E. R. Barnes

An Algorithm for Separating Patterns by Ellipsoids

We give an algorithm for finding the ellipsoid of least volume containing a set of points in a finite-dimensional Euclidean space. Such ellipsoids have been proposed for separating patterns in a feature space.

Introduction

Let $x_1, \cdots, x_N$ be points in n-dimensional Euclidean space E_n representing sample patterns of a certain class. Consider the problem of separating these points from points belonging to other pattern classes. We assume that the sample patterns in each class are normally distributed. It then seems reasonable to try to separate the various classes by ellipsoidal domains. This is the approach taken by Rosen in [1]. He proposed finding the ellipsoid of least volume containing the sample patterns of each class. In this paper we show how this can be done. It suffices to consider the points $x_1, \cdots, x_N$ belonging to one of the pattern classes. We assume that not all of these points lie in a hyperplane of dimension less than n.

Let E denote an ellipsoid containing the points $x_1, \cdots, x_N$. The boundary of E can be described by a set of the form

$$\{x \mid (x - c)^T R (x - c) = 1\},$$

where c is the center of the ellipsoid and R is an $n \times n$ positive definite symmetric matrix. We indicate that R is positive definite by writing $R > 0$. Since the x_j's are in E, we have

$$(x_j - c)^T R(x_j - c) \le 1, \qquad j = 1, \cdots, N. \tag{1}$$

The volume of E is proportional to $(\det R^{-1})^{1/2}$. Thus the problem of determining the ellipsoid of least volume containing the points $x_1, \cdots, x_N$ is equivalent to finding a vector $c \in E_n$ and an $n \times n$ positive definite symmetric matrix R which minimize $\det (R^{-1})$ subject to (1). The determinant $\det (R^{-1})$ is a complicated function of the entries r_{ij} of R. Our pattern separation problem is further complicated by the fact that many highly nonlinear constraints must be placed on the r_{ij}'s to ensure that R is positive definite. These constraints are ignored in [1]. The treatment of the problem in [1] is further simplified by replacing $\det (R^{-1})$ with $Tr(R^{-1})$. In this paper we describe a procedure for minimizing both $\det (R^{-1})$ and $Tr(R^{-1})$ subject to $R > 0$ and the constraints (1). We give the details for minimizing $\det (R^{-1})$. $Tr(R^{-1})$ can be minimized in a similar fashion.

For a given $c \in E_n$ we denote by $V(c)$ the minimum value of $\det (R^{-1})$ in the optimization problem described in the previous paragraph. $V^{1/2}(c)$ is proportional to the volume of the ellipsoid of least volume having center c and containing the points $x_1, \cdots, x_N$. Our problem is to determine a c which minimizes $V(c)$. We show that $V(c)$ can be minimized by a gradient technique. Gradient techniques generally require many function evaluations. And, as we have seen, evaluating V for a given c requires the solution of a nonlinear programming problem. It is therefore important to start the minimization process with a good initial guess of an optimal c. If N is large and if the points $x_1, \cdots, x_N$ are normally distributed, as we are assuming, the average of these points provides a good initial guess for c. This initial guess should require very little adjusting to obtain results satisfactory for pattern separation.

Evaluating $V(c)$

Assume that c is fixed and consider the problem of minimizing $f(R) = \det (R^{-1})$ subject to the constraints (1). We first show that this is a convex programming problem on the class of $n \times n$ positive definite matrices R. Let R_1 and R_2 be $n \times n$ positive definite matrices, and let α satisfying $0 \le \alpha \le 1$ be given. It then follows from Theorem 6 on page 63 of [2] that

$$\det (\alpha R_1 + (1 - \alpha)R_2) \ge (\det R_1)^{\alpha}(\det R_2)^{1-\alpha}.$$

Strict inequality holds if $R_1 \neq R_2$ and $0 < \alpha < 1$. In addition, the inequality

$$x^\alpha y^{1-\alpha} \leq \alpha x + (1-\alpha)y$$

relating the arithmetic and geometric means of x and y holds for $x, y \geq 0$ and $0 \leq \alpha \leq 1$. It follows that

$$\begin{aligned} f(\alpha R_1 + (1-\alpha)R_2) &\leq (\det R_1^{-1})^\alpha (\det R_2^{-1})^{1-\alpha} \\ &\leq \alpha \det (R_1^{-1}) + (1-\alpha) \det (R_2^{-1}) \\ &= \alpha f(R_1) + (1-\alpha) f(R_2). \end{aligned}$$

This shows that f is convex on the class of matrices $R > 0$. Minimizing f subject to (1) and $R > 0$ is therefore a convex programming problem with linear constraints. In what follows we make free use of the fact that such problems have dual formulations in terms of Lagrange multipliers.

