

E. SANVICENTE

En este trabajo se presentan tests, y sus correspondientes algoritmos, para la decisión de hipótesis estadísticas en presencia de mezclas. Después de explicar el modelo, se considera el test óptimo, basado en maximizar la verosimilitud, y se introducen modificaciones que, aunque lo apartan de la optimalidad, lo hacen de implementación más fácil.

### 1. INTRODUCCION

Dadas dos poblaciones gaussianas de diferentes medias e iguales varianzas, es bien conocido /1/, /2/, el problema de decidir con el valor de una muestra la población de la que dicha muestra fue extraída. La situación puede también plantearse en los siguientes términos, más apropiados para nuestro propósito: dados dos números reales,  $m_1$  y  $m_2$ , y observando  $z=m+x$  donde  $m$  es  $m_1$  ó  $m_2$  y  $x$  es una variable aleatoria gaussiana de media cero y varianza  $\sigma^2$ , decidir si en la construcción de  $z$  intervino  $m_1$  ó  $m_2$ . Obviamente la decisión puede llevarse a cabo con  $\Delta$  muestras en lugar de una, teniéndose entonces -- que decidir en base a  $(z_1, \dots, z_\Delta) = m(1, \dots, 1) + (x_1, \dots, x_\Delta)$  donde  $x_1, \dots, x_\Delta$  (errores de las mediciones) se suponen incorreladas y por ello independientes.

En lo anterior no hay nada esencial sobre el vector  $(1, \dots, 1)$  ni sobre el hecho de que  $m$  pueda tomar sólo dos valores. Así, un problema de dificultad esencialmente equivalente es decidir sobre  $m \in \{m_1, \dots, m_\mu\}$ , siendo  $m_1, \dots, m_\mu$  números reales con  $m_1 < \dots < m_\mu$ , cuando se observa  $(z_1, \dots, z_\Delta) = m(\alpha_1, \dots, \alpha_\Delta) + (x_1, \dots, x_\Delta)$  donde  $(\alpha_1, \dots, \alpha_\Delta)$  es un vector fijo. Por conveniencia supondremos que

$$\sum_{i=1}^{\Delta} \alpha_i^2 = 1,$$

pasando cualquier factor de escala a las  $m_1, \dots, m_\mu$ .

- E. Sanvicente de l'Escola Tècnica Superior d'Enginyers de Telecomunicació de la Universitat Politècnica. Baja de S. Pedro, 7. Barcelona 3.
- Article rebut el Març del 1978.

La dificultad aparece, sin embargo, cuando las muestras se mezclan por efecto del proceso de observación (medición). Concretamente supondremos la forma de interacción que se describe a continuación.

Dada la sucesión  $m(1), m(2), \dots, m(k), \dots$  con valores en  $\{m_1, \dots, m_\mu\}$  se observa no ya

$$\sum_{i=1}^{\infty} m(i) \alpha(k-(i-1)\Delta) + x(k), \quad k=1,2,\dots$$

(donde se ha supuesto  $\alpha(j)=0, j \neq 1, \dots, \Delta$ ), sino

$$z(k) = \sum_{i=1}^{\infty} m(i) \tilde{\alpha}(k-(i-1)\Delta) + x(k),$$

con  $(\tilde{\alpha}(1), \dots, \tilde{\alpha}(\tilde{\Delta}))$  conocido y dependiente del proceso de medición. Por convenio  $\tilde{\alpha}(j)=0, j \neq 1, \dots, \tilde{\Delta}$ . Supondremos también que  $\tilde{\Delta} > \Delta$  y que  $(\tilde{\alpha}(1), \dots, \tilde{\alpha}(\tilde{\Delta})) \neq (0, \dots, 0)$ . Obsérvese que, por ser  $\tilde{\Delta} < \infty$ , sólo hay en cada  $z(k)$  mezclas de un número finito de  $m(i)$ . Por ejemplo, si  $\tilde{\Delta} \leq \delta \Delta$  ( $\delta$  natural mayor que 1) en  $z(k)$  influyen además de  $m(k)$  otros  $\delta-1$  valores de la sucesión como mucho.

