191
12/5/79

LA-8093-MS

Informal Report

LA 361

General Least-Squares Fitting Procedures to Minimize the Volume of a Hyperellipsoid

University of California

LOS ALAMOS SCIENTIFIC LABORATORY

Post Office Box 1663 Los Alamos, New Mexico 87545

General Least-Squares Fitting Procedures to Minimize the Volume of a Hyperellipsoid

E. Alan Wadlinger

DISCLAIMER

This document contains information which is being furnished in good faith by the Los Alamos Scientific Laboratory. It is not to be distributed outside the Laboratory without the express written permission of the Laboratory. The information contained herein is not to be used for any purpose other than that for which it was originally intended. The information contained herein is not to be used for any purpose other than that for which it was originally intended. The information contained herein is not to be used for any purpose other than that for which it was originally intended.

GENERAL LEAST-SQUARES FITTING PROCEDURES TO MINIMIZE THE VOLUME OF A HYPERELLIPSOID

by

E. Alan Wadlinger

ABSTRACT

We present several methods for determining the shape parameters, which in two dimensions are the Courant-Snyder parameters, and the volume of an ellipse or hyperellipse that represents a set of phase-space points in a two- or more dimensional hyperspace. The ellipse parameters are useful for matching a beam to an accelerating or transport system and in studies of emittance growth. The fitting procedure minimizes the total volume of a hyperellipse by adjusting the ellipse shape parameters. The total volume is the sum of the individual particle volumes defined by the hyperellipse that passes through a particle's phase-space point. We first deal with a two-dimensional space then generalize to higher dimensions. Computer programs using these techniques have been written.

I. TWO-DIMENSIONAL CASE

The equation for an ellipse may be written

$$\gamma x^2 + 2\alpha xy + \beta y^2 = E/\pi = A \quad , \quad (1)$$

$$\gamma\beta - \alpha^2 - 1 = 0 \quad , \quad (2a)$$

where α , β , and γ are the Courant and Snyder parameters¹ and E is the area of the ellipse. The emittance required for an ellipse centered at the origin to

encompass a particle with phase-space coordinates x_i, y_i is

$$A_i = \gamma x_i^2 + 2\alpha x_i y_i + \beta y_i^2 . \quad (3)$$

If there are N particles and each particle has an emittance, A_i ($i = 1, N$), we define a total summed emittance,

$$I = \sum_{i=1}^N A_i . \quad (4a)$$

The rms ellipse parameters minimize I . These parameters are found most easily by multiplying Eq. (2a) by a Lagrange multiplier λ ; adding the product to Eq. (4a) giving J [Eq. (4b)]; differentiating J by α , β , γ , and λ , respectively; and setting the result to zero. The Lagrange multiplier allows α , β , and γ to be treated as independent variables.

$$J = I + \lambda (\gamma\beta - \alpha^2 - 1) , \quad (4b)$$

$$\frac{\partial J}{\partial \alpha} = 2 \sum x_i y_i - 2\alpha\lambda = 0 , \quad (5)$$

$$\frac{\partial J}{\partial \beta} = \sum y_i^2 + \gamma\lambda = 0 , \quad (6)$$

$$\frac{\partial J}{\partial \gamma} = \sum x_i^2 + \beta\lambda = 0 , \quad (7)$$

$$\frac{\partial J}{\partial \lambda} = \gamma\beta - \alpha^2 - 1 = 0 . \quad (2b)$$

Solving Eqs. (2b), (5), (6), and (7) we get

$$\gamma = \frac{\sum y_i^2}{D} , \quad (8)$$

$$\beta = \frac{\sum x_i^2}{D} , \quad (9)$$

$$\alpha = - \frac{\sum x_i y_i}{D} , \quad (10)$$

and

$$D = \sqrt{\sum x_i^2 \sum y_i^2 - \left(\sum x_i y_i\right)^2} , \quad (11)$$

which is a standard method for fitting an rms ellipse to a distribution.

