CR 134311

N74-27668

(NASA-CR-134311) ERROR DETECTION AND DATA SMOOTHING BASED ON LOCAL PROCEDURES (Rice Univ.) 61 p HC $6.25 CSCL 09B

Unclas
G3/08 43418

ICSA

INSTITUTE FOR COMPUTER SERVICES AND APPLICATIONS

RICE UNIVERSITY

275-025-012

Error Detection & Data Smoothing Based on Local Procedures

by

Victor M. Guerra
Dept. of Mathematical Sciences
Rice University

ABSTRACT

This thesis presents an algorithm which is able to locate isolated bad points and correct them without contaminating the rest of the good data. This work has been greatly influenced and motivated by what is currently done in the manual loft. It is not within the scope of this work to handle small random errors characteristic of a noisy system, and it is therefore assumed that the bad points are isolated and relatively few when compared with the total number of points.

Motivated by the desire to imitate the loftsman we conducted a visual experiment to determine what is considered smooth data by most people. This criterion is used to determine how much the data should be smoothed and to prove that our method produces such data. The method ultimately converges to a set of points that lies on the polynomial that interpolates the first and last points; however convergence to such a set is definitely not the purpose of our algorithm. The proof of convergence is necessary to demonstrate that oscillation does not take place and that in a finite number of steps the method produces a set as smooth as desired.

The amount of work for the method described here is of order n. The one dimensional and two dimensional cases are treated in detail; the theory can be readily extended to higher dimensions.

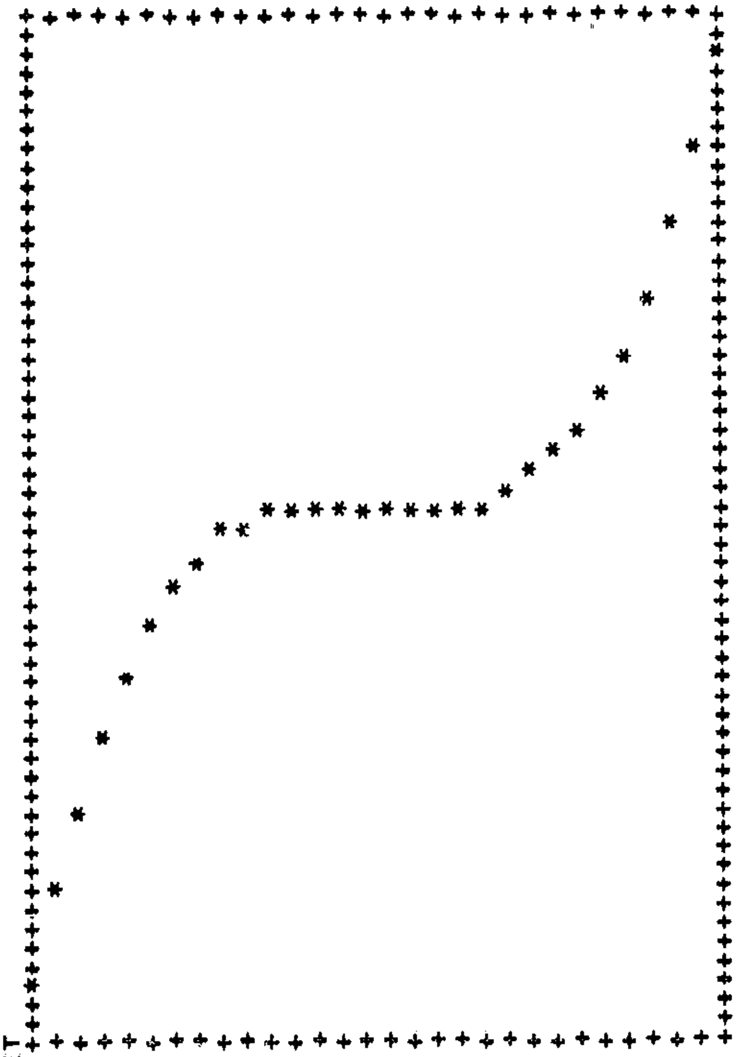
Institute for Computer Services & Applications
Rice University
Houston, Texas 77001

