BARYCENTRIC COORDINATES AND THE ORIENTATION OF THE CLASSICAL MIXTURE SURFACE

William S. RAYENS

Department of Statistics, University of Kentucky, Lexington, KY 40506, USA

Abstract

The signs of the barycentric coordinates of a point exterior to a nondegenerate k-simplex in $\mathbb{R}^p$ contain useful information about how that point is positioned relative to the vertices of that simplex. This relationship is certainly not newly observed, with some of the first ideas dating back to Möbius in 1827. However, this article presents some new geometrical results which further quantify the relationship and focuses on applying these new results to help solve the problem of finding the point on a simplex that is closest to a given exterior point. In particular, it is shown that the signs of the barycentrics can be used to immediately identify a potentially large set of facets that could not contain this closest point. Such results have immediate applications to the problem of identifying the components in a chemical linear mixture. Real PCB mixtures are employed to illustrate the new ideas.

1.   General background

The geometry of a nondegenerate k-simplex, or simplicial mixture surface, in $\mathbb{R}^p$ has been exploited to successfully model linear mixtures of highly structured chemical classes, such as polychlorinated biphenyls (PCBs) [1,2]. Typically, the goal of such modelling is to identify the components in an unknown chemical mixture and to estimate their relative amounts. In the model developed by Burdick and Rayens [1], this estimation requires one to identify the point on the mixture surface that is closest to a given exterior point. This section briefly reviews how the geometrical model arises from the chemistry and how the identification problem arises from the geometry. Polychlorinated biphenyl mixtures provide the illustrations.

PCBs that occur in the environment of the United States originate from one or more of nine industrial products known as Aroclors (registered trademark of the Monsanto Corporation). Each of these nine Aroclors can be characterized by a particular chromatogram. Algebraically, of course, a chromatogram is simply a vector in which the variable entries correspond to the relative concentrations of a particular set of constituents. For PCBs, these constituents differ according to the arrangement of chlorine atoms along the carbon chain associated with a biphenyl molecule. Although in theory there are 209 distinguishable arrangements, far fewer are generally available in practice. In the application that will be discussed later in this article, a particular

Aroclor was identified with a vector that reflected the relative concentrations of 93 congeners. Likewise, an environmental or biological mixture of two or more of these Aroclors will correspond to a particular vector of constituent concentrations in $\mathbb{R}^{93}$. This mixture chromatogram is often taken to be a nonnegative weighted average of the chromatograms associated with the component Aroclors that are present in the mixture. These weights represent the relative presence of the nine component Aroclors in the observed mixture. To estimate these, one must jointly consider the mixture chromatogram and the nine component chromatograms. After the estimation is completed, the unknown mixture is said to be classified.

The estimation of these component proportions can be informal, say based on the chemist's experience with the substances at hand. However, the assumption that the mixture chromatogram is a linear combination of the component chromatograms allows more formal reasoning to be applied. For instance, all nonnegative linear combinations of the nine component chromatograms generate an eight-dimensional simplex in $\mathbb{R}^{93}$, with each of the component chromatograms corresponding to a vertex. This simplicial surface is often called a mixture surface. A mixture that is 50% Aroclor 1016 and 50% Aroclor 1221 should admit a chromatogram that, as a vector in $\mathbb{R}^{93}$, is approximately in the middle of the line segment joining these two vertex chromatograms. An equal mix of all nine of the Aroclors would correspond to the centroid of the mixture surface. Thus, the positioning of the mixture chromatogram relative to the vertex chromatograms suggests which Aroclors are actually present in the mixture and in what relative proportions.

From a deterministic point of view, it is thus clear that estimating the mixing weights is equivalent to determining the spatial position of the unknown relative to the vertices. Indeed, if the simplicial model is exact, with the vertices precisely estimated, and the associated laboratory processes without noise, the chromatogram of an unknown mixture must lie interior to or on the surface of the simplex. In practice, however, the chromatogram of such a mixture may not lie inside the simplex. Thus, it cannot be expressed as a nonnegative linear combination of the component chromatograms. What is one to do with such an observation? If it lies "too far" from the simplex, then the integrity of the model or the laboratory procedures may be suspect [6]. However, if it does not lie too far outside, it is plausible to attribute the aberration to "noise". In this case, one could identify this unknown with the chromatogram on the simplicial surface that is closest to that of the exterior mixture. Of course, the projected chromatogram can be written as a nonnegative linear combination of the vertex chromatograms and, hence, the unknown mixture can be classified.