If we define emittance to be the ellipse area that encompasses all of the beam, then to determine the ellipse parameters that minimize this area, we should use a fitting procedure that most heavily weights those beam particles that lie farthest from the centroid of the distribution. We do this by using A_i^2 instead of A_i in Eq. (4b),

$$J = \sum A_i^2 + \lambda(\gamma\beta - \alpha^2 - 1) , \quad (12)$$

Minimizing J with respect to α , β , γ , and λ gives [see Eq. (3)],

$$\frac{\partial J}{\partial \alpha} = 4 \sum A_i x_i y_i - 2\alpha\lambda = 0 , \quad (13)$$

$$\frac{\partial J}{\partial \beta} = 2 \sum A_i y_i^2 + \gamma\lambda = 0 , \quad (14)$$

$$\frac{\partial J}{\partial \gamma} = 2 \sum A_i x_i^2 + \beta\lambda = 0 , \quad (15)$$

$$\frac{\partial J}{\partial \lambda} = \gamma\beta - \alpha^2 - 1 = 0 . \quad (16)$$

We use Eq. (13) to solve for λ and substitute the result in Eqs. (14) and (15) to get

$$\alpha \sum A_i y_i^2 + \gamma \sum A_i x_i y_i = 0 \quad , \quad (17)$$

$$\alpha \sum A_i x_i^2 + \beta \sum A_i x_i y_i = 0 \quad . \quad (18)$$

Combining Eq. (3) with Eqs. (17) and (18) gives

$$3\alpha\gamma \sum x_i^2 y_i^2 + 2\alpha^2 \sum x_i y_i^3 + \alpha\beta \sum y_i^4 + \gamma^2 \sum x_i^3 y_i + \gamma\beta \sum x_i y_i^3 = 0 \quad , \quad (19)$$

$$\alpha\gamma \sum x_i^4 + 2\alpha^2 \sum x_i^3 y_i + 3\alpha\beta \sum x_i^2 y_i^2 + \gamma\beta \sum x_i^3 y_i + \beta^2 \sum x_i y_i^3 = 0 \quad . \quad (20)$$

Equations (16), (19), and (20) define α , β , and γ . We use an iteration procedure to solve these equations. Define

$$\textcircled{1} = \sum x_i^4, \textcircled{2} = \sum x_i^3 y_i, \textcircled{3} = \sum x_i^2 y_i^2, \textcircled{4} = \sum x_i y_i^3, \textcircled{5} = \sum y_i^4 \quad , \quad (21)$$

and from Eqs. (16), (19), and (20) define

$$K_1 = \gamma\beta - \alpha^2 - 1 \quad , \quad (22)$$

$$K_2 = 3\alpha\gamma \textcircled{3} + 2\alpha^2 \textcircled{4} + \alpha\beta \textcircled{5} + \gamma^2 \textcircled{2} + \gamma\beta \textcircled{4} \quad , \quad (23)$$

$$K_3 = \alpha\gamma \textcircled{1} + 2\alpha^2 \textcircled{2} + 3\alpha\beta \textcircled{3} + \gamma\beta \textcircled{2} + \beta^2 \textcircled{4} \quad . \quad (24)$$

We require $K_1 = K_2 = K_3 = 0$. Given α^n , β^n , and γ^n for the n th iteration, we find α^{n+1} , β^{n+1} , and γ^{n+1} .

$$K_i^{n+1} = K_i^n + \frac{\partial K_i}{\partial \alpha} \bigg|_n (\alpha^{n+1} - \alpha^n) + \frac{\partial K_i}{\partial \beta} \bigg|_n (\beta^{n+1} - \beta^n) + \frac{\partial K_i}{\partial \gamma} \bigg|_n (\gamma^{n+1} - \gamma^n) = 0 \quad , \quad (25)$$

for $i = 1, 2, 3$, and

$$\left. \frac{\partial K_i}{\partial \alpha} \right|_n = \frac{\partial K_i}{\partial \alpha} ,$$

for $\alpha = \alpha^n$, $\beta = \beta^n$, and $\gamma = \gamma^n$.

We then solve the linear system of three equations for α^{n+1} , β^{n+1} , and γ^{n+1} .

$$\left. \frac{\partial K_i}{\partial \alpha} \right|_n \alpha^{n+1} + \left. \frac{\partial K_i}{\partial \beta} \right|_n \beta^{n+1} + \left. \frac{\partial K_i}{\partial \gamma} \right|_n \gamma^{n+1} = -K_i^n + \left. \frac{\partial K_i}{\partial \alpha} \right|_n \alpha^n + \left. \frac{\partial K_i}{\partial \beta} \right|_n \beta^n + \left. \frac{\partial K_i}{\partial \gamma} \right|_n \gamma^n . \quad (26)$$

The starting values for the procedure may be obtained from Eqs. (8-11).