May, 1974

Research supported in part under NASA Contract NAS 9-12776

CUBIC

K = 2   θ = f(θ)   SMOOTHNESS = 0.02   IT = 12

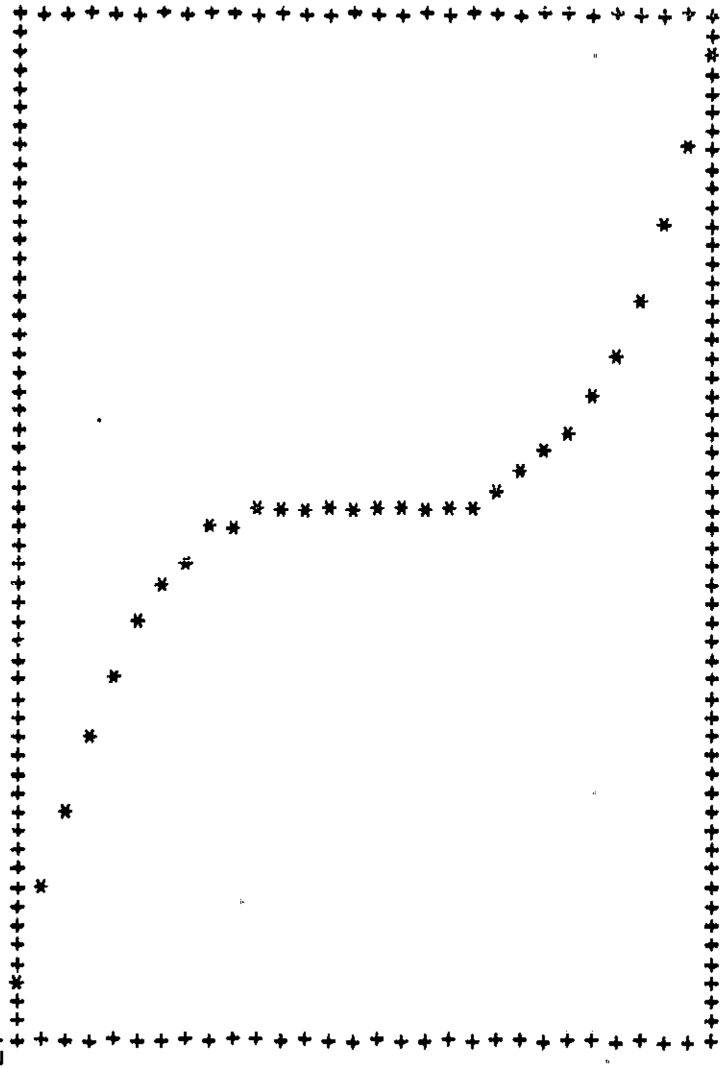


IT=	APPROX	SMOOTHED SET
1	0.0	0.78E+00
2	0.0	0.17E+01
3	0.0	0.25E+01
4	0.0	0.32E+01
5	0.0	0.38E+01
6	0.0	0.43E+01
7	0.0	0.47E+01
8	0.19E-05	0.50E+01
9	0.0	0.52E+01
10	0.0	0.53E+01
11	0.0	0.54E+01
12	0.0	0.55E+01
13	0.0	0.55E+01
14	0.0	0.55E+01
15	0.0	0.55E+01
16	0.0	0.55E+01
17	0.0	0.55E+01
18	0.0	0.54E+01
19	0.12E-01	0.55E+01
20	0.12E-01	0.55E+01
21	0.0	0.56E+01
22	0.0	0.57E+01
23	0.0	0.59E+01
24	0.0	0.62E+01
25	0.0	0.65E+01
26	0.19E-05	0.70E+01
27	0.0	0.75E+01
28	0.0	0.82E+01
29	0.0	0.90E+01
30	0.0	0.99E+01

