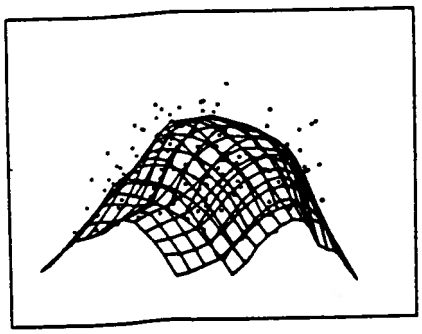
Department Copy

A VARIABLE SPAN SMOOTHER

Jerome H. Friedman

Technical Report No. 5
November 1984

Laboratory for Computational Statistics



Department of Statistics
Stanford University

A VARIABLE SPAN SMOOTHER*

Jerome H. Friedman

Department of Statistics
and
Stanford Linear Accelerator Center
Stanford, California

LCS Technical Report No. 5
SLAC PUB-3477
November 1984

ABSTRACT

A variable span smoother based on linear fits is described. Local cross-validation is used to estimate the optimal span as a function of abscissa value. Computationally efficient algorithms making use of updating formulas are presented.

Submitted to (Journal of American Statistical Association)

---

* Work supported by the Office of Naval Research under contracts ONR N00014-83-K-0472 and ONR N00014-81-K-0340, by the Department of Energy under contracts DE-AC03-76SF00515 and DE-AT03-81-ER10843, and by the U.S. Army Research Office under contract DAAG39-82-K-0056.

1. Introduction

A smoother is a procedure applied to bivariate data $(x_1, y_1) \ldots (x_n, y_n)$ that produces a decomposition

$$y_i = s(x_i) + r_i, \quad i = 1 \ldots n, \tag{1}$$

where s is a smooth function, often simply called the smooth, and the r_i are residuals. It is possible to formally define smoothness, but for our purposes an intuitive notion will be sufficient. Smoothers are used to summarize the association between the predictor variable X and the response Y. It was pointed out by Cleveland (1979) and is a commonly held belief, that when looking at a scatterplot the eye is distracted by the extreme points in the point cloud, i.e., the fuzzy background, and tends to miss structure in the bulk of the data. Augmentation of the plot by a smooth is a possible remedy.

More formally, one can consider a probabilistic framework in which the data are an i.i.d random sample from some joint distribution X,Y. One can define an optimal function f for predicting Y as a function of X that minimizes the expected squared difference between Y and f(X). That is,

$$E_{X,Y}\,[Y - f(X)]^2 = \min_g E_{X,Y}\,[Y - g(X)] \tag{2}$$

where g ranges over all functions. The function f(X) is also the transformation of X that is maximally correlated to Y. The solution function f is

$$f(x) = E\,[Y \mid X = x]\,.$$

Smoothers can be regarded as procedures for estimating the conditional expectation of Y given $X = x$. In many cases, one imagines the joint distribution X,Y to be generated from the process

$$Y = f(X) + \epsilon \tag{3}$$

where f(X) is a smooth function and ϵ is an i.i.d random variable with zero expectation. Clearly, $E\,[Y \mid X = x] = f(x)$, so that the smooth s can be considered an estimate for f.

Recently, smoothers have found new uses in multiple regression algorithms (Friedman and Stuetzle, 1981, Breiman and Friedman, 1984, Hastie and Tibshirani, 1984, and Friedman, 1984). In these procedures, a smoother is used as a primitive operation repeatedly applied to varying projections of the data; the quality of the smooth (2) is used as a figure-of-merit driving the algorithm. In such applications, the smoother must be both very flexible and rapidly computable. This paper describes such a smoother, and is, in fact, the one currently in use with all but one of these algorithms.

2. Basic Concepts

Assume the data are generated according to (3). We are interested in procedures that can approximate f arbitrarily closely, given a dense enough sample. A straightforward estimator of a conditional expectation would be a conditional average

$$\hat{E}(Y \mid x_i) = ave(y \mid x_i) = y_i.$$

Although this estimate is unbiased, it can have high variance. Also, this estimate need not approach f as the sample becomes denser. A more reasonable estimate is based on local averaging. Take $s(x_i)$ to be the average of the responses y for those observations with predictor values x in a neighborhood N_i of x_i:

$$\hat{E}(Y \mid x_i) \equiv s(x_i) = ave(y_j \mid x_j \epsilon N_i). \tag{4}$$

A critical parameter to be chosen is the SPAN, the size of the neighborhood over which averaging takes place. It controls the smoothness of s. The bigger the span, the smoother s will be. To obtain consistency, i.e., to make sure that s gets arbitrarily close to f as the sampling rate increases, one must shrink the diameter of the neighborhood in such a way that the number of observations in the neighborhood still grows to infinity. Shrinking the neighborhood makes the systematic or bias component in the estimation error diminish, while increasing the neighborhood sample size guarantees that the variance component of the error goes to zero as well.

3. A Simple Nonresistant Smoother

With a local averaging smoother (4), the size of the neighborhood is usually specified by the span, the number J of observations to be included in the averaging. We will

assume J to be odd and the abscissas x_i to be in increasing order. The neighborhood can be chosen either symmetrically, containing $J/2$ observations to the left of x_i and the same number to the right, or it can be chosen to contain the J nearest neighbors of x_i, including x_i. (We assume that $J/2$ is computed by integer division.) There are no general results on which of these two possibilities is better. The nearest neighbors approach generalizes to higher dimensions, but the choice of a symmetric neighborhood is computationally simpler in that exactly one point enters and one point leaves the neighborhood as one moves from observation i to observation $i+1$. We will, in the following, use symmetric neighborhoods. Near the boundaries, it is, of course, not possible to keep N symmetric. The average (4) need not be recomputed every time. It can be updated, reducing the computation from nJ to n. Such updating can be done for all the smoothers we will consider, and is highly desirable because in typical applications J is 5% to 50% of n, and thus the savings are substantial.

The simple moving average smoother has some serious shortcomings. One disturbing property is that it does not reproduce straight lines if the abscissa values are not equispaced. Another disturbing feature is bad behavior at the boundaries. If, for example, the slope of the underlying function f is positive at the right boundary, the estimate for observations close to the boundary will be biased downwards; if the slope is negative, the estimate is biased upwards. Both problems can be alleviated by fitting a least squares straight line to the observations in the neighborhood instead of fitting a constant (zero slope) and taking the value of the line at x_i as the smoothed value. (This keeps the bias of the curve estimate strictly proportional to $d^2f \mid dx^2$.) For the computation, again updating formulas can be used. The slope β and intercept α of the least squares straight line through a set of points $(x_1, y_1) \ldots (x_J, y_J)$ are given by

$$\alpha = \bar{y}_J - \beta \bar{x}_J$$
$$\beta = \frac{C_J}{V_J}$$

with (5)

$$\bar{x}_J = \Sigma x_j / J,$$
$$\bar{y}_J = \Sigma y_j / J,$$
$$C_J = \Sigma (x_j - \bar{x}_J)(y_j - \bar{y}_J),$$
$$V_J = \Sigma (x_j - \bar{x}_J)^2.$$

When we want to add an observation (x_{J+1}, y_{J+1}), we can make use of the following easily derived formulas:

$$\begin{aligned}
\bar{x}_{J+1} &= (J\,\bar{x}_J + x_{J+1})/(J+1), \\
\bar{y}_{J+1} &= (J\,\bar{y}_J + y_{J+1})/(J+1), \\
C_{J+1} &= C_J + \tfrac{J+1}{J}(x_{J+1} - \bar{x}_{J+1})(y_{J+1} - \bar{y}_{J+1}), \\
V_{J+1} &= V_J + \tfrac{J+1}{J}(x_{J+1} - \bar{x}_{J+1})^2.
\end{aligned}$$

Analogous formulas can be used for removal of an observation from the set.