As has been mentioned above, such insights were exploited by Burdick and Rayens. They developed a methodology that entwined discriminant analysis with elementary convex geometry, and arrived at a statistical model which gave formal interpretations to the estimates found as a result of this closest-point identification [1,2,5]. They did not work with a simplex in the original variable space ($\mathbb{R}^{93}$, for the PCB example). Rather, they created another simplex from this original one

which had the vertices "best separated" in a rigorous sense. Their estimates of the mixing weights were provided by the barycentric coordinates of the unknown mixture with respect to the particular mixture surface. Indeed, these coordinates provide a formal orientation of the unknown mixture with respect to the component vertices that were discussed above. Since they are negative when and only when a vector is exterior to that surface, Burdick and Rayens identified any exterior vector with the correspondingly closest vector or point on the surface. Hence, the problem of finding the point on a simplex that is closest to a given exterior point arises naturally and critically in this context.

This article suggests that the signs of the barycentric coordinates of a point can be used to better understand the relative orientation of this point to the simplex. As a result, the problem of finding the point on the simplex that is closest to a given exterior point can be greatly simplified. In particular, it is shown that the signs of the barycentrics can be used to immediately discard a potentially large set of facets that could not contain this closest point. These observations and the remainder of the paper are organized as follows. First, some basic definitions and notations are presented and a straightforward nonlinear programming routine for solving this closest-point problem is discussed. Next, barycentric coordinates are specifically addressed and results pertaining to their signs are proved. It is then shown how these results can be utilized to improve the efficiency of the aforementioned nonlinear programming routine (or any similar routine). Finally, real PCB data are employed in the context of the above-mentioned classification problem to illustrate the improvements that are possible when these new results are invoked.

2. Definitions and a standard algorithm

Intuitively, a k-simplex is a higher-dimensional version of a triangle (2-simplex) and a tetrahedron (3-simplex). A k-simplex S can be defined formally as the convex hull determined by $k + 1$ linear independent points in $\mathbb{R}^p$, $p \geq k$. This set of linearly independent points, say $\{v_1, v_2, \ldots, v_{k+1}\}$, is usually called the vertex set of S. Naturally associated with a k-simplex in $\mathbb{R}^p$ and a point $z \in \mathbb{R}^p$ ($p \geq k$) is a vector of barycentric coordinates given by $\boldsymbol{\beta}^\mathrm{T} = (\beta_1, \ldots, \beta_{k+1})$ where, intuitively, the ith barycentric coordinate β_i represents the "influence" vertex v_i has on the point z. One can calculate β by defining $\mathbf{U}$ to be the $p \times k$ matrix $(v_2 - v_1, \ldots, v_{k+1} - v_1)$, and solving the equation $(z - v_1) = \mathbf{U} (\beta_2, \ldots, \beta_{k+1})^\mathrm{T}$. Since $\mathbf{U}$ is assumed to have full column rank, the solution is, of course, $(\mathbf{U}^\mathrm{T}\mathbf{U})^{-1} \mathbf{U}^\mathrm{T}(z - v_1) = (\beta_2, \ldots, \beta_{k+1})^\mathrm{T}$, with

$$\beta_1 \equiv 1 - \sum_{i=2}^{k+1} \beta_i.$$

The barycentric coordinates of a point are uniquely determined by the simplex in question; and a point z is interior to or on the surface of S if and only if all of the

barycentric coordinates of z relative to S are nonnegative. Indeed, such observations are quite old and the reader is referred to Kelly and Weiss [3] if more details are needed.

Algorithms to find the point on a simplex that is closest to a given exterior point are not difficult to construct. For example, a class of routines, known as "gradient projection methods" can be created by employing the Kuhn–Tucker theorem. This theorem gives necessary conditions for the local minima of a nonlinear (continuously differentiable) function of several variables, subject to certain linear constraints. If, in addition, the function is convex, the conditions are also sufficient to guarantee that any local minimum is an absolute minimum. In the context of this paper, the nonlinear function is simply the squared distance between z and z^*, say $F(z)$, where $z \in S$. Since this function is convex, the Kuhn–Tucker theorem can be used to identify the absolute minimum of F, subject to the linear constraints imposed by requiring z to be in the closure of S.