En estas condiciones se trata de ver cómo decidir sobre la sucesión  $m(1), m(2), \dots, m(k), \dots$  que originó la  $z(1), z(2), \dots, z(k), \dots$

## 2. EL TEST DE DECISION

Supondremos que la longitud de la sucesión es finita:  $m(1), \dots, m(L)$ . A dicha sucesión le corresponde otra  $z(k)$  de longitud  $\Delta L + (\tilde{\Delta} - \Delta)$ , ya que aunque  $z(k)$  pueda considerarse infinita, para  $k > \Delta L + \tilde{\Delta} - \Delta$  es  $z(k) = x(k)$ , que no contiene información sobre los posibles  $m(1), \dots, m(L)$ . Llamemos  $m = (m(1), \dots, m(L))$  y  $z = (z(1), \dots, z(\Delta L + \tilde{\Delta} - \Delta))$ .  $m$  puede ser una de  $\mu^L$  posibles sucesiones,  $m^1, \dots, m^{\mu^L}$  y se trata de decidir sobre cuál de ellas se usó en la construcción de  $z$ .

Llamando verosimilitudes a las  $\mu^L$  funciones de densidad condicionadas  $f_{z|M}(z|m^i)$ ,  $i=1, \dots, \mu^L$ , se decidirá (una vez obtenida  $z$ ) -- aquella  $m$  que maximiza la verosimilitud. Es decir, llamando  $\hat{m}$  a la decisión,

$$\hat{m} = \arg \max_{m^1, \dots, m^{\mu^L}} f_{z|M}(z|m),$$

donde por  $\arg \max$  se denota, con  $z$  fijo, el argumento que maximiza  $f_{z|M}(z|m)$  para los  $\mu^L$  posibles valores de  $m$ .

Además de ser razonable, el test anterior -- da la máxima probabilidad de decidir correctamente si se suponen  $m^1, \dots, m^{\mu^L}$  equiprobables.

En efecto; designemos por  $p_T(c)$  la probabilidad de decidir correctamente usando el -- test y comparémosla con la  $p_{T'}(c)$  obtenida con otro test. Como

$$p_T(c) = \int \dots \int_z p_T(c|z) f_z(z) dz$$

será  $p_T(c) \geq p_{T'}(c)$  si  $p_T(c|z) \geq p_{T'}(c|z)$ .

Pero  $p_T(c|z) = p(m_T^{i_T(z)}|z)$ ,  $p_{T'}(c|z) = p(m_T^{i_{T'}(z)}|z)$  siendo  $i_T(z)$  (respectivamente  $i_{T'}(z)$ ) la decisión del test  $T$  (resp.  $T'$ ) al recibir  $z$ . Para que sea

$$p(m_T^{i_T(z)}|z) \geq p(m_T^{i_{T'}(z)}|z)$$

basta tomar

$$\begin{aligned} \hat{m} &= \arg \max p(m^i|z) = \\ &= \arg \max \{f_{z|M}(z|m^i) p(m^i)\} \end{aligned}$$

QÜESTIÓ - v.2, n°2 (juny 1978)

Por ello adoptaremos dicho test aquí, y lo llamaremos óptimo a sabiendas de que las  $m^1, \dots, m^{\mu^L}$  pueden no ser equiprobables.

## 3. ALGORITMOS DE DECISION PARA EL TEST OPTIMO

Llamemos  $y = (y(1), \dots, y(\Delta L + \tilde{\Delta} - \Delta))$  con

$$y(k) = \sum_{i=1}^L m(i) \tilde{\alpha}(k - (i-1)\Delta).$$

Entonces  $z = y + x$  y se tiene  $f_{z|M}(z|m) = f_{z|Y}(z|y)$ , donde  $y$  es el (único) vector que corresponde a  $m$ . Como  $f_{z|Y}(z|y) = f_{X|Y}(z-y|y)$ , ya que el jacobiano

$$\int \begin{Bmatrix} z(1), \dots, z(\Delta L + \tilde{\Delta} - \Delta) \\ y(1), \dots, y(\Delta L + \tilde{\Delta} - \Delta) \end{Bmatrix} = 1,$$

suponiendo  $x$  y  $m$  independientes, se puede poner  $f_{z|M}(z|m) = f_X(z-y)$ .

