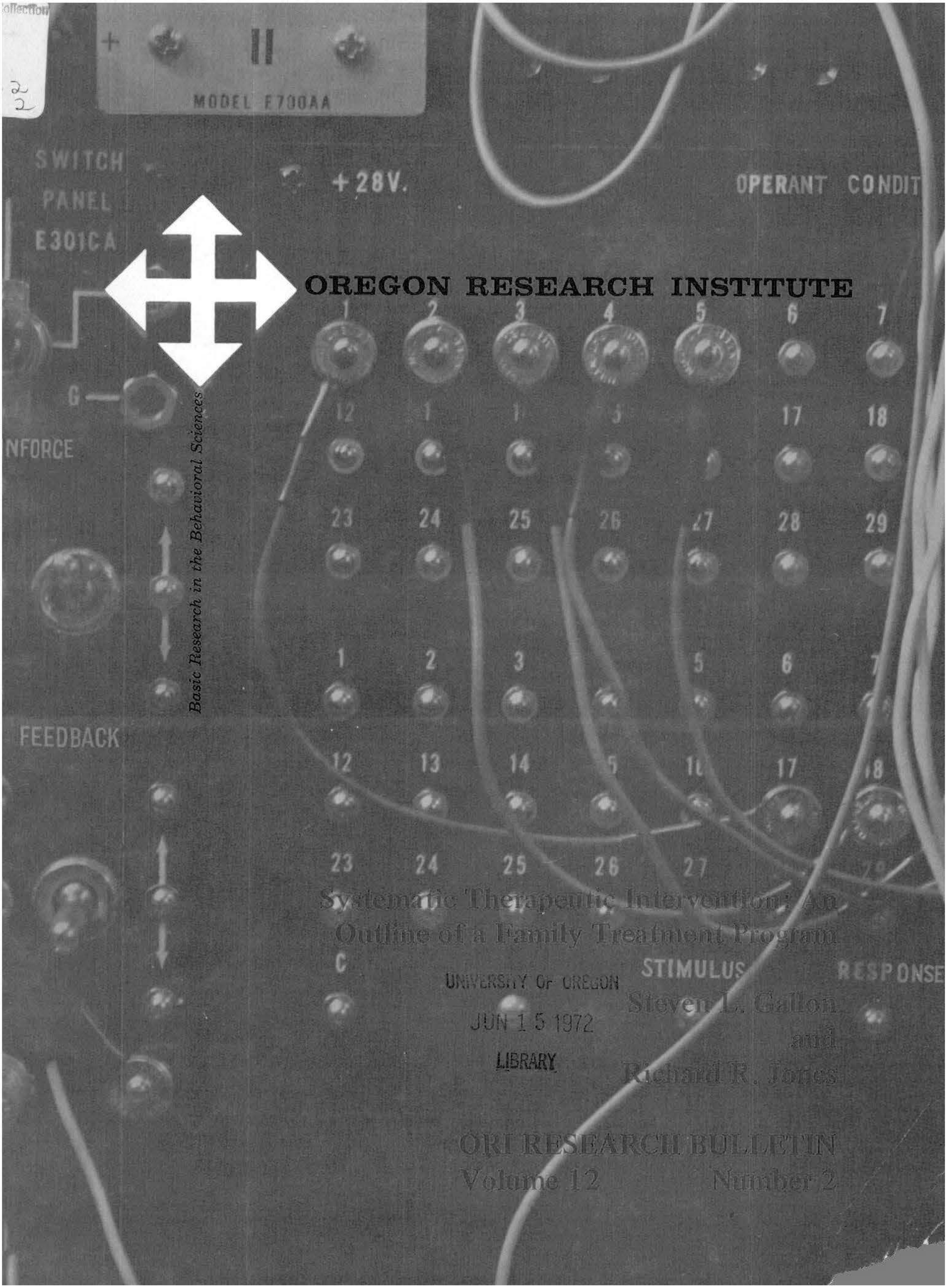


OREGON RESEARCH INSTITUTE

Basic Research in the Behavioral Sciences

Systematic Therapeutic Intervention: An Outline of a Family Treatment Program

Steven L. Gallon
and
Richard R. Jones

ORI RESEARCH BULLETIN
Volume 12 Number 2

OREGON RESEARCH INSTITUTE

Systematic Therapeutic Intervention: An Outline of a Family Treatment Program

Steven L. Gallon and Richard R. Jones

Oregon Research Institute
Eugene, Oregon

Research Bulletin
Vol. 12 No. 2

March 1972

Systematic Therapeutic Intervention: An Outline of a Family Treatment Program[1]

Steven L. Gallon and Richard R. Jones

Oregon Research Institute

As part of the Social Learning Project at Oregon Research Institute, there has evolved over the past several years a set of procedures which has been systematically applied in the treatment of families of aggressive boys. The therapeutic techniques are based on the body of psychological principles known as social learning theory (e.g., Skinner, 1953; Gewirtz, 1969). In a classification of clinical therapies, these treatment techniques would be appropriately categorized under the rubric of behavior modification.

Heretofore the clinical and research findings from this extensive project have been presented in several monographs (e.g., Patterson, Cobb, & Ray, 1972; Patterson, 1969; Patterson, Ray, & Shaw, 1968) which have discussed in varying detail the actual activities undertaken in the treatment of disturbed families. There has not been available, however, in one easily accessible source, a concise outline of the sequence of steps in the treatment process and a description of the activities involved at each stage of family intervention. This report is an attempt to fill this need for an overview of the therapeutic procedures developed in the Social Learning Project. Hopefully, this overview is presented in a manner such that the reader can rapidly gain an appreciation for the detail, complexity, and orderliness of these treatment activities.

Components of Family Treatment

The system analysis of the family treatment program shown in Figure 1 outlines the numerous steps employed by the Social Learning Project in treating families of deviant boys. These treatment stages, and the acocmpanying description in the text, apply to all families treated by the project, although minor exceptions to these techniques obviously, and inevitably, occur.

The treatment system involves different participants--therapists, family, and observers--shown in three columns in Figure 1. For each participant, the treatment activities are presented in sequential flow, as they occur during the course of the therapeutic process. By reading down the entries in all three columns, the steps are revealed as they occur in time. The dotted lines linking boxes between columns indicate interactions between participant systems. The approximate amount of time required to complete the different phases is referenced in the left-most column.

Description of Treatment Stages

The treatment steps depicted in the System Analysis (Figure 1, fold-outs) have been combined into stages and explained more fully below. Each stage is identified by a title phrase plus the codes corresponding to the steps in Figure 1. It should be noted that families accepted into the project during 1972 will experience a somewhat revised version of the treatment procedures employed prior to that time. The changes known at this writing are footnoted in Figure 1 and described more fully in the text below.

Steps leading to the intake interview (T1, T2-F1, T3, T4, T5). During referral telephone calls with both referring professionals (T1) and troubled parents (T2-F1),[2] the "Basic Criteria for Acceptance" form is completed by a therapist. Based on these data, the therapist decides either to refer the parents to another agency or, if the family seems to fit the criteria for acceptance, submit the case for staff screening (T3). The acceptance criteria are:

1. presence of at least two children and one parent in the family
2. deviant child is a male, between the ages of four and 14
3. family lives within twenty minutes commuting time of ORI

4. family will permit observations to be made in the home and, if necessary, in the school
5. deviant child and/or parents are not known to be psychotic
6. deviant child is not grossly "brain damaged," and
7. for cases begun prior to 1972, the general problems characterizing the deviant child were aggressive, hyperactive, "out-of-control" behaviors. (See Patterson et al., 1972, for a more thorough description of this criterion.) For cases accepted during 1972, the deviant child will be characterized as a "high rate stealer," whose frequency of stealing is at least once every two weeks.

Staff discussion (T4) centers on the family's eligibility (re above criteria) for inclusion in the project, availability of staff to work with the family, and ability of the staff to successfully treat the reported problem behaviors. A positive decision indicates an acceptance for intake (T5), while a negative decision leads to a referral elsewhere (T3.1).

The intake interview (T6-F2). The intake interview is held at ORI. Typically, the parents meet with one therapist while the target child meets with the other member of the therapy team. (1) Parent interview--the therapist first explains the experimental nature of the project and that it can provide no therapeutic guarantees. Project procedures, including home observations, length of treatment and followup, data collection, etc., are described and a statement summarizing observation requirements is given the parents. Parents are also requested to sign release-of-information slips for social service agencies with which the family has been involved previously. A behavior checklist for the problem child, on which the parents rate the child's behavior on a seven-point desirability scale, is completed jointly by parents and therapist and used later for research purposes. The therapist then sketches the general steps involved in treatment and outlines the parents' responsibilities from intake through followup. Parent cooperation in fulfilling this "contract" represents the treatment fee. In addition, beginning in 1972, an outline of a series of deposits refundable during various phases of the program is explained. The parents are asked to complete a personality inventory (MMPI) at home and to bring the answer sheet to the "pinpoint" interview (T14-F11). (2) Child interview--the other therapist interviews the child by eliciting his reactions to a set of standardized settings and having him complete the reading and arithmetic sections of the Jastek Wide Range Achievement Test. As the child works, the interviewer or another staff member observes the child and completes the Behavior Check List.[3] The co-therapists complete a "Family Face Sheet" and create a "Behavioral Profile" on the problem child before the staff conference (T7). Before the next staff conference (T7), the therapists contact other social agencies with which the family has dealt to collect additional information.

Staff conference following intake interview (T7, T8). In the staff conference (T7), the therapists involved in intake present their observations of the parents, the children, and reports collected from other agencies. The "socially aggressive" or "stealing" child behaviors and parent behaviors toward the children are stressed and those most appropriate to treatment are noted. Prior to 1972, it was at this conference that the final acceptance decision (T10) was made. In the current year, however, a case is to be given tentative acceptance (T8) at the conference and arrangements made for three pre-baseline observations in the home (F3-01, F4-02). A staff decision not to accept a case results in a referral elsewhere (T3.1).

Pre-baseline "standard observations" (F3-01, F4-02). The lead observer telephones the family, informs them of the tentative nature of their acceptance into the project and schedules three pre-baseline "standard observations." Such observations are made in the family's home at the same time (just prior to, during, or following supper) on three different weekdays. During the "standard observation" each family member is observed twice, each time for five minutes. Observations are encoded on a "Behavior Rating Sheet," using a 29-item behavior code (Patterson, Ray, Shaw, & Cobb, 1969). Following each "standard observation" and each intervention "probe" throughout the project, the parents complete a "Parent Referral Sheet" on which they are asked to record the problem behaviors that occurred during that day. Immediately following the observation, the observer records anecdotal notes from the observation in a notebook designated for that purpose. These pre-baseline observation give the family an opportunity to experience being observed before they commit themselves to the project. They also give the therapy team information on how the family interacts during observation and more data upon which to base a final acceptance decision (T10).