It should be noted that the linear constraints imposed by the simplex are usually given in the form $a_i^{\mathrm{T}} z + b_i \le 0$, for $i = 1, 2, \ldots, k+1$, and not in the form of a specified vertex set. However, it is simple to translate the vertex specification to these inequality constraints. That is, suppose $\boldsymbol{\beta}^{\mathrm{T}} = (\beta_1, \ldots, \beta_{k+1})$ are the barycentric coordinates of z with respect to S. Following the above notation, it is clear that the requirement that $\beta_j \ge 0$ for $j = 2, \ldots, k+1$ is equivalent to the requirement that $(\mathbf{U}^{\mathrm{T}}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{T}}(z - v_1) \ge 0$, where v_1 is an arbitrary choice of origin. Setting $\mathbf{W} = (\mathbf{U}^{\mathrm{T}}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{T}}$, the requirements (for $i = 2, \ldots, k+1$) translate to $w_{i-1}^{\mathrm{T}}(z - v_i) \ge 0$, where w_j^{T} is the jth row of $\mathbf{W}$, $j = 1, 2, \ldots, k$. For $i = 1$,

$$(1 - (\beta_2 + \ldots + \beta_{k+1})) \ge 0 \quad \text{iff}$$

$$[1 - w_i^{\mathrm{T}}(z - v_1) - \ldots - w_k^{\mathrm{T}}(z - v_1)] \ge 0 \quad \text{iff}$$

$$[w_1^{\mathrm{T}}(z - v_1) + \ldots + w_k^{\mathrm{T}}(z - v_1) - 1] \le 0.$$

Hence, the linear constraints that define the k-simplex can be written in the form $a_i^{\mathrm{T}} z + b_i \le 0$, where

$$a_1^{\mathrm{T}} = \left(w_1^{\mathrm{T}} + \ldots + w_k^{\mathrm{T}} \right) \equiv w_+^{\mathrm{T}}, \quad \text{and} \quad b_1 = -w_+^{\mathrm{T}} v_1 - 1;$$

and for $i \ge 2$,

$$a_i^{\mathrm{T}} = -w_{i-1}^{\mathrm{T}}, \quad \text{and} \quad b_i = w_{i-1}^{\mathrm{T}} v_i.$$

Such a constraint is said to be *active* at z if $a_i^{\mathrm{T}} z + b_i = 0$. Otherwise, the constraint is said to be *inactive*. For a specific z, let $I(z)$ denote the set of all indices corresponding to active constraints at z. Having established this notation, the Kuhn–Tucker theorem can be stated as follows:

THEOREM (KUHN–TUCKER)

Suppose S is a region in $\mathbb{R}^p$ described by $k + 1$ inequalities of the form $a_i^T z + b_i \leq 0$. Suppose, further, that F is a convex continuously differentiable function in S. Then, $F(\hat{z}) \leq F(z)$ for all $z \in S$ if and only if there exist scalars $\mu_i \geq 0$ such that

$$\nabla F + \sum_{i \in I(z)} \mu_i a_i^T = 0^T,$$

where ∇F denotes the gradient of F (taken, by convention, to be a row vector).

The following (useful, though not surprising) note verifies that the form of the constraints just obtained is reasonable and provides a geometrical understanding of the conditions of Kuhn–Tucker.

THEOREM 1

(a)　The vector a_i, as defined above, is perpendicular to the facet of S opposite the ith vertex.

(b)　For any $\lambda > 0$ and $j \in I(z_0)$, $z_0 + \lambda(a_j)$ admits $\beta_j < 0$.

Proof

Some observations are necessary. The simplex S is the convex hull of $\{v_1, \ldots, v_{k+1}\}$. This same simplex, translated to the origin, is represented by the convex hull of $\{v_2 - v_1, \ldots, v_{k+1} - v_1\} \equiv \{u_1, \ldots, u_k\}$. For $i \neq 1$, the facet of S opposite vertex v_i is the convex hull of $\{v_j\}_{j \neq i}$, hence, parallel to the convex hull of $\{u_j\}_{j \neq i-1}$. The facet opposite vertex v_1 is the convex hull of $\{v_j\}_{j=2}^{k+1}$, parallel to the convex hull of $\{v_j - v_2\}_{j=3}^{k+1}$. Notice that $v_{j+1} - v_2 = u_j + v_1 - v_2 = u_j - u_1$. Hence, the facet opposite vertex v_1 is parallel to the span $\{u_j - u_1\}_{j=1}^k$.