Así

$$\begin{aligned} \hat{m} &= \arg \max_{m^1, \dots, m^{\mu^L}} f_{z|M}(z|m) = \\ &= \arg \max_{m^1, \dots, m^{\mu^L}} f_X(z-y) = \\ &= \arg \max_{y^1, \dots, y^{\mu^L}} \left( \frac{1}{2\pi\sigma} \right)^{\frac{\Delta L + \tilde{\Delta} - \Delta}{2}} \exp \left\{ -\frac{1}{2} \sum_{k=1}^{\Delta L + \tilde{\Delta} - \Delta} \left( \frac{z(k) - y(k)}{\sigma} \right)^2 \right\} \\ &= \arg \min_{y^1, \dots, y^{\mu^L}} \|z-y\|^2 \end{aligned}$$

donde  $\|\cdot\|$  denota la norma euclídea.

La anterior se puede aún poner como

$$\arg \max_{y^1, \dots, y^{\mu^L}} \{ \langle z, y \rangle - \frac{1}{2} \|y\|^2 \}$$

donde  $\langle \cdot, \cdot \rangle$  es el producto escalar ordinario

El algoritmo de decisión es así:

- 1) Calcular (off-line)  $\frac{1}{2} \|y\|^2$  para los  $\mu^L$  posibles vectores  $y$
- 2) Obtenido  $z$ , calcular  $\langle z, y \rangle$

- 3) Calcular  $\langle z, y \rangle - \frac{1}{2} \|y\|^2$
- 4) Calcular el máximo de los  $\mu^L \langle z, y \rangle - \frac{1}{2} \|y\|^2$
- 5) Si el máximo se obtiene para  $y^i$ , decidir  $m^i$ .

El algoritmo anterior se puede modificar equivalentemente como se dirá a continuación.

Para ello empecemos por poner

$$\sum_{k=1}^{\Delta L + \tilde{\Delta} - \Delta} z(k) y^i(k) =$$

$$= \sum_{k=1}^{\Delta L + \tilde{\Delta} - \Delta} z(k) \sum_{j=1}^L m^i(j) \tilde{\alpha}(k - (j-1)\Delta) =$$

$$= \sum_{j=1}^L m^i(j) \sum_{k=1}^{\Delta L + \tilde{\Delta} - \Delta} z(k) \tilde{\alpha}(k - (j-1)\Delta) =$$

$$= \sum_{j=1}^L m^i(j) \zeta(j)$$

donde

$$\zeta(j) = \sum_{k=1}^{\Delta L + \tilde{\Delta} - \Delta} z(k) \tilde{\alpha}(k - (j-1)\Delta)$$

$$= \sum_{n=1}^{\tilde{\Delta}} z(n + (j-1)\Delta) \tilde{\alpha}(n)$$

ya que

$$\tilde{\alpha}(n) = 0, \quad n \neq 1, \dots, \tilde{\Delta}.$$

Con ello el algoritmo anterior puede ponerse en la siguiente forma:

- 1) Como antes.
- 2') Obtenido  $z$ , descomponerlo en los  $L$  vectores (solapados):  $(z(1), \dots, z(\tilde{\Delta}))$ ,  $(z(\Delta), \dots, z(\Delta + \tilde{\Delta}))$ ,  $\dots$ ,  $(z((L-1)\Delta), \dots, z((L-1)\Delta + \tilde{\Delta}))$  y multiplicar escalarmente  $(\tilde{\alpha}(1), \dots, \tilde{\alpha}(\tilde{\Delta}))$  por cada segmento para obtener  $(\zeta(1), \dots, \zeta(L))$ .
- 2'') Multiplicar escalarmente  $\zeta = (\zeta(1), \dots, \zeta(L))$  por cada uno de los  $\mu^L m^i$ .

Obsérvese que 2') 2'') equivalen al 2) anterior.

rior.

- 3) Como antes.
- 4) Como antes.
- 5) Como antes.

Ahora también, hasta que no se dispone de  $\zeta(L)$  no se puede llegar a una decisión. Esto es un grave inconveniente cuando  $L$  es elevado y nos sugiere la búsqueda de un test subóptimo, más fácil de implementar.