Staff conference following pre-baseline "standard observations" (T9, T10). This staff conference (T9) focuses on the data collected during the pre-baseline observations. If the staff feels that the criteria for acceptance have been met, that the treatment program can help the parent bring the problem behaviors under social control, and that the family has been cooperative during observations, the, contingent upon the family's desire to continue, the case is formally accepted for treatment.

The baseline observation study (T11-F5-03, F6-04). Once accepted for treatment, the parents again come to ORI to meet with the therapists and the lead observer. Discussion centers on home observations and whether the family wishes to commit itself to the project. Family responsibilities during the observations (all family members present, movement restricted to two rooms, no telephone calls, no television watching, no cancellation of observation except in emergency) are reviewed, and parents are asked to pay a deposit which is refunded, contingent upon their keeping observation appointments. If the parents agree to continue, the lead observer outlines the baseline observation schedule and a starting date is established. Prior to 1972, six to ten "standard observations" over a two-week period comprised the baseline study, during which no treatment was given the family. For 1972, however, new baseline procedures have been instituted. Each family is to be put on one of two observation schedules. As families are accepted into the project, they are alternately assigned to Schedule 1 or Schedule 2 as follows:

Schedule 1: "standard observations" are made on 12 different days over a three to four week period.

Schedule 2: an alternating program of four hours of observation one day and a "standard observation" the next for 12 different days over a three to four week period. For the six days on which a four-hour observation occurs, the following schedule pertains:

a. one hour before school in the morning

b. one hour after school (at least one parent need be present)

c. one hour before supper

d. one hour at suppertime ("standard observation")

For "a," "b," and "c" observations at least one parent (both if possible) and the target child must be present; the siblings may come and go as they please. For "d" observations, all family members must be present. If the family operates on a schedule in which the "b" and "c" observations overlap, "c" may occur after the suppertime "standard observation."

On Schedule 1 all family members are observed twice for a total of ten minutes per family member per session. On Schedule 2 "a." "c." and "c," observers code only the target child, while on "d" all family members are obserbed as in Schedule 1, for a total of ten minutes. The same "Behavior Rating Sheet," 29-item behavioral observation code and "Parent Referral Sheet" used in the pre-baseline observations (F3-01, F4-02) are used in the baseline observations. The five observers employed by the project are rotated between cases during the baseline study and throughout the treatment and followup process. Periodically two observers collect data during a single observation as a check on the reliability of data collected between observers.

Assignment of programmed text (F7-05). On the last day of the baseline study, the observer gives the parents each a copy of a programmed text outlining the principles of behavior treatment. Living with Children (Patterson & Gullion, 1968), used prior to 1972, has been replaced by Families: Applications of Social Learning to Family Life (Patterson, 1971) for 1972 cases. The parents are either told that one of the therapists will telephone them in four or five days to check on their reading progress or asked to call the therapists when they have completed the book.

Therapist check on reading progress (T12-F8, T13). When the therapist telephones parents to check on their reading progress or when the parents call to report they have finished the book, any questions they might have regarding social learning principles are answered. When the parents have completed the book, the therapist notified the observers (T13) that the parents are ready for two more home observations (F9-06). In 1972, if the parents find Families difficult reading, they are asked to pick up Living with Children. The therapist telephones again in four or five days.

The telephone call and all others made between the therapists and the family is recorded and summarized in the "Therapist-Family Intervention Log" kept by the therapists. The time elapsed during the call is noted, as it is after all other therapist-client contacts while the family is being treated. The total professional time required for treatment is used as one indicator of treatment success, the project striving to maximize its effect while utilizing the least amount of professional time.

The book "probe" (F9-06, F10-07). The lead observer schedules two observations at suppertime on consecutive nights with the family as a "probe" to determine if any changes in behavioral interaction have occurred since reading the programmed text. As yet there has been no therapist intervention or treatment. This "probe," and those during formal intervention (08), are identical to baseline observations except that the observers code the deviant child's behavior for a total of 20 minutes per session (two ten-minute periods) and all other family members for one five-minute period per session.

The pinpoint interview (T14-F11). When the book "probe" has been completed, the parents are invited to ORI by the therapists to pinpoint the problem behaviors and discuss intervention (treatment) procedures. First a brief examination of the social learning concepts and principles presented in the programmed text is given the parents. If understanding is not adequate the therapists and parents review critical frames from the text. Parents are asked if they think they could apply such principles to a behavior change program. Target behaviors for a treatment program are then discussed. The therapists use all the data collected thus far to facilitate making decisions as to which behaviors will be dealt with first. Clear descriptions of these behaviors are developed and the therapists present the parents with methods for their observation. As practice in observing, the parents are asked to count and record a particular therapist behavior for the remainder of the interview. The mother and father, or whichever is present in the case of single-parent families, are asked to collect data separately on one target behavior for at least three days, approximately two hours per day. They are told that a therapist will call each day to record the data they collect and that after three to five days of collection, they will be invited to a parent group meeting, which has already been described during the intake interview (T6-F2) and is outlined here in section T17. Cases begun in 1972 are tentatively scheduled to not participate in weekly group meetings but instead will receive individualized treatment from the therapy team.

Parent data collection (T15-F12, T16-F13). Parents collect data for three to five days after the pinpoint interview. Therapists telephone parents daily to record and graph the data. When at least three days of baseline data have been reported, parents are asked to begin thinking of a program to change the behavior they have been observing and to bring their ideas to the parent group meeting. Invitation to that meeting is contingent upon completion of the data collection.

Weekly parent group meetings (T17-F14). These meetings are applicable to cases begun prior to 1972. Six to eight parents (couples and/or singles) and the therapists comprise a parent group. Each week each family represented in the group is given thirty minutes to focus discussion on current behavioral problems of their respective children. The goal of these meetings is to teach parents to strengthen socially adaptive behaviors and weaken deviant ones. Parents learn to set up programs to change their child's behaviors. The parent(s) tells the child of the program which normally involves one target behavior and specific consequences for its occurrence. When the initial behavior change program has been implemented successfully, another is established. The number of programs handled by parents at any one time varies from family to family. Parents' involvement in these meetings usually runs for eight to 12 weeks (see Section T20).

The reader is reminded that for cases beginning in 1972, the current plan is that parents will not attend group meetings but will instead receive individualized treatment from the therapy team.

Therapist telephone calls to parents (T18). During the time the family is working through its behavior contracts, the therapists make daily five to 10 minute telephone calls to the parents. These calls have several purposes:

1. to record data collected by parents
2. to discuss parent use of treatment procedures (social reinforcers,

timeout, etc.)

3. to redefine the problem behaviors, if necessary, and

4. to provide parents with reinforcement for their efforts.

After the parents satisfactorily complete one behavior change program, these telephone calls are gradually reduced to one per week by the eighth week of intervention.

Intervention "probes" (08). Beginning four weeks after the parents attend their first parent meeting, the observers begin again making behavior observations in the home. These "probes" occur every four weeks during treatment and at termination of intervention. A "probe" consists of a suppertime observation on each of two consecutive week nights. The target child is observed for a total of 20 minutes (two 10-minute periods) while every other family member is observed once for a total of five minutes. The same "Behavior Rating Sheet" and 29-item behavioral observation code that are used in "standard observations" (see F3-01, F4-02) are used in "probes" as well. On each day of observation the parents complete a "Parent Referral Sheet" as they did following baseline observation and the book "probe."

Commencement of school program (T19). When the targeted deviant behaviors appear to be coming under control (after a minimum of four weeks home intervention), therapists make arrangements to begin the school intervention program. No attempt will be made here to describe the procedures employed in school intervention. The interested reader is referred to Patterson, Cobb, & Ray (1972), Ray, Shaw, & Cobb, (1970), and Cobb & Ray, (1970).

Decision to continue or terminate treatment (T20-09). After eight weeks of intervention, the therapists decide whether the family should continue group treatment, begin individual treatment, or terminate treatment and begin followup procedures. To facilitate this decision, an observation probe is made during the eighth week (09). If the family seems to be making steady progress but has not yet reached the criteria for successful intervention described below, the parents will continue in group treatment. If progress has been minimal and the problem behaviors persist or have worsened, the therapists may choose to take the parents out of the group and treat their case individually. If, however, the parents have (1) reduced all deviant behaviors by at least 50 percent (this figure varies from therapist to therapist but the lower limit of success is assumed to be 50 percent), (2) successfully completed one intervention program of their own design, and (3) satisfied the therpist that they can handle future programs independently, then treatment may be terminated (T21-F16).

Family continues treatment (F15). If at the eighth week of intervention it is decided that the parents will continue in group treatment, they do so until either the criteria for successful intervention (T20) have been satisfied or they reach the fifteenth week of treatment. At that point, if the criteria have not been met, they enter individualized treatment with the therapy team. If already in individualized treatment, the parents continue until they satisfy the criteria. Several modes of individual treatment are offered. When first placed in individualized treatment, the therapist may observe family interaction *in vivo* or via video tape. From these observations, he attempts through discussion with the parents and problem child to improve the treatment program. If this fails, the therapist utilizes direct intervention techniques (individual interviews, modeling, etc.) in the home.

Termination of intervention, initiation of followup (T21-F16-010). When the family has met the criteria for successful intervention the parents again complete the personality inventory and behavior check list completed during the intake interview (T6-F2). Followup lasts for a minimum of 12 months. "Standard observations" are made on two consecutive evenings during each of the first six months and for months eight, 10, and 12. "Parent Referral Sheets" are collected on each observation. Therapist assistance is still available to the family during followup with the addition of two contingencies:

1. the family must go six months without therapist contact before followup is considered complete;
2. if more than two hours of therapist time are used in a single month, the followup will be restarted.

"Booster Shot" Program (T22-F17). When necessary, a "Booster Shot" program of intervention may be instituted during followup. Normally it occurs under one of two conditions. Either the parents complain of a serious behavior disturbance in the home or the therapists, in the analysis of followup observation data, note a marked increase in total deviant behavior for a given "probe" (two consecutive observations). If the "Booster Shot" is deemed necessary, the family is sent a copy of "Vest Pocket Refresher," a review of social learning principles, and is asked to reread the programmed text. The parents are asked by telephone to pinpoint several problem behaviors (old or new) and then requested to either attend a parent group for one session, see the therapist in his office or at their home, or design a treatment program of their own and report the data to the therapist by phone. Followup telephone calls are then made to the parents as a check on progress.