(a) Suppose $i \neq 1$. Then, a_i is perpendicular to the facet opposite v_i if a_i is perpendicular to each of the generators in the set $\{u_j\}_{j \neq i-1}$. Clearly,

$$(U^T U)^{-1} U^T U = I(k) = \begin{bmatrix} w_1^T u_1 & \cdots & w_1^T u_k \\ \vdots & & \vdots \\ w_k^T u_1 & \cdots & w_k^T u_k \end{bmatrix},$$

where $I(k)$ is the $k \times k$ identity matrix. It follows that

$$a_i^T u_j = -w_{i-1}^T u_j = \begin{cases} 0 & \text{if } j \neq i-1, \\ 1 & \text{if } j = i-1. \end{cases}$$

Suppose $i = 1$. Then

$$a_1^T(u_j - u_1) = \sum_{i=1}^{k} w_i^T u_j - \sum_{i=1}^{k} w_i^T u_1$$

$$= (0 + \ldots \underset{(i=j+1)}{-1 +} \ldots 0) - (0 - 1 + 0 + \ldots + 0)$$

$$= 0.$$

Hence, a_1 is perpendicular to a linear subspace parallel to the facet opposite v_1, and thus perpendicular to this facet.

(b) This part is trivial:

$$\beta_j = -a_j^T[z_0 + \lambda a_j)] - b_j$$

$$= a_j^T z_0 + b_j) + \lambda(a_j^T a_j)$$

$$= -(a_j^T z_0 + b_j) - \lambda \|a_j\|^2$$

$$= -\lambda \|a_j\|^2 \quad (\text{since } j \in I(z_0))$$

$$< 0. \qquad \qquad \qquad \square$$

Hence, as expected, the a_i^T vectors are to be interpreted as outward normal directions. Further, the point on S, say $\hat{z}$, which is closest to a given point z_0 can be characterized as the unique point at which the negative gradient of F can be written as a nonnegative linear combination of these normal directions. A trivial calculation establishes the negative gradient of F and any point z^* to be a scalar multiple of the vector from z^* to z_0.

This geometrical intuition can be used in conjunction with a standard proof to Kuhn–Tucker to develop a numerical algorithm for finding this closest point. Russell [8], for example, outlines a method for solving this problem for general convex polytopes, and his ideas are easily adapted to the special case of a k-simplex. For instance, a convergent procedure can be constructed by producing a sequence of points in S, $\{z_j\}$ generated by the recursion relation $z_{j+1} = z_j + \lambda_j r_j$, where r_j is a *feasible direction* at z_j and λ_j is the *step parameter* which determines how far one should move in the feasible direction. It can be shown that a direction is feasible at z_j if $z_j + \lambda_j r \in S$ and $F(z_j + \lambda_j r) < F(z_j)$ for small positive values of λ_j. While the choice of λ_j is not always an easy task, it is actually straightforward in the case of a linear–quadratic objective function F, as is the squared-distance function. The reader is referred to Russell for details as to how r_j and λ can be effectively chosen. Note that each iteration in this algorithm requires essentially the same type of calculation, namely the determination of r and λ. Clearly, this procedure will try to identify points that move successively in the direction of the projection

z

vertex 2

KT satisfied

Start

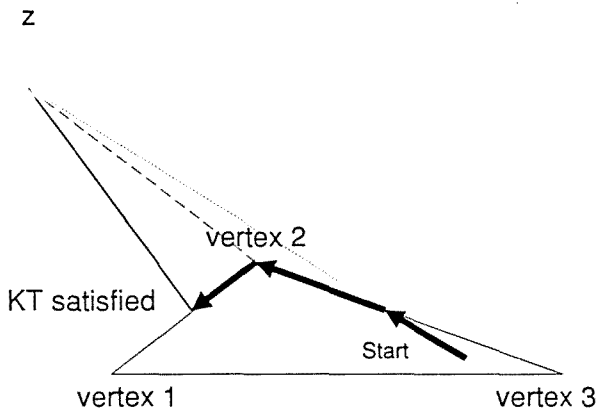vertex 1   vertex 3

Fig. 1. Geometrical basis for algorithm.

of the negative gradient onto the facet of the simplex that z_j is on (hence the term "gradient projection technique"). The arrows in the diagram of fig. 1 indicate how the technique would locate a typical closest point in three iterations.

3.   The signs of the barycentrics

Certainly, the task of trying to find this closest point could be simplified by having some formal method for determining how the exterior point is positioned relative to the vertices of the simplex. That is, if a point is clearly identifiable as isolated from certain facets, then it makes no sense to allow any procedure to look there for the closest point. To a useful extent, this idea of orientation can be properly quantified. The following result, which is quite old, helps to provide the necessary motivation.

RESULT

Suppose z_0 admits barycentric coordinates $(\beta_1, \ldots, \beta_{k+1})$ with a certain subset of these, say $\{\beta_1, \beta_2, \ldots, \beta_q\}$, without loss of generality, all being negative. Denote by H_i the hyperplane determined by the facet of the simplex opposite the vertex v_i, $i = 1, 2, \ldots, q$. Each of these hyperplanes will divide $\mathbb{R}^k$ into half-spaces and z_0 will be separated from the simplex in the sense that z_0 and S will lie in opposite half-spaces.

