SLAC PUB-2504
STAN-CS-80-799
May 1980
(M)

MULTIDIMENSIONAL ADDITIVE SPLINE APPROXIMATION*

Jerome H. Friedman
Stanford Linear Accelerator Center
Stanford, California 94305

Eric Grosse
Computer Science Department
Stanford University
Stanford, California 94305

Werner Stuetzle
Stanford Linear Accelerator Center
and
Department of Statistics, Stanford University
Stanford, California 94305

ABSTRACT

We describe an adaptive procedure that approximates a function of many variables by a sum of (univariate) spline functions s_m of selected linear combinations $a_m \cdot x$ of the coordinates

$$\varphi(x) = \sum_{1 \le m \le M} s_m(a_m \cdot x)$$

The procedure is nonlinear in that not only the spline coefficients but also the linear combinations are optimized for the particular problem. The sample need not lie on a regular grid, and the approximation is affine invariant, smooth, and lends itself to graphical interpretation. Function values, derivatives, and integrals are cheap to evaluate.

*Work partially supported by the Department of Energy under contract DE-AC03-76SF00515, and by the National Science Foundation under grant MCS 78-17697.
(Submitted to the SIAM Journal on Scientific and Statistical Computing)

1. Introduction

Multidimensional surface approximation is recognized as an important problem for which several methodologies have been developed. The aim is to construct an approximation $\phi(\mathbf{x})$ to a p-dimensional surface $y = f(\mathbf{x})$ on the basis of (possibly noisy) observations $\{(y_i, \mathbf{x}_i)\}_{1 \le i \le n}$. Most existing methods, such as tensor product splines, kernels, and thin plate splines (for a survey, see Schumaker [1976]), are linear in that

$$\phi(\mathbf{x}) = \sum_{1 \le i \le n} w_i y_i,$$

where the weights $\{w_i\}$ depend only on $\mathbf{x}$ and $\{\mathbf{x}_i\}_{1 \le i \le n}$, but not on $\{y_i\}_{1 \le i \le n}$. These methods have the advantage that they are straightforward to compute and their theory is tractable. In practice, however, they are limited because they cannot take advantage of special properties of the surface. Due to the inherent sparsity of high-dimensional sampling, procedures successful in high dimensions must be adaptive and thus nonlinear.

In this paper we describe an adaptive procedure that approximates $f(\mathbf{x})$ by a sum of (univariate) spline functions s_m of selected linear combinations $\mathbf{a}_m \cdot \mathbf{x}$ of the coordinates

$$\phi(\mathbf{x}) = \sum_{1 \le m \le M} s_m(\mathbf{a}_m \cdot \mathbf{x}). \tag{1}$$

The procedure is nonlinear in that not only the spline coefficients but also the linear combinations are optimized for the particular problem.

2. The algorithm

The spline function s_m along $\mathbf{a}_m \cdot \mathbf{x}$ is represented as a sum of j_m B-splines [de Boor, 1979] of order q

$$s_m(\mathbf{a}_m \cdot \mathbf{x}) = \sum_{1 \le j \le j_m} \beta_{mj} B_{mj}(\mathbf{a}_m \cdot \mathbf{x}). \tag{2}$$

The approximation $\phi(\mathbf{x})$ (given by equations (1) and (2)) is specified by the directions $\{\mathbf{a}_m\}_{1\le m\le M}$, the knot sequences along $\mathbf{a}_m \cdot \mathbf{x}$ for $1 \le m \le M$, and the B-spline coefficients $\{\beta_{mj}\}_{1\le m\le M, 1\le j\le j_m}$. For particular $\{\mathbf{a}_m\}$, the knots are placed heuristically and then the $\{\beta_{mj}\}$ are determined by (linear) least squares. The residual sum of squares from this fit is taken to be the inverse figure of merit for $\{\mathbf{a}_m\}_{1\le m\le M}$.

Following Friedman, Jacobson, and Stuetzle [1980], the approximation is constructed in a stepwise manner: given $\{\mathbf{a}_m\}_{1\le m\le M-1}$, find $\mathbf{a}_M$ to optimize the figure of merit of $\{\mathbf{a}_m\}_{1\le m\le M}$. Terminate when the figure of merit is not significantly improved.

3. Implementation

The most difficult part of the algorithm is finding each direction $\mathbf{a}_m$. We perform a numerical search using a Rosenbrock method [Rosenbrock, 1966] modified for the unit sphere, starting at the best coordinate direction. On any given search, there is no guarantee that the global optimum will be found. If the local optimum is insignificant, the search is restarted at random directions. This guards against premature termination. If the local optimum is significant but not identical to the global optimum, no great harm is done because a new search is performed in the next iteration on an object function for which the previous optima have been deflated. Each iteration of the optimizer requires 3.5 p function evaluations, on the average, where p is the dimension of $\mathbf{x}$. Two iterations are nearly always sufficient.

For high dimensionality, the computation is dominated by the evaluations of the object function. Since it is not crucial to find the precise optimum, considerable savings can be achieved by substituting a similar, but much less expensive figure of merit during the search for a new direction. For this figure of merit not only the previously found directions but also the corresponding spline coefficients are held fixed. The new direction can thus be found by considering the residuals from the model of the previous iteration. For a given direction, the

residuals are modelled by a smooth based on local linear fits [Cleveland, 1979], [Friedman, Jacobson, and Stuetzle, 1980]. The characteristic bandwidth (that is, the average fraction of observations used in each local fit) is taken to be inversely proportional to the number of knots. The residual sum of squares from the smooth is the figure of merit used for the smooth. Solving the least squares problem for the original figure of merit requires

$$O\left(n\left(\sum_{1\le m\le M} j_m\right)^2\right)$$

operations, while the new figure of merit can be evaluated in roughly n operations using updating formulas for the moving average. The least squares problem has to be solved only once for each iteration to determine the new model after $\mathbf{a}_m$ has been found.

To solve the least squares problem, we form the normal equations and use a pseudo-inverse, since the design matrix might not have full rank. The singularity which arises form the inclusion of a constant term for each direction is remedied by simply dropping one column per direction from the design matrix. Higher order singularities caused, for example, by the linear terms for three co-planar directions, are not explicitly taken care of, but are handled by the pseudo-inverse.

Our knot placement procedure is motivated by the sequential nature of the algorithm. At each iteration, the knot positions are required for the least squares fit, after the new direction has been found. Our model at this point is the spline fit of the previous iteration, plus the moving average smooth along the newly found direction. The knot placement is based on the residuals $\{r_i\}$ from this model. Multidimensional structure in these residuals due to incompleteness of the model manifests itself as high local variability in the scatterplots of r_i against $\mathbf{a}_m \cdot \mathbf{x}_i$. In order to preserve the ability of fitting this structure in further iterations, it is important to avoid accounting for it by spurious fits along existing directions. For this reason we place fewer knots in regions of higher local variability. Since the residuals change, the knots are replaced along all directions at each iteration.

