Synthesis of Nonlinear Decision Boundaries by Cascaded Threshold Gates

JAMES A. CADZOW, MEMBER, IEEE

Abstract—This paper presents a technique for generating a nonlinear decision surface that separates two finite classes of elements in n-dimensional real space. It incorporates the principle of cascaded threshold gates in evolving the required separation surface. At each stage of the design, an additional cascaded gate is added which will separate at least one new element. Necessary and sufficient conditions are developed whereby one may readily determine if two new elements may be properly separated by the addition of one cascaded threshold gate. The results are illustrated by means of numerical examples.

Index Terms—Cascaded threshold gates, classification of two misclassified elements, linear separable sets, nonlinear decision boundaries.

I. INTRODUCTION

CLASSIFICATION techniques for separating patterns into two classes have recently received much attention. There exist many algorithms which generate satisfactory classification systems when the two classes are finite and linearly separable. Unfortunately, most classification problems do not lend themselves to linear separation. In such cases, a nonlinear decision boundary is required to effect the desired classification. Efficient and simple means for obtaining the nonlinear decision boundary are essential.

It has been demonstrated that it is always possible to generate a nonlinear decision boundary which will properly separate two classes, each of which contains a finite number of elements. The nonlinear decision boundary is obtained by cascading a number of standard threshold gates. This method is attractive since it yields a classification system which may be instrumented by simple threshold gates. Its main drawback is that, in the worst case, one threshold gate is required for each point to be separated. This worst case will not be met in general, but that possibility motivates a search for more efficient synthesis techniques.

This paper presents a method, using the structure of cascaded threshold gates, which extends the concepts presented in Akers. It is shown that if it is possible to properly classify two additional points by the addition of one threshold gate, the parameters characterizing this gate may always be obtained directly. By this method, it is possible to synthesize a nonlinear decision boundary which requires relatively few threshold gates. Numerical examples are included to demonstrate the efficiency of this method in comparison to that of Akers.

II. SEPARATION PROBLEM

The elementary decision making process of linear separation may be easily implemented by standard threshold gates. In this paper a threshold gate will be defined as a mapping from a real n-tuple \((x_1, x_2, \ldots, x_n)\) to the set \((1, -1)\) as follows:

\[
G(x_1, x_2, \ldots, x_n) = \begin{cases} 1 & \text{if and only if } \sum_{j=1}^{n} w_j x_j > T \\ -1 & \text{if and only if } \sum_{j=1}^{n} w_j x_j < T \\ \text{undefined} & \text{if } \sum_{j=1}^{n} w_j x_j = T \end{cases}
\]

where the \(w_i\) and \(T\) are real numbers called weights and threshold, respectively.

Linear separation requires the determination of a set of weights \((w_i)\) and threshold \(T\) such that the threshold gate separates the two sets \((X_i)\) and \((Y_i)\), that is

\[
G(X_i) = 1 \quad \text{for } i = 1, 2, \ldots, m \\
G(Y_i) = -1 \quad \text{for } i = 1, 2, \ldots, p
\]

where

\[
X_i = (x_{i1}, x_{i2}, \ldots, x_{in}) \quad \text{for } i = 1, 2, \ldots, m \\
Y_i = (y_{i1}, y_{i2}, \ldots, y_{in}) \quad \text{for } i = 1, 2, \ldots, p
\]

If it is possible to find a set of \(w_i\) and \(T\) such that (2) is satisfied then the two sets \((X_1, X_2, \ldots, X_m)\) and \((Y_1, Y_2, \ldots, Y_p)\) are said to be linearly separable. Unfortunately, such a separation is not always possible. This paper employs the characteristics of standard threshold gates so that such devices may be used to separate sets which are not linearly separable.

To aid this development, some well-known results are reviewed.

It will be beneficial to consider the linear separability problem from a geometrical viewpoint. The \(n\)-tuples which are to be separated (3) are in fact points in \(n\)-dimensional space. If the two sets are linearly separa-
rable, then it is possible to find a hyperplane
\[ \sum_{j=1}^{n} w_j x_{ij} = T \] (4)
in this space such that the set of points \((X_1, X_2, \ldots, X_n)\) lie on one side of hyperplane (4) and the set \((Y_1, Y_2, \ldots, Y_p)\) lie on the opposite side.