FIGURE 15

CUBIC

K = 2   θ = 0.5   SMOOTHNESS = 0.2   IT = 26

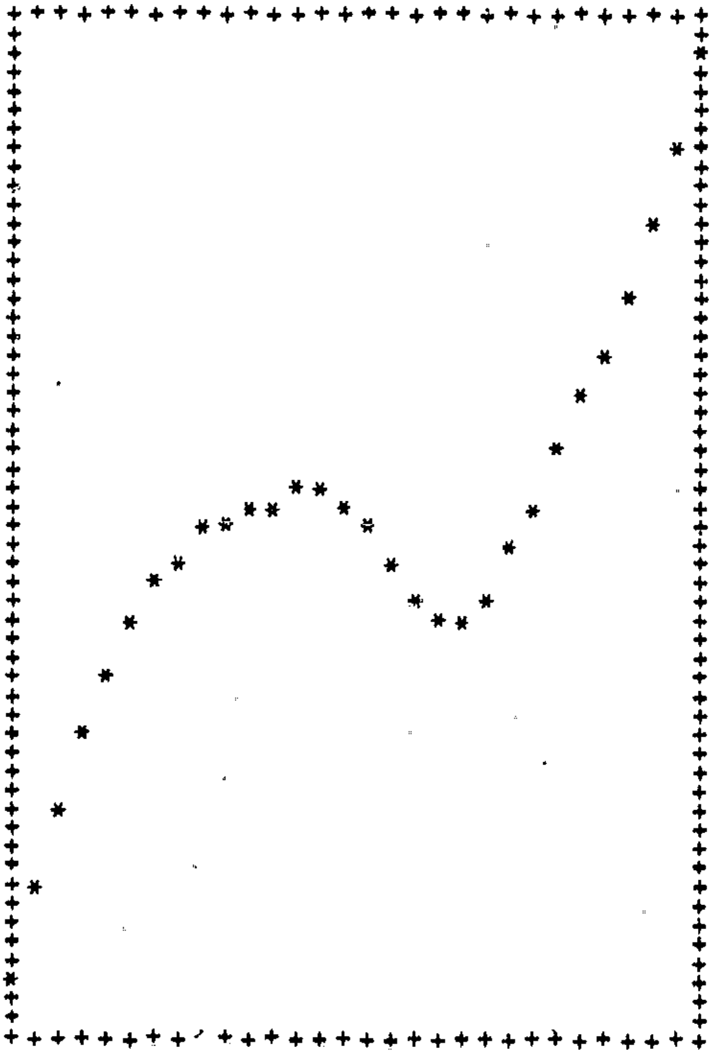


IT=	APPROX	SMOOTHED SET
1	0.0	0.78E+00
2	0.0	0.17E+01
3	0.0	0.25E+01
4	0.0	0.32E+01
5	0.0	0.38E+01
6	0.0	0.43E+01
7	0.0	0.47E+01
8	0.20E-01	0.50E+01
9	0.0	0.52E+01
10	0.0	0.53E+01
11	0.0	0.54E+01
12	0.0	0.55E+01
13	0.0	0.55E+01
14	0.0	0.55E+01
15	0.0	0.55E+01
16	0.0	0.55E+01
17	0.0	0.55E+01
18	0.0	0.54E+01
19	0.12E-01	0.55E+01
20	0.12E-01	0.55E+01
21	0.0	0.56E+01
22	0.0	0.57E+01
23	0.0	0.59E+01
24	0.0	0.62E+01
25	0.0	0.65E+01
26	0.20E-01	0.70E+01
27	0.0	0.75E+01
28	0.0	0.82E+01
29	0.0	0.90E+01
30	0.0	0.99E+01

FIGURE 16

CUBIC

K = 2   θ = 1.0   SMOOTHNESS = 0.02   IT = 62

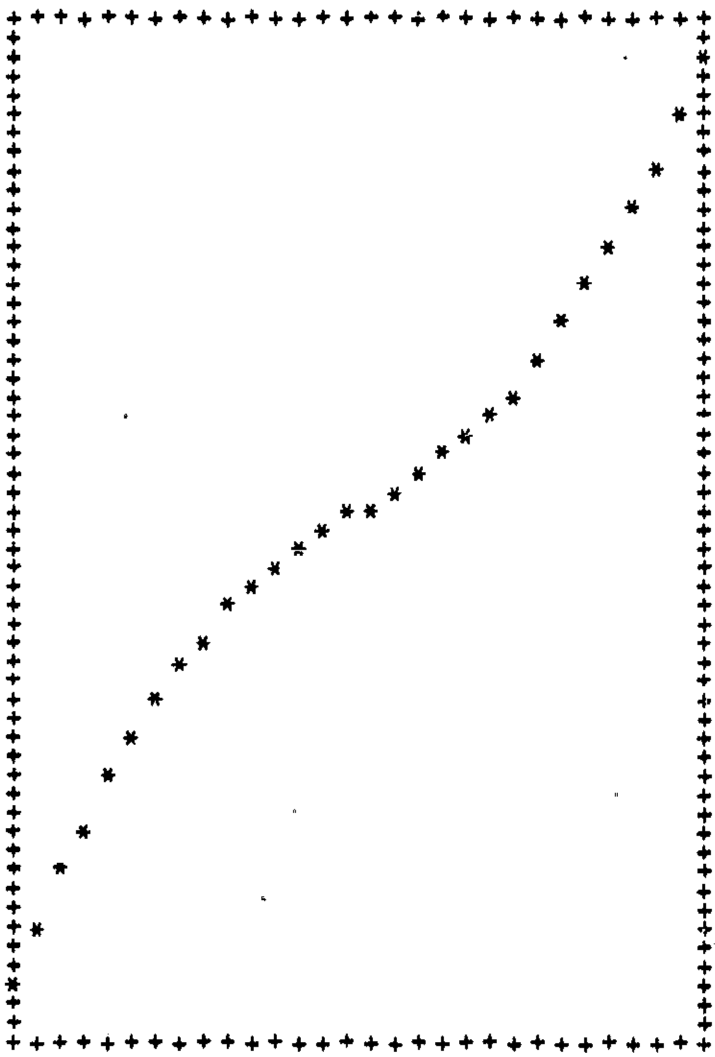


IT=	APPROX	SMOOTHED SET
1	0.0	0.78E+00
2	0.0	0.17E+01
3	0.0	0.25E+01
4	0.0	0.32E+01
5	0.0	0.38E+01
6	0.0	0.43E+01
7	0.0	0.47E+01
8	0.19E-05	0.50E+01
9	0.0	0.52E+01
10	0.0	0.53E+01
11	0.0	0.54E+01
12	0.0	0.55E+01
13	0.22E-01	0.55E+01
14	0.27E-01	0.55E+01
15	0.42E-01	0.55E+01
16	0.25E+00	0.52E+01
17	0.56E+00	0.49E+01
18	0.91E+00	0.45E+01
19	0.12E+01	0.43E+01
20	0.12E+01	0.43E+01
21	0.11E+01	0.45E+01
22	0.74E+00	0.50E+01
23	0.39E+00	0.55E+01
24	0.14E+00	0.61E+01
25	0.27E-01	0.65E+01
26	0.19E-05	0.70E+01
27	0.0	0.75E+01
28	0.0	0.82E+01
29	0.0	0.90E+01
30	0.0	0.99E+01