Figure 1 compares the results of fitting an ellipse to the given set of points using both methods described above. For this distribution the weighted fitting method (dashed line) gives an ellipse with a smaller area.

Finally, we have assumed that the particle distribution is centered about the origin. This deficiency can be corrected by substituting $(x_i - x_0)$ for

Fig. 1.

Weighted and rms ellipses that encompass all particles.
The inner ellipse (dashed line) resulted from the "weighted" fit.

x_i and $(y_i - y_0)$ for y_i in Eq. (3) then minimizing Eq. (4b) or Eq. (12) with respect to x_0 and y_0 , where x_0, y_0 is the origin of the ellipse.

II. HIGHER DIMENSIONAL HYPERELLIPSOID

The equation of a hyperellipse in an n -dimensional hyperspace is

$$\underline{X}^T \underline{A} \underline{X} = 1 \quad , \quad (27)$$

where $\underline{A}$ is an $(n \times n)$ dimensional symmetric matrix and $\underline{X}$ is an n -dimensional column vector. Given a set of k points (particles) represented by $\underline{X}_i$ ($i = 1, k$) in this hyperspace, we find the hyperellipse defining matrix $\underline{A}$ that best represents this set of points. We again fit by minimizing the volume enclosed by the hyperellipse characterized by the data. (We do not consider the technique² of minimizing the function,

$$I = \sum_{i=1}^k (\underline{X}_i^T \underline{A} \underline{X}_i - 1)^2 \quad , \quad (28)$$

with respect to the elements of the symmetric matrix $\underline{A}$. This technique minimizes the distances between the data points and the surface of the hyperellipsoid defined by $\underline{A}$.)

We first obtain an explicit relationship between the volume of the hyperellipsoid and the shape-defining parameters of the ellipse. Crandall³ showed that the volume of an n -dimensional upright ellipsoid,

$$\sum_{i=1}^n \frac{x_i^2}{d_i^2} = 1 \quad , \quad (29)$$

with semiaxes of length d_i is

$$V_n = K(n) \prod_{i=1}^n d_i \quad , \quad (30)$$

$$K(n) = \frac{n/2}{(n/2)!} , \quad n \text{ even} , \quad (31a)$$

$$K(n) = \frac{(n-1)/2}{(n+1)!} 2^{n+1} \left(\frac{n+1}{2} \right)! , \quad n \text{ odd} . \quad (31b)$$

Note that the sum in Eq. (29) and the product in Eq. (30) are over the n -dimensional vector space, not over the particle coordinates as in Eq. (28). Now let $\underline{U}$ be an orthogonal matrix that diagonalizes $\underline{A}$, Eq. (27), let $\underline{D}$ be the resulting diagonal matrix,

$$\underline{D} = \underline{U} \underline{A} \underline{U}^T , \quad (32)$$

and let $\underline{Y}$ be the transformed vector,

$$\underline{Y} = \underline{U} \underline{X} . \quad (33)$$

Then

$$\underline{Y}^T \underline{D} \underline{Y} = 1 . \quad (34)$$

Equation (34) is in the form of Eq. (29) with components of $\underline{D}$,

$$\underline{D}_{ij} = \frac{\delta_{ij}}{d_j^2} , \quad (35)$$

where δ_{ij} is the Kronecker delta function. Using the determinant of $\underline{D}$, symbolized by $|\underline{D}|$, we find that

$$|\underline{D}| = \left[\prod_{i=1}^n d_i^2 \right]^{-1} , \quad (36)$$

$$= \left(\frac{K(n)}{V_n} \right)^2 . \quad (37)$$

The determinant value is invariant under an orthogonal transformation.
Therefore,

$$|\underline{A}| = \left(\frac{K(n)}{V_n} \right)^2 . \quad (38)$$

We define a new matrix $\underline{B}$ whose elements are

$$\underline{B}_{ij} = \left(\frac{V_n}{K(n)} \right)^{2/n} \underline{A}_{ij} ; \quad (39)$$

then,

$$|\underline{B}| = \left[\left(\frac{V_n}{K(n)} \right)^{2/n} \right]^n |\underline{A}| = 1 . \quad (40)$$