Some modifications are now made without loss of generality. Add to each of the points in \((X_1, Y_1)\) an \(n+1\)st component which is set equal to one. The problem of finding the threshold value \(T\) is then replaced by that of finding the weight \(w_{n+1}\) (i.e., \(w_{n+1} = -T\)) such that
\[ \sum_{j=1}^{n+1} w_j x_{ij} > 0, \quad i = 1, 2, \ldots, m \] (5)
and
\[ \sum_{j=1}^{n+1} w_j y_{ij} < 0, \quad i = 1, 2, \ldots, p \] (6)
with \(x_{i,n+1} = 1\) and \(y_{i,n+1} = 1\).

Inequality (6) may be changed to
\[ \sum_{j=1}^{n+1} w_j (-y_{ij}) > 0, \quad i = 1, 2, \ldots, p. \] (7)

Now reformulate the separation problem; two sets \((X_1, X_2, \ldots, X_n)\) and \((Y_1, Y_2, \ldots, Y_p)\) are linearly separable if there exists a set of weights \((w_1, w_2, \ldots, w_{n+1})\) such that
\[ G(\overline{X}_i) = +1 \quad \text{for} \quad i = 1, 2, \ldots, m, \]
\[ G(-\overline{Y}_i) = +1 \quad \text{for} \quad i = 1, 2, \ldots, p. \]

Thus, without any loss of generality, the original problem of separating two sets of points \(\{X_i\}, \{Y_i\}\) in \(n\)-dimensional space is equivalent to the determination of a set of weights \((w_1, w_2, \ldots, w_{n+1})\) for which
\[ \sum_{j=1}^{n+1} w_j x_{ij} > 0 \quad \text{for} \quad i = 1, 2, \ldots, m + p, \] (8)
where
\[ X_{m+i} = -Y_i \quad \text{for} \quad i = 1, 2, \ldots, p \]
and
\[ x_{i,m+1} = \begin{cases} 1 & \text{for} \ i = 1, 2, \ldots, m \\ -1 & \text{for} \ i = m+1, m+2, \ldots, m+p. \end{cases} \]

Suppose there exists a set of weights \((w_i)\) for which (8) is satisfied. Let
\[ \xi = \min \sum_{i=1}^{m+p} w_j x_{ij} \]
\footnote{The \((n+1)\)-dimensional vectors \(\overline{X}_i, \overline{Y}_i\) are related to the original \(n\)-dimensional vectors \(X_1, Y_1\) by
\[ \overline{X}_i = (X_1, 1) = (x_{1i}, x_{2i}, \ldots, x_{ni}, 1) \]
\[ \overline{Y}_i = (Y_1, 1) = (y_{1i}, y_{2i}, \ldots, y_{ni}, 1). \]}

then the inequalities (8) may be written as
\[ \sum_{j=1}^{n+1} w_j^* x_{ij} \geq 1 \quad \text{for} \quad i = 1, 2, \ldots, m + p. \] (9)

The line separation problem is equivalent to that of finding a hyperplane
\[ \sum_{j=1}^{n+1} w_j x_j = 1 \]
which separates the origin from the set of points \((\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_{m+p})\) in \((n+1)\)-dimensional space. If such a hyperplane exists, the two classes \(\{X_i\}\) and \(\{Y_i\}\) are linearly separable.

The changes made in the original separation problem, leading to inequalities of the form (9), were made for mathematical convenience and will be employed hereafter.

III. Nonlinear Separation Problem

As was stated previously, it may not always be possible to separate two classes of points by a hyperplane. In this case, more sophisticated decision surfaces are required to accomplish the desired separation. Perhaps the most basic extension of the linear decision surface is the piecewise linear decision surface. Such decision surfaces consist of segments of a number of hyperplane surfaces which when taken together form a piecewise linear surface. Since each segment of the decision surface is a portion of a hyperplane, each segment may be generated by a threshold gate. It will now be shown that any two finite classes of points may be separated by a nonlinear decision surface and it is always possible to cascade standard threshold gates to perform the desired separation. The modifications, discussed at the end of Section II, which lead to the separation criterion (9) are assumed to have been made.