The knots along a direction $\mathbf{a}_m$ are placed as follows: the smooth described above is applied to $\{(r_i, \mathbf{a}_m \cdot \mathbf{x}_i)\}_{1 \le i \le n}$ and the local variability v_i at each point is taken to be the average squared residual from its local linear fit. The Winsorized local variabilities are defined by

$$w_i = \begin{cases} 2\bar{v} & \text{if } v_i > 2\bar{v} \\ \frac{1}{2}\bar{v} & \text{if } v_i < \frac{1}{2}\bar{v} \\ v_i & \text{otherwise} \end{cases}$$

(where $\bar{v} = \frac{1}{n}\sum_{1 \le i \le n} v_i$), and then are scaled so that $\sum_{1 \le i \le n} \frac{1}{w_i} = 1$. The knots $\{t_l\}$ are placed to divide the line into intervals with equal content of $\frac{1}{w_i}$:

$$\text{for each } l, \qquad \frac{1}{j_m - q + 1} = \sum_{\mathbf{a}_m \cdot \mathbf{x}_i \in [t_l, t_{l+1}]} \frac{1}{w_i}.$$

4. Examples

In this section we present and discuss the results of applying the Multidimensional Spline Approximation method (MASA) to four examples. (A FORTRAN program implementing MASA is available from the authors.) The first three examples were suggested elsewhere for testing surface approximation procedures. The function in the fourth example was studied in connection with a problem in mathematical genetics.

The first example is taken from Friedman [1979]. In this example uniformly distributed random points $\{x_i \mid 1 \le i \le 200\}$ were generated in the six-dimensional hypercube $[0,1]^6$. Associated with each point x_i was a surface value

$$y_i = 10\sin(\pi x_i(1)x_i(2)) + 20[x_i(3) - 0.5]^2 + 10x_i(4) + 5x_i(5) + 0x_i(6) + \epsilon_i,$$

where the $\{\epsilon_i\}$ were independent identically distributed standard normal. The inverse figures of merit for the approximation with $M = 1, \ldots, 4$ terms were $6.71, 4.29, 1.87, 0.97$. In three restarts, the figure of merit did not decrease below

0.86, so $M = 4$ was chosen. The four linear combinations and the corresponding spline functions are shown in figures 1.1–1.4. (For completeness, the program parameters are also listed; see the program comments for a detailed explanation.) The spline along the first linear combination (figure 1.1) is seen to model the linear part of the surface. The second term in the approximation (figure 1.2) models the additive quadratic dependence on $x(3)$. The final two terms (figures 1.3, 1.4) model the interaction between $x(1)$ and $x(2)$. The L_2 norm of the error $\|f - \phi\|_2$ was 0.57.

Although the full advantages of MASA compared to other procedures are realized in higher dimensional or noisy settings, we applied it to two bivariate examples used by Franke [1979] to compare a number of interpolatory surface approximation schemes. For both examples 100 uniformly distributed random points in the unit square $[0,1]^2$ were generated. The function in Franke's first example is

$$\begin{aligned} f(x,y) = \; & 0.75\exp[-\frac{(9x-2)^2+(9y-2)^2}{4}] \\ & + 0.75\exp[-\frac{(9x+1)^2}{49} - \frac{9y+1}{10}] \\ & + 0.5\exp[-\frac{(9x-7)^2+(9y-3)^2}{4}] \\ & + 0.2\exp[-(9x-4)^2-(9y-7)^2]. \end{aligned}$$

Considerations similar to those in the previous example led to an approximation with three terms. The linear combinations and corresponding spline functions are shown in figures 2.1–2.3.

The function in Franke's second example is

$$f(x,y) = \frac{1}{9}[\tanh(9y-9x)+1].$$

For this case the approximation used only one term, shown in figure 3.1.

Since different random points were used in Franke's and our tests, precise comparisons are not possible. On the first example, MASA gave roughly an order

of magnitude larger errors than the best methods in Franke's trials (global basis function methods) while on the second example, MASA gave an order of magnitude smaller errors than the best methods. These results are not surprising since the peak-shaped basis functions of the global basis methods are especially suited for representing the peaks of the first example, whereas the ridge-shaped basis functions of MASA are especially suited to the second example. Unfortunately, peak-shaped basis functions are not appropriate for moderate or higher dimensionality. The difficulty is that in order to achieve a smooth fit, the width of the basis peaks needs to be comparable to the distance between data points. For n uniformly distributed random points in a p-dimensional hypercube $[0,1]^p$, the typical nearest neighbor distance is $(\frac{1}{n})^{\frac{1}{p}}$. In particular for $n = 1000$ and $p = 10$, this distance is 0.5, and for $p = 20$ is 0.7. Thus variation of the surface over distances small compared to such large interpoint distances cannot be well approximated with these global basis functions methods.

Our final example is a 19-dimensional function encountered by Carmelli and Cavalli [1979]. An important question is the structure of this function near its minimum. We sampled the function at 200 points uniformly distributed in a small hypercube centered at the minimum found by numerical optimization and applied MASA. The inverse figure of merit for the best constant fit was 13.3. The inverse figure of merit for $M = 1$ was 0.78. In 30 restarts, the figure of merit did not decrease below 0.42. Figure 4.1 gives the linear combination and corresponding spline function. This picture shows considerable structure that was not revealed in the original study.

5. Discussion

MASA can be expected to work well to the extent that the surface can be approximated by a function of the form (1). Of course in the limit $M \to \infty$ all smooth surfaces can be represented by (1), but even for moderate M functions of this form constitute a rich class.

As seen in the previous section, an advantage of using essentially one-dimensional basis functions is the possibility of graphical interpretation. The

entire model can be represented by graphing $s_m(\mathbf{a}_m \cdot \mathbf{x})$ against $\mathbf{a}_m \cdot \mathbf{x}$ and by specifying $\{\mathbf{a}_m\}_{1 \le m \le M}$ (perhaps graphically for $p = 2$ or 3). Additionally, adequacy of the knot placement can be judged using the M plots of the residuals from the final model against $\mathbf{a}_m \cdot \mathbf{x}$. Proper termination of the algorithm can be assured by monitoring at each iteration the plot of the residuals from the model of the previous iteration along the newly found direction.

The problem of sparse sampling in high dimensions is not encountered, since MASA is fitting one-dimensonal projections of the entire sample. The sample need not lie on a regular grid, and the approximation is affine invariant and smooth. Function values, derivatives, and integrals are cheap to evaluate. In addition, since the approximation is locally quadratic for $q = 3$, optimization algorithms can be expected to converge rapidly.

References

Carl de Boor

[1978] *A practical guide to splines*, Springer-Verlag.