Therapists write case summary, organize data (T23). When retraining programs are complete and followup nears conclusion, the therapists write a case summary, outlining all procedures used and data collected in approximately four pages.

Conclusion of followup, case termination (F18-O11). When all followup observations have been completed, the case is formally terminated, hopefully to the satisfaction of all concerned.

References

Cobb, J. A., & Ray, R. S. Manual for coding discrete behaviors in the school setting. Unpublished manuscript, Oregon Research Institute, 1970.

Gewirtz, J. L. Mechanisms of social learning: Some roles of stimulation and behavior in early human development. In D. A. Goslin (Ed.), Handbook of socialization theory and research. Chicago: Rand McNally, 1969.

Patterson, G. R. Behavioral intervention procedures in the classroom and in the home. In A. E. Bergin and S. L. Garfield (Eds.), Handbook of psychotherapy and behavior change. New York: Wiley, 1969.

Patterson, G. R. Families: Applications of social learning to family life. Champaign, Ill.: Research Press, 1971.

Patterson, G. R., Cobb, J. A., & Ray, R. S. Direct intervention in the classroom: A set of procedures for the aggressive child. In F. W. Clark, D. R. Evans, and L. A. Hamerlynck (Eds), Implementing behavioral programs in educational and clinical settings. Champaign, Ill: Research Press, 1972.

Patterson, G. R., Cobb, J. A., & Ray, R. S. A social engineering technology for retraining the families of aggressive boys. In H. E. Adams and I. P. Unikel (Eds.) Issues and trends in behavior therapy. Springfield, Ill.: Charles C. Thomas, 1972, in press.

Patterson, G. R., & Gullion, M. E. Living with children: New methods for parents and teachers. Champaign, Ill: Research Press, 1968.

Patterson, G. R., Ray, R. S., & Shaw, D. A. Direct intervention in families of deviant children. Oregon Research Institute Research Bulletin, 1968, 8, No. 9.

Patterson, G. R., Ray, R. S., Shaw, D. A., & Cobb, J. A. Manual for coding of family interactions. Unpublished manuscript, Oregon Research Institute, 1969.

Ray, R. S., Shaw, D. A., & Cobb, J. A. The Work Box: An innovation in teaching attentional behavior. The School Counselor, 1970, 18, 15-35.

Skinner, B. F. Science and human behavior. New York: MacMillan, 1953.

Footnotes

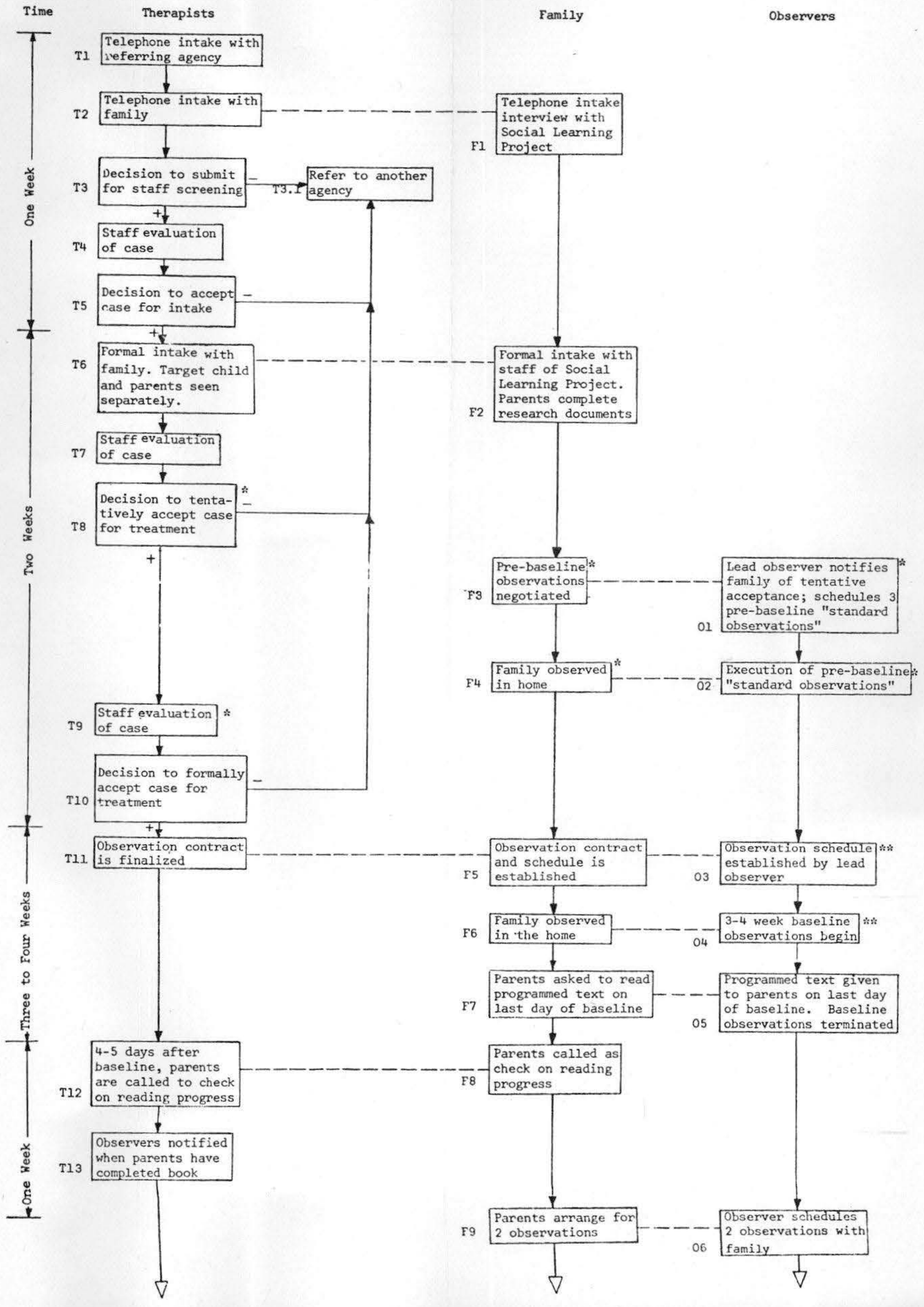
1. This research was supported by grant number 5 R01 MH15985 from the National Institute of Mental Health, U. S. Public Health Service.

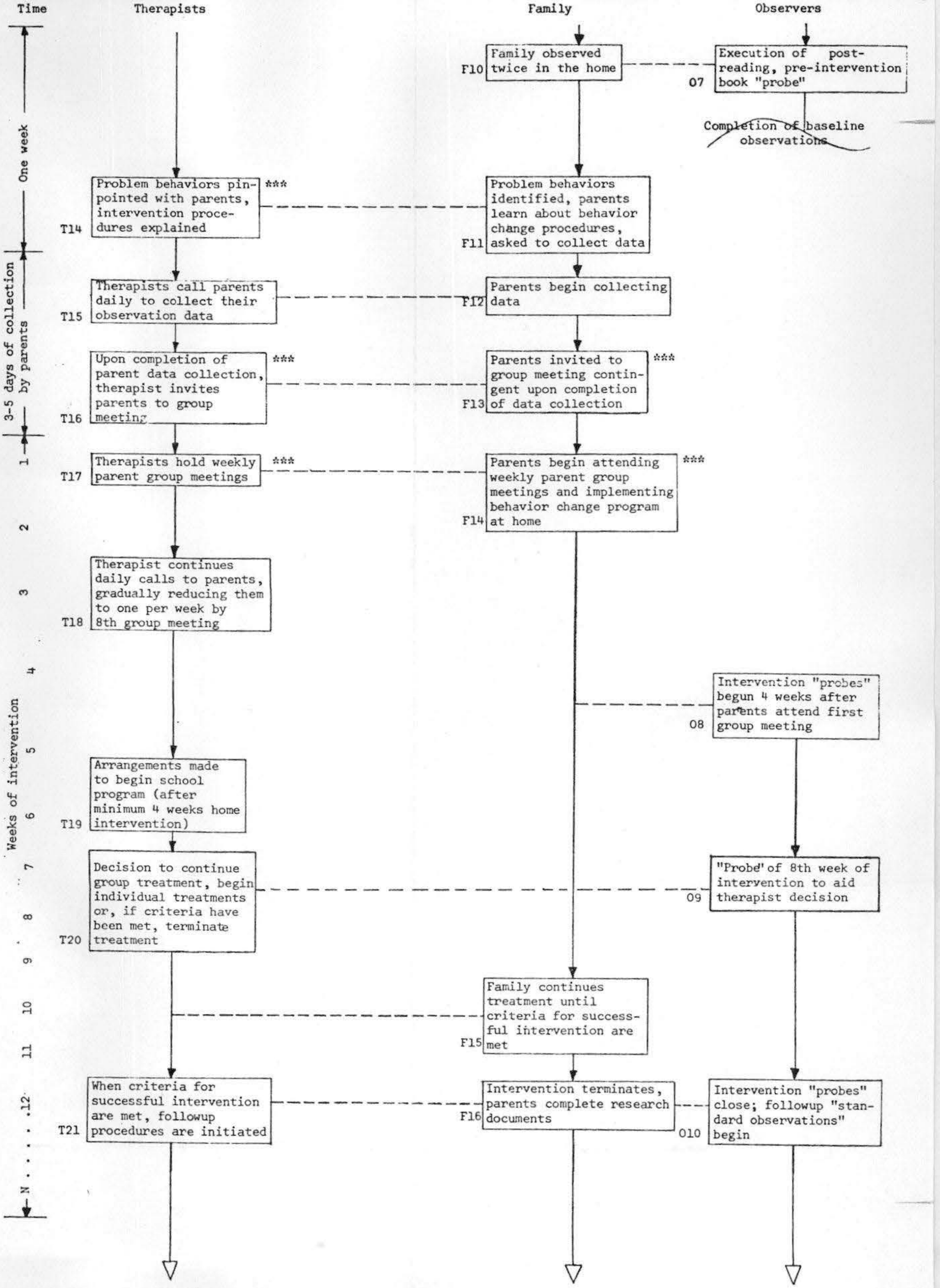
We are especially grateful for the assistance given us by G. R. Patterson, Director of the Social Learning Project, R. S. Ray, J. A. Cobb, J. B. Reid, and L. Garber, lead observer. Without their cooperation this paper could not have been possible. Inquiries should be addressed to the second author.