FIGURE 17

LINEAR

K = 1   θ = 0.5   SMOOTHNESS = 0.02   IT = 390



IT=	APPROX	SMOOTHED SET
1	0.0	0.78E+00
2	0.39E+00	0.13E+01
3	0.68E+00	0.19E+01
4	0.89E+00	0.23E+01
5	0.10E+01	0.28E+01
6	0.11E+01	0.32E+01
7	0.11E+01	0.36E+01
8	0.10E+01	0.39E+01
9	0.97E+00	0.42E+01
10	0.85E+00	0.45E+01
11	0.73E+00	0.47E+01
12	0.55E+00	0.49E+01
13	0.40E+00	0.51E+01
14	0.24E+00	0.53E+01
15	0.89E-01	0.54E+01
16	0.23E-01	0.55E+01
17	0.16E+00	0.56E+01
18	0.32E+00	0.58E+01
19	0.49E+00	0.59E+01
20	0.66E+00	0.62E+01
21	0.79E+00	0.64E+01
22	0.90E+00	0.66E+01
23	0.99E+00	0.69E+01
24	0.10E+01	0.72E+01
25	0.10E+01	0.76E+01
26	0.98E+00	0.80E+01
27	0.86E+00	0.84E+01
28	0.66E+00	0.88E+01
29	0.38E+00	0.93E+01
30	0.0	0.99E+01

FIGURE 18

Example 7.5. In this example we apply the (k,h)-order smoothing algorithm to the set

$$z(x,y) = \exp(-(x-10)^2 - (y-20)^2) \qquad x,y=1,2,\ldots,30.$$

For $k=h=2$ this set is 0.0001-smooth. We introduce errors to 90 points on the set. Figure 19 represents this set, the asterisk represents an error, each error is equal to 12.0 cm. . Table 3 represents a list of the points that have errors and the sign of the error. Figure 20 represents the smoothed set. After 363 iterations we get the desired smoothness of 0.02 cm. . The mean of the difference of the points in the smoothed set and the original set (equation (7.2)) is 0.0004 cm. .

CONTAMINATED DATA WITH 66 ERRORS

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	*	0	0	0	0	0	0	0	0	0	0	0	0	0	*	*	0	0	0	0	0	0	0
4	0	0	0	*	0	0	*	0	*	0	0	*	0	0	*	0	*	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	*	*	0	0	0	0	0	*	0	0	0	*	0	0	0	0	*	0	0	0	0	0	0	0
6	0	0	0	0	0	*	0	0	*	0	0	0	0	0	0	0	*	0	0	0	0	0	0	0	*	0	0	0	0	0
7	0	0	0	0	*	0	0	0	0	0	0	*	1	1	1	1	1	*	1	1	1	1	*	1	1	1	1	0	0	0
8	0	0	0	0	0	0	0	*	0	1	1	1	1	1	1	1	*	1	1	*	1	1	1	*	1	1	1	1	1	1
9	0	0	0	*	0	0	0	1	1	1	1	1	1	1	2	2	2	2	*	2	2	2	2	2	2	1	*	1	1	1
10	0	0	0	0	*	1	1	*	1	1	1	2	2	2	2	*	2	2	2	2	2	2	2	2	2	*	2	2	1	1
11	0	0	0	1	1	1	1	1	1	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2
12	0	0	1	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	4	3	3	3	3	2	2
13	0	1	1	1	1	1	2	2	*	3	3	3	4	4	4	4	4	5	5	5	5	5	4	4	4	4	*	3	3	3
14	0	1	1	1	1	2	2	2	3	3	4	4	4	5	5	5	*	5	5	6	5	5	*	5	5	5	4	4	4	3
15	1	1	1	1	2	2	2	3	3	4	4	4	5	*	5	6	6	6	6	6	6	6	*	6	5	5	5	4	4	4
16	1	1	1	2	2	2	3	3	*	4	5	5	5	6	6	6	7	7	7	7	7	7	7	6	*	6	5	5	5	4
17	1	1	1	*	2	3	3	3	4	4	5	*	6	6	7	7	7	7	8	8	8	7	*	7	7	6	6	5	5	4
18	1	1	2	2	2	3	3	*	4	5	5	6	6	7	7	7	*	8	8	*	8	8	8	*	7	7	6	6	5	5
19	1	1	2	2	2	3	3	4	4	5	5	6	6	7	*	8	8	8	*	8	8	8	8	8	7	7	6	6	5	5
20	1	1	2	2	2	3	3	*	4	5	6	6	7	7	7	*	8	8	8	9	*	8	8	8	7	*	7	6	6	5
21	1	1	2	2	2	3	3	4	4	5	5	6	6	7	7	8	8	8	8	8	8	8	8	8	7	7	6	6	5	5
22	1	1	2	2	2	3	3	4	4	5	5	6	6	7	7	7	8	8	8	8	8	8	8	7	7	7	6	6	5	5
23	1	1	1	2	2	3	3	*	*	4	5	5	6	6	7	7	7	7	8	*	8	7	7	7	7	6	6	5	5	4
24	1	1	1	*	2	2	3	3	4	4	5	*	5	6	6	6	7	7	7	7	7	7	*	*	6	6	5	5	5	4
25	1	1	1	1	2	2	*	3	3	4	4	4	5	5	5	6	6	*	6	6	6	6	6	6	5	5	5	4	4	4
26	0	1	1	1	1	*	2	2	3	3	4	4	4	5	5	5	*	5	5	6	5	5	5	5	*	5	4	4	4	3
27	0	1	1	1	*	1	2	2	2	3	*	3	4	4	*	4	4	*	5	5	*	5	4	4	4	4	4	3	3	3
28	0	0	1	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	4	3	3	3	3	2	2
29	0	0	0	1	1	1	1	1	1	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2
30	0	0	0	0	0	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1