4. Choice of Span

The most important choice in the use of a local averaging smoother is the choice of the span value. If the smoother is regarded as an estimator for $f(x)$ (3), then the span controls the trade off between bias and variance of the estimate. We illustrate this for the case of a simple moving average smoother (4). In this case, the smoothed value at point x_i is given by

$$s(x_i) = \frac{1}{J} \sum_{i-J/2}^{i+J/2} y_j.$$

If we assume that the errors ϵ_i are i.i.d. with expected value zero and variance σ^2, then the expected squared error at point x_i is

$$e^2(x_i \mid J) = (f(x_i) - \frac{1}{J} \sum_{i-J/2}^{i+J/2} f(x_j))^2 + \frac{1}{J}\sigma^2. \tag{6}$$

Increasing the span J will (if $d^2f/dx^2 \neq 0$) increase the first term, the bias component of the estimation error and decrease the second term, the variance component; decreasing the span will have the opposite effect. Stated more geometrically, a larger span makes the smooth appear less wiggly by more strongly damping high frequency components of the series (x_i, y_i).

One can estimate the optimal span value in a particular situation as that value that minimizes an estimate for

$$e^2(J) = E_{X,Y}\,[Y - s(X \mid J)]^2.$$

Using the average squared residual of the data from the smooth

$$\hat{e}^2(J) = \frac{1}{n} \sum_{i=1}^{n} [y_i - s(x_i \mid J)]^2$$

for this purpose is not appropriate since this is always minimized by the span value $J = 1$. A better estimate is provided by a method referred to as "cross-validation" (M. Stone, 1974) or "predictive sample reuse" (Geisser, 1975). Each observation is in turn deleted and the value of the smooth $s_{(i)}(x_i \mid J)$ at x_i is calculated from the other $n-1$ observations. The cross-validated estimate of the integrated square error is

$$\hat{e}^2_{cv}(J) = \frac{1}{n} \sum_{i=1}^{n} \left[y_i - s_{(i)}(x_i \mid J) \right]^2 . \tag{7}$$

Clearly, $E[\hat{e}^2_{cv}]$ equals the expected squared error obtained by applying the procedure to a sample of $n-1$ observations from the same distribution. The cross-validated estimate for the optimal span value is taken to be the value J_{cv} that minimizes (7),

$$\hat{e}^2_{cv}(J_{cv}) = \min_{0<J\leq N} \hat{e}^2_{cv}(J).$$

Model selection through cross-validation has been remarkably successful in a wide variety of situations (see M. Stone, 1974, Geisser, 1975, Craven and Wahba, 1979, C. Stone, 1981).

For the moving average smoothers discussed above, the cross-validated residuals

$$r_{(i)}(J) = y_i - s_{(i)}(x_i \mid J)$$

are simply related to the ordinary residuals

$$r_i(J) = y_i - s(x_i \mid J)$$

owing to the fact that these smoothers are linear. A linear smoother is one for which the value of the smooth for a particular observation is a linear combination of the y values for all of the observations, i.e.,

$$s(x_i \mid J) = \sum_{j=1}^{n} H_{ij}(J) y_j .$$

The linear combination H_{ij} may be different for each observation i and depends on J. (Note that if x_j is not in the neighborhood of x_i, $H_{ij}(J) = 0$.) For linear smoothers, the cross-validated residual is given by

$$r_{(i)}(J) = r_i(J)/(1 - H_{ii}(J)).$$

For the local straight line smoother discussed above, it is straightforward to calculate

$$H_{ii}(J) = \frac{1}{J} + \frac{(x_i - \bar{x}_J)^2}{V_J}$$

with $\bar{x}_J$ and V_J given by (5). Therefore,

$$\hat{e}^2_{cv}(J) = \frac{1}{n}\sum_{i=1}^{n} [y_i - s(x_i \mid J)]^2 \Big/ \left[1 - \frac{1}{J} - \frac{(x_i - \bar{x}_J)^2}{V_J}\right]^2 .$$

For small to moderate changes in J, $\hat{e}^2_{cv}(J)$ changes very little so that it is adequate to evaluate it for several (3 to 5) discrete values of J in the range $[0 < J < n]$. The value of J corresponding to the smallest of these $\hat{e}^2_{cv}(J)$ values is then used. This can be accomplished by maintaining several running average smoothers - one for each span value - in the pass over the data, thus keeping the computational cost linear in n.

5. Variable Span Smoother

So far, we have been assuming that the (number of counts in the) span remains constant over the whole range of predictor x values. This is not optimal if either the variance of the random component and/or the second derivative of the underlying function f change over the range of predictor values. A local increase in error variance would call for an increase in span, whereas an increase in second derivative of f would require a decrease. It is, therefore, desirable to allow the span value to adapt to these changing conditions. This requires that the optimal span value be chosen locally rather than using a single global value.

More formally, one can estimate an optimal span value for each x, as well as the corresponding optimal smooth value, by minimizing an estimate for

$$e^2(s, J) = E_{X,Y}\,[Y - s(X \mid J(X))]^2$$

with respect to both functions $s(x)$ and $J(x)$. The resulting function $s(x)$ is then taken as our smooth. Re-expressing this criterion as

$$e^2(s, J) = E_X E_Y [(Y - s(X \mid J(X))^2 \mid X],$$

we see that $s(x)$ (and $J(x)$) can be found by minimizing

$$e^2(s, J \mid x) = E_Y [(Y - s(x \mid J))^2 \mid x] \tag{8}$$

with respect to s and J for each value of x. This will result in smaller e^2 than constraining $J(x)$ to be constant. (This is not necessarily true for the estimates however. The decrease in bias associated with the variable span may be more than offset by the increased variance associated with estimating the additional function $J(x)$.)

As with the constant span case, we begin by applying the local linear smoother several times with several discrete values of J in the range $0 < J < n$. In our implementation, we use three values $J = 0.05n$, $0.2n$, and $0.5n$. These are intended to reproduce the three main parts of the frequency spectrum of $f(x)$ and are referred to as the tweeter, midrange, and woofer smoothers respectively. It is then necessary to estimate (8) at each data value x_i for each smoother. Simply using the cross-validated residual

$$r_{(i)}(J) = [y_i - s(x_i \mid J)] \Big/ \left(1 - \frac{1}{J} - \frac{(x_i - \bar{x}_J)^2}{V_J}\right) \tag{9}$$

results in estimates with too much variance since each estimate is based on only one observation. Better estimates can be obtained by smoothing $r^2_{(i)}(J)$ against x_i (with the midrange smoother) and using the smoothed values as the estimates $\hat{e}^2(s, J \mid x_i)$. For stability reasons, it turns out to be a little better to smooth $|r_{(i)}(J)|$ against x_i using the resulting estimates $\hat{e}(s, J \mid x_i)$ to select the best span value:

$$\hat{e}(s, J_{cv}(x_i) \mid x_i) = \min_J \hat{e}(s, J \mid x_i) \tag{10}$$

where J takes on the tweeter, midrange and woofer span values. The smoothed response value $s^*(x_i)$ at each x_i can then be taken as the smoother (tweeter, midrange, or woofer) value associated with the optimal span estimate

$$s^*(x_i) = s(x_i \mid J_{cv}(x_i)).$$

When obtained in this manner, the optimal span (and curve) estimates can have unnecessarily high variance. This is because the estimated span value $J_{cv}(x_i)$ is not constrained to vary smoothly from one observation to the next (as ordered on x_i). It is possible that two (or more) smoothers can have very similar e values in a region of x, but different values of s. Due to variance in the estimates $\hat{e}(s, J \mid x_i)$, different span (and curve) values can be choosen for neighboring x_i. Better optimal span (and resulting curve) estimates are obtained by smoothing the values $J_{cv}(x_i)$ (10) against x_i (again with the midrange smoother). The result is an estimated span for each observation with a value between the tweeter and woofer values. The resulting curve estimate is obtained by interpolating between the two (out of the three) smoothers with closest span values.