Let $\lambda_1, \cdots, \lambda_N$ be trial Lagrange multipliers associated with the constraints (1). Since these constraints are inequalities, the λ_j's are ≥ 0. The Lagrangian associated with our convex programming problem is then

$$\det(R^{-1}) + \sum_{j=1}^{N} \lambda_j [(x_j - c)^T R(x_j - c) - 1] . \tag{2}$$

If the λ_j's were chosen properly, we could find the R which solves our convex programming problem by minimizing this expression with respect to $R > 0$. Let λ denote the vector $(\lambda_1, \cdots, \lambda_N)$. The proper choice of λ is the one which maximizes the dual objective

$$g(\lambda) = \min_{R>0} \left\{ \det (R^{-1}) + \sum_{j=1}^{N} \lambda_j [(x_j - c)^T R(x_j - c) - 1] \right\} \tag{3}$$

subject to $\lambda \geq 0$. Maximizing g turns out to be an easy problem. To see this we first find a simple expression for g. For this we require the Hoffman-Wielandt inequality from [3]. According to this inequality, if A and B are two real $n \times n$ symmetric matrices with eigenvalues $\alpha_1 \geq \cdots \geq \alpha_n$ and $\beta_1 \geq \cdots \geq \beta_n$, respectively, then

$$\sum_{i=1}^{n} \alpha_{n-i+1} \beta_i \leq Tr(AB) \leq \sum_{i=1}^{n} \alpha_i \beta_i . \tag{4}$$

Let A denote the $n \times N$ matrix whose jth column is $\sqrt{\lambda_j}(x_j - c)$. We assume that not all the points $\sqrt{\lambda_j}(x_j - c)$ lie in a hyperplane of dimension less than n so that A has rank n. We can then write (2) as

$$\begin{aligned} &\det (R^{-1}) + Tr(A^T RA) - \sum_{j=1}^{N} \lambda_j \\ &= \det (R^{-1}) + Tr(RAA^T) - \sum_{j=1}^{N} \lambda_j . \end{aligned}$$

Let $0 < \sigma_1 \leq \cdots \leq \sigma_n$ denote the eigenvalues of AA^T in increasing order. Let $r_1 \geq \cdots \geq r_n > 0$ denote the eigenvalues of R in decreasing order. It then follows from (4) that

$$\begin{aligned} &\det (R^{-1}) + Tr(RAA)^T - \sum_{j=1}^{N} \lambda_j \\ &\geq (r_1 \cdots r_n)^{-1} + \sum_{i=1}^{n} r_i \sigma_i - \sum_{j=1}^{N} \lambda_j . \end{aligned} \tag{5}$$

Moreover, if

$$AA^T = U \Sigma U^T, \Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_n),$$

is the spectral decomposition of AA^T, equality holds in (5) for

$$R = UDU^T, D = \mathrm{diag}(r_1, \cdots, r_n). \tag{6}$$

Now consider the problem of minimizing the expression on the right side of the inequality in (5) with respect to the variables $r_1, \cdots, r_n$. At the minimum we have

$$-\frac{(r_1 \cdots r_n)^{-1}}{r_i} + \sigma_i = 0, \qquad i = 1, \cdots, n,$$

so that

$$\begin{aligned} r_i &= \frac{(\sigma_1 \cdots \sigma_n)^{1/n+1}}{\sigma_i} \\ &= \frac{(\det AA^T)^{1/n+1}}{\sigma_i}, \qquad i = 1, \cdots, n. \end{aligned} \tag{7}$$