Figure 2 illustrates this relationship between the signs of the barycentrics of a point and its relative position to a 2-simplex (triangle). The separating property that characterizes this relationship, flagged by the presence of negative barycentric coordinates, can be quite useful. For example, the following theorem can now be proved.
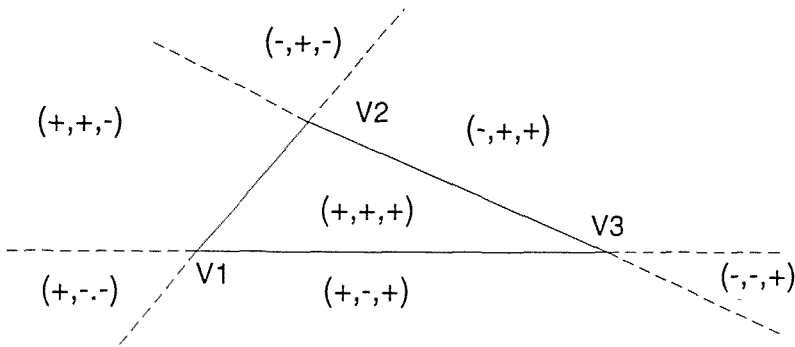
Fig. 2. Relationship of signs to orientation.

THEOREM 2

Suppose $z \in \mathbb{R}^p \sim S$ with barycentric coordinates $(\beta_1, \ldots, \beta_{k+1})$, and suppose without loss of generality that $\beta_1, \beta_2, \ldots, \beta_t \geq 0$ and $\beta_{t+1}, \ldots, \beta_{k+1} < 0$. If $\mathbf{F}_j$ denotes the $(g-1)$-dimensional facet opposite vertex v_j, then the point on S that is closest to z must lie on an $\mathbf{F}_j$ where $t+1 \leq j \leq k+1$.

Proof

It will be necessary to set up some notation first. Let $\mathbf{H}_r$ denote the unique hyperplane determined by the $(g-1)$-dimensional facet $\mathbf{F}_r$ opposite vertex v_r (called a "face-hyperplane"). Use $\mathbf{\Delta}^-$ to denote the set of all face-hyperplanes that are opposite vertices that admit a negative barycentric. That is, $\mathbf{\Delta}^- \equiv \{\mathbf{H}_{t+1}, \mathbf{H}_{t+2}, \ldots, \mathbf{H}_{k+1}\}$. Finally, let c be any point in the closure of S, and let $\mathbf{S}$ denote the line segment form z to c, parameterized as $\{\mathbf{Q}_\alpha = (1-\alpha)z + (\alpha)c : \alpha \in [0, 1]\}$. $\mathbf{S}$ must cross all of the face-hyperplanes in $\mathbf{\Delta}^-$ before reaching c; denote $\mathbf{S} \cap \mathbf{H}_j$ by $\mathbf{Q}_{\alpha_j}$. In fact, it will be shown that if $\mathbf{H}_i \in \mathbf{\Delta}^-$ and $0 \leq \alpha_j \leq \alpha_i \leq 1$ for all j between $t+1$ and $k+1$, then $\mathbf{S} \cap \mathbf{H}_i \equiv z^* \in S$.

Suppose $z^* \notin S$. Then there exists a face-hyperplane which separates z^* from S; call this hyperplane $\mathbf{H}_m$. It is clear that $\mathbf{H}_m \neq \mathbf{H}_i$. Hence, there exists an $\alpha_m \in [0, 1]$ so that $z_m \equiv \mathbf{Q}_{\alpha_m} = \mathbf{S} \cap \mathbf{H}_m$ and $\alpha_i < \alpha_m \leq 1$. Also, $\mathbf{S} \notin \mathbf{H}_m$, it can intersect $\mathbf{H}_m$ at most once; thus, it must be that z and z^* are in the same (closed) half-space determined by $\mathbf{H}_m$ (recall, $z = \mathbf{Q}_0$, $z^* = \mathbf{Q}_{\alpha_i}$, and $z_m = \mathbf{Q}_{\alpha_m}$, with $0 \leq \alpha_i < \alpha_m \leq 1$). Hence, $\mathbf{H}_m \in \mathbf{\Delta}^-$, separating z from S, with $\alpha_i < \alpha_m$. This is a contradiction, so it must be that $z^* \in S$. □

This result notes, of course, that any line segment from a point $z \notin S$ to a point $c \in S$ clearly has to cross all hyperplanes opposite vertices where the barycentric coordinates of z corresponding to those vertices are negative. The result then shows that the point where the last such hyperplane it crosses is, in fact, a point on the simplex. The following lemma is simply a special case of this theorem.