A further modification will now be made. Namely, an \((n+2)\)th component \(x_{i,n+2}\) is added to each of the \((n+1)\)-dimensional points \(\{\xi_1, \xi_2, \ldots, \xi_{m+p}\}\) so that
\[ \sum_{j=1}^{n+2} x_{ij}^2 = R^2 \quad \text{for} \quad i = 1, 2, \ldots, m + p. \] (10)
that is, the \( m+p \) points lie on a hypersphere of radius \( R \) in \( (n+2) \)-dimensional space. This is always possible since

\[
x_{in+2} = \sqrt{R^2 - \sum_{j=1}^{n+1} x_{ij}^2}
\]  

(11)

and \( R \) may be selected large enough so that \( x_{in+2} \) as given by (11) is real.

After the modifications specified above and in Section II have been made, it will simplify the notation if the resulting set of points are denoted by \( X_1, X_2, \ldots, X_{m+p} \). Further, they will be considered as \( n \)-dimensional rather than \( (n+2) \)-dimensional vectors, again a notational simplification.

To explain the philosophy of this approach, assume that a threshold gate \( G_1 \) has been determined which separates at least one of the \( m+p \) points \( \{X_i\} \), that is

\[
\sum_{j=1}^{n} w_j x_{ij} \geq 1 \quad \text{for some } 1 \in (1, 2, \ldots, m+p). \]  

(12)

Now add \( G_1(X_i) = \pm 1 \) as the \( (n+1) \)st measurement for each point \( X_i \). Repeat this process. In this manner, a system of cascaded threshold gates is generated (see Fig. 1).

A theorem, first presented by Akers [1], will now be given.

**Theorem 1:** Given \( k \) points \( X_1, X_2, \ldots, X_k \) where

\[
X_i = (x_{i1}, x_{i2}, \ldots, x_{in}) \quad \text{for } i = 1, 2, \ldots, k
\]

and

\[
\sum_{j=1}^{n} x_{ij}^2 = R^2 \quad \text{for } i = 1, 2, \ldots, k
\]  

(13)

if there exists a decision function \( W(\cdot) \) such that

\[
X(X_i) = +1 \quad \text{for } i = 1, 2, \ldots, k - 1
\]

\[
W(X_k) = -1
\]

then a linear decision \( W^*/(w_1^*, w_2^*, \ldots, w_k^*, \omega^*) \) can always be found such that

\[
\sum_{j=1}^{n} w_j x_{ij} + \omega^* W(X_i) \geq 1 \quad \text{for } i = 1, 2, \ldots, k. \]  

(14)

In other words, if a decision function (linear or non-linear), \( W(\cdot) \), exists which separates points \( X_1, X_2, \ldots, X_{k-1} \) but not \( X_k \), then a linear decision function can be found, having as inputs \( x_{i1}, x_{i2}, \ldots, x_{in} \), plus \( W(X_i) \) which will separate all \( k \) points.

**Proof:** A proof similar to Akers [1] will be given. Let

\[
w_j^* = \frac{2}{L + D} x_{kj} \quad \text{for } j = 1, 2, \ldots, n
\]  

(15)

\[\omega^* = \frac{L - D}{L + D}\]

(16)

where \( D = \min_{i=1,2,\ldots,k-1} \left[ \sum_{j=1}^{n} x_{ij}x_{kj} \right] \)

and

\[L = X_k \cdot X_k = R^2,\]

then for \( i = 1, 2, \ldots, k - 1 \) (14) becomes

\[
\frac{2}{L + D \sum_{j=1}^{n} x_{ij}x_{kj}} \geq \frac{2D}{L + D} + \frac{L - D}{L + D} = 1
\]

where use of the fact that \( L+D>0 \) (from Holders inequality) has been made.

For \( i = k \), (14) becomes

\[
\frac{2}{L + D \sum_{j=1}^{n} x_{kj}^2} = \frac{L - D}{L + D} = 1.
\]

Therefore, if the \( w_i^* \) and \( \omega^* \) are selected according to (15) and (16), the desired inequalities are guaranteed. Q.E.D.

**Example 1:** Consider the four points

\[
X_1 = (0, 5) \quad X_3 = (4, 3)
\]

\[
X_2 = (-3, -4) \quad X_4 = (3, -4)
\]

which are to be separated from the origin in 2-dimensional space by a decision boundary.