FIGURE 19

SMOOTHED DATA

K = 2 θ = 3/4 SMOOTHNESS = 0.02 IT = 363

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
8	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	0	0	0	0	0	0	0	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1
10	0	0	0	0	0	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1
11	0	0	0	1	1	1	1	1	1	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2
12	0	0	1	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	4	3	3	3	3	2	2
13	0	1	1	1	1	1	2	2	2	3	3	3	4	4	4	4	4	5	5	5	5	5	4	4	4	4	4	3	3	3
14	0	1	1	1	1	2	2	2	3	3	4	4	4	5	5	5	5	5	5	6	5	5	5	5	5	5	4	4	4	3
15	1	1	1	1	2	2	2	3	3	4	4	4	5	5	5	6	6	6	6	6	6	6	6	6	5	5	5	4	4	4
16	1	1	1	2	2	2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	7	7	7	6	6	6	5	5	5	4
17	1	1	1	2	2	3	3	3	4	4	5	5	6	6	7	7	7	7	8	8	8	7	7	7	7	6	6	5	5	4
18	1	1	2	2	2	3	3	4	4	5	5	6	6	7	7	7	8	8	8	8	8	8	8	7	7	7	6	6	5	5
19	1	1	2	2	2	3	3	4	4	5	5	6	6	7	7	8	8	8	8	8	8	8	8	8	7	7	6	6	5	5
20	1	1	2	2	2	3	3	4	4	5	6	6	7	7	7	8	8	8	8	9	8	8	8	8	7	7	7	6	6	5
21	1	1	2	2	2	3	3	4	4	5	5	6	6	7	7	8	8	8	8	8	8	8	8	8	7	7	6	6	5	5
22	1	1	2	2	2	3	3	4	4	5	5	6	6	7	7	7	8	8	8	8	8	8	8	7	7	7	6	6	5	5
23	1	1	1	2	2	3	3	3	4	4	5	5	6	6	7	7	7	7	8	8	8	7	7	7	7	6	6	5	5	4
24	1	1	1	2	2	2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	7	7	7	6	6	6	5	5	5	4
25	1	1	1	1	2	2	2	3	3	4	4	4	5	5	5	6	6	6	6	6	6	6	6	6	5	5	5	4	4	4
26	0	1	1	1	1	2	2	2	3	3	4	4	4	5	5	5	5	5	5	6	5	5	5	5	5	5	4	4	4	3
27	0	1	1	1	1	1	2	2	2	3	3	3	4	4	4	4	4	5	5	5	5	5	4	4	4	4	4	3	3	3
28	0	0	1	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	4	3	3	3	3	2	2
29	0	0	0	1	1	1	1	1	1	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2
30	0	0	0	0	0	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1

FIGURE 20

I	J	$Z_{I,J}$
-5	23	-9.315
-15	23	-3.505
6	6	10.372
26	6	12.253
6	9	10.542
16	9	14.105
-6	17	-9.051
-26	17	-4.261
6	25	10.876
16	25	16.634
7	5	10.436
27	5	11.684
7	12	10.976
17	12	15.906
7	18	11.318
27	18	15.083
7	23	11.285
17	23	17.775
8	8	10.867
18	8	14.188
8	17	11.703
18	17	18.225
-8	20	-8.219
-18	20	-1.396
-8	24	-8.356
-18	24	-2.058
9	4	10.641
19	15	17.854
9	19	12.296
19	19	18.855
9	27	11.806
-10	5	-9.051
24	23	17.187
-10	26	-7.559
-20	26	-2.483
-10	8	-8.578
-20	8	-5.619
-10	16	-7.303
-20	16	-1.692