It is often known (or suspected) that the underlying true curve $f(x)$ (3) is very smooth. When this is, in fact, the case, more accurate curve estimates can be obtained by biasing the span selection procedure toward larger span values. Even when this is not the case, people often find smoother curves more visually pleasing and are willing to sacrifice a degree of accuracy for an estimate that is less rough. We, therefore, need a method for enhancing the low frequency (bass) component of the smoother output. For this purpose, we introduce a bass (tone) control.

The idea is to increase the span value selected at each x_i in inverse proportion to the increase in predicted-absolute-error $\hat{e}$ associated with the span increase. Let $J_{cv}(x_i)$ be the estimated optimal span and J_w the woofer span. The span value for each x_i is taken to be

$$J(x_i) = J_{cv}(x_i) + (J_w - J_{cv}(x_i))R_i^{10-\alpha} \tag{11}$$

with

$$R_i = \left[\frac{\hat{e}(J_{cv}(x_i)|x_i)}{\hat{e}(J_w|x_i)}\right].$$

Here $0 \leq \alpha \leq 10$ is a user specified parameter (tone control). The value $\alpha = 0$ corresponds to $J(x_i) \simeq J_{cv}(x_i)$ (very little bass enhancement) while $\alpha = 10$ corresponds to $J(x_i) = J_w$ (maximum bass). Values of α between these extremes cause different degrees of bass enhancement. For a given value of α, the amount of bass increase is controlled by the ratio R_i. The larger this ratio, the smaller the loss in increasing the span, and thus, the more it is increased. This tone control is applied before the spans

are smoothed. Note that the amount of bass enhancement is highly nonlinear in the parameter α. Increases for small values of α have much less effect than the same sized increases at larger α values. Figure 1 shows the amount of bass enhancement as a function of R_i for several values of α.

The resulting variable span smoother makes nine passes over the data:

1. Primary data smooths with tweeter, midrange, and woofer spans.
2. Smooth cross-validated absolute residuals (9) for each of the primary smooths with midrange span.
3. Select best span as minimizing the output of Step 2 for each observation. (Apply low frequency bass enhancement if desired.)
4. Smooth best span estimates with midrange span.
5. Use smoothed span estimates to interpolate between primary smoother values.

It is important to note that using cross-validated residuals as a basis for choosing span value is highly sensitive to lack of independence among the ϵ_i (3) as ordered on x. If there is a large positive (negative) correlation among observations with similar x values, substantial under (over) estimates will result. In situations where a high degree of auto-correlation is suspected, these span selection procedures should be used with caution.

6. An Example

In this section, we present a simulated example intended to illustrate a situation where variable span is important. The data for this example consist of $n = 200$ pairs (x_i, y_i) with the x_i drawn randomly (i.i.d) from a uniform distribution in the interval [0,1]. The y_i are obtained from

$$y_i = sin(2\pi(1 - x_i)^2) + x_i\epsilon_i \tag{12}$$

with the ϵ_i i.i.d standard normal. This example simulates a situation in which the curvature of f decreases and the variance of the random component increases with increasing x. In the first set of examples, no bass enhancement was used. Figure 2a shows a scatterplot of these data with the resulting variable span smooth $s(x)$ superimposed.

Figure 2b shows the individual tweeter, midrange, and woofer smooths. Figure 2c shows the estimated optimal span $J(x) = J_{cv}(x)$ as a function of x.

In the low noise high curvature region ($x < 0.2$), the tweeter span is selected. In the high noise low curvature region $x > 0.8$, the span increases rapidly to the woofer value. In the region where both curvature and noise are moderate, the selected span averages just below the midrange value. The resulting composite smooth $s(x)$ (Fig. 2a) is seen to be much better than any of the individual (tweeter, midrange, or woofer) smooths (Fig. 2b).

In order to see to what extent these results reflect general behavior, 1000 data sets were generated, all with identical set of x_i, but each with a different random set ϵ_i. The y_i were constructed as in (12). Figure 2d shows the estimated optimal span function $J(x)$ averaged over these 1000 runs. This $\overline{J}(x)$ reflects similar behavior to that of the first run, $J(x)$. The span is seen to rise a bit more rapidly in the region of middle x values, but not to as high a value for large x. Figure 2e shows the average accuracy of the composite smooth, as well as each of the three primary smooths, as a function of x. The absolute error

$$e(x_i) = |s(x_i) - sin\,[2\pi(1 - x_i)^2]|$$

was averaged over the 1000 runs for each x_i. (The points for each smoother are connected by straight lines.) The composite variable span smooth is again seen to be much better than any of the three constant span primary smooths. It incurs none of the (very large) bias associated with the midrange and woofer spans for low x values, and its absolute error is about one-half that of the tweeter for the larger x values. Over the entire range of x values, the variable span smoother has performance comparable to the best of the primary smoothers at each x value. Only for the very largest x values ($x > 0.7$), the woofer smoother incurs about 20% less error. Figure 2e also illustrates the problems associated with end effects. The average error for points near the very edges of the x interval is about twice that for close-by interior points.

Figures 3a–3e show the corresponding results for data generated as above but with $n = 100$. The results for this smaller sample size reflect the same general behavior described above. The average absolute error is somewhat higher, especially in the high variance (large x) region.

Figure 4a shows the same data as that of Figure 2a, but the superimposed smooth is the result of applying some bass enhancement, $\alpha = 5$ (11). The result is visually more pleasing in that it is less wiggly in the high variance region ($x > 0.5$). There appears to be an increase in bias, however, in that the curve seems to lie above the data near $x = 0.1$ and undershoot the data near $x = 0.5$. These suspicions are verified in Figure 4b where the average absolute error (over 1000 runs) of the composite variable span smoother, as well as the three primary smoothers, are shown. Although the error is reduced to that of the woofer for $x > 0.6$, it is dramatically increased in the high curvature regions $0.05 \leq x \leq 0.20$ and $0.35 \leq x \leq 0.60$. Figure 4c shows the average span function $\overline{J}(x)$. Except for the very low noise high curvature region ($x < 0.1$), the selected span value is generally larger than the estimated optimal span $J_{cv}(x)$ (Figure 2d).

This example was deliberately constructed to be difficult and to test the variable span aspect of the smoothing procedure. It shows that the method can readily adapt to changing circumstances (function curvature and/or error variance). Not all situations encountered in practice are this dramatic and in less dramatic situations the gain using variable span will be correspondingly less. In some settings, the additional variance encountered in estimating the two functions $s(x)$ and $J(x)$ can more than offset the decrease in bias so that using an optimally estimated constant span will incur less absolute error. This becomes more likely for small sample sizes ($n < 40$). Even in these cases, however, the variable span smoother is usually almost as good as the best single span smoother, especially if some bass enhancement is employed.

7. Discussion

Cleveland (1979) suggested a smoother also based on local linear fits. It differs from the one described in this report mainly in three respects:

- It does not automatically choose the span by cross-validation.
- It does not use variable span.
- In the fit of the local straight line determining the smooth $s(x_i)$ for predictor value x_i, the observations are weighted according to their distance from x_i; observations towards the extremes of the span receive lower weights than observations

with predictor values close to x_i. Asymptotic calculations suggest that assigning unequal weights should reduce the error of the curve estimate, but there is no evidence that it makes a substantial difference for sample sizes occurring in practice. It does, however, produce a smoother looking estimate.