The threshold gate shown in Fig. 2 separates points \( X_1 \) and \( X_3 \) but not \( X_2 \) or \( X_4 \). This is illustrated in Fig. 3. Let us now try to separate point \( X_4 \) from the origin.

\[\text{The inner-product notation}
\]

\[
X_i \cdot X_k = \sum_{j=1}^{n} x_{ij}x_{kj}
\]

is used throughout this paper.
by adding an additional thresholding network using the concepts developed in Theorem 1. Now

\[
D = \min_{i=1,2} \left[ \sum_{j=1}^{2} x_{ij} x_{2j} \right] = -24
\]

\[
L = \sum_{j=1}^{2} x_{2j}^2 = 25.
\]

Therefore, from (15)

\[
w_{1*} = \frac{2}{L + D} x_{21} = 8
\]

\[
w_{2*} = \frac{2}{L + D} x_{22} = 6
\]

\[
\omega* = \frac{L - D}{L + D} = 49.
\]

The resultant interconnection of this additional cascaded threshold gate is shown in Fig. 4. One may easily verify that the two cascaded threshold gates separate points \(X_1, X_2,\) and \(X_3,\) but not \(X_4,\) from the origin as shown in Fig. 5. In order to separate all four points a third threshold gate could be added.

IV. EXTENSION OF BASIC NONLINEAR SEPARATION THEOREM

An obvious question arises from Example 1. Is it possible by the addition of one cascade threshold gate to separate two or more additional points? The following theorem gives the necessary and sufficient conditions which must be satisfied in order to separate two additional misclassified points.

**Theorem 2:** Given \(k\) points \(X_1, X_2, \ldots, X_k\) where

\[
X_i = (x_{i1}, x_{i2}, \ldots, x_{in})
\]

and

\[
\sum_{j=1}^{n} x_{ij}^2 = R^2 \quad \text{for } i = 1, 2, \ldots, k
\]

(i.e., all \(k\) points lie on a hypersphere of radius \(R\) centered at the origin), if there exists a decision function, \(W(\cdot)\) such that

\[
W(X_i) = 1 \quad \text{for } i = 1, 2, \ldots, k - 2
\]

and

\[
W(X_{k-1}) = W(X_k) = -1
\]
then a necessary and sufficient condition that a linear
decision function \( W^* = \langle w_1^*, w_2^*, \ldots, w_n^*, \omega^* \rangle \) exist
such that
\[
\sum_{j=1}^{n} w_j^* x_{ij} + \omega^* W(X_i) \geq 1 \quad \text{for } i = 1, 2, \ldots, k
\]
(17)
is that either
\[
L + D_{\min} > 0
\]
or
\[
L + D_{\max} < 0
\]
where
\[
L = (X_{k-1} + X_k) \cdot X_{k-1} = (X_{k-1} + X_k) \cdot X_{k}
\]
and
\[
D_{\min} = \min_{i=1,2,\ldots,k-2} (X_{k-1} + X_k) \cdot X_i
\]
\[
D_{\max} = \max_{i=1,2,\ldots,k-2} (X_{k-1} + X_k) \cdot X_i.
\]

Proof:

Sufficiency: Assume that \( L + D_{\min} > 0 \)

let
\[
w_j^* = \frac{2}{L + D_{\min}} (x_{kj} + x_{k-1j})
\]
\[
\omega^* = \frac{L - D_{\min}}{L + D_{\min}}
\]
(18)

then for \( i = 1, 2, \ldots, k-2 \) (17) becomes
\[
\sum_{j=1}^{n} \frac{2}{L + D_{\min}} (x_{kj} + x_{k-1j}) x_{ij} + \frac{L - D_{\min}}{L + D_{\min}} \geq \frac{2D_{\min}}{L + D_{\min}} + \frac{L - D_{\min}}{L + D_{\min}} = 1,
\]
and for \( i = k-1, k \) (17) becomes
\[
\sum_{j=1}^{n} \frac{2}{L + D_{\min}} (x_{kj} + x_{k-1j}) x_{ij} = \frac{2L}{L + D_{\min}} - \frac{L - D_{\min}}{L + D_{\min}} = 1.
\]

Thus, if \( w_j^* \) and \( \omega^* \) are defined by (18) and (19), the desired inequalities are ensured. For the case when \( L + D_{\max} < 0 \), replacing \( D_{\min} \) by \( D_{\max} \) in (18) and (19) leads to a similar result.