From Eq. (27) we get

$$X^T \underline{B} X = \left(\frac{V_n}{K(n)} \right)^{2/n} , \quad (41)$$

with

$$|\underline{B}| = 1 . \quad (42)$$

As a check, we consider the ($n = 2$) case. From Eq. (41) we obtain

$$X^T \underline{B} X = \left(\frac{V_n}{\pi} \right)^{2/2} = \frac{E}{\pi} , \quad (43)$$

which is our old friend [see Eq. (1)],

$$\gamma X^2 + 2\alpha XY + \beta Y^2 = \frac{E}{\pi} ,$$

with $\gamma\beta - \alpha^2 = 1$ using Eq. (42).

The hyperellipse volume, determined from Eq. (41), is

$$V_n = K(n) \left[X^T \underline{B} X \right]^{n/2} . \quad (44)$$

The hyperellipse volume (or some power of the volume) is minimized for k particles in an n -dimensional hyperspace. We define

$$I = \sum_{i=1}^k V_{n_i}^m + \lambda (|\underline{B}| - 1) , \quad (45)$$

where i is the sum over k particles, m is an exponent, n is the hyperspace dimension, and λ is a Lagrange multiplier. We require that $\underline{B}$ be symmetric. Substituting Eq. (44) into Eq. (45) and absorbing the $K(n)$ term into the Lagrange multiplier, we have

$$I = \sum_{i=1}^k \left(X_i^T \underline{B} X_i \right)^{nm/2} + \lambda (|\underline{B}| - 1) . \quad (46)$$

The matrix $\underline{B}$ is determined by minimizing I with respect to each element $\underline{B}_{ij}$ and λ , giving the series of equations

$$\left. \begin{aligned} \partial I / \partial \underline{B}_{ij} &= 0 \quad (i \leq j = 1, \dots, n) , \\ \partial I / \partial \lambda &= 0 . \end{aligned} \right\} \quad (47)$$

The restriction $(i \leq j = 1, \dots, n)$ accounts for $\underline{B}$ being symmetric ($\underline{B}_{ij} = \underline{B}_{ji}$). We use Eq. (47) to solve for $\underline{B}$. In all but the simplest cases, Eq. (47) will be nonlinear and will have to be solved by an iterative approximation scheme. If the distribution has a nonzero centroid, we can substitute $(X_i - X_o)$ for X_i in Eq. (46) and minimize Eq. (46) with respect to X_o .

Given the ellipse in Eq. (41), we can determine the projected area in the two-dimensional plane defined by $\hat{\ell}$ and $\hat{m}$. We take the gradient of Eq. (41) to obtain

$$\nabla(X^T \underline{B} X) = 2 \sum_i \hat{i} \sum_j \underline{B}_{ij} X_j, \quad (48)$$

where i and j are vector, not particle, indices.

The maximum projection of Eq. (41) occurs where the coefficients of $\hat{i}$ ($i \neq \ell, m$) are zero. This gives $(n - 2)$ relations of the form

$$\sum_j \underline{B}_{ij} X_j = 0 \quad (i \neq \ell, m), \quad (49)$$

which can be used to solve for X_j ($j \neq \ell, m$) in terms of X_ℓ and X_m . We substitute these relations in Eq. (41) to get

$$(X_\ell, X_m) R \begin{pmatrix} X_\ell \\ X_m \end{pmatrix} = \left(\frac{V_n}{K(n)} \right)^{2/n}. \quad (50)$$

The area of this ellipse is

$$A = \pi \left(\frac{V_n}{K(n)} \right)^{2/n} |R|^{-1/2}. \quad (51)$$

ACKNOWLEDGMENTS

I thank K. Crandall, R. Jameson, and R. Mills for their useful suggestions and criticisms.

REFERENCES

1. E. D. Courant and H. S. Snyder, "Theory of the Alternating-Gradient Synchrotron," *Ann. Phys.* 3, 1-49 (1958).

2. A. Law, "A General Hyperellipsoid Least-Squares Fitting Routine," Los Alamos Scientific Laboratory, internal memorandum, June 1977.
3. K. R. Crandall, Los Alamos Scientific Laboratory, personal communication, January 1978.