Updating formulas cannot be used in this scheme, making it comparatively expensive in terms of computing. To reduce computation, Cleveland suggests evaluating the smooth only for every ($\ell \leq\leq n$) predictor value. The smoothing procedure described in this report was developed because the best span value is usually not known in advance, a variable span is often important, and because the use of updating formulas dramatically reduces computation. This is critical when the smoother is repeatedly applied as a primitive operation in more complicated algorithms.

Another class of procedures suggested for smoothing are based on splines. A spline function s of order ℓ with knots at $z_1 \ldots z_k$ is a function satisfying the following two conditions:

- In each of the intervals $(-\infty, z_1), (z_1, z_2) \ldots (z_{k-1}, z_k), (z_k, \infty)$, s is a polynomial of degree $\ell - 1$;
- s has $\ell - 2$ continuous derivatives.

One way to use spline functions in smoothing is to fit a spline function with knots $z_1 \ldots z_k$ to the data $(x_1, y_1) \ldots (x_n, y_n)$, either by least squares or by some resistant method. The degree of smoothness is determined by the number and position of the knots. A major disadvantage of this method is that $k + 1$ parameters must be chosen: the number and the positions of the knots. Usually some heuristic procedure is used to place the knots once k has been fixed (Jupp, 1978). This leaves the number of knots to be determined. This number plays the role of the span in determining the degree of smoothing. Unfortunately, the output of the smoother can depend on k in a very nonlinear way; it is easy to construct examples where the addition of one more knot substantially decreases the residual sum of squares, whereas further knots hardly make any difference. This makes k more difficult to choose than the span in a local averaging smoother. Furthermore, least squares fit of splines is substantially slower so that choosing k through cross-validation is usually too expensive.

Another way is to use smoothing splines in the sense of Reinsch (1967). A smoothing

spline s of order 2ℓ for smoothing parameter λ is the function that minimizes

$$\sum (y_i - f(x_i))^2 + \lambda \int_{x_1}^{x_n} f^{(\ell)2}(x)dx$$

among all functions f with ℓ derivatives. The solution turns out to be a spline function of order 2ℓ with knots $x_1 \ldots x_n$; the name is thus justified. The larger λ is chosen, the smoother s becomes; thus, λ here plays the role of the span. Computation of the spline for given λ requires the solution of a banded $n * n$ linear system. A drawback of the method, as described here, is that it is impossible to obtain an intuitive feeling for the choice of λ in a given example. So, one usually fixes not λ, but the residual sum of squares around the smooth. The corresponding value of λ then has to be found iteratively by repeatedly solving the minimization problem. This substantially increases the necessary amount of computation. Algorithms to determine the optimal λ by cross-validation usually require computation of the singular value decomposition of an $n * n$ matrix; they are expensive and infeasible for sample sizes larger than 200-300. An approximate method has recently been proposed (Silverman, 1984), however, that is much faster, thereby extending the use of smoothing splines to larger samples.

To summarize, the local averaging smoother described in this report has two desirable properties that set it apart from other smoothers: it is both very fast to compute and the value of the parameter that controls the amount of smoothing is automatically optimized locally (through cross-validation), allowing it to adapt to the response function over the range of predictor values. Listing of a FORTRAN program implementing the procedure described herein is available from the author.

REFERENCES

Breiman, L. and Friedman, J.H. (1984). "Estimating optimal transformations for multiple regression and correlation." J. Amer. Statis. Assn. (to appear).

Cleveland, W.S. (1979). "Robust locally weighted regression and smoothing scatterplots," J. Amer. Statist. Assoc., 74, 828-836.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. Estimating the correct degree of smoothing by the method of generalized cross-validation. Numerische Mathematik 31, 317-403.

Friedman, J. H. and Stuetzle, W. (1981). "Projection pursuit regression," J. Amer. Statist. Assoc. 76, 817-823.

Friedman, J.H. (1984). "Classification and multiple response regression through projection pursuit," Dept. of Statistics Tech. Report LCM006, Stanford University.

Geisser, S. (1975). The predictive sample reuse method with applications, J. Amer. Statist. Assoc. 74, 153-160.

Hastie, T. and Tibshirani, R. (1984). "Generalized Additive Models," Dept. of Statistics Tech. Report LCM02, Stanford University.

Jupp, D. L. (1978). "Approximation to data by splines with free knots," SIAM J. Numer. Anal. 15, 328-343.

Reinsch, C.H. (1967). "Smoothing by spline functions." Numer. Math. 10, 177-183.

Silverman, B.W. (1984). A fast and efficient cross-validation method for smoothing parameter choice in spline regression. J. Amer. Statist. Assn. 19, 584-589.

Stone, C.J. (1981). Admissible selection of an accurate and parsimonious normal linear regression model. Ann. Stat. 9, 475-485.

Stone, H.M. (1974). "Cross-validatory choice and assessment of statistical predictions." J. Roy. Statist. Soc. B-36, 111-147.

FIGURE CAPTIONS

Figure 1: Bass amplification factor as a function predictive-absolute-error ratio for various tone control settings.

Figure 2a: Scatterplot of data with composite smooth superimposed.

Figure 2b: Individual tweeter, midrange and woofer smooths.

Figure 2c: Selected span $J_{cv}(x)$.

Figure 2d: Expected estimated optimal span $\mathcal{J}_{cv}(x)$.

Figure 2e: Expected absolute error of three primary smooths and composite variable span smooth.

Figure 3a: Scatterplot of data with composite smooth superimposed.

Figure 3b: Individual tweeter, midrange and woofer smooths.

Figure 3c: Selected span $J_{cv}(x)$.

Figure 3d: Expected estimated optimal span $\mathcal{J}_{cv}(x)$.

Figure 3e: Expected absolute error of three primary smooths and composite variable span smooth.

Figure 4a: Scatterplot of data with composite smooth superimposed.

Figure 4b: Expected absolute error of three primary smooths and composite variable span smooth.

Figure 4c: Expected chosen span $\mathcal{J}(x)$.