Substituting these values into (5) we obtain

$$\begin{aligned} &\det (R^{-1}) + Tr(RAA^T) - \sum_{j=1}^{N} \lambda_j \\ &\geq (n+1)(\sigma_1 \cdots \sigma_n)^{1/n+1} - \sum_{j=1}^{N} \lambda_j \\ &= (n+1)(\det AA^T)^{1/n+1} - \sum_{j=1}^{N} \lambda_j \end{aligned}$$

with equality if R is given by (6) and (7). Note that this R depends on the λ_j's since the eigenvalues and eigenvectors of AA^T depend on the λ_j's. To indicate this dependence we write $R = R(\lambda)$. It follows that

$$g(\lambda) = (n+1)(\det AA^T)^{1/n+1} - \sum_{j=1}^{N} \lambda_j \tag{8}$$

and

$$R(\lambda) = (\det AA^T)^{1/n+1}(AA^T)^{-1}. \tag{9}$$

Now consider the problem of maximizing g. Since

$$AA^T = \sum_{j=1}^{N} \lambda_j (x_j - c)(x_j - c)^T,$$

a change in one component of λ produces a rank-one change in AA^T. It is easy to compute the changes in $\det (AA^T)$ and

$(AA^T)^{-1}$ due to a rank-one change in AA^T. It is therefore easy to compute the changes in $g(\lambda)$ and $R(\lambda)$ due to a change in one component of λ. All this suggests maximizing g by varying one component of λ at a time. Given an initial value of λ we produce a new value by varying λ_1 to maximize g. We then increase g as much as possible by varying λ_2 in the new value of λ. Continuing in this way we maximize g with respect to each component of λ, choosing the components in order and cyclically. Thus λ_1 follows λ_n. The fact that this procedure converges is proved in [4]. Before we give a mathematical statement of the algorithm we show how to maximize g with respect to one component, say λ_j, of λ.

Suppose λ_j is changed to $\lambda_j + \alpha$, where α is chosen such that $\lambda_j + \alpha \geq 0$. Let $\bar{\lambda}$ and $\bar{A}$ denote the corresponding changes in λ and A. We then have

$$\bar{A}\bar{A}^T = AA^T + \alpha(x_j - c)(x_j - c)^T.$$

In order to evaluate $g(\bar{\lambda})$ by the formula (8) we require $\bar{A}\bar{A}^T$ to be nonsingular. We have

$$\det(\bar{A}\bar{A}^T) = (\det AA^T)[1 + \alpha(x_j - c)^T(AA^T)^{-1}(x_j - c)]. \tag{10}$$

Thus $\bar{A}\bar{A}^T$ is nonsingular if AA^T is nonsingular and

$$\alpha > -\frac{1}{(x_j - c)^T(AA^T)^{-1}(x_j - c)}. \tag{11}$$

We shall see that the value of α which maximizes $g(\bar{\lambda})$ satisfies this condition.

From (8) we see that

$$\begin{aligned} g(\bar{\lambda}) &= (n+1)(\det \bar{A}\bar{A}^T)^{1/n+1} - \sum_{j=1}^{N} \lambda_j - \alpha \\ &= (n+1)(\det AA^T)^{1/n+1} \\ &\quad \times [1 + \alpha(x_j - c)^T(AA^T)^{-1}(x_j - c)]^{1/n+1} \\ &\quad - \sum_{j=1}^{N} \lambda_j - \alpha. \end{aligned}$$

An easy calculation shows that the maximum of $g(\bar{\lambda})$ with respect to α, subject to $\lambda_j + \alpha \geq 0$, occurs at

$$\alpha = \max \begin{cases} \dfrac{[(x_j - c)^T R(x_j - c)]^{n+1/n} - 1}{(x_j - c)^T(AA^T)^{-1}(x_j - c)}, \\ -\lambda_j, \end{cases} \tag{12}$$

where

$$R = (\det AA^T)^{1/n+1}(AA^T)^{-1}.$$

This value of α satisfies (11) and so $\bar{A}\bar{A}^T$ is nonsingular, provided AA^T is nonsingular.

From (12) we see that in order to maximize g with respect to another component of λ we need expressions for det $(\bar{A}\bar{A}^T)$ and $(\bar{A}\bar{A}^T)^{-1}$. The expression for det $(\bar{A}\bar{A}^T)$ is given by (10). The inverse of $\bar{A}\bar{A}^T$ is given by

$$(\bar{A}\bar{A}^T)^{-1} = (AA^T)^{-1} - \frac{\alpha(AA^T)^{-1}(x_j - c)(x_j - c)^T(AA^T)^{-1}}{1 + \alpha(x_j - c)^T(AA^T)^{-1}(x_j - c)}.$$

We can now give a formal statement of our algorithm for maximizing g:

1. Choose a nonnegative vector $\lambda = (\lambda_1, \cdots, \lambda_N)$, and let A_0 denote the $n \times N$ matrix whose jth column is $\sqrt{\lambda_j}(x_j - c)$. Assume that $A_0A_0^T$ is nonsingular. Define $\Delta_0 = \det(A_0A_0^T)$ and $M_0 = (A_0A_0^T)^{-1}$.
2. For $j = 1, 2, \cdots, N$, define recursively

$$\beta_j = (x_j - c)^T M_{j-1}(x_j - c),$$

$$\gamma_j = \Delta_{j-1}^{1/n+1}\beta_j,$$

$$\alpha_j = \max \begin{cases} \dfrac{\gamma_j^{n+1/n} - 1}{\beta_j}, \\ -\lambda_j, \end{cases}$$

$$\Delta_j = \Delta_{j-1}(1 + \alpha_j\beta_j),$$

and

$$M_j = M_{j-1} - \left(\frac{\alpha_j}{1 + \alpha_j\beta_j}\right) M_{j-1}(x_j - c)(x_j - c)^T M_{j-1}.$$

3. Let $\bar{\lambda}_j = \lambda_j + \alpha_j$, $j = 1, \cdots, N$, and define $\bar{\lambda} = (\bar{\lambda}_1, \cdots, \bar{\lambda}_N)$. Compare

$$g(\bar{\lambda}) = (n+1)\Delta_N^{1/n+1} - \sum_{j=1}^{N} \bar{\lambda}_j$$

with

$$g(\lambda) = (n+1)\Delta_0^{1/n+1} - \sum_{j=1}^{N} \lambda_j.$$

If $g(\bar{\lambda}) - g(\lambda)$ is smaller than some preassigned small positive number ε, go to step 4; otherwise set $\lambda = \bar{\lambda}$, $\Delta_0 = \Delta_N$, $M_0 = M_N$, and go to step 2.
4. The algorithm has converged. $V(c)$ is approximately equal to $\Delta_N^{1/n+1}$ and

$$R(\bar{\lambda}) = \Delta_N^{1/n+1} M_N$$

is an approximate solution of min det (R^{-1}) subject to $R > 0$ and (1).

Minimizing $V(c)$

We are going to propose a gradient technique for minimizing the function $V(c)$. First we give a result from convex programming theory which will be useful in evaluating directional derivatives of V.

Let A be an $m \times n$ matrix, and let f be a convex function defined on an open subset X of E_n. Let b be a vector in E_m. Let $\Phi(A)$ denote the value of the minimum in the convex programming problem

$$\text{minimize} \quad f(x)$$
$$\text{subject to } Ax \le b \qquad x \in X. \tag{13}$$

We are going to study how $\Phi(A)$ changes as A changes in a given direction ΔA. For this we need to know that $\Phi(A + \alpha\Delta A)$ is defined for all sufficiently small nonnegative values of α. We make some simple assumptions to guarantee that this holds. It suffices to assume that f is strictly convex on X and that (13) has a solution x^*, and that the region defined by the constraints in (13), with A replaced by $A + \alpha\Delta A$, has a nonempty interior for all $\alpha \ge 0$ and sufficiently small. These assumptions guarantee that x^* is unique and that there is a Lagrange multiplier $\lambda^* \in E_m$ associated with x^*. With these assumptions the following theorem holds.

• *Theorem 1*

If H is any real $m \times n$ matrix, the limit

$$\lim_{\alpha \to 0^+} \frac{\Phi(A + \alpha H) - \Phi(A)}{\alpha}$$

exists and is equal to

$$\max_{\lambda \in \Lambda} \lambda^T H x^*,$$

where Λ is the set of Lagrange multipliers corresponding to x^*. The analog of this theorem for linear programming problems, without the uniqueness assumptions on x^*, is proved in [5]. The extension of that proof to the present situation is straightforward.