#### 4. EL TEST SUBÓPTIMO Y SU ALGORITMO DE DECISION

Obsérvese que  $\zeta$  es el vector que nos permite decidir. El, pues, debe llevar dentro de sí la información sobre  $m(1), \dots, m(L)$ . Veamos en qué modo.

$$\zeta(j) = \sum_{k=1}^{\Delta L + \tilde{\Delta} - \Delta} z(k) \tilde{\alpha}(k - (j-1)\Delta) =$$

$$= \sum_{k=1}^{\Delta L + \tilde{\Delta} - \Delta} (y(k) + x(k)) \tilde{\alpha}(k - (j-1)\Delta) =$$

$$= \sum_{k=1}^{\Delta L + \tilde{\Delta} - \Delta} \left( \sum_{i=1}^L m(i) \tilde{\alpha}(k - (i-1)\Delta) + x(k) \right) \cdot$$

$$\tilde{\alpha}(k - (j-1)\Delta) = \sum_{i=1}^L m(i) \sum_{k=1}^{\Delta L + \tilde{\Delta} - \Delta} \tilde{\alpha}(k - (i-1)\Delta) \cdot$$

$$\tilde{\alpha}(k - (j-1)\Delta) + \sum_{k=1}^{\Delta L + \tilde{\Delta} - \Delta} x(k) \tilde{\alpha}(k - (j-1)\Delta) =$$

$$= \sum_{i=1}^L m(i) R(i-1, j-1) + \xi(j)$$

donde

$$R(i-1, j-1) = \sum_{k=1}^{\Delta L + \tilde{\Delta} - \Delta} \tilde{\alpha}(k - (i-1)\Delta) \tilde{\alpha}(k - (j-1)\Delta)$$

$$\xi(j) = \sum_{k=1}^{\Delta L + \tilde{\Delta} - \Delta} x(k) \tilde{\alpha}(k - (j-1)\Delta)$$

Matricialmente se puede poner:

$$\begin{pmatrix} \zeta(1) \\ \cdot \\ \cdot \\ \cdot \\ \zeta(L) \end{pmatrix} = \begin{pmatrix} R(0,0) \cdots R(L-1,0) \\ \cdot \\ \cdot \\ \cdot \\ R(0,L-1) \cdots R(L-1,L-1) \end{pmatrix} \begin{pmatrix} m(1) \\ \cdot \\ \cdot \\ \cdot \\ m(L) \end{pmatrix} + \begin{pmatrix} \xi(1) \\ \cdot \\ \cdot \\ \cdot \\ \xi(L) \end{pmatrix}$$

donde la matriz  $L \times L$  anterior es no-singular, hecho este que tendrá utilidad luego.

La matriz  $R$  es positiva definida, y por ello no-singular, como se comprueba a continuación.

Siendo  $v = (v(1), \dots, v(L))$  un vector cualquiera de  $L$  dimensiones, es

$$\langle v, Rv \rangle = \sum_{i=1}^L \sum_{j=1}^L v(i) v(j) R(i-1, j-1) =$$

$$= \sum_{i=1}^L \sum_{j=1}^L v(i) v(j) \sum_{k=1}^{\Delta L + \tilde{\Delta} - \Delta} \tilde{\alpha}(k - (i-1)\Delta) \cdot$$

$$\cdot \tilde{\alpha}(k - (j-1)\Delta) =$$

$$= \sum_{k=1}^{\Delta L + \tilde{\Delta} - \Delta} \left( \sum_{i=1}^L v(i) \tilde{\alpha}(k - (i-1)\Delta) \right)^2$$

que nunca es negativa. Para que sea nula deberá tenerse

$$\sum_{i=1}^L v(i) \tilde{\alpha}(k - (i-1)\Delta) = 0$$

para todo  $k$ , lo que implica que  $v(1)=0$  (recuérdese que  $(\tilde{\alpha}(1), \dots, \tilde{\alpha}(\Delta)) \neq (0, \dots, 0)$ ) y, reiterando el razonamiento,  $v(2)=\dots=v(L)=0$ .

Volviendo al problema del test subóptimo, la forma en que  $(m(1), \dots, m(L))$  influye en  $(\zeta(1), \dots, \zeta(L))$  sugiere, como alternativa al test de decisión óptimo, estimar  $m$  linealmente por mínimos cuadrados. Suponiendo  $E(m)=0$  por conveniencia y sin restringir la generalidad, se tiene entonces  $\hat{m} = K_{m\zeta}^{-1} K_{\zeta\zeta} \zeta$ , /3/, - supuesta  $K_{\zeta\zeta}^{-1}$  existente (ver luego).  $K_{m\zeta}$  y  $K_{\zeta\zeta}$  son:

$$K_{m\zeta} = E \{ m\zeta' \}$$

$$K_{\zeta\zeta} = E \{ \zeta\zeta' \}$$

donde los vectores se escriben como columnas y el acento denota transpuesta.