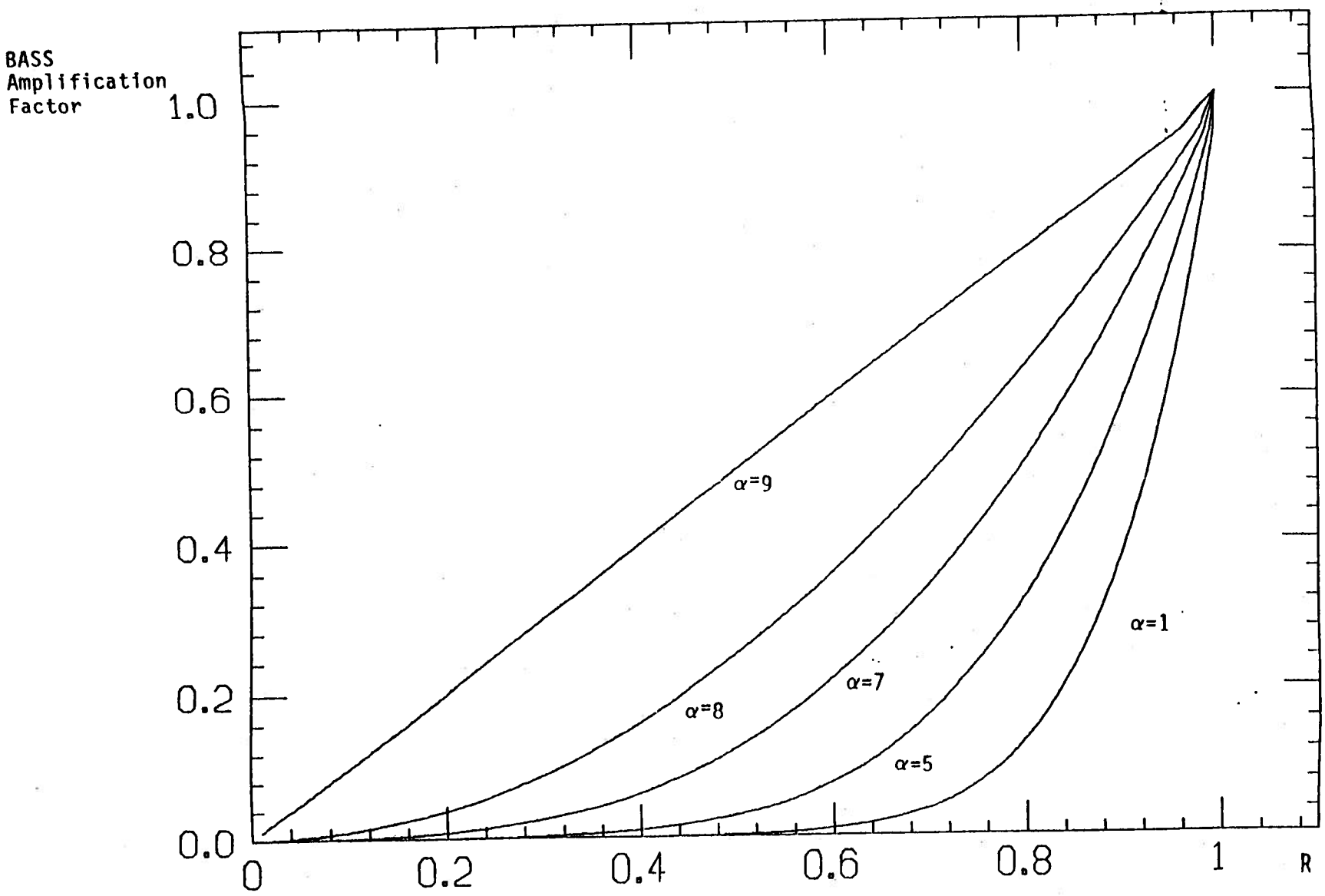

BASS
Amplification
Factor
1.0
0.8
0.6
0.4
0.2
0.0
0
0.2
0.4
0.6
0.8
1
R
α=9
α=8
α=7
α=5
α=1

FIGURE 1

FIGURE 2a

Tweeter
Woofer
Midrange
Tweeter
Woofer
Midrange
Midrange
Tweeter
Woofer
Midrange
Woofer
2
1
0
-1
-2
Y
0
0.2
0.4
0.6
0.8
1
X

