeley, California

Contract No. W-7405-eng-48

LINEAR COMPUTATION OF FUNCTION APPROXIMATION WITH APPLICATION TO FITTING AND EDITING TWO-DIMENSIONAL DATA

Jonathan D. Young

March 26, 1963

Printed in USA. Price 50 cents. Available from the
Office of Technical Services
U. S. Department of Commerce
Washington 25, D.C.

LINEAR COMPUTATION OF FUNCTION APPROXIMATION WITH APPLICATION TO FITTING AND EDITING TWO-DIMENSIONAL DATA

Jonathan D. Young

Lawrence Radiation Laboratory
University of California
Berkeley, California

March 26, 1963

ABSTRACT

Linear computation of level functional approximations for a discrete function is applied to fitting and editing two-dimensional data with particular emphasis on the case in which the approximating function involves two fewer parameters than the number of points.

LINEAR COMPUTATION OF FUNCTION APPROXIMATION WITH APPLICATION TO FITTING AND EDITING TWO-DIMENSIONAL DATA

Jonathan D. Young

Lawrence Radiation Laboratory
University of California
Berkeley, California

March 26, 1963

INTRODUCTION

The approximation of a discrete function, (x_i, y_i) for $i = 1, \cdots n$, by a continuous function $g(x)$ is a common problem in numerical analysis. One familiar example is polynomial approximation. The advantages of using the continuous function rather than the discrete data are that (a) it may be defined in fewer than n parameters, and (b) processes such as interpolation, numerical integration, and differentiation for the discrete function may be approximated analytically on the continuous function. Most existing methods are restrictive as to the choice of approximating function — for example, to polynomials or to a finite sum of trigonometric terms. In what follows, we discuss a method that offers considerable freedom in the choice of the approximating function. An application of the method to data editing is also discussed.

LINEAR APPROXIMATION

From the set of data (x_i, y_i) for $i = 1, \cdots n$, we desire to approximate the function of $y(x)$ whose value at x_i is y_i by a sum

$$s = \sum_{j=1}^{m} a_j f_j, \quad m \leq n$$

where the f_j are functions of x whose numerical values f_{ij} at the points x_i are readily obtainable. The set $\{f_j\}$ must be linearly independent for dimension m, that is, for every subset $\{x_k\}$ of m points from $\{x_i\}$, we have

$$\sum_{j=1}^{m} b_j f_{kj} = 0$$

for all k if and only if all $b_j = 0$. The theoretical treatment of this problem is given in reference 1. The rigorous exposition given there is essential to establish the theory, the simpler development presented here is sufficient to outline computational methods for fitting, editing, and smoothing data.

THE UNIQUE FIT

For $m = n$, the coefficients a_j are uniquely determined as the components of a vector a which satisfies

$$Fa = y ,$$

where F is an $m \times m$ matrix whose elements are the f_{ij}, and y is a vector with components y_i.

THE LEVEL FIT

For $m = n - 1$, we define an additional function f_n with the property

$$f_n(x_i) = (-1)^i .$$

Now if the set $f_1, \cdots f_m, f_n$ is independent, we find the unique fit

$$y = s = \sum_{j=1}^{n} a_j f_j$$

as above. Let

$$g = \sum_{j=1}^{m} a_j f_j ;$$

then we have

$$y_i = g_i + (-1)^i a_n ,$$

so that the function g has the property that the difference $y_i - g_i$ alternates in sign over i, but has constant magnitude $|a_n|$ for all i. The function g is called the level fit of y with respect to $f_1 \cdots f_m$.

OPTIONAL LEVEL FIT

For $m = n - 2$, let $p = n - 1$, so that $m = p - 1$. Any reference subset $\{x_k, y_k\}$; $k = 1, \cdots p$ from the original set $\{x_i, y_i\}$; $i = 1, \cdots n$ may be selected and a level fit obtained as above. There are n such subsets, each characterized by the omission of one point (x_ℓ, y_ℓ) in its calculation. The integer ℓ (as a superscript) can be used to index these level fits as follows:

g^ℓ is the level function obtained when (x_ℓ, y_ℓ) is omitted

$|a^\ell| = |g_k^\ell - y_k|$ is the level difference

$|d^\ell| = |g_\ell^\ell - y_\ell^\ell|$ is the external difference.

As shown in reference 1, a minmax relation exists whereby the level difference has its maximum value at that ℓ for which the external difference is a minimum, and vice-versa. The choice of fit may be predicated on the situation. The most useful cases are described below.

A. Optimal Level Fit

In this case, our confidence is in the data, $\{x_i, y_i\}$, and we simply wish to choose ℓ so that max $|g_i^\ell - y_i|$ is a minimum. For each ℓ;

$\ell = 1, \cdots n$ we need only compute $|a^{\ell}|$ and $|d^{\ell}|$ and select that ℓ for which max $\{|a^{\ell}|, |d^{\ell}|\}$ has its least value.

B. Edited Level Fit

In this case, our confidence is in the fitting rather than the data. We suspect that in the set $\{x_i, y_i\}$ there is a bad point whose inclusion in the optimal fit would distort the approximation. The desired fit here is the one for which this point is omitted. Contrary to (a) above, the ℓ selected is that for which $|d^{\ell}|$ is maximum.

MULTIPLE OPTIONAL FIT

For $m \leq n - 2$ we have a generalization of the above. We now have $\binom{n}{m+1}$ choices for reference subsets $\{x_k, y_k\}$, and each such reference involves omission of $n - m + 1$ points (x_{ℓ}, y_{ℓ}) from the calculation of the level fit. These level fits are now indexed on L, where each L corresponds to the omission of some particular set S_L of $n - m + 1$ points. We define

g^L as the level function where the set S_L is omitted,

$|a^L| = |g_k^L - y_k|$ as the level difference,

and $|d^L| = \max\limits_{\ell \in s_{\ell}} |d^{\ell}|$ as the maximum external difference.

The optimal and edited fits may be obtained now by selecting that L for which $|a^L|$ and $|d^L|$ meet the requirements specified earlier.

* * *

This work was done under the auspices of the U. S. Atomic Commission.

REFERENCE

1. E. L. Stiefel, "Numerical Methods of Tchebycheff Approximation," in On Numerical Approximation, R. E. Langer, Editor, (University of Wisconsin Press, Madison, Wisc., 1959).