LEMMA

Suppose $z \in \mathbb{R}^p \sim S$ with barycentric coordinates $(\beta_1, \ldots, \beta_{k+1})$, and suppose also that $\beta_i \geq 0$ for all $i \neq j$, and $\beta_j < 0$. If $\mathbf{F}_j$ denotes the $(g - 1)$-dimensional facet of S that is opposite v_j, then the point on S that is closest to z must lie on $\mathbf{F}_j$.

The point of the lemma is obvious. If z admits a vector of barycentric coordinates in which only β_j is negative, then it is certain that the closest point lies on the facet of S opposite v_j. Hence, one would want to start looking on this facet (as opposed to any of the other k facets). Theorem 2 is simply a generalization of this idea, and its use if equally clear. The facet which contains the closest point must correspond to the convex hull of a set of k vertices with exactly t of these admitting nonnegative barycentrics for z. As an example, suppose the barycentric coordinates of a point z relative to some 4-simplex with vertices $\{v_1, v_2, v_3, v_4\}$ admits the signs $(+, -, -, +)$. Since v_1 and v_4 are positive, it is certain that the point on the simplex which is closest to z must lie on one of the two facets formed by the vertices $\{v_1, v_2, v_4\}$ or $\{v_1, v_3, v_4\}$.

Clearly, theorem 2 is most useful when there are a lot of vertices involved and only a few negative barycentrics have been admitted. In such a case, one can eliminate a large number of facets from the search for the closest point. Even when the set of "candidate facets" for containing the closest point is large, one can, at the very least, intelligently choose a facet on which to start the search. It should be pointed out that it is not true that the closest point has to lie on the lower-dimensional facet (or "edge") determined by only the vertices that admit positive barycentrics. Counter-examples are easy to construct. However, it will be seen below that such edges are often wise places to look.

4. Application and some comparisons

As was discussed in the introduction, an irregular k-simplex arises naturally in chemometrics when one is trying to decompose a physical mixture having $k + 1$ potential constituents. The methodology that has been mentioned, due to Burdick and Rayens, modelled the unknown compound x as an observation from a multivariate normal population having mean $\sum_{i=1}^{g} \alpha_i \mu_i$ and covariance matrix $\mathbf{\Sigma}$, where μ_i is the (true) unknown mean of the ith group, $\mathbf{\Sigma}$ is a covariance matrix common to all g groups, and the α_i's are the unknown mixing proportions that one desires to estimate. The basic steps involved in the construction of the Burdick/Rayens simplex model can be roughly summarized as follows:

(i) On the basis of some training set $\mathbf{Y}_{n \times p}$, perform a standard linear discriminant analysis, producing a $(g - 1)$-simplex defined by the vertex set $\{\bar{z}_i\}_{i=1}^{g}$, where $\bar{z}_i$ is the sample mean of the ith group of discriminant scores (calculated from $\mathbf{Y}$).

(ii)   Given an unknown observation y, calculate the corresponding discriminant score z. Find the point, say $\hat{z}$, on the simplex that is closest to z. Properly quantified, the barycentric coordinates of $\hat{z}$ relative to the simplex will represent maximum likelihood estimates of the α_i's. Also, once the α_i's are estimated, the mixture is said to be classified.

Although there are some admittedly naive assumptions that accompany this model, it has been applied quite successfully to real PCB and dioxin data. For details on the model development, formal interpretations of the estimated mixing proportions and real-data applications, the reader is referred to Clayton et al. [2], Burdick and Rayens [1], and Rayens [5–7]. It is clear from (ii) that the closest-point problem has to be solved every time this model is invoked to classify an unknown mixture. Hence, insight into the orientation of an unknown relative to the mixture surface would prove particularly helpful to users of that methodology. In their PCB application, Burdick and Rayens were dealing with $g = 9$ groups (corresponding to the nine Aroclors), and the simplex that resulted in step (ii) was an 8-simplex in $\mathbb{R}^8$. That is, it was the convex hull of nine linearly independent vertices, each of which was a point in $\mathbb{R}^8$. These vertices are to be identified with Aroclors 1016, 1221, 1232, 1242, 1248, 1254, 1260, 1262, and 1268, respectively. Hence, there are nine eight-dimensional facets on this simplex.