This theorem can be used to evaluate the directional derivative of $V(c)$ in any direction Δc. To see this let R be the matrix determined by our algorithm for evaluating $V(c)$. Let λ be the corresponding Lagrange multiplier. R is given by (9) with

$$AA^T = \sum_{j=1}^{N} \lambda_j (x_j - c)(x_j - c)^T.$$

Since the optimum value of a convex programming problem is equal to the optimum value of its dual, we have

$$g(\lambda) = \det R^{-1} = (\det AA^T)^{1/n+1}.$$

It follows from (8) that

$$\sum_{j=1}^{N} \lambda_j = n(\det AA^T)^{1/n+1}.$$

Let I denote the set of indices j for which $(x_j - c)^T \times R(x_j - c) = 1$. Let $\eta = (\eta_1, \cdots, \eta_N) \ge 0$ be any nonnegative vector satisfying $\eta_j = 0$ if $j \notin I$ and the equations

$$\sum_{j \in I} \eta_j (x_j - c)(x_j - c)^T = AA^T. \tag{14}$$

We have

$$n(\det AA^T)^{1/n+1} = TrAA^TR$$
$$= \sum_{j \in I} \eta_j (x_j - c)^T R(x_j - c) = \sum_{j=1}^{N} \eta_j .$$

It follows from (8) that $g(\eta) = g(\lambda)$. This argument shows that the set of Lagrange multipliers corresponding to R is precisely the set of nonnegative vectors satisfying (14) with

$$\eta_j = 0 \quad \text{if } j \notin I. \tag{15}$$

This set of vectors is denoted by $\Lambda(R)$.

To evaluate the directional derivative of $V(c)$ in the direction Δc, it is convenient to write the constraint $(x_j - c)^T \times R(x_j - c) \le 1$ as

$$Tr(x_j - c)(x_j - c)^T R \le 1.$$

This shows that the coefficient of R in this constraint is $(x_j - c)(x_j - c)^T$. If we replace c by $c + \alpha\Delta c$, this coefficient becomes

$$(x_j - c)(x_j - c)^T$$
$$- \alpha[(x_j - c)(\Delta c)^T + (\Delta c)(x_j - c)^T] + O(\alpha^2).$$

It follows from Theorem 1 that

$$\lim_{\alpha \to 0^+} \frac{V(c + \alpha\Delta c) - V(c)}{\alpha}$$
$$= \max_{\eta \in \Lambda(R)} - \sum_{j=1}^{N} \eta_j Tr[(x_j - c)(\Delta c)^T + (\Delta c)(x_j - c)^T]R$$
$$= 2 \max_{\eta \in \Lambda(R)} -(\Delta c)^T R \sum_{j=1}^{N} \eta_j (x_j - c).$$

This formula suggests a procedure for choosing Δc such that

$$V(c + \alpha\Delta c) < V(c) \tag{16}$$

for $\alpha > 0$ and sufficiently small.

First we solve the quadratic programming problem:

$$\text{minimize} \left[\sum_{j \in I} \eta_j (x_j - c)\right]^T R \sum_{j \in I} \eta_j (x_j - c) = \eta^T X^T R X \eta, \tag{17}$$

where X is the $n \times |I|$ matrix whose columns are the vectors $x_j - c, j \in I$, and η is the vector with components $\eta_j, j \in I$. The minimization is taken subject to the constraints

$$\sum_{j \in I} \eta_j (x_j - c)(x_j - c)^T = AA^T, \tag{18}$$
$$\eta_j \ge 0, \quad j \in I.$$

If η^* is a solution of this quadratic programming problem, we take

$$\Delta c = \sum_{j \in I} \eta_j^* (x_j - c) = X\eta^*.$$

It then follows that

$$\lim_{\alpha \to 0^+} \frac{V(c + \alpha \Delta c) - V(c)}{\alpha} = -2(\Delta c)^T R \Delta c. \tag{19}$$

Thus if $\Delta c \neq 0$, (16) holds for $\alpha > 0$ and sufficiently small. The gradient technique we propose for minimizing $V(c)$ is the following:

1. Let c_0 be the mean of the points $x_1, \cdots, x_N$.
2. Given c_k, $k \geq 0$, let

   $$c_{k+1} = c_k + \alpha_k \Delta c_k,$$

   where Δc_k is a descent direction determined by solving the quadratic programming problem (17)–(18) with $c = c_k$, and α_k is the largest number of the form $\gamma(\beta)^m$ such that

   $$V(c_k) - V(c_k + \alpha_k \Delta c_k) \geq \frac{1}{2} (\Delta c_k)^T R_k (\Delta c_k) \alpha_k.$$

   β is a fixed number in the interval (0, 1), γ is a fixed positive number, and m ranges over the integers ≥ 0. R_k is the optimal coefficient matrix for the ellipsoid with center c_k. Equation (19) shows that α_k is well-defined. This rule for selecting α_k is due to Armijo [6].