Dorit Carmelli and L L Cavalli-Sforza

[1979] *The genetic origin of the Jews: a multivariate approach*, Human Biology 51, 41–61.

William S Cleveland

[1979] *Robust locally weighted regression and smoothing scatterplots*, J Amer Stat Assoc 74, 829–836.

Richard Franke

[1979] *A critical comparison of some methods for interpolation of scattered data*, Naval Postgraduate School NPS-53-79-003.

Jerome H Friedman, Mark Jacobson, and Werner Stuetzle

[1980] *Projection pursuit regression*, submitted to J Amer Stat Assoc.

H H Rosenbrock

[1960] *An automatic method for finding the greatest or least value of a function*, Computer J **3**, 175–184.

Larry L Schumaker

[1976] *Fitting surfaces to scattered data*, in Approximation Theory, (G G Lorentz, C K Chui, and L L Schumaker, Eds.) III, 203–268.

```
SPLINE FUNCTION AND KNOTS ALONG DIRECTION NR   1
  0.4260   0.4295   0.0416   0.7118   0.3541  -0.0146

PLOT STATISTICS =   0.   0.   0.   /   0.   9.   0.   /   0.   0.   0.
```

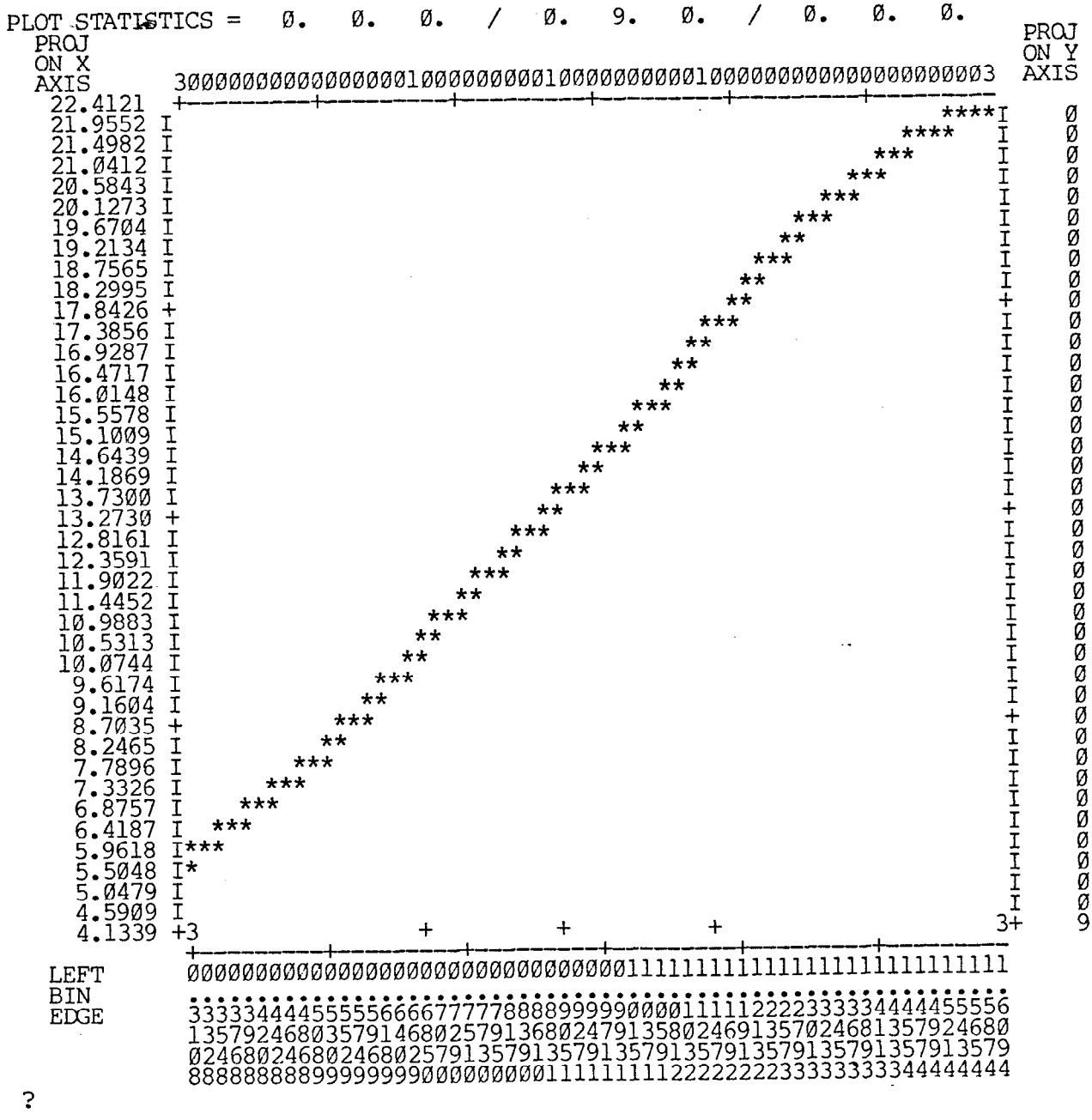


figure 1.1

```
SPLINE FUNCTION AND KNOTS ALONG DIRECTION NR   2
 -0.0211   0.0897   0.9956  -0.0136   0.0       0.0

PLOT STATISTICS =   0.   0.   0.   /   0.   9.   0.   /   0.   0.   0.
```