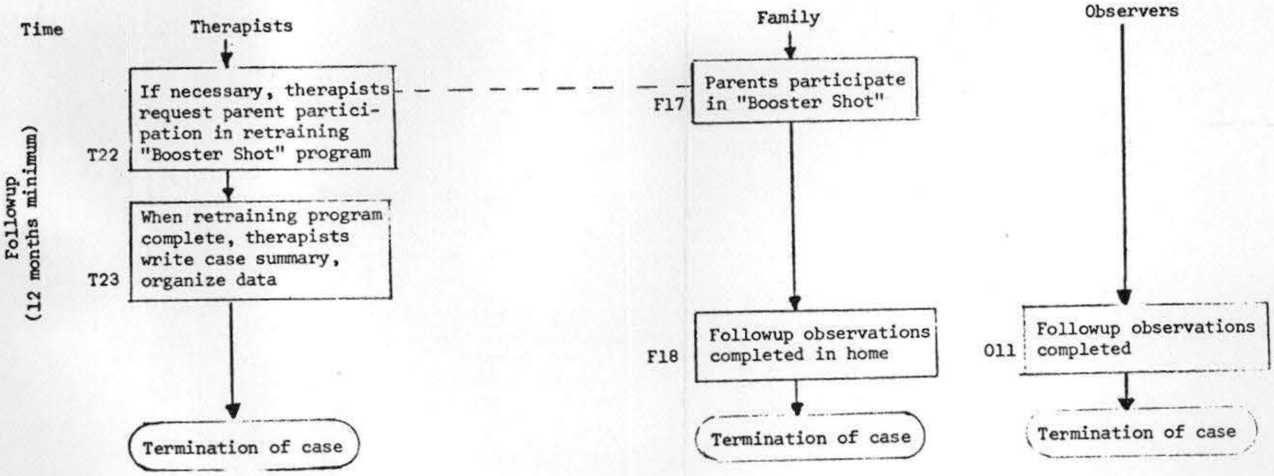
2. Although reference is made to "parents," the reader should note that many families treated by the Social Learning Project are single-parent families. The presence of both parents in the home is not a requirement for treatment.

3. The Behavior Check List used for child observation and the check list used in obtaining parent descriptions of the problem behaviors are different instruments. Both are used in preparing a composite profile of the child.

Figure 1: System analysis of the family treatment program




Time
Therapists
Family
Observers
One week
3-5 days of collection by parents
Weeks of intervention
1
2
3
4
5
6
7
8
9
10
11
12
N 12
F10 Family observed twice in the home
07 Execution of post-reading, pre-intervention book "probe"
Completion of baseline observations
T14 Problem behaviors pin-pointed with parents, intervention procedures explained ***
F11 Problem behaviors identified, parents learn about behavior change procedures, asked to collect data
T15 Therapists call parents daily to collect their observation data
F12 Parents begin collecting data
T16 Upon completion of parent data collection, therapist invites parents to group meeting ***
F13 Parents invited to group meeting contingent upon completion of data collection ***
T17 Therapists hold weekly parent group meetings ***
F14 Parents begin attending weekly parent group meetings and implementing behavior change program at home ***
T18 Therapist continues daily calls to parents, gradually reducing them to one per week by 8th group meeting
08 Intervention "probes" begun 4 weeks after parents attend first group meeting
T19 Arrangements made to begin school program (after minimum 4 weeks home intervention)
T20 Decision to continue group treatment, begin individual treatments or, if criteria have been met, terminate treatment
09 "Probe" of 8th week of intervention to aid therapist decision
F15 Family continues treatment until criteria for successful intervention are met
T21 When criteria for successful intervention are met, followup procedures are initiated
F16 Intervention terminates, parents complete research documents
010 Intervention "probes" close; followup "standard observations" begin



* These steps have been added to clinical procedures for 1972 cases. During and preceding 1971, formal acceptance into the project (T10) was decided upon during the staff conference described in T7.

** Baseline observation schedules have been altered for 1972 cases. For a description of the changes, see page 8.

*** Cases beginning in 1972 will not participate in weekly group meetings. They will instead receive individualized treatment from ther therapy team.

OREGON RESEARCH INSTITUTE

FLANGE DETECTION CLUSTER ANALYSIS

E. Mark Gold and Paul J. Hoffman

Oregon Research Institute
P.O. Box 3196
Eugene, Oregon

Research Bulletin
Vol. 12 No. 3
April 1972

CONTENTS

Submitted to Biometrics

1. SUMMARY

The subject of this paper is cluster analysis of a sample of vector observations drawn from a population assumed to be a mixture of a primary population and an unknown number of deviant populations. The primary population is assumed to have a roughly normal distribution. Each deviant population is assumed to have a smaller population than the primary population and to have a distribution which extends out from that of the primary population in an unknown direction, so that the total population consists of a principle mode with an unknown number of flanges sticking out in unknown directions.

The problem considered here is to determine the number of flanges and identify each by means of a linear scale, that is, a linear function $s(\vec{x})$ of the components of a vector observation $\vec{x}$:

$$s(\vec{x}) = \vec{s}\cdot\vec{x} - s_o \tag{1}$$

Thus, a linear scale is defined by its weight vector $\vec{s}$ and center s_o. For the purpose of defining a flange the linear scale is to be chosen so that the scale values of the deviant population are as large as possible subject to the constraint that the scale values of the primary population have mean 0 and variance 1. Such a scale will be called a flange scale.

Two approaches to the flange detection problem are proposed: main cluster sphericalization and outlier maximization. Most clustering techniques

simultaneously partition the sample points into clusters. It is characteristic of the two approaches to flange detection proposed in this paper that they start by determining the <u>main cluster</u>, the largest concentration of sample points, without regard to the distribution of outlying clusters, flanges, or unclassifiables. Then the flange directions are determined, in which there are a significant number of outlying sample points which are distant from the main cluster mean compared to the main cluster standard deviation i this direction.

Both flange detection approaches are invariant with respect to linear transformations of sample space. This is because in both approaches the method used to determine the main cluster is invariant, and then the main cluster determines the metric which defines the flanges.

<u>Main cluster sphericalization</u> consists of three steps:

1) The mean and covariance matrix of the primary population are estimated iteratively starting from an initial iterate consisting of the mean and covariance matrix of the entire sample. At each iteration the <u>main cluster</u> is defined to consist of those sample points whose Mahalanobis distance (with respect to the current estimate of the mean and covariance of the primary population) is less than the median of the chi square distribution. The next iterate for the estimated mean and covariance of the primary population is taken to be the mean and twice the covariance matrix of the main cluster. That is, an ellipsoidal window is moved around in observation space until the number of sample points inside it is maximum and its shape coincides with the covariance matrix of the points inside.

2) A linear transformation of observation space is used to make the estimated covariance matrix of the primary population spherical, and a translation is performed to put its estimated mean at the origin.

3) The covariance matrix of all points, inside and outside the window (which

is now spherical) is determined and it is diagonalized. The principal eigenvectors are the estimated flange scales.

The main cluster sphericalization algorithm has successfully detected the flanges in manufactured data consisting of a sample space of 5 dimensions with 50 sample points from a normally distributed primary population and 8 sample points from each of two orthogonal flanges, one of r.m.s. length 18 and the other of r.m.s. length 4.5, where flange length is measured relative to the standard deviation of the primary population. The first step, iteratively determining the main cluster of the sample, always converged stably and rapidly. In 15 dimensions neither flange was detectable, but increasing the number of primary population sample points to 100 made it possible to detect the longer flange.

In the <u>outlier maximization</u> approach, an estimate for a single flange scale is iterated, starting from an arbitrary initial iterate. At each iteration the scale values of the entire sample are determined, the mean and variance of the main cluster of scale values is determined in 1 dimension in a manner similar to that used in main cluster sphericalization, and a measure of the total "weight" of outliers from the main cluster of scale values is computed. A standard gradient technique is used to iterate the flange scale in order to maximize the outlier weight. Tested on manufactured data, this approach shows some promise, but more development is needed.

2. THE FLANGE DETECTION PROBLEM

It is common in clustering to assume that the sample is drawn from a multimodal mixture of populations. Therefore, the sample consists of a set of clusters which are spatially separable, and it is this separation which defines the clusters. In this investigation, on the other hand, it is assumed that the clusters overlap, for instance as shown in Figs. 1, 2, and 3. In Fig. 1 the two clusters might be distinguished by the irregular shape of the total distribution, if one were to assume that the clusters were roughly ellipsoidal. In Figs. 2 and 3 the overlapping clusters can be distinguished only by a difference in density. In both of these cases, if the population of each cluster of the pair is proportional to its volume, then there is no basis for distinguishing them.

In the particular problem for which the techniques described herein were intended, a primary population is assumed to be distributed ellipsoidally and an unknown number of deviant populations are assumed to be distributed as flanges appended to the main cluster in various directions. The flanges are assumed to be distinguished from the main cluster by their smaller population and by their location, extended from the main cluster in various directions for distances which are large compared to the variance of the main cluster in the flange direction. For instance, in Fig. 3 flange A extends a small distance from the main cluster compared to the length of the major axis of the main cluster (in the original metric), but this distance is significantly larger than the width of the main cluster in the flange direction.

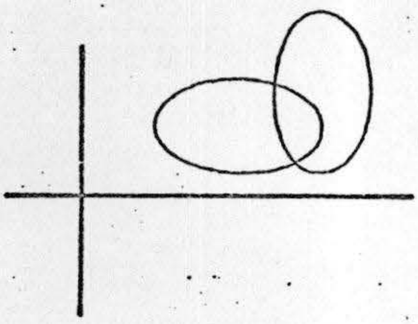


Fig. 1. Example of overlapping clusters distinguished by the fact that the total distribution is irregular.

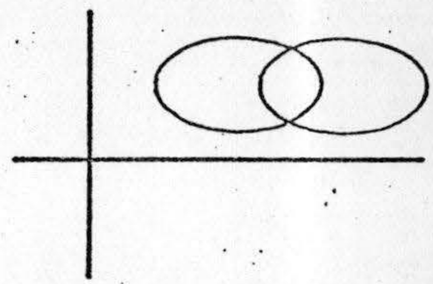


Fig. 2. Example of overlapping clusters which can only be distinguished if the populations are significantly different.