Necessity: In order to prove necessity, use will be made of the fact that if there exists a linear decision function \( W = \langle w_1, w_2, \ldots, w_n, \omega \rangle \) such that
\[
\sum_{j=1}^{n} w_j x_{ij} + \omega \geq 1 \quad \text{for } i = 1, 2, \ldots, k-2
\]
(20)
and
\[
\sum_{j=1}^{n} w_j x_{ij} - \omega \geq 1 \quad \text{for } i = k-1, k
\]
(21)
then there exists a decision function \( W^* = \langle w_1^*, w_2^*, \ldots, w_n^*, \omega^* \rangle \) such that
\[
\sum_{j=1}^{n} w_j^* x_{ij} + \omega^* \geq 1 \quad \text{for } i = 1, 2, \ldots, k-2
\]
(22)
and
\[
\sum_{j=1}^{n} w_j^* x_{ij} - \omega^* = 1 \quad \text{for } i = k-1, k.
\]
(23)
In addition
\[
\sum_{j=1}^{n} w_j^* x_{pj} + \omega^* = 1
\]
(24)
where the vector \( X_p \) is determined either by
\[
\min_{i=1,2,\ldots,k-2} (X_{k-1} + X_k) \cdot X_i = (X_{k-1} + X_k) \cdot X_p,
\]
(25)
i.e., \( X_p \) is vector furthest from vector \( X_{k-1} + X_k \), or by
\[
\max_{i=1,2,\ldots,k-2} (X_{k-1} + X_k) \cdot X_i = (X_{k-1} + X_k) \cdot X_p,
\]
(26)
i.e., \( X_p \) is vector closest to vector \( X_{k-1} + X_k \).

In other words, the \( k \) inequalities of (20) and (21) may always be changed to 3 equalities and \( k-3 \) inequalities as given by (22) through (26).

The required parameters \( w_j^* \) are of the form
\[
w_j^* = \gamma (x_{k-1j} + x_{kj}) \quad \text{for } j = 1, 2, \ldots, n.
\]
(27)

These properties follow from the lemma proven in the Appendix.

To prove the necessity portion of Theorem 2, insert (27) into (22) and (23) giving
\[
\sum_{j=1}^{n} \gamma (x_{k-1j} + x_{kj}) x_{ij} + \omega^* \geq 1 \quad \text{for } i = 1, 2, \ldots, k-2
\]
(28)
and
\[
\sum_{j=1}^{n} \gamma (x_{k-1j} + x_{kj}) x_{ij} - \omega^* = 1 \quad \text{for } i = k-1, k.
\]
(29)
Now (29) becomes
\[
\gamma L - \omega^* = 1
\]
therefore
\[
\omega^* = \gamma L - 1.
\]
(30)

Inserting (30) into (28) gives
\[
\gamma \left[ \sum_{j=1}^{n} (x_{k-1j} + x_{kj}) x_{ij} + L \right] \geq 2
\]
(31)
for \( i = 1, 2, \ldots, k-2 \).

If \( \gamma \) turns out to be greater than zero, then relationship (31) is most difficult to satisfy for that \( X_i \) for which
\[ D_{\text{min}} = \min_{t=1,2,\ldots,k-2} \sum_{j=1}^{n} (x_{k-t} + x_{kj})x_{ij} \]

that is

\[ \gamma[D_{\text{min}} + L] = 2. \quad (32) \]

Solving (30) and (32) yields

\[ \gamma = \frac{2}{L + D_{\text{min}}} > 0 \]

\[ \omega^* = \frac{L - D_{\text{min}}}{L + D_{\text{min}}}. \]

Similarly, if \( \gamma < 0 \), one finds

\[ \gamma = \frac{2}{L + D_{\text{max}}} < 0 \]

\[ \omega^* = \frac{L - D_{\text{max}}}{L + D_{\text{max}}} \]

where

\[ D_{\text{max}} = \max_{t=1,2,\ldots,k-2} \sum_{j=1}^{n} (x_{k-t} + x_{kj})x_{ij} \]

which concludes the proof.

**Example 2:** Consider the set of four points given in Example 1. Assume that the threshold gate shown in Fig. 2 is given. Recall that this gate separates points \( X_1 \) and \( X_3 \) from the origin but not \( X_2 \) or \( X_4 \). Is it possible to cascade one additional threshold gate so as to separate all four points?