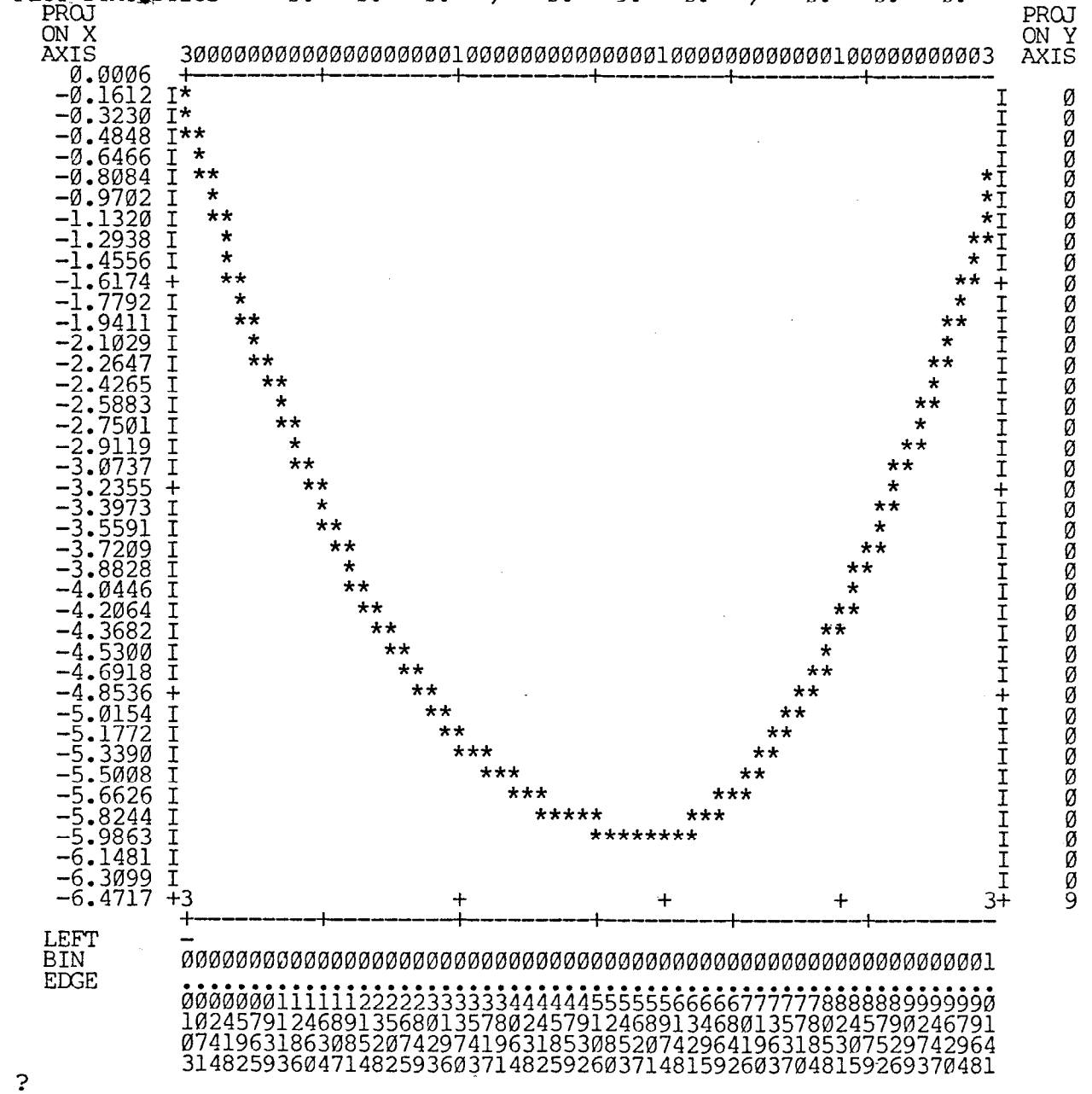
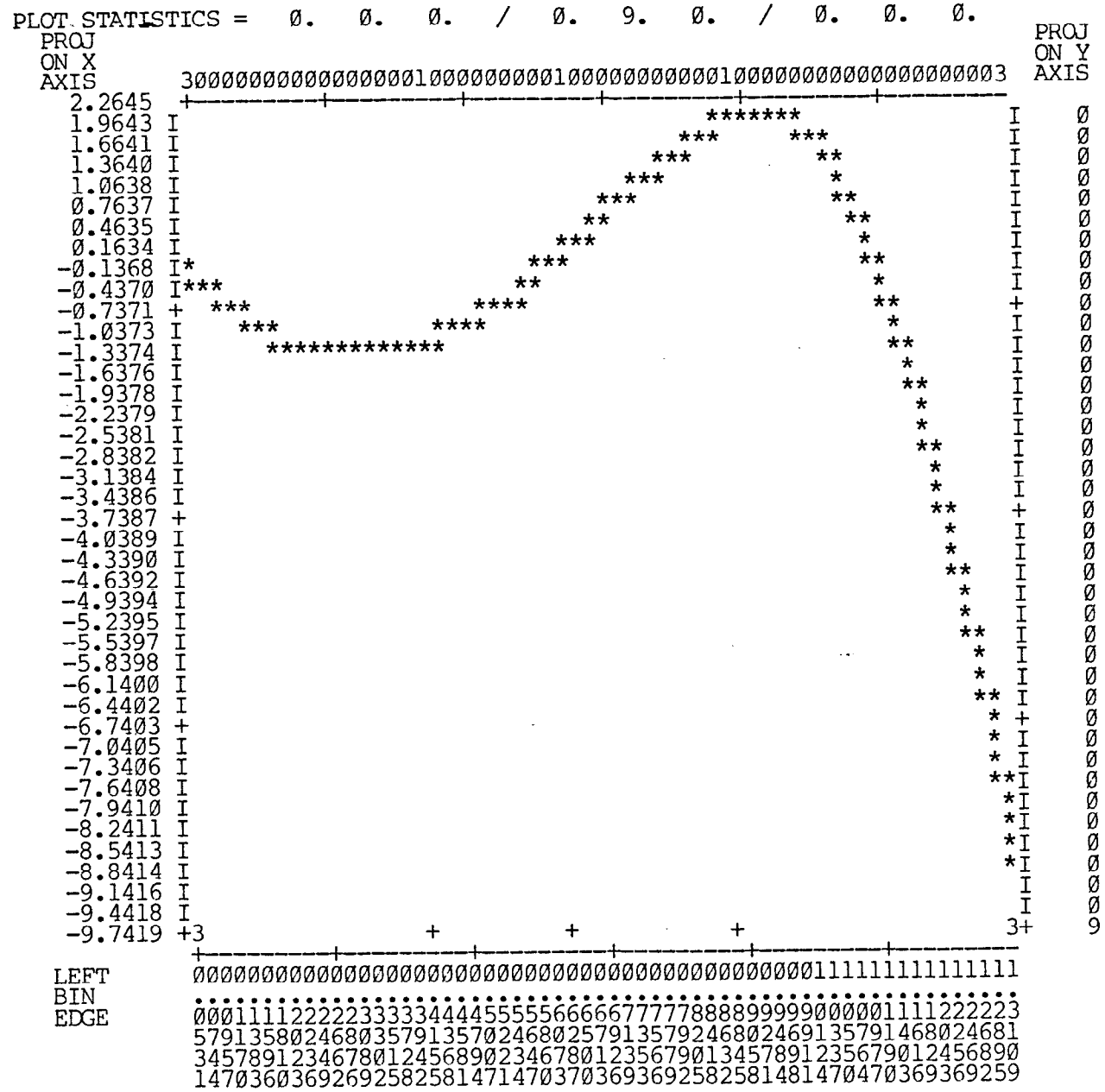


figure 1.2

```
SPLINE FUNCTION AND KNOTS ALONG DIRECTION NR    3
  Ø.6587   Ø.7512   Ø.Ø251  -Ø.Ø34Ø   Ø.Ø       Ø.ØØ83

PLOT STATISTICS =    Ø.    Ø.    Ø.   /   Ø.    9.    Ø.   /   Ø.    Ø.    Ø.
```



?