Sin embargo, aunque  $\hat{m}$  minimiza  $E \|m - \hat{m}\|^2$ , - la fórmula  $K_{m\zeta} K_{\zeta\zeta}^{-1} \zeta$  no produce, en general, un elemento del conjunto  $m^1, \dots, m^{\mu L}$ . La dificultad se solventa discretizando  $\hat{m}$  componente a componente como se indica a continuación. Con  $R_j$  definido como el subconjunto de la recta real

$$R_j = \left\{ x \in R: \frac{m_{j-1} - m_j}{2} \leq x - m_j < \frac{m_{j+1} - m_j}{2} \right\},$$

si  $\hat{m}(i) \in R_j$ , tómesese  $\hat{m}(i) = m_j$ . Así, se toma el  $m^i$  más cercano, en distancia euclídea, a  $\hat{m}$ .

Para poner la fórmula que da  $m$  más directamente en función de los datos, recordemos - que  $\zeta = Rm + \xi$ . Entonces, /4/:

$$K_{m\zeta} = K_{mm} R,$$

$$K_{\zeta\zeta} = R K_{mm} R + K_{\xi\xi},$$

puesto que  $R$  es simétrica.

$E \{ \xi(i) \xi(j) \}$ , elemento  $(i, j)$  de  $K_{\xi\xi}$ , vale:

$$E \{ \xi(i) \xi(j) \} = E \left\{ \sum_{k=1}^{\Delta + \tilde{\Delta} - \Delta} x(k) \tilde{\alpha}(k - (i-1)\Delta) \cdot \right.$$

$$\left. \sum_{l=1}^{\Delta L + \tilde{\Delta} - \Delta} x(l) \tilde{\alpha}(l - (j-1)\Delta) \right\} =$$

$$= \sum_{k=1}^{\Delta L + \tilde{\Delta} - \Delta} \sum_{l=1}^{\Delta L + \tilde{\Delta} - \Delta} \tilde{\alpha}(k - (i-1)\Delta) \cdot$$

$$\cdot \tilde{\alpha}(l - (j-1)\Delta) \sigma^2 \delta_{kl} =$$

$$= \sigma^2 R(i-1, j-1)$$

ya que

$$E \{ x(k) x(l) \} = \sigma^2 \delta_{kl},$$

donde  $\delta_{kl} = 1$  si  $k=l$  y  $\delta_{kl} = 0$  si  $k \neq l$ .

Matricialmente es, pues:  $K_{\xi\xi} = \sigma^2 R$ .

Para calcular  $K_{mm}$  recordemos que el test óptimo suponía  $m^1, \dots, m^{\mu L}$  equiprobables, y - eso es también lo que se supondrá ahora. - Ello equivale a decir que las variables - aleatorias  $m(1), \dots, m(L)$  son independientes y toman los valores  $m_1, \dots, m_{\mu}$  con probabilidades  $1/\mu$ . Por lo tanto:

$$E\{m(i) m(j)\} = \begin{cases} 0, & j \neq i \\ \frac{1}{\mu} (m_1^2 + \dots + m_\mu^2), & j = i \end{cases}$$

Matricialmente se pondrá:  $K_{mm} = \sigma_*^2 I$ , donde  $I$  es la matriz identidad y  $\sigma_*^2 = \frac{1}{\mu} (m_1^2 + \dots + m_\mu^2)$ .

Así se tiene:

$$\hat{m} = \sigma_*^2 R (R \sigma_*^2 R + \sigma^2 R)^{-1} = (R + \frac{\sigma^2}{\sigma_*^2} I)^{-1}$$

donde la inversa de  $\sigma_*^2 R^2 + \sigma^2 R$  existe por ser la suma de dos matrices positivas definidas.

Este test subóptimo, en ausencia de errores de medición, da  $\hat{m} = R^{-1} \zeta$ , como debiera. Por otra parte, como  $\langle z, y \rangle = \langle \zeta, m \rangle$  (ver 2") y en ausencia de mezclas es  $\|y\|^2 = \|m\|^2$ , se tiene para el test óptimo:

$$\begin{aligned} \max_y \left\{ \langle z, y \rangle - \frac{1}{2} \|y\|^2 \right\} &= \max_m \left\{ \langle \zeta, m \rangle - \frac{1}{2} \|m\|^2 \right\} \\ &= \min_m \|\zeta - m\|^2 \end{aligned}$$

por lo que, si  $\zeta(i) \in R_j$  es  $\hat{m}(i) = m_j$ .