In order to test the effectiveness of their model, Burdick and Rayens had access to 38 runs on a pseudo-unknown mixture of Aroclors 1016, 1254, and 1260. Owing to variability in the data, and imperfections in the modelling, all 38 of these unknowns fell outside the simplex mentioned in (i). The actual distribution of the signs of the barycentrics for these 38 runs is shown in table 1. It is clear, for example, that the point on the simplex closest to the first observed mixture is on one of the four 8-dimensional facets partially formed by vertices 1, 2, 6, 7, and 8. In fact, the overall distribution, in conjunction with theorem 2, suggests that any routine that seeks to solve the closest-point problem for these unknowns should only look on these four facets, and ignore the other five, which is a significant reduction in the scope of the search. In lieu of developing a new routine, however, one could, at the very least, use this new information to effectively initialize an existing routine. For example, since the replications on the PCB mixture appear to be isolated from the vertices corresponding to Aroclors 1232, 1242, 1248, and 1268, such a routine should be started away from the five facets that contain all of these vertices. In fact, a "correct" start alone can significantly reduce the total number of iterations required by this procedure to "classify" the 38 runs. Since Burdick and Rayens used the quadratic programming procedure that was discussed above to identify these points, it will be used to construct some comparisons. Specifically, this algorithm was started at each of six different places:

   S1: the centroid of the simplex (used by Burdick and Rayens).

   S2: a randomly chosen point in the closure of the simplex.

Table 1

Distribution of signs

	Aroclor								
Run	1016	1221	1232	1242	1248	1254	1260	1262	1268
1	+	+	−	−	−	+	+	+	−
2	+	+	−	+	−	+	−	+	−
3	+	+	−	−	−	+	−	+	−
4	+	+	−	+	−	+	+	+	−
5	+	−	+	−	−	+	−	+	+
6	+	+	−	−	−	+	+	+	−
7	+	+	−	+	−	+	+	+	−
8	+	+	−	+	−	+	+	+	−
9	+	+	−	+	−	+	+	+	−
10	+	+	−	−	−	+	+	+	−
11	+	−	+	−	−	+	−	+	−
12	+	+	−	−	−	+	+	+	−
13	+	−	+	−	−	+	+	+	−
14	+	+	−	+	+	+	+	+	+
15	+	+	−	−	−	+	+	+	−
16	+	+	−	+	−	+	+	+	−
17	+	+	−	+	−	+	+	+	−
18	+	+	−	−	−	+	+	+	−
19	+	+	−	−	−	+	+	+	−
20	+	+	−	−	−	+	+	+	−
21	+	+	−	−	−	+	+	+	−
22	+	+	−	−	−	+	+	+	−
23	+	+	−	−	−	+	+	+	−
24	+	+	−	−	−	+	+	+	−
25	+	+	−	−	−	+	+	+	−
26	+	+	−	−	−	+	+	+	−
27	+	+	−	−	−	+	+	+	−
28	+	+	−	−	−	+	+	+	−
29	+	+	−	−	−	+	−	+	−
30	+	+	−	−	−	+	+	+	−
31	+	+	−	−	−	+	+	+	−
32	+	+	−	+	−	+	+	+	−
33	−	+	−	+	+	+	+	+	+
34	−	+	−	+	+	+	+	+	+
35	+	−	+	−	+	+	+	+	−
36	−	+	−	+	+	+	+	−	+
37	−	+	−	+	−	+	+	−	+
38	−	+	−	+	−	+	+	−	+
Total minuses	5	4	34	24	33	0	5	3	31

Table 2

Number of steps to convergence

Run	S1	S2	S3	S4	S5	S6
1	7	7	3	3	6	7
2	6	8	2	5	7	9
3	6	7	2	3	5	6
4	6	7	2	3	5	6
5	6	6	4	5	7	7
6	6	6	2	2	5	6
7	6	6	2	3	5	6
8	7	7	3	4	6	7
9	7	7	3	4	6	7
10	6	6	2	2	5	6
11	7	9	3	5	7	7
12	6	6	2	2	5	6
13	6	6	2	4	6	6
14	4	4	4	3	3	6
15	6	6	2	2	5	6
16	6	6	2	4	6	6
17	6	7	2	3	5	6
18	6	7	2	2	5	6
19	6	7	2	2	5	6
20	6	7	2	2	5	6
21	6	7	2	2	5	6
22	6	7	2	2	5	6
23	7	7	3	3	6	7
24	6	7	2	2	5	6
25	6	7	2	2	5	6
26	6	6	2	2	5	6
27	7	7	3	3	6	7
28	7	8	3	3	6	7
29	6	8	2	3	5	6
30	7	7	3	3	6	7
31	7	7	3	3	6	7
32	7	7	3	4	6	7
33	2	2	4	5	1	11
34	2	2	3	4	1	12
35	6	6	2	5	6	7
36	4	4	5	2	3	11
37	3	3	3	2	2	8
38	3	3	3	5	2	7
Avg.	5.8	6.2	2.6	3.1	5.0	6.9

S3: a point on a lower-dimensional facet determined by considering those vertices which admitted the highest concentration of pluses over all 38 of the replications. This starting point was then used for all of the 38 points.