figure 1.3

```
SPLINE FUNCTION AND KNOTS ALONG DIRECTION NR    4
 -0.4506    0.8818    0.1241   -0.0169    0.0605    0.0072

PLOT STATISTICS =    0.    0.    0.   /   0.    9.    0.   /   0.    0.    0.
```

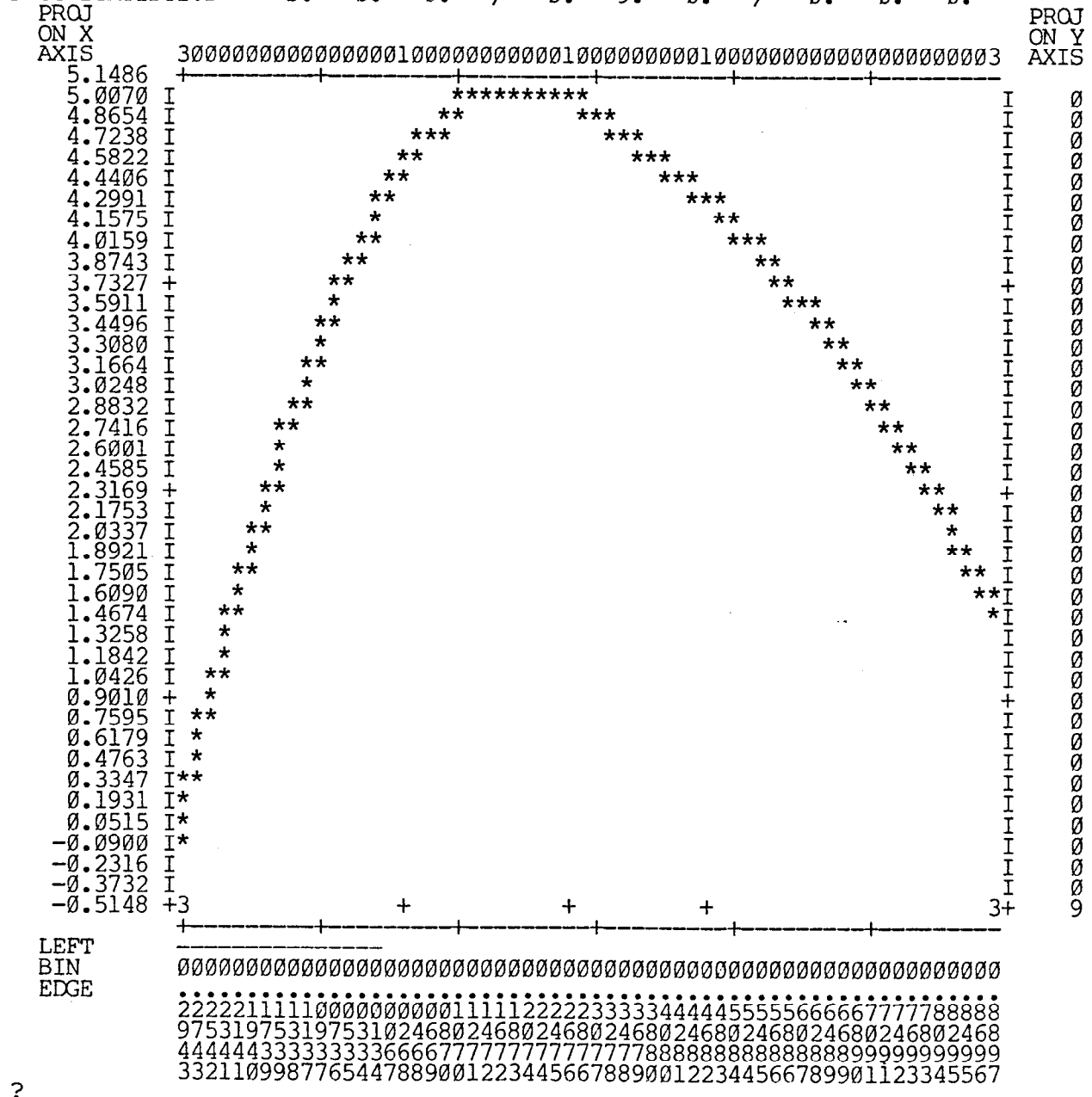


figure 1.4

```
MULTIDIMENSIONAL ADDITIVE SPLINE APPROXIMATION (4/19/80)
PARAMETERS FOR THIS RUN
NOBS            200
NPRED             6
MODE              2
MAXTRY            4
MAXPRO            7
PPCONV        .150000
MAXIT             4
KORDER            3
MAXKNO            9
BANFAC        2.00000
IPRINT            2
NPRINT            1
PLOTRM            .0
AVERAGE SQUARED RESIDUAL AROUND THE MEAN    26.7495
?
```

```
SPLINE FUNCTION AND KNOTS ALONG DIRECTION NR   1
  0.6398   0.7685

PLOT STATISTICS =   0.   0.   0.   /   0.  11.   0.   /   0.   0.   0.
```