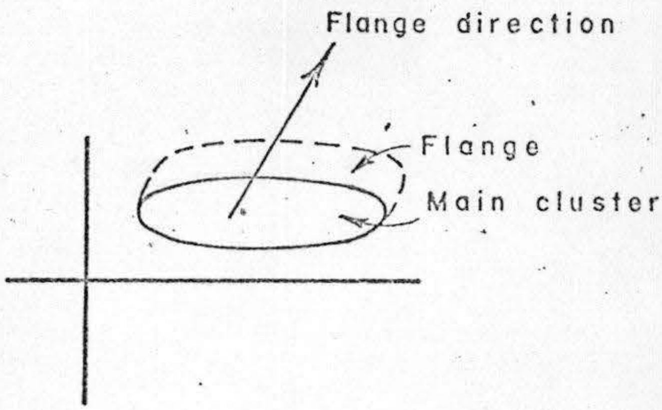


Fig. 3. Example of overlapping clusters which can only be distinguished if the population densities of the main cluster significantly exceed that of the flange.

Consider the variation of the total population density along the line from the center of the primary population in the direction of a flange. The population density falls rapidly at the edge of the primary population, but then, rather than continuing to fall rapidly to 0, it has a long tail, as in Fig. 4. This tail is the sign that this direction is a flange direction. The methods described herein for flange detection do not measure the skewness of the total distribution of points. Rather, they are sensitive to the number of points which appear to lie outside the main cluster on either side. Therefore, they are expected to detect also the double-tailed situation depicted in Fig. 5.

Presumably, the methods described herein are applicable to the determination of clusters of equal density if the clusters are either separated or have distinctly different orientations as in Fig. 1. In this case, the flange detection algorithms would probably identify one of the clusters at random as the main cluster, and then consider the others as flanges. However, the discussion of the problem and the algorithms for solving it will here be restricted to the specific situation of a main cluster with high density and flanges with low density.

The objective of flange detection is to determine the number r of flanges and define each by means of a flange scale:

$$(\vec{s},s_o)_1, \cdots, (\vec{s},s_o)_r \tag{2}$$

A sample point $\vec{x}$ is then classified by computing its scale value on each of the flange scales. If one of these scale values has a large magnitude compared to 1, then $\vec{x}$ is considered a member of this flange. Otherwise it is classified as a member of the primary population. In the main cluster sphericalization

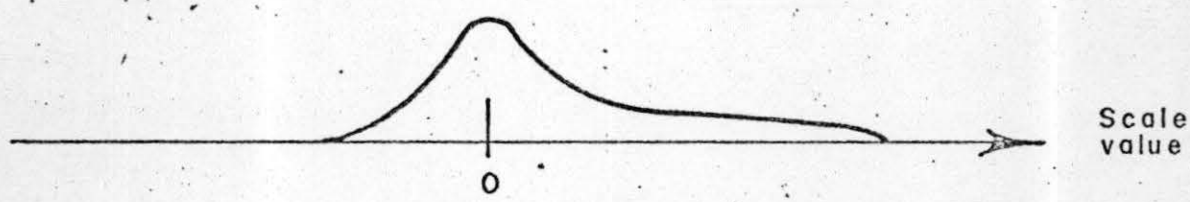


Fig. 4. Density of individuals vs. scale value for a scale which distinguishes a weirdo type (flange). Main cluster with weirdos on one side.

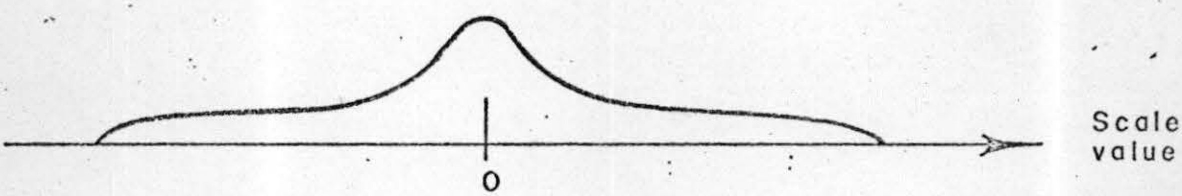


Fig. 5. Density of individuals vs. scale value for a scale which distinguishes a flange. Main cluster in the center with weirdos on both sides.

approach the covariance matrix of the primary population is estimated (via the covariance matrix of the main cluster). This makes it possible to introduce another classification. Namely, if the Mahalanobis distance of $\vec{x}$ is large, but none of its scale values are of large magnitude, then it can be considered <u>unclassifiable</u>.

For the i-th flange, the direction in which this subpopulation deviates from the mean of the primary population will be called its <u>flange direction</u>. The direction of the scale weight vector $\vec{s}_i$ of the linear scale which is the optimal scale for identifying members of this flange will be called its <u>scale direction</u>. Note that the flange direction is not the same as the scale direction unless the distribution of the primary population is spherical.

When a flange detection algorithm has proposed a number of possible flange scales, for instance the main cluster sphericalization algorithm automatically yields one for each dimension of sample space, it is straightforward to decide how many actually define flanges. Namely, if a significant number of sample points do not have large scale values on a scale, then the scale is discarded.

Two assumptions are made concerning the population distribution: (1) The distribution of the primary population is roughly Gaussian (no assumptions are made concerning the distributions of the flanges). (2) The flanges lie in orthogonal directions from the primary population in the coordinate system in which the covariance matrix of the primary population is spherical. This assumption makes it possible for the i-th flange scale to distinguish the members of the i-th flange not only from the primary population, but also from the members of the remaining flanges. Due to the orthogonality assumption the other flanges will merge with the main cluster on the i-th flange scale. This assumption is best justified in problems for which the sample space is of very high dimensionality --, compared with the expected number of flanges. In this case if the flange directions are randomly distributed, there is a high probability that they are perpendicular.

If the flange directions are not orthogonal, then it may still be possible to determine the flanges one at a time by excising the outlying sample points on each flange scale before determining a new flange scale. This is, presumably, the necessary approach to obtain more than one flange scale using the outlier maximization technique, but it will not be pursued in this paper.

3. COMPARISON OF CLUSTERING METHODS

The plethora of proposed clustering techniques can mostly be assigned to two categories, 1) agglomerative clustering (also called hierarchical clustering) e.g. Wishart [1969] provides a recent example with further references and 2) the use of an objective function, which approach has two variations, 2a) objective function clustering, e.g. Scott and Symons [1971] and Friedman and Rubin [1967], and objective function parameter estimation, see Wolfe [1967].

AGGLOMERATIVE APPROACH: This approach starts with a symmetric distance matrix defining the amount of difference between each pair of the N sample points. Initially the sample points are partitioned into N clusters, each containing one point. The distance matrix for these clusters is defined to be the same as that for the points. This approach consists of N-1 steps, each of which consists of clumping together the two closest clusters to form a single cluster. The variations of this approach differ in the method of defining the distance from a clumped cluster to another cluster as a function of the distances from each of its two components to the other cluster, in order to define the new distance matrix at the end of each agglomerative step.

In practice, the most popular variation of agglomerative clustering defines the distance from the union of two clusters to another cluster to be the minimum of the distance from each of the original clusters to the other. The various stages of clustering obtained by this variation can be defined equivalently as follows: Choose a cutoff and conceptually link together any two points such that the distance between them is less than this cutoff. The sample points and links form a graph. Take the connected components of the graph to be the clusters. The use of different cutoffs gives the different degrees of clustering of the agglomerative approach.

This interpretation of this variation of the agglomerative approach makes its properties clear: This method of clustering can find clusters of any shape (assuming a Euclidean sample space), so long as the clusters are separated by a region in which there are no sample points. That is, the clusters will, at an appropriate stage in the agglomeration represent the modes of the total population distribution.

<u>OBJECTIVE FUNCTION APPROACH</u>: In this approach it is assumed that the sample space is Euclidean and the sample points are drawn from a given number k of populations.

<u>In objective function clustering</u> the sample points are partitioned into k clusters in such a way that some objective function is minimized. This objective function represents, in some sense, the total of the volumes spanned by the k clusters, usually measured in terms of the covariance matrices of the clusters. That is, the sample points are partitioned in such a way that the clusters are as compact as possible, with "compact" defined by the objective function.

In objective function parameter estimation the total population is assumed to consist of a mixture of subpopulations, each having a distribution of a specified form with unspecified parameters and the mixing proportions unspecified. Typically, the subpopulations are assumed normally distributed and the unspecified parameters of the total distribution are determined by maximum likelihood. If it is desired to cluster the sample points, this can be done after the distribution parameters are determined via the probability of a given sample point being in each subpopulation conditioned on its position.

COMPARISON: The properties of the objective function approach can be compared to those of the agglomerative approach as follows: The objective function approach can separate overlapping clusters or subpopulations, possibly even when the subpopulations do not have distinct modes, but this approach assumes that the subpopulations have roughly ellipsoidal distributions (in assuming normal distributions or in measuring cluster volume by means of its covariance matrix) and also that every sample point can be associated with one of the clusters or subpopulations. The agglomerative approach can identify oddly shaped clusters which are separated and is not affected by distant unclassifiable sample points. The objective function approach can identify overlapping subpopulations if they are roughly ellipsoidal and we know how many there are, and is little affected by nearby unclassifiable sample points.

In both approaches it is difficult to decide how many clusters or subpopulations are appropriate, since both define a clustering for any specified number of clusters.

Flange detection is designed for the situation in which the subpopulations overlap, we do not know how many there are, and the primary population has an ellipsoidal distribution. No assumptions are made about the deviant distributions except that they deviate in a specified direction from the mean of the primary population, and their sizes are small compared to the size of the primary population. Also, if there is more than one flange, the main cluster sphericalization algorithm, in its present form, assumes that the flanges deviate from the primary population in orthogonal directions.

4. MAIN CLUSTER DETERMINATION IN THE MAIN CLUSTER SPHERICALIZATION APPROACH

The main cluster signifies the set of sample points which are drawn from the inner region of the primary population. In the main cluster sphericalization approach the method used to identify the main cluster was closely related to the NORMIX algorithm proposed by John Wolfe [1967]. John Wolfe assumed a specified number of Gaussian clusters and iteratively adjusted a multimodal Gaussian distribution until it best fit (in the sense of maximum likelihood) the entire sample. We have constructed a similar algorithm which may be thought of as iteratively adjusting a unimodal Gaussian distribution until it best describes the main cluster. The estimate of the set of points which constitute the main cluster is simultaneously adjusted. The estimated Gaussian distribution of the primary population is described by its mean $\vec{x}_o$ and covariance matrix P. A sample point $\vec{x}$ is considered to be a member of the main cluster if

$$(\vec{x}-\vec{x}_o)^t \, P \, (\vec{x}-\vec{x}_o) \leq d - .5 \tag{3}$$

where d is the dimensionality of sample space. The quantity on the left side of (3) is the Mahalanobis distance of x. The cutoff of d-1/2 was chosen because it is roughly the median of the chi square distribution.