FIGURE 2b

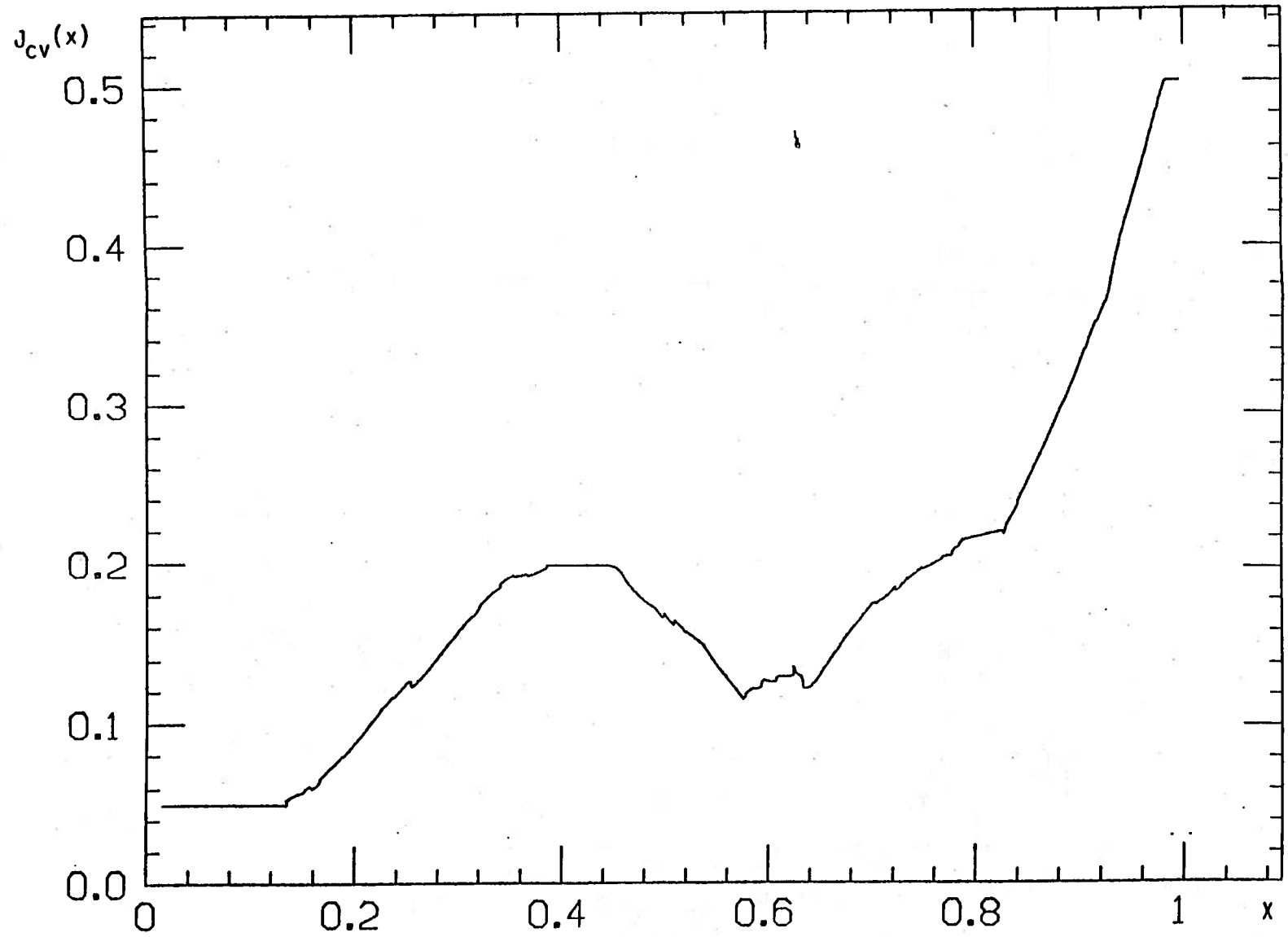


FIGURE 2c

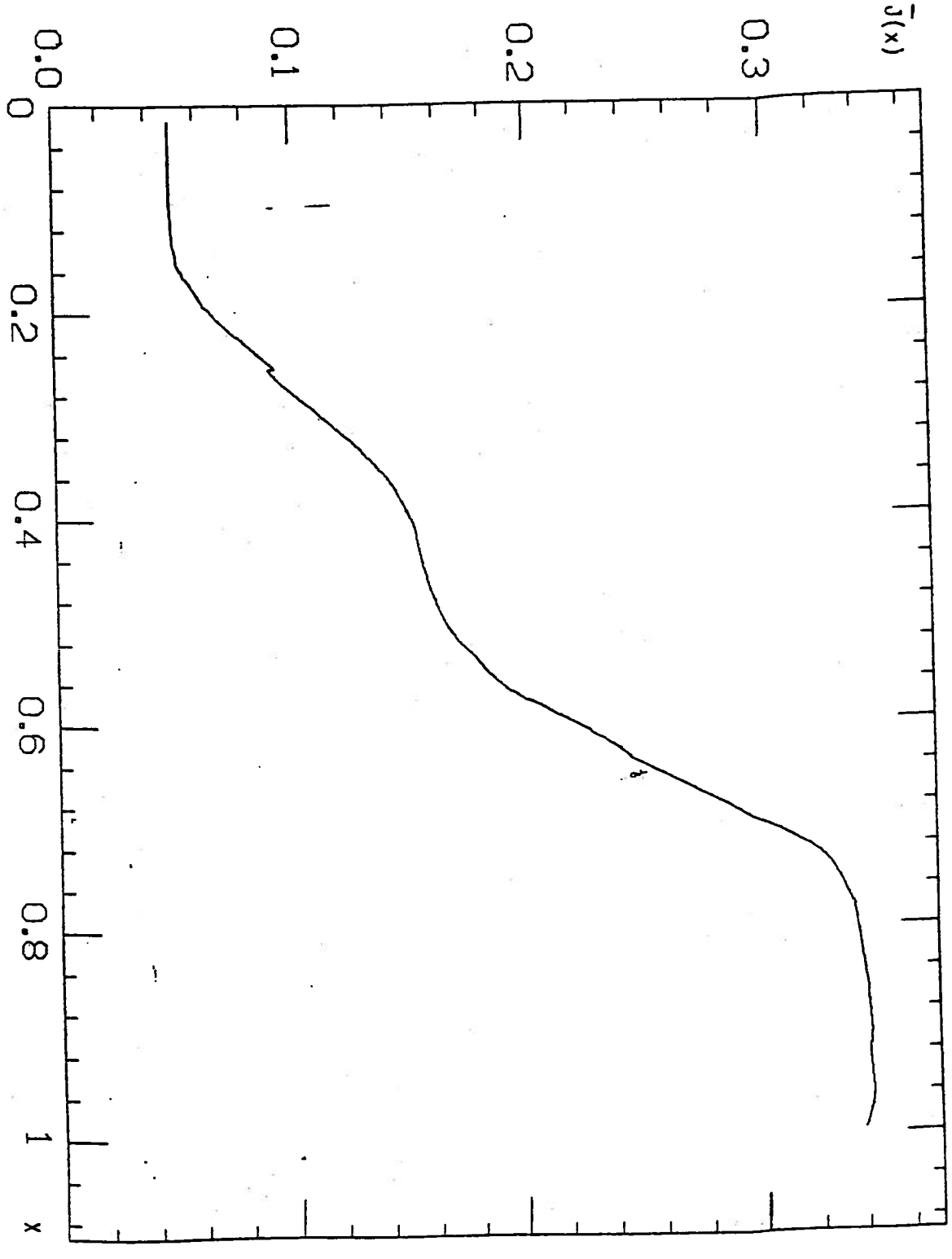


FIGURE 2d

e(x)
0.3
0.2
0.1
0.0
0
0.2
0.4
0.6
0.8
1
x
Woofer
Tweeter
Midrange
Midrange
Variable Span

FIGURE 2e

FIGURE 3a

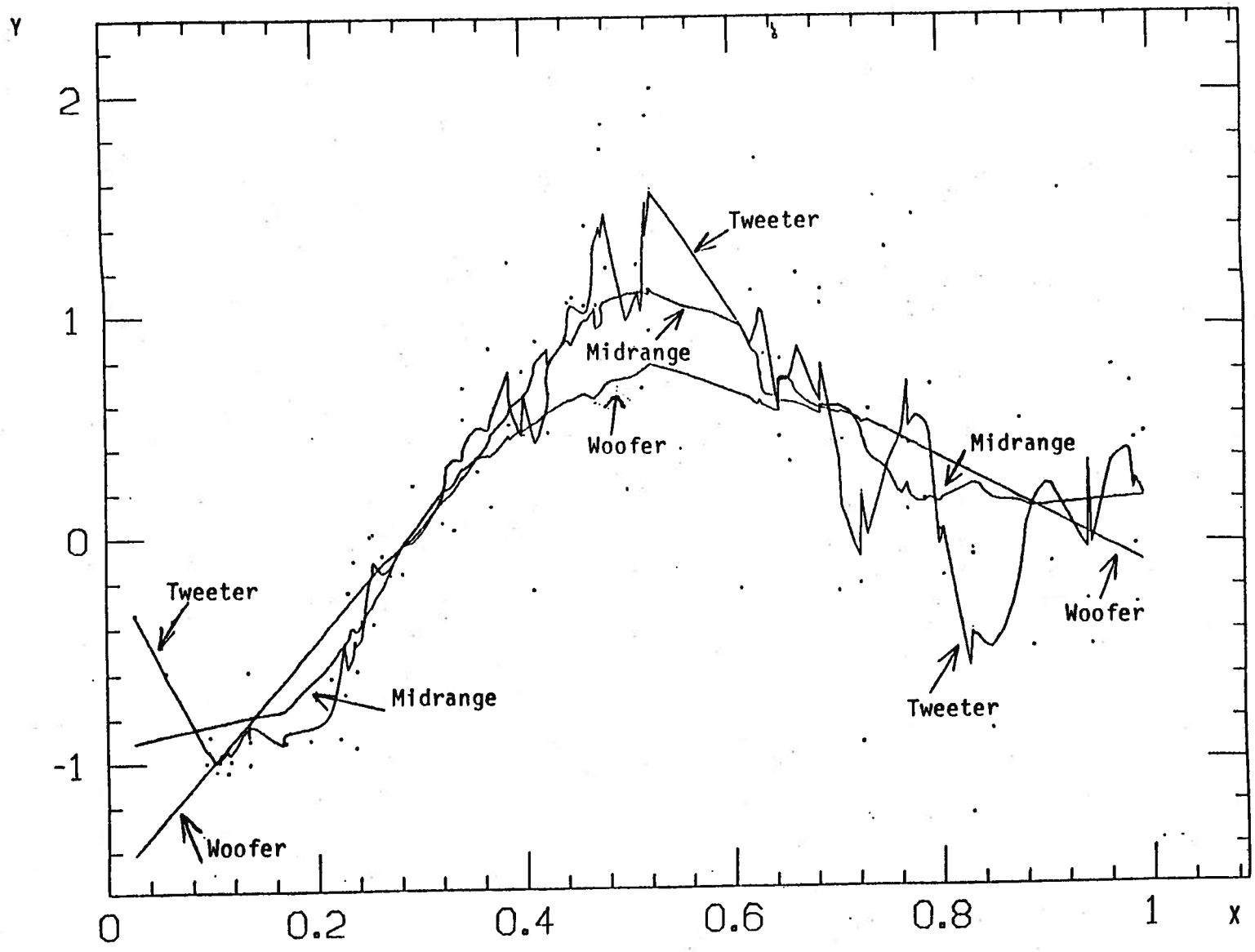

Y
2
1
0
-1
0
0.2
0.4
0.6
0.8
1
X
Tweeter
Midrange
Woofer
Midrange
Tweeter
Woofer
Tweeter
Midrange
Woofer
Tweeter