I	J	$Z_{I,J}$
-5	7	-9.692
5	8	10.349
-3	8	-9.830
-23	8	-6.041
-13	9	-7.168
-26	25	-4.703
24	24	16.939
-20	21	-1.045
27	21	15.160
-17	4	-7.739
27	11	13.459
-27	15	-5.423
13	27	14.059
23	9	14.441
23	20	18.133
3	22	10.342
33	22	10.333
4	4	10.140
24	4	12.090
-4	7	-9.783
4	9	10.276
-4	12	-9.633
-24	12	-4.541
4	15	10.446
14	23	15.739
4	17	10.483
14	17	15.739
-5	7	-19.692
-25	7	-7.082
-5	14	-9.402
-15	14	-4.326
-5	18	-9.298
-25	18	-3.341

TABLE 3

APPENDIX 1

In this section we describe the experiment used to find the lower bound of the error detectabl- by human beings. From experimental psycology we take the next definition and assumption: The differential threshold is defined as the point at which a difference can be detected 75% of the time. We assume that the data obtained from this experiment behaves like a normal distribution function with mean equal to zero.

In this experiment each subject is going to look at 88 targets that are divided in two equall groups. The first group is shown at 4ft. from the subject and the secong group is shown at 16ft. from the subject. Each target is a $8\frac{1}{2}$" x 11" white cardboard infront of a white wall, this targets contain 5 or 6 dots centered in the cardboard so they form a line or a semi-circle respectively. The dots are 1/16" in diameter which is above the minimum size to be recognizable. The targets are shown one by one to each subject and there is no time limit to look at them. All the people that took this test claim to have normal sight (wearing glases if necessary).

In the linear pattern one of the two center points is above the line that interpolates the rest, so that every subject dicides whether the left or right center point contains the error. The case when point with error is under the the interpolating line is not considered since it is assumed to be symmetrical to our case. And in the semicircular pattern all the points belong to a semicircle (105°) enxcept the center point which is either above or below the semicircle. In the cardboards the ordinates are 1" apart and the semicircle has 3" of radius. In Figure 21 we represent these patterns.

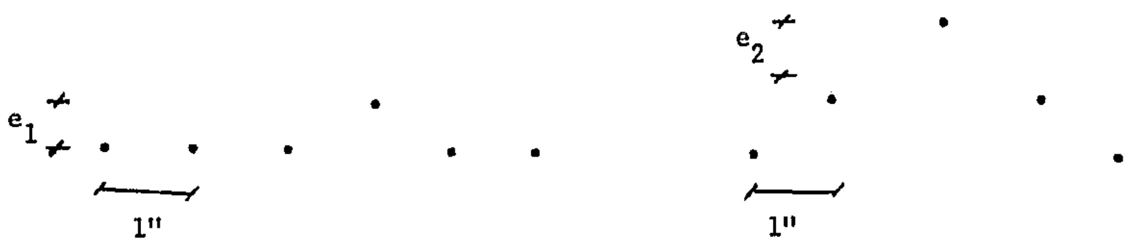
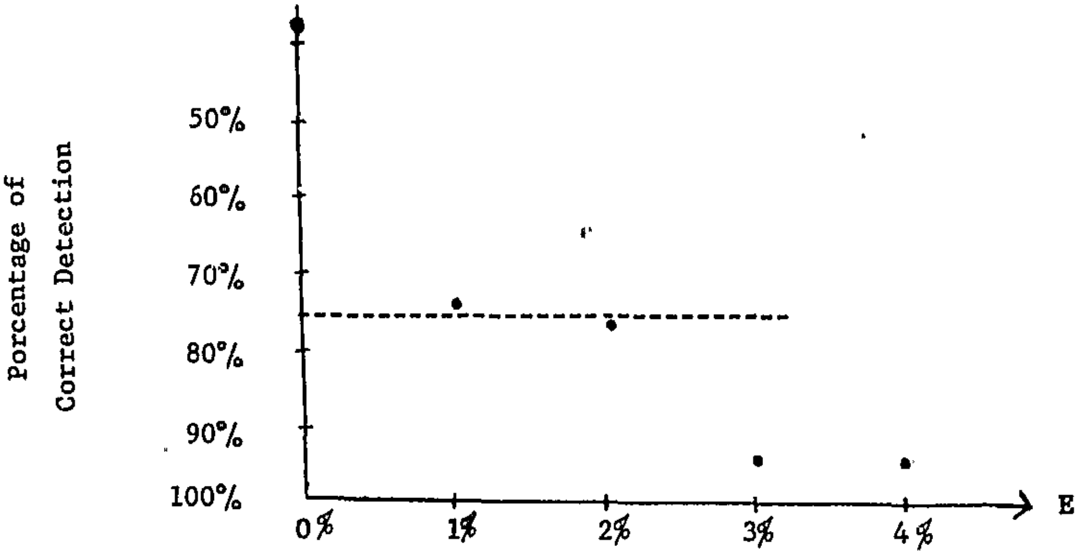