The inequality (3) defines an ellipsoidal window in sample space. This ellipsoidal window simultaneously defines the estimated main cluster and the estimated distribution of the primary population. The algorithm for determining the main cluster of the sample iteratively adjusts this window by means of two operations: Centering adjusts $\vec{x}_o$ and shaping adjusts P.

In the centering operation the mean $\vec{y}_o$ of the points within the window is determined and the window is moved to this center by replacing $\vec{x}_o$ by $\vec{y}_o$. The end result of repetitively performing this operation is that the center of the window coincides with the mean of the points inside. This is the same as saying that the window is centered to the point at which the sample population it contains is maximum, if the population is properly weighted, namely by any constant minus square distance from the window center.

In shaping, the covariance matrix P in the window definition is adjusted according to the following rule: Let M be the covariance matrix of the points inside the window (the main cluster). Then replace P by 2M. This is repeated until the P which defines the window is just twice the covariance of the points inside. The reason for the factor of 2 is that M is only intended to estimate the covariance matrix of the inner half of the primary population. The outer half will, presumably, have the same shape, but its covariances will obviously be larger. The factor of 2 can be thought of as an approximation to the ratio of the expected value of the chi square distribution to the same expected value with the integral which defines it truncated at the median of the chi square distribution. Actually, this factor was chosen empirically to give stable convergence for manufactured data sampled from a normal distribution.

The possible instability in main cluster determination using this algorithm is that the window might collapse to 0. It can't expand to ∞ because P can't be larger than twice the covariance matrix of all the sample points.

5. FLANGE DETERMINATION IN THE MAIN CLUSTER SPHERICALIZATION APPROACH

In the main cluster sphericalization approach, when the main cluster has been determined an affine transformation of sample space is performed to bring the estimated mean of the primary population to the origin and make its estimated covariance matrix equal to the identity matrix. It is now assumed that the covariance matrix of all the sample points drawn from the primary population is spherical. The covariance matrix of the entire sample is determined and its principal eigenvectors are taken as the estimated flange scales in this transformed space.

That is, the estimated flange scales are taken to be the principal eigenvectors normalized to unit length with scale centers equal to 0. Because of the way sample space was transformed the estimated mean and variance of the primary population scale values for such scales are 0 and 1. In this transformed space the flange directions and scale directions are the same.

Let the <u>primary sample</u> denote those sample points which come from the primary population. The main cluster is intended to be the central part of the primary sample. No attempt is made to determine the outer part of the primary sample because it may be mixed with the inner parts of the flange sample points. Sphericalizing the estimated primary population covariance matrix means, in practice, sphericalizing the covariance matrix of the main cluster. The procedure given in this section for estimating flange scales assumes that sphericalizing the main cluster results in sphericalizing the entire primary sample. In other words, it is assumed that the covariance matrix of the outer part of the primary sample is proportional to the covariance matrix of the inner part. If this fails to be the case, then the estimated flange scale directions will be pulled away from their correct directions by the bulges in the outer part of the primary sample, compared to the inner part.

The outer part of the primary sample may fail to have a covariance matrix proportional to that of the inner part for two reasons: 1) The outer part of the actual primary population distribution may have a covariance matrix which is not proportional to that of the inner part. In this case the main cluster sphericalization approach to flange detection is not applicable. 2) The covariance matrix of the main cluster, i.e. the inner part of the primary sample, will differ from the covariance matrix of the inner part of the primary population and the same is true of the outer parts. This is statistical noise due to the finiteness of the inner and outer primary samples. Note that the amount of statistical noise depends on the ratio N_p/d, where N_p is the size of the primary sample and d the dimensionality of sample space. Our empirical results on manufactured data indicate that N_p/d should exceed $15L_f^{-1/2}$ where L_f is the flange length.

The flange scale estimates determined by the main cluster sphericalization approach can be defined in four equivalent ways, summarized below. The first is the definition given at the beginning of this section. The first two are algorithmic. The third definition provides the motivation for estimating flange scales in this way. Namely, it says that the flange scales are to be chosen so that as many sample points as possible will have large scale values after normalizing the scale with respect to the estimated primary population distribution of scale values. The fourth definition is called characteristic vectors of T in the metric of P by Anderson [1958], where the distribution of the roots and vectors is derived, assuming spherical, normal distributions.

NOTATION: P is the estimated covariance matrix of the primary population. T is the covariance matrix of the entire sample. $\vec{x}_{op}$ is the estimated mean of the primary population. $\vec{x}_i$ is a sample point. $\vec{x}$ denotes an arbitrary point of the sample space, $\vec{x}'$ its representation in the space transformed in order to sphericalize the main cluster. The scale value s of $\vec{x}$ is defined by the scale weights $\vec{s}$ and the scale center s_o:

$$s(\vec{x}) = \vec{s}\cdot\vec{x} - s_o \tag{4}$$

DEFINITION 1:

$$\vec{x}_i' = P^{-1/2}(\vec{x}_i - \vec{x}_{op}) \tag{5}$$

$$T' = \Sigma_i \vec{x}_i'\vec{x}_i'^{\,t} = P^{-1/2}TP^{-1/2} \tag{6}$$

In X'-space the flange scale weight vectors are defined to be the eigenvectors $\vec{e}'$ of T':

$$T'\vec{e}' = \lambda\vec{e}' \tag{7}$$

normalized to unit length, and the scale centers are 0.

In X'-space the flange directions are the same as the flange scale weight vectors. The scale value of $\vec{x}$ is thus defined to be

$$s(\vec{x}) = \vec{e}\,'\cdot\vec{x}\,' = \vec{e}\,'\cdot P^{-1/2}(\vec{x}_i - \vec{x}_{op}) \tag{8}$$

so that the scale weight vectors and centers in X-space are

$$\vec{s} = P^{-1/2}\vec{e}\,' \qquad s_o = \vec{s}\cdot\vec{x}_{op} \tag{9}$$

The flange directions in X-space are obtained by transforming the flange directions $\vec{e}\,'$ from X'-space to X-space:

$$\vec{f} = P^{1/2}\,\vec{e}\,' = P\vec{s} \tag{10}$$

DEFINITION 2: In the original X-space the flange scale weight vectors are defined to be the eigenvectors of $P^{-1}T$:

$$P^{-1}T\vec{s} = \lambda\vec{s} \tag{11}$$

normalized by requiring the estimated primary population variance to be unity:

$$\vec{s}^{\,t}P\vec{s} = 1 \tag{12}$$

The flange scale centers are the scale values of the primary population mean on the uncentered scales:

$$s_o = \vec{s}\cdot\vec{x}_{op} \tag{13}$$

From definition 1 we know that the flange directions (with respect to the main cluster mean) are

$$\vec{f} = P\vec{s} \tag{14}$$

DEFINITION 3: The first flange scale weight vector, up to a normalizing constant, is defined by the requirement that the ratio

$$(\vec{s}^{\,t}T\vec{s})/(\vec{s}^{\,t}P\vec{s}) \tag{15}$$

of total variance to estimated primary population variance be maximum. s is normalized and s_o and $\vec{f}$ defined as in definition 2. Subsequent flange scale weight vectors are defined by maximizing the variance ratio subject to the requirement that they be orthogonal to the previous scale weight vectors.

DEFINITION 4: The flange scale weight vectors, up to a normalizing constant, are defined by the requirement that for some λ

$$T\vec{s} = \lambda P\vec{s} \tag{16}$$

so that the possible values of λ are the solutions of

$$|T-\lambda P| = 0 \tag{17}$$

This definition derives from the previous one by the well known route of maximizing $\vec{s}^{\,t}T\vec{s}$ subject to the constraint that $\vec{s}^{\,t}P\vec{s}=1$, and defining the solutions via Lagrange multipliers λ. The flange scale normalizing constant, center s_o, and flange direction $\vec{f}$ are defined as in definition 2.

6. COMPUTATIONAL ALGORITHM FOR THE MAIN CLUSTER SPHERICALIZATION APPROACH

The algorithm which was used computationally is shown in Fig. 6. In estimating the mean and covariance matrix of the primary population the centering operation is repeated until it converges before each shaping operation is performed. There are two reasons: (1) Centering requires less computational effort. (2) The centering operation will probably converge to the correct position even when the shape is wrong. This is because centering determines the window position for which the population inside the window is maximum. This is, roughly, the position at which the population density is maximum, which is independent of the window shape.

In the "determine main cluster" subroutine p_i is set at 1 if $\vec{x}_i$ is inside the window and 0 otherwise. The convergence criterion for both the inner loop and outer loop is determined by the total number of points which move either from the outside to the inside of the window or vice versa during one operation of centering or shaping. This criterion was arbitrarily set at 2.

In the last section two alternative algorithms were proposed for determining the flange scales, namely the first two of the four equivalent definitions. The first algorithm finds the flange scales in the transformed sample space in which the main cluster is spherical by finding the eigenvectors of P'. The second algorithm finds the flange scales in the original sample space by determining the eigenvectors of $P^{-1}T$. The first algorithm was used computationally because it involves diagonalizing a symmetric matrix. In practice, this algorithm required two diagonalizations, the first in order to determine the square root of P in order to transform sample space to the sphericalized coordinates.

$$\vec{x}_o = \sum_i^N \vec{x}_i / N$$

TOTAL SHAPE (initial iterate for P)

$$P = \sum_i (\vec{x}_i - \vec{x}_o)(\vec{x}_i - \vec{x}_o)^t / (N-1)$$

DETERMINE MAIN CLUSTER

$$r_i^2 = (\vec{x}_i - \vec{x}_o)^t \, P^{-1} (\vec{x}_i - \vec{x}_o)$$

$$p_i = \begin{cases} 1 \text{ if } r_i^2 \leq d-.5 \\ 0 \text{ if } r_i^2 > d-.5 \end{cases}$$

MAIN CLUSTER CENTER (iterate x_o)

$$\vec{x}_o = \sum_i p_i \vec{x}_i / \sum_i p_i$$

DETERMINE MAIN CLUSTER

$$\sum_i |\Delta p_i| \leq 2 \ ?$$

NO

YES

MAIN CLUSTER SHAPE (iterate P)

$$P = 2 \sum_i p_i (\vec{x}_i - \vec{x}_o)(\vec{x}_i - \vec{x}_o)^t / (\sum_i p_i - 1)$$

DETERMINE MAIN CLUSTER

$$\sum_i |\Delta p_i| \leq 2 \ ?$$

NO

YES

SPHERICALIZE MAIN CLUSTER

$$\vec{x}_i' = P^{-1/2} \vec{x}_i$$

$$\vec{x}_o' = P^{-1/2} \vec{x}_o$$

DETERMINE TOTAL FLANGE COVARIANCE IN TRANSFORMED SPACE

$$T' = \sum_i (\vec{y}_i - \vec{y}_o)(\vec{y}_i - \vec{y}_o)^t/(N-1)$$

DIAGONALIZE COVARIANCE MATRIX

$$T' = R^{-1} \Lambda R$$

where Λ is diagonal (the flange scale eigenvalues) & columns of R are flange scales

TRANSFORM FLANGE SCALES BACK TO ORIGINAL SAMPLE SPACE

$$S = P^{-1/2} R$$

where columns of S are flange scale weight vectors

DETERMINE SCALE CENTERS

$$S_o = S^t \vec{x}_o$$

FIGURE 6 (2 pages). Algorithm for flange detection via main cluster cluster sphericalization. Vectors are columns. N designates total number of sample points, i = 1, ···, N. d is the dimensionality of sample space.