FIGURE 3b

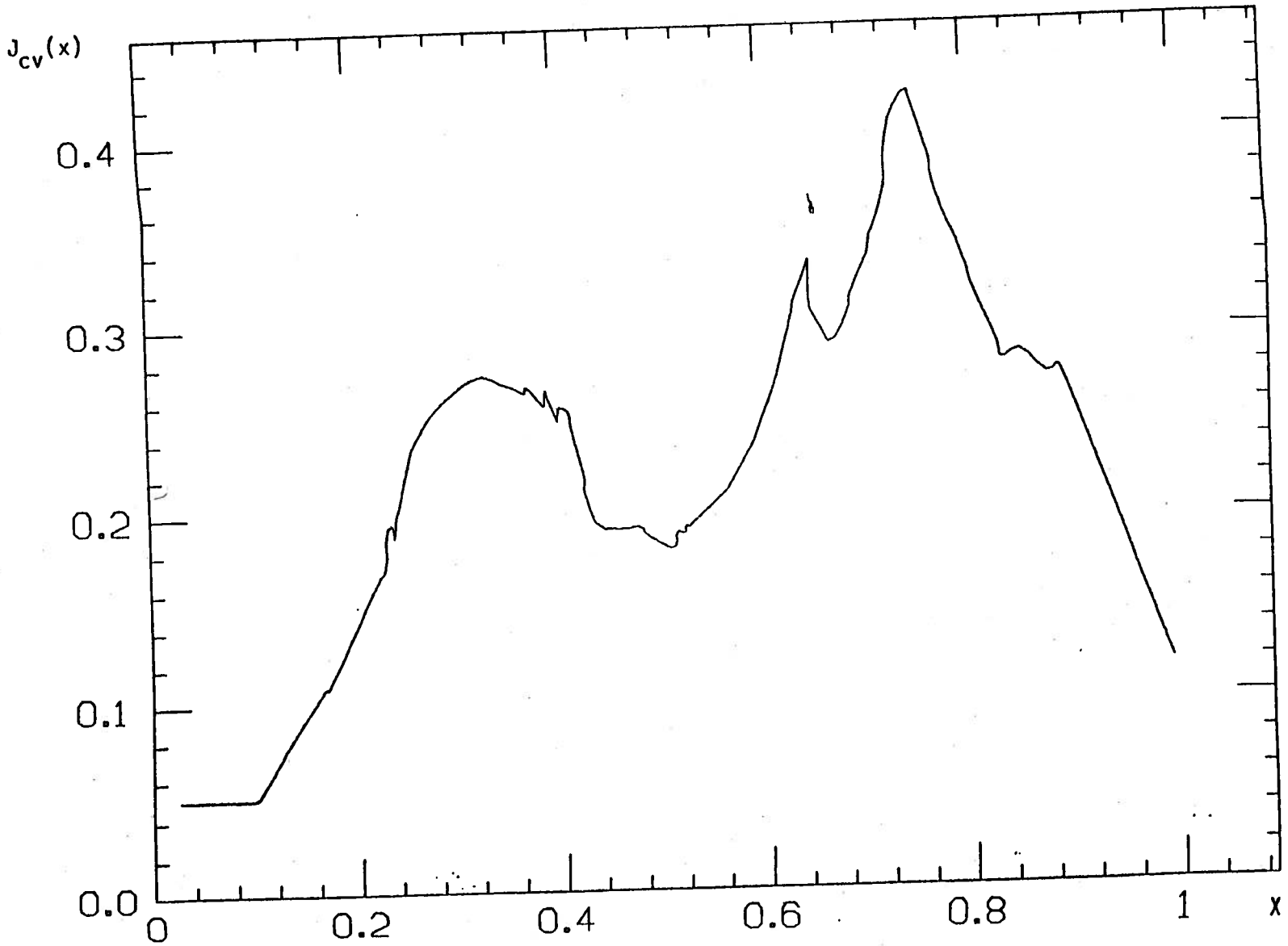


FIGURE 3c

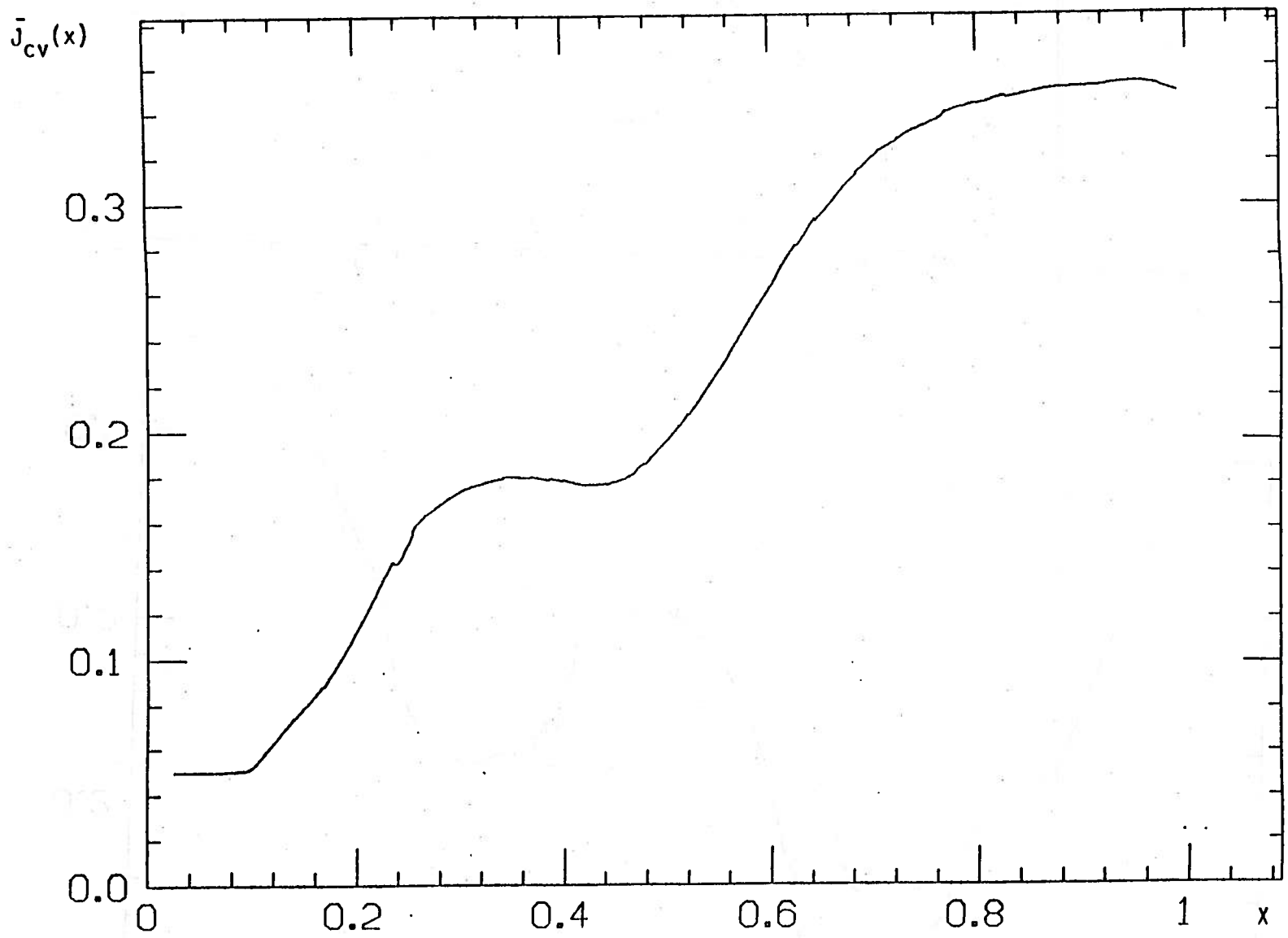


FIGURE 3d

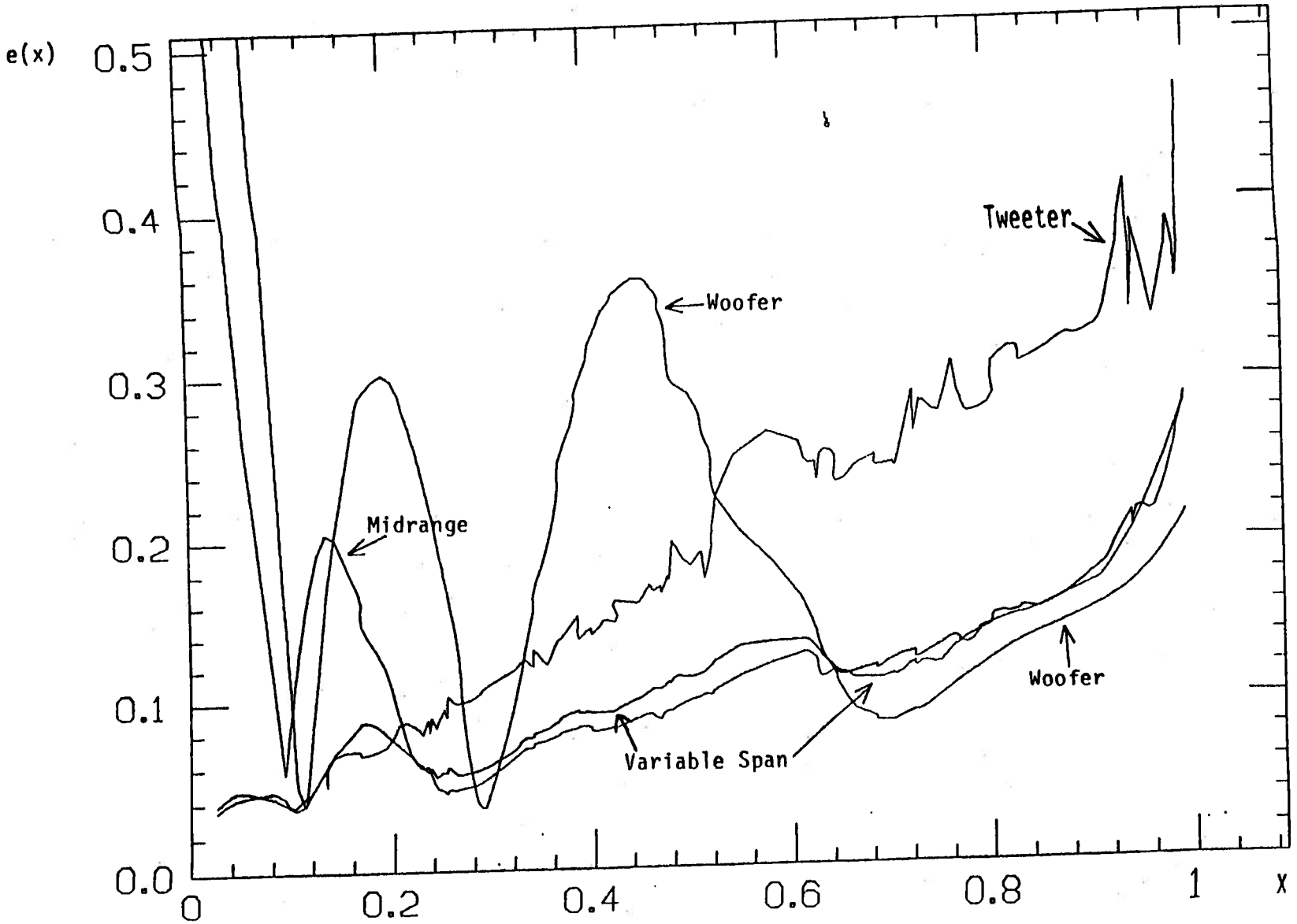

e(x)
0.5
0.4
0.3
0.2
0.1
0.0
0
0.2
0.4
0.6
0.8
1
X
Tweeter
Woofer