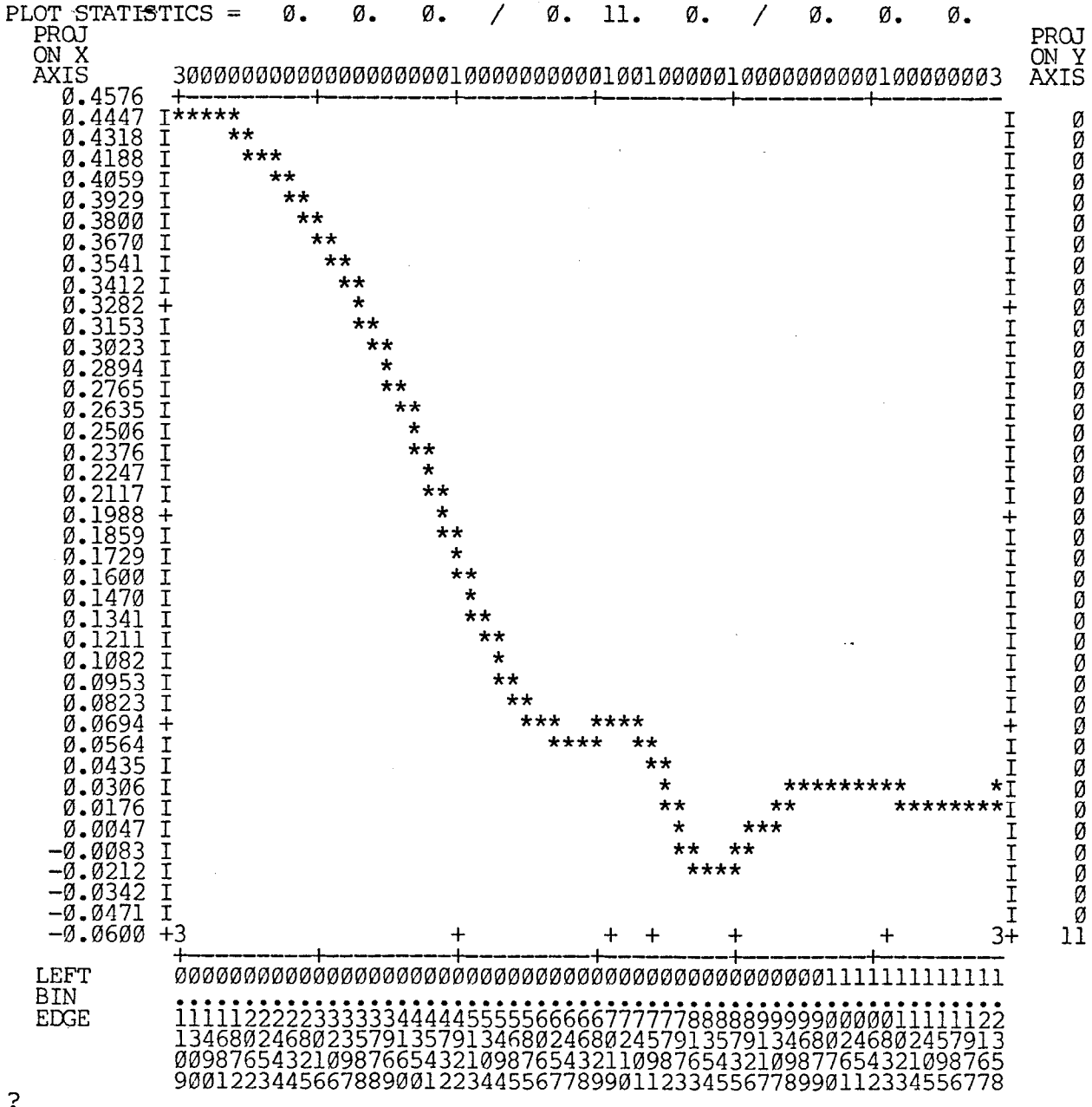
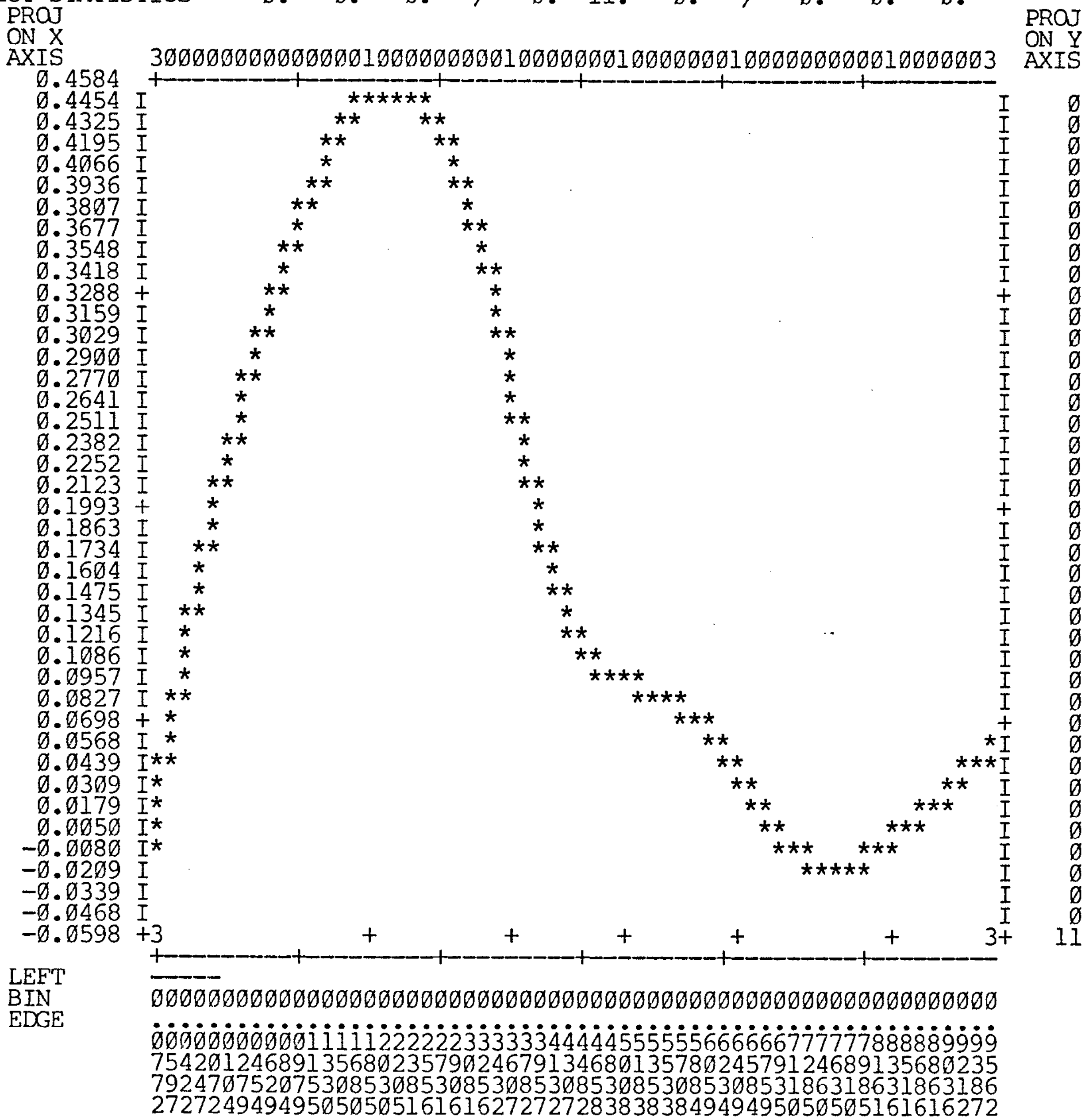


figure 2.1

SPLINE FUNCTION AND KNOTS ALONG DIRECTION NR 2
-Ø.1465 Ø.9892

PLOT STATTSTICS = Ø. Ø. Ø. / Ø. 11. Ø. / Ø. Ø. Ø.