S4: a point on a lower-dimensional facet determined by considering only those vertices which admitted positive barycentrics for a given replication. Hence, the decision as to where the procedure should be started was made individually for each run.

S5: a point on the eight-dimensional facet corresponding to the greatest concentration of pluses.

S6: an isolated point (i.e. a poor choice).

The distribution of +'s and −'s in table 1 suggests that for starting S3 one could choose a point on the lower-dimensional facet determined by vertices 1, 2, 6, 7, and 8 (Aroclors 1016, 1221, 1254, 1260, and 1262). If the procedure is initiated as directed by S4, then it would start on the edge formed by these same vertices for the first observation, but would choose the edge formed by the vertices 1, 2, 4, 6, and 8 for observation 2, etc. To start at S5, one might choose the facet determined by vertices 1, 2, 6, 7, and 8 in conjunction with 4, 5, and 9 (leaving 3 out since it admitted the most negatives, overall). Likewise, it is clear from table 1 that starting on the "edge" formed by vertices 3, 4, 5, and 9 would constitute a poor place to start, since this edge is isolated away from the data.

The gradient projection routine outlined above was initiated at each of these six places and executed for the 38 runs. In each case the number of iterations needed until convergence was recorded. The results are shown in table 2. It is clear from the table that starting with the centroid of the simplex is not a bad idea. Since the above-mentioned procedure will follow the negative gradient of F to the surface of the simplex, then one is somewhat assured of being oriented correctly after the second step. Certainly this can be a better strategy than unintelligently choosing a place on the surface to start, as the average under the "Poor" column supports. Also, while beginning on the eight-dimensional facet that omitted vertex 3 yielded a better performance than starting at S1, S2, or S5, it is clear that starting on the lower-dimensional edges suggested by the positives is even more efficient. It is possible to make some probabilistic statements about why such an edge tends to be a good place to start, even though it is not true that the closest point has to lie on that edge. However, the basic intuition is clear in their absence.

5.   Conclusions

The principal purpose of this article was to point out that there is information contained in the signs of the barycentric coordinates that is a potentially useful addition to one's understanding of the simplicial mixture surface. This new information

is obtained at little or no expense, since calculating the barycentric coordinates for N data points amounts to the inversion of one symmetric matrix. The practical need to address this problem arose from the author's involvement in the application presented in the paper, namely the classification of linear mixtures. Although this application is of particular interest to chemometricians, using the results proved in this article to efficiently start existing nonlinear routines for achieving this classification does not exhaust the potential worth of the ideas. One could, presumably, construct an algorithm which uses only the signs of the barycentrics to solve the closest point problem. It is also worth noting that a procedure for solving this problem that is of current interest to some numerical analysts, due to Iusem and De Pierro [7], relies on simultaneous projections on all of the face-hyperplanes. The results in this paper suggest that these ideas could be improved upon by dismissing from their algorithm those face-hyperplanes that cannot contain the closest point. However, the purpose of this paper was not to develop a completely new routine for solving the closest-point problem, but rather to highlight the information that is in the signs of barycentric coordinates for practitioners who routinely work with a linear mixture surface.

References

[1]   D. Burdick and W. Rayens, J. Chemometrics 1(1986)157.
[2]   C. Clayton, E. Pellezzari, W. Rayens and D. Burdick, RTI Final Report – EPA contract 68-01-6826 (1986).
[3]   P. Kelly and M. Weiss, *Geometry and Convexity* (Wiley, New York, 1979).
[4]   A. Iusem and A. De Pierro, SIAM J. Control Optim. 2(1987)231.
[5]   W. Rayens, Ph.D. Thesis, Department of Mathematics, Duke University (1986).
[6]   W. Rayens, J. Chemometrics 2(1988)121.
[7]   W. Rayens, J. Chemometrics 4(1990)159.
[8]   D. Russell, *Optimization Theory* (Benjamin, New York, 1970).