Midrange
Woofer
Variable Span

FIGURE 3e

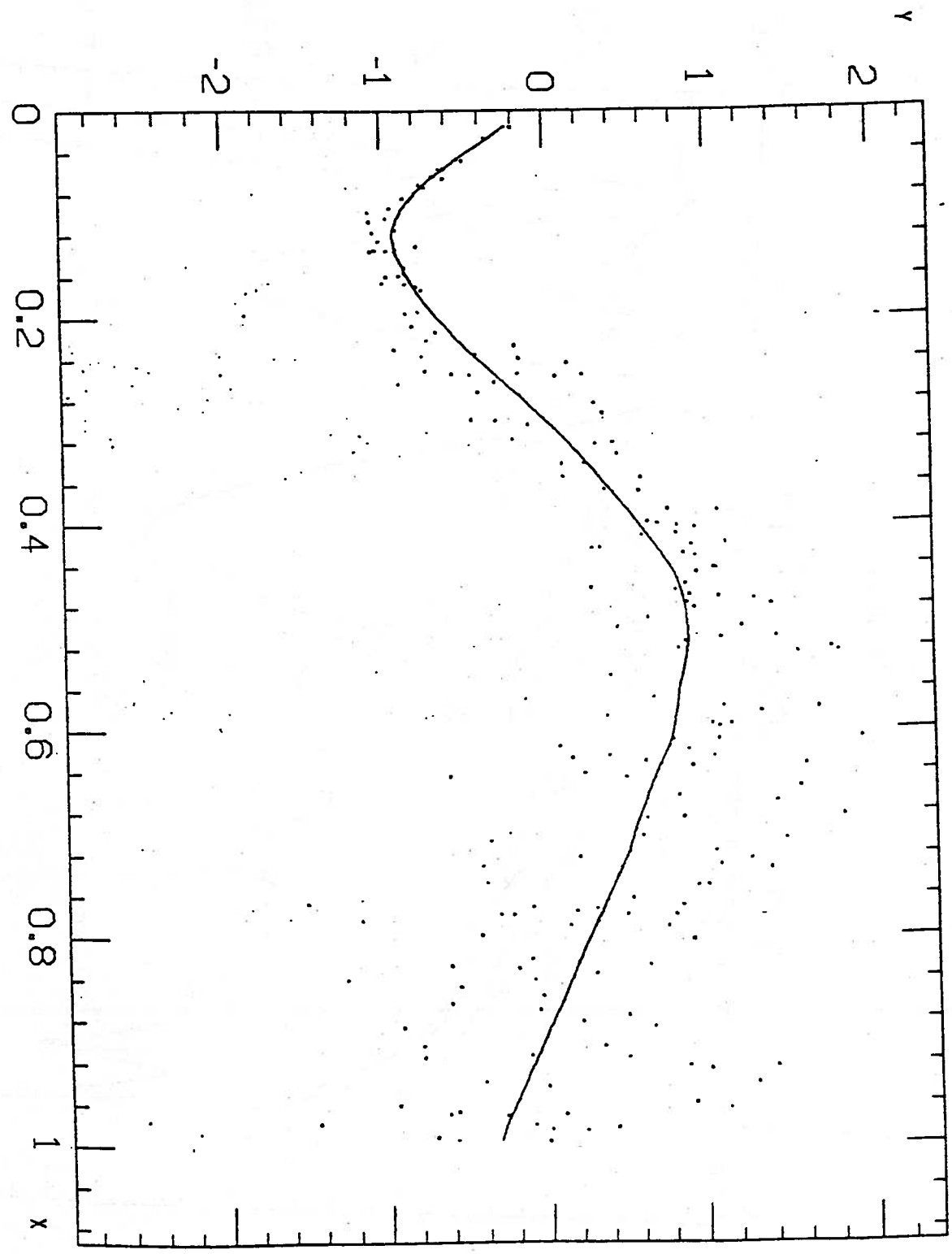


FIGURE 4a

e(x)
0.3
0.2
0.1
0.0
0
0.2
0.4
0.6
0.8
1
x
Woofer
Variable Span
Midrange
Midrange
Variable Span
Tweeter

FIGURE 4b

FIGURE 4c