Let V^* denote the minimum value of V. We then have

$$\begin{aligned} V(c_0) - V^* &\geq \sum_{k=0}^{\infty} \{V(c_k) - V(c_{k+1})\} \\ &\geq \frac{1}{2} \sum_{k=0}^{\infty} (\Delta c_k)^T R_k (\Delta c_k) \alpha_k. \end{aligned}$$

It follows that the sequence $\{(\Delta c_k)\alpha_k\}$ converges to zero. Also, some subsequence of the sequence $\{c_k\}$ converges. However, it sometimes happens in gradient techniques of the type we have described that the sequence of step sizes $\{\alpha_k\}$ converges to zero causing the sequence $\{c_k\}$ to converge to a nonoptimal point. This is called jamming. The following theorem gives a way of testing for convergence.

• *Theorem 2*

If the sequence $\{\Delta c_k\}$ converges to zero, the sequence $\{V(c_k)\}$ converges to V^* and the corresponding sequence of ellipsoids converges to the ellipsoid of least volume containing the points $x_1, \cdots, x_N$.

• *Proof*

Let

$$(x - \hat{c})^T S (x - \hat{c}) \leq 1$$

be any ellipsoid containing the points $x_1, \cdots, x_N$. We need only show that, for any positive number ε, we have $V(c_k) = \det R_k^{-1} < \det S^{-1} + \varepsilon$ for k sufficiently large.

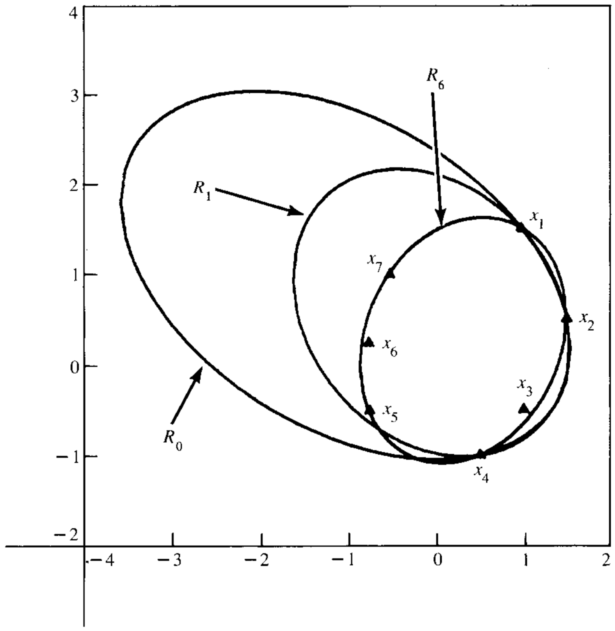

Figure 1 Converging ellipsoids.

Let λ be the Lagrange multiplier occurring in the definition of Δc_k. Then, since $\lambda \geq 0$,

$$\begin{aligned} \det(S^{-1}) &\geq \det(S^{-1}) + \sum_{j \in I} \lambda_j [(x_j - \hat{c})^T S (x_j - \hat{c}) - 1] \\ &= \det(S^{-1}) + \sum_{j \in I} \lambda_j [(x_j - c_k)^T S (x_j - c_k) - 1] \\ &\quad + 2(\Delta c_k)^T S (\hat{c} - c_k) \\ &\quad + (\hat{c} - c_k)^T S (\hat{c} - c_k) \sum_{j \in I} \lambda_j. \end{aligned}$$

Now, if k is so large that

$$2(\Delta c_k)^T S (\hat{c} - c_k) + (\hat{c} - c_k)^T S (\hat{c} - c_k) \sum_{j \in I} \lambda_j > -\varepsilon,$$

we have

$$\begin{aligned} &\det(S^{-1}) \\ &> \min_{\mathcal{J} > 0} \left\{ \det \mathcal{J}^{-1} + \sum_{j \in I} \lambda_j [(x_j - c_k)^T \mathcal{J} (x_j - c_k) - 1] \right\} - \varepsilon \\ &= \det R_k^{-1} + \sum_{j \in I} \lambda_j [(x_j - c_k)^T R_k (x_j - c_k) - 1] - \varepsilon \\ &= \det R_k^{-1} - \varepsilon = V(c_k) - \varepsilon. \end{aligned}$$

This completes the proof.