Applying Theorem 2,

\[ X_3 + X_4 = (7, -1) \]

so that

\[ L = (7, -1) \begin{pmatrix} 4 \\ 3 \end{pmatrix} = 25 \]

\[ D_{\text{min}} = \min_{t=1,2} (7, -1) \cdot x_t = -17. \]

Thus \( L + D_{\text{min}} > 0 \), and by Theorem 2 it is possible to separate \( X_1, X_2, X_3, \) and \( X_4 \) by the addition of one threshold gate. From (18) and (19)

\[ (w_1^*, w_2^*) = \frac{2}{25 - 17} \begin{pmatrix} 7 \\ -1 \end{pmatrix} = \begin{pmatrix} \frac{7}{4} \\ -\frac{1}{4} \end{pmatrix} \]

\[ \omega^* = \frac{25 + 17}{25 - 17} = \frac{21}{4}. \]

The complete network needed to effect the desired separation is shown in Fig. 6 and the decision boundary in Fig. 7. A comparison of the decision boundaries shown in Figs. 5 and 7 reveals that the threshold network generated by Theorem 2 separates all four points while that generated by Theorem 1 separates only three points even though each uses two cascaded threshold gates.

**V. Conclusions**

A technique for generating a pattern classifier which will properly classify two finite nonlinearly separable classes has been presented. It was shown that if it is possible to properly classify two additional points by the addition of one threshold gate, the parameters characterizing this gate may always be obtained directly. The illustrative separation problem treated in this paper demonstrates the possible savings in threshold gate hardware if this more efficient technique is used to generate the desired nonlinear decision boundary.
Lemma: If there exists a linear decision function
\[ W = (w_1, w_2, \ldots, w_n, \omega) \]
such that
\[ \sum_{j=1}^{n} w_j x_{ij} + \omega \geq 1 \quad \text{for } i = 1, 2, \ldots, k - 2 \quad (33) \]
and
\[ \sum_{j=1}^{n} w_j x_{ij} - \omega \geq 1 \quad \text{for } i = k - 1, k \quad (34) \]
then there exists a decision function \( W^* = (w_1^*, w_2^*, \ldots, w_n^*, \omega^*) \) such that
\[ \sum_{j=1}^{n} w_j^* x_{ij} + \omega^* \geq 1 \quad \text{for } i = 1, 2, \ldots, k - 2 \quad (35) \]
and
\[ \sum_{j=1}^{n} w_j^* x_{ij} - \omega^* = 1 \quad \text{for } i = k - 1, k. \quad (36) \]
In addition
\[ \sum_{j=1}^{n} w_j^* x_{ij} + \omega^* = 1 \quad (37) \]
where the vector \( X_p \) is determined either by
\[ \min_{l=1,2,\ldots,k-2} (X_{k-l} + X_k) \cdot X_i = (X_{k-l} + X_k) \cdot X_p \quad (38) \]
(i.e., \( X_p \) is vector furthest from vector \( X_{k-l}+X_k \)), or by
\[ \max_{l=1,2,\ldots,k-2} (X_{k-l} + X_k) \cdot X_i = (X_{k-l} + X_k) \cdot X_p \quad (39) \]
(i.e., \( X_p \) is vector closest to vector \( X_{k-l}+X_k \)).

To demonstrate this fact, first assume that \( \omega \) as given in (33) and (34) is positive. Since \( \omega > 0 \), it immediately follows that all \( k \) points \( X_1, X_2, \ldots, X_k \) lie on one side of the hyperplane defined by the set of vectors \( X = (x_1, x_2, \ldots, x_n) \) such that
\[ H_{w^1} = \left\{ X : \sum_{j=1}^{n} w_j x_{ij} = 1 - \omega \right\}. \quad (40) \]
In addition, the hyperplane \( H_{w^2} \) defined by
\[ H_{w^2} = \left\{ X : \sum_{j=1}^{n} w_j x_{ij} = 1 + \omega \right\} \quad (41) \]
is parallel to \( H_{w^1} \) and lies on the same side of \( H_{w^1} \) as do the \( k \) points \( \{X_i\} \).