?

figure 2.2

```
SPLINE FUNCTION AND KNOTS ALONG DIRECTION NR   3
  0.9997   0.0259

PLOT STATISTICS =   0.   0.   0.   /   0.  11.   0.   /   0.   0.   0.
```

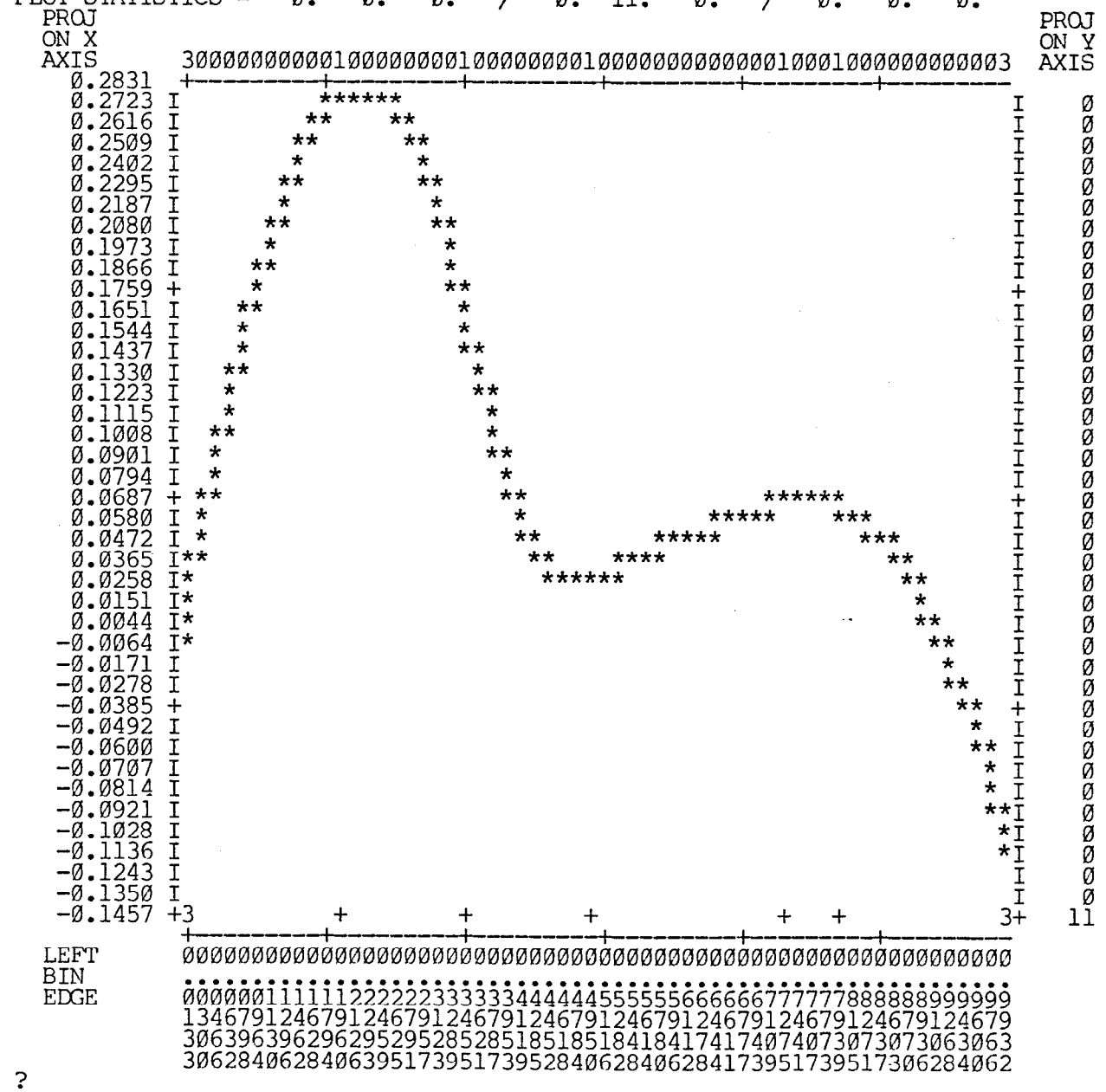
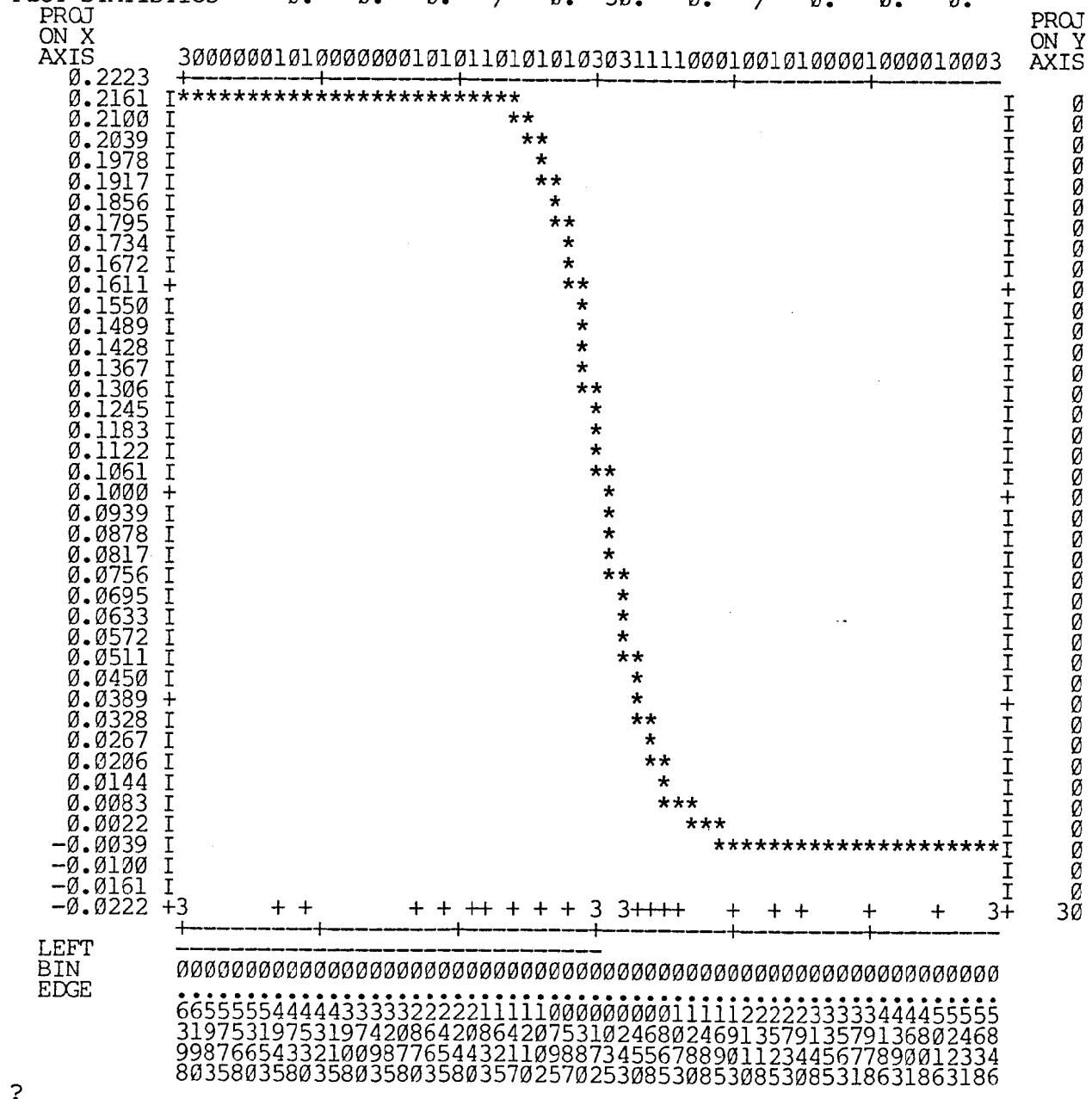