FIGURE 21

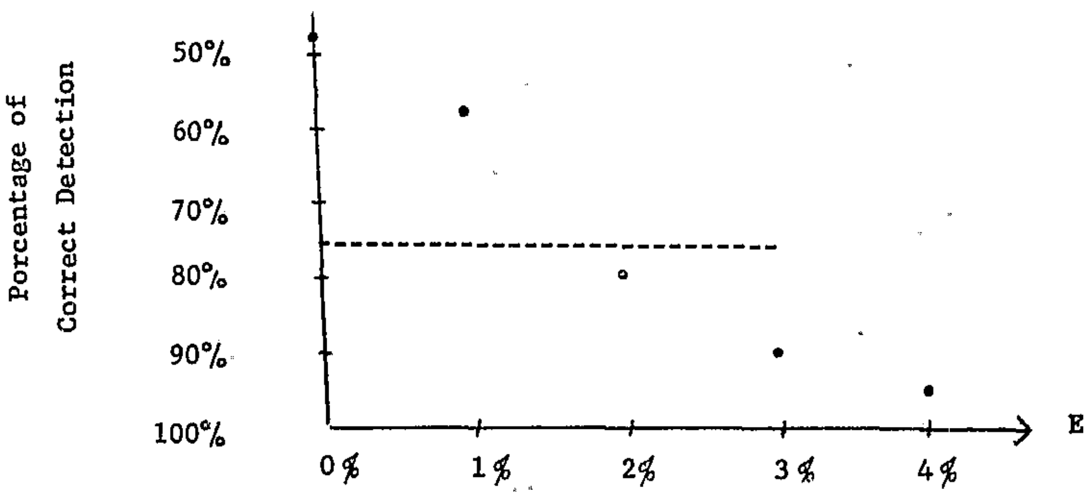
The quantities e_1 and e_2 represent the distance between the points with error and the line or semicircle that interpolates the rest of the points. The distance e_2 may be positive or negative.

From the preliminary tests we learned that was easier to detect an error in a line than in a circle, therefore the variations in the errors for the line are smaller than those of the semicircle. Let these variations be e_1= 0.0, 1/100. 2/100, 3/100 and 4/100 of an inch and e_2= -10/100, -8/100, -6/100, -4/100, -2/100, 0.0, 2/100, 4/100, 6/100, 8/100, 10/100 and 12/100 of an inch. The zero errors are introduced to give consistency to the experiment, that is, we expect half of the people to give one answer and the other half to give the opposite answer in these cases. The targets will be ordered randomly.

The results for the line case are sumarizad in the next graphs.



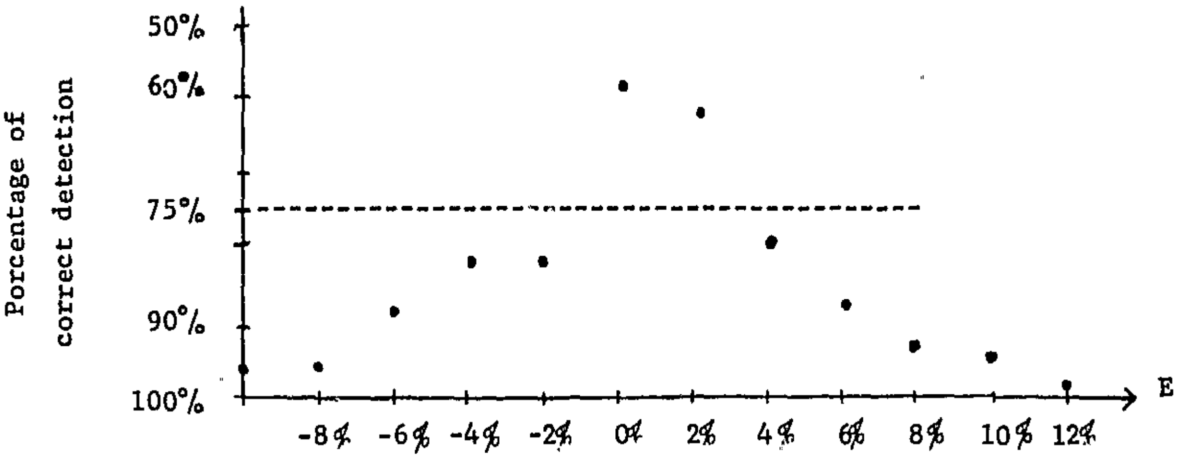
Porcentage of the error at 4 ft.