7. RESULTS ON MANUFACTURED DATA USING MAIN CLUSTER SPHERICALIZATION APPROACH

Sample vectors were generated in either 5 or 15 dimensions. The primary population was spherical Gaussian, mean 0 and covariance matrix 1. The first flange had the same distribution stretched out along the first axis by adding a random variable with expected value 16 to its first component. Therefore, the first flange had an r.m.s. length of about 18.5. The second flange was stretched out along the second axis by adding a r.v. with expected value 4 to the second component of the sample vector so that the r.m.s. length of the second flange was about 4.5. The increment added to the flange sample points in the flange direction was either generated with an exponential distribution (exponential flanges) or with a rectangular distribution (uniform flanges). For instance, in the case of the first flange, first a vector with spherical normal distribution was generated, then a r.v. Δx_1 was added to its first component. Δx_1 was either generated between 0 and ∞ with probability density $\exp(-\Delta x_1/16)/16$ in the case of an exponential flange, or else was generated between 0 and 32 with density 1/16 in the case of a uniform flange.

Small numbers of sample points were used in order to show the situation when the flanges are on the verge of being lost in the statistical noise of the main cluster.

Table 1 shows the results of the main cluster sphericalization program for flange detection applied to manufactured data of this type consisting of

TYPE OF DATA	First 2 estimated flange scale weight vectors		Eigenvalue of each flange scale
Exponential flanges	0.756	-0.039	3.909
	-0.080	0.354	1.621
	0.130	1.266	1.323
5 dimensions	0.060	0.774	0.885
	-0.283	0.118	0.881
Exponential flanges	0.714	-0.013	8.045
	0.125	0.815	2.748
	0.016	0.237	0.929
5 dimensions	0.171	-0.048	0.773
	-0.058	0.016	0.712
Exponential flanges	0.730	-0.019	3.469
	0.221	0.666	1.296
	0.048	0.147	1.094
5 dimensions	0.122	-0.319	0.997
	-0.218	-0.198	0.851
Uniform flanges	0.785	-0.041	4.825
	0.113	0.666	1.319
	0.351	0.102	1.185
5 dimensions	-0.114	-0.047	0.907
	0.264	0.055	0.775

Table 1. Results of applying main cluster sphericalization program for flange detection to manufactured data. Sample space: 5 dimensions. Primary population: 50 sample points. Long flange: r.m.s. length 18.5; correct scale weight vector $(1,0,0,\cdots,0)$; 8 sample points. Short flange: r.m.s. length 4.5; correct scale weight vector $(0,1,0,\cdots,0)$; 8 sample points.

the following: 5 dimensional sample space, 50 sample points from the primary population, 8 sample points from each flange. The flange scale for the longer flange of r.m.s. length 18.5 was always determined accurately. The flange scale for the shorter flange of r.m.s. length 4.5 was determined approximately in 3 out of 4 trials. It is not shown in Table 1, but in the first trial the third estimated flange scale was a very good estimate for the flange scale of the shorter flange, even though the second estimated flange scale was determined entirely by the statistical noise of the main cluster.

It is important to note that increasing the number of dimensions (variables) while holding the number of sample points fixed increases the statistical noise in the main cluster, that is, in the estimate of the covariance matrix of the primary population, thereby making the flanges more difficult to detect. In Table 2 the sample was generated in the same way as in Table 1, except that the dimensionality of sample space was 15. Now both flanges were lost in the statistical noise of the main cluster covariance matrix. In Table 3 the sample was generated in the same way as in Table 2 except that the number of points generated from the primary population was increased to 100. The numbers of sample points generated from the flange populations were not increased. However, now the longer flange is accurately determined.

It should be noted that there was no loss in generality in manufacturing data with a spherical primary population. If one were to take the data which was generated and subject it to a linear transformation, then the

TYPE OF DATA	First 2 estimated flange scale weight vectors		Eigenvalue of each flange scale
Exponential flanges	0.252	-0.023	2.610
	-0.043	-0.627	1.610
	0.135	1.046	1.360
	-0.193	0.003	1.260
	-0.316	-0.044	1.111
	0.139	0.341	1.083
	0.203	0.302	1.014
	0.015	-0.275	0.849
	0.296	-0.086	0.806
	0.224	0.157	0.767
15 dimensions	0.465	0.371	0.730
	0.111	0.458	0.640
	0.203	1.164	0.639
	0.260	-0.305	0.608
	0.183	-0.135	0.599
Uniform flanges	0.522	-0.036	3.423
	-0.283	-0.472	1.562
	0.521	0.126	1.492
	0.516	-0.264	1.246
	0.113	-0.141	1.125
	0.198	0.286	1.089
	-0.118	0.371	0.979
	0.789	0.902	0.867
	0.356	0.110	0.845
15 dimensions	-0.107	0.223	0.761
	-0.123	-0.729	0.717
	-0.468	0.365	0.663
	0.297	-0.182	0.611
	0.279	-0.006	0.599
	-0.237	1.240	0.561

Table 2. Same as Table 1 except sample space has 15 dimensions.

TYPE OF DATA	First 2 estimated flange scale weight vectors		Eigenvalue of each flange scale
Exponential flanges	0.579	0.009	4.351
	0.040	0.254	1.272
	0.110	0.180	1.124
	0.041	-0.228	1.062
	-0.158	0.462	0.957
	0.040	-0.113	0.902
	0.100	0.498	0.884
	0.063	-0.164	0.759
	0.087	0.251	0.739
	0.123	-0.615	0.714
15 dimensions	-0.214	0.389	0.684
	-0.095	0.004	0.675
	-0.014	0.389	0.646
	0.045	0.283	0.633
	-0.117	0.108	0.629
Uniform flanges	0.719	-0.001	4.415
	0.290	0.785	1.471
	0.043	0.005	1.219
	0.062	0.227	1.031
	0.157	-0.375	0.922
	-0.201	-0.270	0.875
	0.057	-0.809	0.856
	0.037	-0.239	0.822
	-0.009	-0.405	0.760
	0.076	-0.943	0.692
15 dimensions	0.021	0.134	0.680
	0.020	0.022	0.664
	0.125	-0.014	0.637
	-0.018	0.160	0.625
	0.115	-0.103	0.613

Table 3. Same as Table 2, except number of sample points from primary population increased from 50 to 100.

results would be exactly the same. The same main cluster would be determined. After sphericalizing the main cluster the coordinates of the sample points would be precisely the same as if the linear transformation had not been performed on the original data.

8. OUTLIER MAXIMIZATION APPROACH

In the outlier maximization approach a vector $\vec{s}^*$ which defines the scale direction is adjusted iteratively in order to find that $\vec{s}^*$ which appears to have the largest outlier "weight". That is, the number of points which seem to lie outside the main cluster on the scale is maximized. Standard function minimization techniques are used to find the $\vec{s}^*$ with the greatest outlier weight, starting from an arbitrary first guess. The outlier weight is defined as follows:

A given scale direction $\vec{s}^*$ assigns an unnormalized scale value s_i^* to each sample point $\vec{x}_i$, according to the following rule:

$$s_i^* = \vec{s}^* . \vec{x}_i \tag{18}$$

These s_i^* are distributed in a 1-dimensional space. The main cluster of this distribution is determined by the method specified below. The result of this 1-dimension main cluster analysis is an estimated primary population mean s_o^* and standard deviation σ^*. Now the scale value of each sample point is defined by

$$s_i = (s_i^* - s_o^*)/\sigma^* \tag{19}$$

Now some outlier measure $m(s_i)$ is chosen to define the extent to which the i-th sample point lies outside the main cluster on the linear scale with direction $\vec{s}^*$. The total outlier weight for $\vec{s}^*$ is defined to be

$$M = \sum_i m(s_i) \tag{20}$$

In outlier maximization, that scale direction $\vec{s}^*$ is found which maximizes

M. Then the estimated linear flange scale is defined to consist of the following vector and center:

$$(\vec{s}, s_o) = (\vec{s}^*/\sigma^*, s_o^*/\sigma^*) \tag{21}$$

<u>DETERMINATION OF MAIN CLUSTER IN 1-DIMENSION</u>: In the outlier maximization program, a technique for finding the main cluster in 1 dimension was used which differed somewhat from that used to determine the main cluster in the d dimensional sample space in the "main cluster sphericalization" program. The primary difference was that the windows which are used in the "outlier maximization" program are smooth rather than having sharply defined boundaries. This makes it possible to define the primary population mean and variance by means of 2 simultaneous equations with smooth derivatives which can be solved by standard iterative techniques.

Basically, the mean and standard deviation of the primary population are estimated as follows: In order to estimate the mean suppose that the standard deviation is given. Construct a window of width equal to 2 standard deviations and slide it along the scale until the greatest number of points are seen through it. This is the estimated primary population mean s_o^*. In order to determine the standard deviation σ^* suppose that s_o^*, is given. Construct 2 concentric windows around s_o^*, a smaller window and a larger window enclosing it, with both window widths proportional to the estimated value of σ^* with fixed proportionality factors. Now adjust σ^* until the number of points visible through the larger window is double the number of points visible through the smaller.

Let $\vec{s}^*$ be given. For any s_o^* and σ^* two functions $F(s_o^*, \sigma^*)$ and $G(s_o^*, \sigma^*)$ must be defined so that s_o^* and σ^* are determined by the equations

$$F(s_o^*, \sigma^*) = G(s_o^*, \sigma^*) = 0 \tag{22}$$

F and G will be defined in terms of the scale values s_i which, in turn, according to (19) depend on s_o^* and σ^*.