figure 2.3

```
MULTIDIMENSIONAL ADDITIVE SPLINE APPROXIMATION (4/19/80)
PARAMETERS FOR THIS RUN
NOBS            100
NPRED             2
MODE              2
MAXTRY            2
MAXPRO            1
PPCONV       .150000
MAXIT             2
KORDER            3
MAXKNO           11
BANFAC       1.60000
IPRINT            3
NPRINT            1
PLOTRM           .0
AVERAGE SQUARED RESIDUAL AROUND THE MEAN    .109703
?
```

```
SPLINE FUNCTION AND KNOTS ALONG DIRECTION NR   1
  0.7088  -0.7054

PLOT STATISTICS =   0.   0.   0.   /   0.  30.   0.   /   0.   0.   0.
```



?

figure 3.1

```
MULTIDIMENSIONAL ADDITIVE SPLINE APPROXIMATION (4/19/80)
PARAMETERS FOR THIS RUN
NOBS            100
NPRED             2
MODE              2
MAXTRY            2
MAXPRO            1
PPCONV       .150000
MAXIT             4
KORDER            3
MAXKNO           30
BANFAC       .800000
IPRINT            3
NPRINT            1
PLOTRM       .0
AVERAGE SQUARED RESIDUAL AROUND THE MEAN    .972118E-02
?
```

```
SPLINE FUNCTION AND KNOTS ALONG DIRECTION NR   1
 -0.0025   0.0       0.7222  -0.2242  -0.3383  -0.5495   0.0351
 -0.0242   0.0       0.0556  -0.0384  -0.0141   0.0       0.0
  0.0     -0.0039  -0.0660  -0.0153  -0.0235

PLOT STATISTICS =   0.   0.   0.   /   0.  11.   0.   /   0.   0.   0.
```

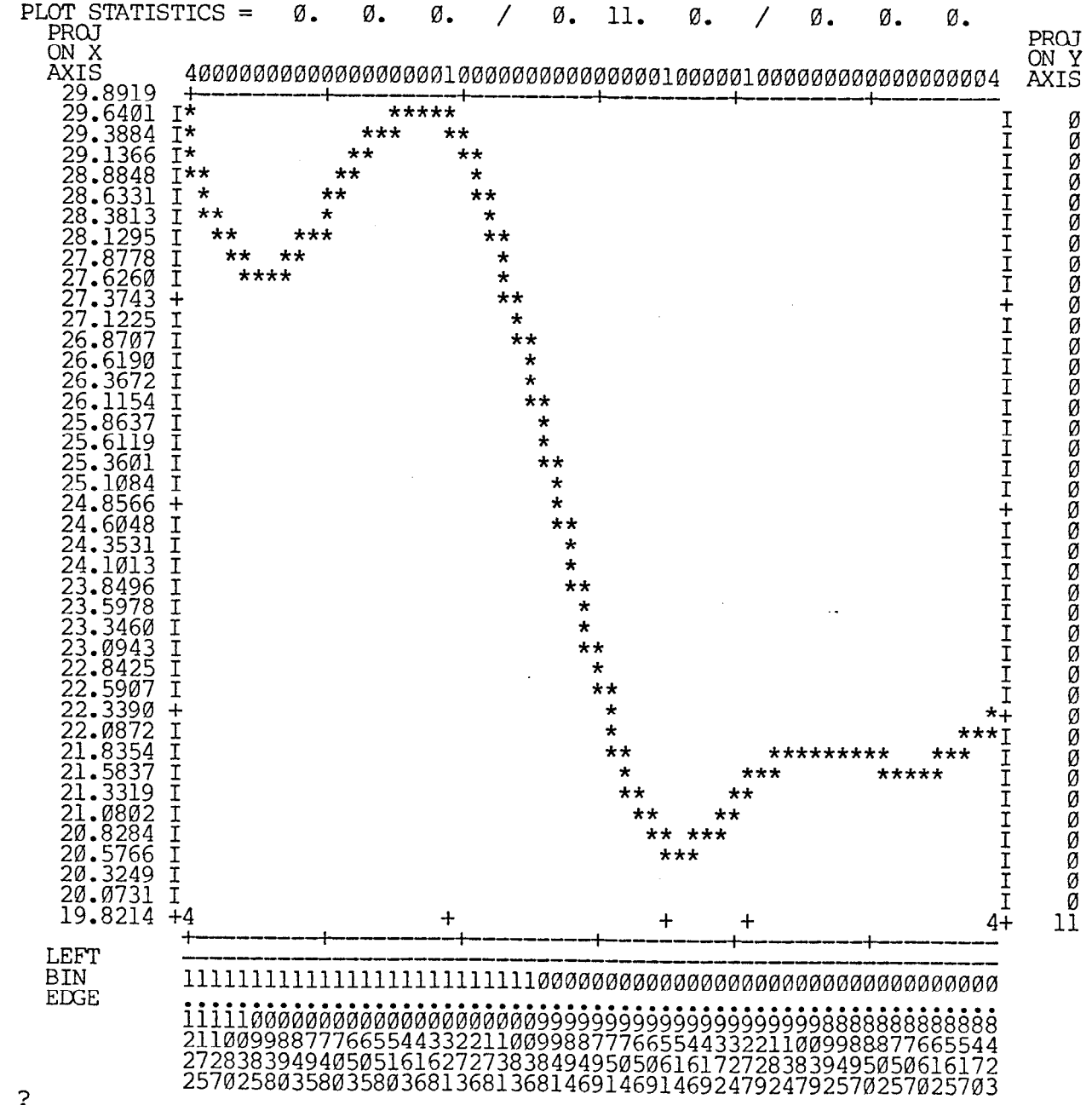


figure 4.1

```
MULTIDIMENSIONAL ADDITIVE SPLINE APPROXIMATION (4/19/80)
PARAMETERS FOR THIS RUN
NOBS            200
NPRED            19
MODE              2
MAXTRY            2
MAXPRO            2
PPCONV       .150000
MAXIT             2
KORDER            4
MAXKNO           11
BANFAC       1.50000
IPRINT            3
NPRINT            1
PLOTRM       .0
AVERAGE SQUARED RESIDUAL AROUND THE MEAN    13.2975
?
```