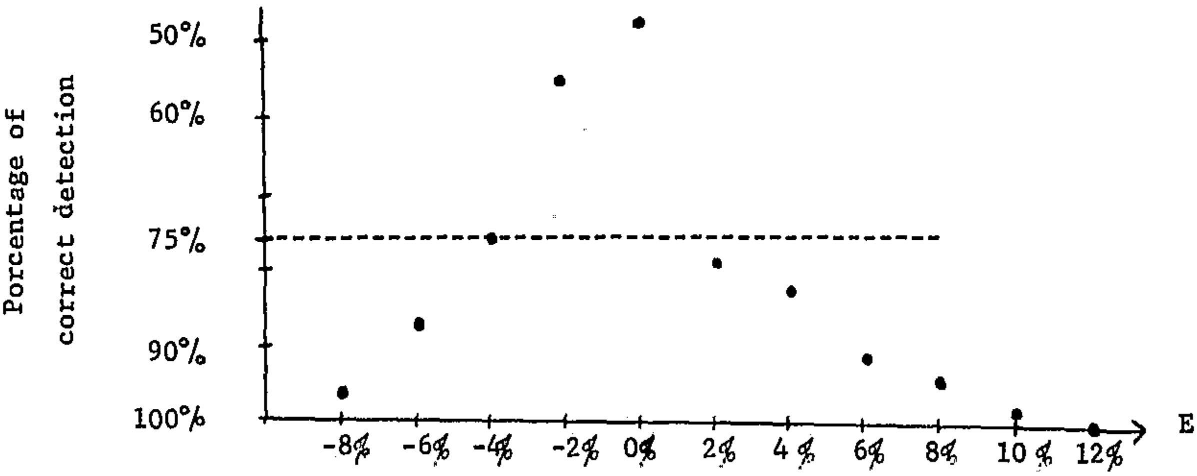
Porcentage of the error at 16 ft.

The quantity E represents the porcentage of the error, that is, the quotient between the point error and the spacing of the points.

The results for the semicircle experiment are sumarized in the next graphs.



Porcentage of the error at 4 ft.



Porcentage of the error at 16 ft.

The porcentage of the error is possitive when the bad point is above the semicircle and is negative if the bad point is under the semicircle.

To interpolate a normal density function with mean equal to zero we compute the variances, then we find the error that corresponds to the differential threshold (i.e., 75% of correct detection).

	LINE	LINE	CIRCLE	CIRCLE
DISTANCE	4	16	4	16
MEAN	0	0	0	0
VARIANCE	2.14	2.98	1.7	1.6
E	1.0%	1.1%	2.8%	2.7%

This result means that the human eye is capable of recognizing errors with respect to the spacing of the points. Also it gives a lower bound below which it is impossible to decide whether one point is perfectly smooth or not, and an upper bound above which we can recognize rough points.

APPENDIX 2

Here we present a table with the coefficients of the first eight divided differences when the order is even.

n-th difference	k	... y_{i-4}	y_{i-3}	y_{i-2}	y_{i-1}	y_i	y_{i+1}	y_{i+2}	y_{i+3}	y_{i+4}
0	0	0	0	0	0	1	0	0	0	0
2	1	0	0	0	1	-2	1	0	0	0
4	2	0	0	1	-4	6	-4	1	0	0
6	3	0	1	-6	15	-20	15	-6	1	0
8	4	1	-8	28	-56	70	-56	28	-8	1
.	.	.	.	.	.	.	.	.	.	.
n	k=2n	.	.	.	$(-1)^n \frac{(-1)^j (2n)!}{(n-j)!(n+j)!}$			.	.	.

REFERENCES

1. M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions, National Bureau of Standards applied Mathematics Series 55 (1970).
2. S.A. Berger, W.C. Wbster, R.A. Tapia and D.A. Atkins, Mathematical Ship Lofting, Jour. of Ship Research, 10 (1966), 203-214.
3. S.H. Bartley, Principles of Perception, Harper & Row Publishers, New York, 1969.
4. G.E.P. Box and G.M. Jenkins, Time Series Analysis Forecasting and Control, Holden-Day, San Francisco, 1970.
5. T.N.E. Greville, Introduction to Spline Functions. T.N.E. Greville Editor, Academic Press, New York, 1969 pp.1-135.
6. P. La Fata and J.B. Rosen, An Interactive Display for Approximation by Linear Programing, Comm. of A.C.M., 13 (1970), 651-659.
7. C. Reinsch, Smoothing by Spline Functions, Numer. Math., 10 (1967), 177-183.
8. C. Reinsch, Smoothing by Spline Functions II, Numer. Math., 16 (1971), pp.451-454.
9. I.J. Schoenberg, Spline Functions and the Problem of Graduation, Proc. Nat. Acad. Sci., U.S.A. 52 (1964), pp.947-950.
10. F. Theilheimer and W. Starkweather, The Fairing of Ship Lines on High-Speed computer, David Taylor Model Basin, Applied Mathematics Laboratory Report 1474, Jan 1961.