For defining s_o^* the window transparency p was chosen to be

$$p(s_i) = \exp(-s_i^2/2) \tag{23}$$

Given σ^*, s_o^* is defined by the requirement

$$P = \sum_i P(s_i) = \max \tag{24}$$

For P to be maximum with respect to s_o^* its partial derivative with respect to s_o^* must be 0. This can be expressed by

$$F = \sum_i f(s_i) = 0 , \tag{25}$$

where
$$f(s_i) = s_i p(s_i) . \tag{26}$$

For defining σ^* a central window p_1 and a peripheral window p_2 (meant to be the difference between the larger and smaller windows) are defined to have the following transparencies:

$$p_1(s_i) = \int_{|s_i|}^{\infty} e^{-x^2/2} dx \tag{27}$$

$$p_2(s_i) = |s_i| \; p(s_i) . \tag{28}$$

The window transparencies p, p_1, and p_2 are compared in Fig. 7.

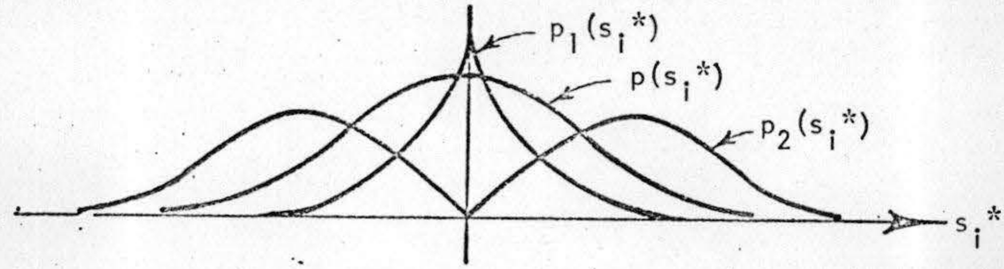


Fig. 7. Comparison of the windows used in estimating the mean and standard deviation of the primary population of the unnormalized scale values.

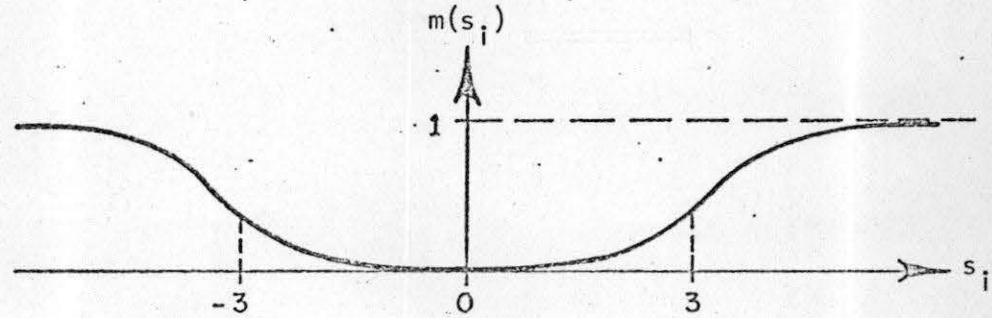


Fig. 8. The outlier measure for a point whose distance from the estimated primary population mean on the scale is s_i standard deviations.

Therefore, σ^* is defined by

$$G = \sum_i g(s_i) = 0 \tag{29}$$

where $$g(s_i) = p_2(s_i) - Cp_1(s_i) \tag{30}$$

where C is defined to be the ratio of the expectation of p_2 to the expectation of p_1 assuming a normalized Gaussian distribution:

$$\begin{aligned} C &= \int_{-\infty}^{\infty} p_2(s_i)p(s_i)ds_i \bigg/ \int_{-\infty}^{\infty} p_1(s_i)p(s_i)ds_i \\ &= 2/\pi \end{aligned} \tag{31}$$

For the outlier weight $m(s_i)$, the following function was used:

$$m(s_i) = 1 - \exp\left(-s_i^2/18\right) \tag{32}$$

This function is shown in Fig. 8. Roughly, the total outlier measure $M=\Sigma_i m(s_i)$ of a scale vector counts the number of points which lie beyond 3 standard deviations from the main cluster center on this scale.

If, instead, one uses

$$m(s_i) = s_i^2 \tag{33}$$

then the flange scale which is obtained by outlier maximization will be the same as the first eigenvector of the total covariance matrix after main cluster sphericalization in sample space.

9. COMPUTATIONAL ALGORITHM FOR THE OUTLIER MAXIMIZATION APPROACH

On each scale direction iterate $\vec{s}^*$, in order to determine the outlier measure (and its partial derivatives for the purpose of maximization), the first thing which must be done is to solve F=G=0 for the primary population mean s_o^* and standard deviation σ^* on this scale. This was done by the standard ol' Newton-Raphson technique. That is, at each iterate of (s_o^*,σ^*) the functions F and G are linearized with respect to s_o^* and σ^* and the resulting 2 linear equations in 2 unknowns are solved for the next iterate of (s_o^*,σ^*). The initial guess for (s_o^*,σ^*) was obtained by taking the mean and standard deviation of all the unnormalized scale values s_i^*.

It was found that the first iteration of the Newton-Raphson technique tended to overshoot the solution so badly that σ became negative, which was of course meaningless. This put F and G in a region of potholes and ruts where the iterations wandered about aimlessly. Therefore, the first few steps were shortened.

In order to maximize M with respect to $\vec{s}$, the partial derivatives of M with respect to the components of $\vec{s}$ are needed. Let $\vec{\nabla}$ denote partial derivatives with respect to the components of $\vec{s}$. Then

$$\vec{\nabla}M = \sum_i m'(s_i)[\vec{\nabla}s_i + (\partial s_i/\vec{\nabla}s_o^*)\ \vec{\nabla}s_o^* + (\partial s_i/\partial\sigma^*)\vec{\nabla}\sigma^*] \tag{35}$$

where prime denotes derivative with respect to s_i. From the definition of s_i, we obtain

$$\vec{\nabla}s_i = \vec{x}_i/\sigma^* \tag{36}$$

$$\partial s_i/\partial\sigma^* = -s_i/\sigma^* \tag{37}$$

$$\partial s_i/\partial s_o^* = -1/\sigma^* \tag{38}$$

In order to determine $\vec{\nabla}\sigma^*$ and $\vec{\nabla}s_o^*$, we take the partial derivatives of the equations which define σ^* and s_o^*:

$$0 = \vec{\nabla}F = \sum_i f'(s_i)[\vec{\nabla}s_i + (\partial s_i/\partial s_o^*)\vec{\nabla}s_o + (\partial s_i/\partial\sigma^*)\vec{\nabla}\sigma^*] \tag{39}$$

$$0 = \vec{\nabla}G = \sum_i g'(s_i)[\vec{\nabla}s_i + (\partial s_i/\partial s_o^*)\vec{\nabla}s_o + (\partial s_i/\partial\sigma^*)\vec{\nabla}\sigma^*] \tag{40}$$

Using the definitions (26) and (30) of f and g together with (36-38), Eqs. (39) and (40) can be solved for $\vec{\nabla}s_o^*$ and $\vec{\nabla}\sigma^*$. These, in turn, can be substituted into (35) to give the required partial derivatives of M. In solving (39) and (40), $\vec{\nabla}s_o^*$ and $\vec{\nabla}\sigma^*$ are expressed in terms of summations. Substituting these into (35) expresses M as a sum of 3 terms, of which 2 are double summations. Reversing the order of summation makes it possible to express the required partial derivatives of M in a form which is particularly convenient computationally: Let

$$F_1 = \partial F/\partial s_o^* = -\sum_i f'(s_i)/\sigma^* \tag{41}$$

$$F_2 = \partial F/\partial\sigma^* = -\sum_i s_i f'(s_i)/\sigma^* \tag{42}$$

$$G_1 = \partial G/\partial s_o^* = -\sum_i g'(s_i)\sigma^* \tag{43}$$

$$G_2 = \partial G/\partial\sigma^* = -\sum_i s_i g'(s_i)/\sigma^* \tag{44}$$

$$D = G_1F_2-F_1G_2 \tag{45}$$

$$M' = \sum_i m'(s_i) \tag{46}$$

$$R' = \sum_i s_i m'(s_i) \tag{47}$$

$$E_i = m'(s_i)\sigma^* + (M'/D\sigma^*)\ (F_2 g'(s_i) - G_2 f'(s_i))$$
$$+ (R'/D\sigma^*)(G_1 f'(s_i) - F_1 g'(s_i)) \tag{48}$$

Then

$$\vec{\nabla}M = \sum_i E_i \vec{x}_i \tag{49}$$

The advantage of this formulation is that (49) is the only operation over a 2-dimensional array.

10. RESULTS ON MANUFACTURED DATA USING THE OUTLIER MAXIMIZATION APPROACH

Manufactured data was generated to test the outlier maximization program in the same way that it was generated for the main cluster sphericalization program. With 50 points in the main cluster and 8 in each of the 2 flanges, the main cluster determination in 1 dimension failed. Namely, the estimate of the main cluster standard deviation dropped to a very low and unstable value. With 80 points in the main cluster and 20 in each flange, the algorithm appeared to correctly identify the main cluster on every scale direction iterate. The scale direction converged to a vector which gave primary weight to the direction of the shorter flange, but enough weight to the direction of the longer flange so that both sets of flange points had a significant probability of lying beyond 3 standard deviations of the main cluster on this scale. Presumably, this result was due to the fact that the particular outlier measure which was used roughly counted the number of outliers beyond 3 standard deviations on the scale without caring much how far beyond 3 standard deviations they were.

PROGRAM AVAILABILITY

The programs and associated subroutines described in this paper are available from the Oregon Research Institute Computing Center, P. O. Box 3196, Eugene, Oregon 97403. Ask for the Archive Dectape titled Flange Detection.

REFERENCES

Anderson, T. W. [1958] An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, Inc., Sect. 13.2.

Friedman, H. P. and Rubin, J. [1967] "On some invariant criteria for grouping data," J. Amer. Statistical Assoc. 62, 1159-1158.

Scott, A. J. and Symons, M. J. [1971] "Clustering methods based on likelihood ratio criteria," Biometrics 27, 387-397.

Wishart, David [1969] "An algorithm for hierarchical classifications," Biometrics 25, 165-170.

Wolfe, J. H. [1967] "Normix: computational methods for estimating the parameters of multivariate normal mixtures of distributions," Research Memorandum #SRM68-2 of the Naval Personnel Bureau, San Diego, California.

FOOTNOTE

1. This research was supported by GRS Grant No. RR05612 from the Division of Research Resources of the National Institutes of Health.

FOOTNOTE

1. This research was supported by GRS Grant No. RR05612 from the Division of Research Resources of the National Institutes of Health.