Table 1 Experimental results.

k	c_k	$V(c_k)$	Δc_k	r_{11}	r_{12}	r_{22}	λ
0	$(-1, 1)$	22.54	$(79.38, -34.56)$	0.1856	0.092	0.2846	(21.39, 1.54, 0, 22.26, 0, 0, 0)
1	$(-0.03, 0.58)$	5.88	$(10.02, -9.31)$	0.4315	0.096	0.4152	(1.87, 4.56, 0, 5.32, 0, 0, 0)
2	$(0.35, 0.21)$	2.66	$(-0.81, 3.88)$	0.7338	−0.1043	0.5270	(1.95, 0, 0, 0, 0.93, 0, 2.42)
3	$(0.33, 0.30)$	2.53	$(-0.92, -3.11)$	0.7722	−0.1481	0.5382	(0, 1.36, 0, 2.39, 1.32, 0, 0)
4	$(0.31, 0.24)$	2.44	$(0.77, 4.36)$	0.7524	−0.140	0.5685	(2.14, 0.78, 0, 0, 0, 0, 1.97)
5	$(0.32, 0.26)$	2.42	$(1.63, 1.31)$	0.7496	−0.1278	0.5723	(0.72, 2.13, 0, 1.97, 0, 0, 0)
6	$(0.32, 0.27)$	2.41	$(-0.007, -0.02)$	0.7605	−0.1343	0.5678	0, 1.45, 0, 2.3, 1.07, 0, 0)

An example

To demonstrate our procedure we take $n = 2$, $N = 7$ and the points $x_1 = (1, 3/2)$, $x_2 = (3/2, 1/2)$, $x_3 = (1, 1/2)$, $x_4 = (1/2, -1)$, $x_5 = (-3/4, -1/2)$, $x_6 = (-3/4, 1/4)$, $x_7 = (-1/2, 1)$. See Fig. 1. The points are chosen so that no one of them is a convex combination of the others. Also, we have chosen c_0 outside the convex hull of the points so that we can observe the algorithm in a situation where a large correction in c_0 has to be made. The first few c's and the corresponding ellipsoids are shown in Fig. 1. In each update of c_k we used

$$\gamma = \frac{5}{\|\Delta c_k\|}, \qquad \beta = 0.8.$$

For the purpose of computing Δc_k, points nearly on the boundary of the kth ellipsoid are treated as being on the boundary. The iterates are shown in Table 1. Since the sequence $\{\Delta c_k\}$ is converging to zero, the corresponding sequence of ellipsoids is converging to the ellipsoid of least volume containing the points $x_1, \cdots, x_N$, by Theorem 2.

Acknowledgments

The author wishes to thank P. Wolfe for several helpful comments and A. A. Munshi for assisting with the numerical results.

References

1. J. B. Rosen, "Pattern Separation by Convex Programming," *J. Math. Anal. Applications* **10,** 123–134 (1965).
2. E. F. Beckenback and R. Bellman, *Inequalities*, Springer-Verlag, Berlin, 1961.
3. A. J. Hoffman and H. W. Wielandt, "The Variation of the Spectrum of a Normal Matrix," *Duke Math. J.* **20,** 37–39 (1953).
4. D. A. D'Esopo, "A Convex Programming Procedure," *Naval Res. Logistics Quarterly* **6,** 33–42 (1959).
5. H. D. Mills, "The Marginal Values of Matrix Games and Linear Programs," *Linear Inequalities and Related Systems,* H. W. Kuhn and A. W. Tucker, Eds., Princeton University Press, Princeton, NJ, 1956.
6. L. Armijo, "Minimization of Functions Having Continuous Partial Derivatives," *Pacific J. Math.* **16,** 1–3 (1966).

Received March 5, 1982; revised June 29, 1982

E. R. Barnes *IBM Research Division, P.O. Box 218, Yorktown Heights, New York 10598.* Dr. Barnes is a research staff member at the Thomas J. Watson Research Center. Since joining IBM in 1968, his research projects have included assignments in pattern recognition and signal processing. Currently he is working on graph-partitioning algorithms for laying out circuits on VLSI chips. Dr. Barnes received his B.S. in mathematics from Morgan State College, Baltimore, Maryland, in 1964 and his Ph.D. in mathematics from the University of Maryland in 1968. He has held visiting teaching and research positions in mathematics and operations research at the Massachusetts Institute of Technology, Columbia University, and Wisconsin University. Dr. Barnes is a member of the Institute of Electrical and Electronics Engineers and the Society for Industrial and Applied Mathematics.