Para el test subóptimo ( $\tilde{\alpha} = \alpha$ ,  $R = I$ ) se tiene --

$$\hat{m} = \left( 1 + \frac{\sigma^2}{\sigma_*^2} \right)^{-1} \zeta,$$

que prácticamente coincide con el óptimo -- ( $= \zeta$ ) en circunstancias usuales de medición -- ( $\sigma_* \gg \sigma$ ).

El algoritmo para el test subóptimo queda entonces así:

2') Como antes.

2") Como antes.

3S) Calcular  $(R + \frac{\sigma^2}{\sigma_*^2} I)^{-1}$  (off-line)

4S) Calcular  $\hat{m} = (R + \frac{\sigma^2}{\sigma_*^2} I)^{-1} \zeta$

5S) Si  $\hat{m}(i) \in R_j$ , tómesese  $\hat{m}(i) = m_j$ .

Los pasos 3S), 4S) y 5S) son subóptimos. Ob-

QUESTIÓ - v.2, n°2 {juny 1978}

sérvase que se ha evitado el paso de comparación del algoritmo óptimo. La ventaja principal de este algoritmo subóptimo se verá, sin embargo, en el apartado siguiente.

## 5. EL ALGORITMO SUBOPTIMO EN REGIMEN ESTACIONARIO

Con el algoritmo subóptimo se puede ir decidiendo sobre los valores de la sucesión  $m(k)$  sin necesidad de conocer el vector  $z$  en su totalidad, lo cual es básico cuando  $L$  es grande. Para ver esto, notemos que --

$$R + \frac{\sigma^2}{\sigma_*^2} I$$

es una matriz de estructura esencialmente -- diagonal, obtenida desplazando hacia abajo un vector (generador) de dimensión  $2\delta+1$ , simétrico con respecto a su componente central. La matriz inversa participa de esas mismas cualidades, aún cuando el vector que la genera tenga dimensión  $2\delta_*+1$  con  $\delta_* > \delta$ . Sea el el vector generador  $g = (g(\delta_*), \dots, g(1), g(0), g(-1), \dots, g(-\delta_*))$ .

El algoritmo de decisión es:

2') Como antes.

2") Como antes.

3'S) Calcular

$$\left( R + \frac{\sigma^2}{\sigma_*^2} I \right)^{-1}$$

(off-line) para obtener el vector generador  $g$ .

4'S)  $m(k) = \sum_{l=-\delta_*}^{\delta_*} g(l) \zeta(k-l)$ , con  $\zeta(j) = 0$

para  $j \neq 1, \dots, L$ .

Obsérvese que la decisión sobre  $m(1)$  se puede hacer al tener  $\zeta(1), \dots, \zeta(1+\delta_*)$  y no hay que esperar a disponer de  $\zeta(L)$ .

Nótese también que aunque 4'S) y 4S) coinciden en régimen estacionario (a partir de  $k > 1+\delta_*$ ) dan resultados algo diferentes para

los primeros (y últimos!) k's.

5S) Como antes.

## 6. CONCLUSION

Se han presentado diversos algoritmos para - decidir entre un número finito de hipótesis en presencia de mezclas. De todos ellos, el último es el más eficaz, aún cuando no sea - óptimo, al simplificar el tipo de decisión - requerida por el algoritmo óptimo y permitir realizar decisiones secuencialmente.

## 7. REFERENCIAS

- /1/ HOGG, R.V., CRAIG, A.T. "Introduction to Mathematical Statistics". McMillan, 1970 pp. 277.
- /2/ RIOS, S. "Métodos estadísticos". Ed. del Castillo, 1967, pp. 365.
- /3/ RHODES, I.B. "A Tutorial Introduction to Estimation and Filtering", I.E.E.E. -- Trans. Aut. Control, vol. Ac-16, n° 6, - Dic. 1971.
- /4/ LUENBERGER, D.G. "Optimization by Vector Space Methods". John Wiley, 1969, pp. 88.