Holding the coefficients \( w_i \) fixed, alter \( \omega \) until hyperplane \( H_{w^1} \) is just tangent to the hypersphere of radius \( R \) centered at the origin. This is always possible since the distance from the origin to this hyperplane is given by
\[ d_1 = \frac{1 - \omega}{\sum_{j=1}^{n} w_j^2} \]
and letting
\[ \omega = 1 + R \sum_{j=1}^{n} w_j^2 > 0 \quad (42) \]
gives
\[ d_1 = R. \]

Hyperplane \( H_{w^2} \) is located
\[ d_2 = \frac{1 + \omega}{\sum_{j=1}^{n} w_j^2} = R + \frac{2}{\sum_{j=1}^{n} w_j^2} \]
units from the origin, and is on the opposite side of the specified hypersphere. It will be further noted that all \( k \) points \( \{X_i\} \) are located between the hyperplanes \( H_{w^1} \) and \( H_{w^2} \). This follows since hyperplane \( H_{w^1} \) is just tangent to the specified hypersphere while hyperplane \( H_{w^2} \), which is parallel to \( H_{w^1} \), is on the opposite side of the hypersphere and does not intersect it (since \( d_2 > R \)).

Now holding this value of \( \omega \) fixed, rotate the vector \( (w_1, w_2, \ldots, w_n) \), maintaining \( \sum_{j=1}^{n} w_j^2 \) constant, until it is aligned with the line passing through the point \( X_{k-1}+X_k \) and the origin. The rotation is not unique since the vector \( (w_1, w_2, \ldots, w_n) \) could be in the direction of vector \( X_{k-1}+X_k \) or its negative. The appropriate rotation is one which guarantees that at least one point of the set \( \{X_1, X_2, \ldots, X_{k-1}\} \) is closer to the rotated hyperplane \( H_{w^1} \) than vectors \( X_{k-1} \) and \( X_k \).

The resulting hyperplanes \( H_{w^1} \) and \( H_{w^2} \) are still parallel and have all \( k \) points \( \{X_i\} \) between them. Further, these hyperplanes are perpendicular to the vector \( X_{k-1}+X_k \). The new values of \( \omega \) and \( w_j \) are denoted by \( \omega^* \) and \( w_j^* \). Now alter \( \omega^* \) and \( w_j^* \) in the following manner: let
\[ w_j^* = \alpha w_j^* \quad \text{for } j = 1, 2, \ldots, n \]
\[ \omega^* = \beta \omega^* \]
where \( \alpha \) and \( \beta \) are positive scalars selected so that the hyperplane
\[ H_{w^1} = \left\{ X : \sum_{j=1}^{n} w_j^* x_{ij} = 1 - \omega^* \right\} \]
contains the vector \( X_1 \) contained in the set \( \{X_1, X_2, \ldots, X_{k-2}\} \) which is closest to the hyperplane \( H_{w^1} \), while the

\[ 1 + \omega = 2 + R \sum_{j=1}^{n} w_j^2 \quad \text{and} \quad 1 - \omega = -R \sum_{j=1}^{n} w_j^2 < 0. \]

This is easily shown since for the selected value of \( \omega \) one finds that
If the original $\omega$ in (33) and (34) were negative, then all $k$ points $\{X_i\}$ lie on the same side of hyperplane $H_{\omega^2}$ where

$$H_{\omega^2} = \left\{ X : \sum_{j=1}^{n} w_j x_j = 1 + \omega \right\}.$$ 

Hyperplane

$$H_{\omega^1} = \left\{ X : \sum_{j=1}^{n} w_j x_j = 1 - \omega \right\}$$

is parallel to $H_{\omega^2}$ and lies on the same side of $H_{\omega^2}$ as do the $k$ points $\{X_i\}$.

Fixing the coefficients $w_j$, let

$$\omega^a = 1 + R \sum_{j=1}^{n} w_j^2 > 0$$

so that the hyperplane $H_{\omega^1}$ is tangent to the hypersphere of radius $R$ centered at the origin. With this new value of $\omega$, all $k$ points $\{X_i\}$ satisfy the inequality

$$\sum_{j=1}^{n} w_j x_{ij} \geq 1 - \omega^a$$

since $1 - \omega \geq 1 - \omega^a$ for negative $\omega$. Holding this value fixed, proceed in the manner used for positive initial values of $\omega$. This completes the proof of the lemma.