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

E. Alan Wadlinger
Los Alamos Scientific Laboratory
Los Alamos, NM 87545

Summary

Several methods are presented for determining the shape parameters, which in two dimensions are the Courant-Snyder parameters, and the volume of an ellipse or hyperellipse that represent 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. A two-dimensional space is treated first, then generalized to higher dimensions. Computer programs using these techniques have been written.

Two-Dimensional Case

The equation for an ellipse may be written

$$\gamma x^2 + 2\alpha xy + \beta y^2 = E/\pi = A \quad , \tag{1}$$

$$\gamma\beta - \alpha^2 - 1 = 0 \quad , \tag{2}$$

where α, β, and γ are the Courant and Snyder parameters[1] 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

$$A_i = \gamma x_i^2 + 2\alpha x_i y_i + \beta y_i^2 \quad . \tag{3}$$

If there are N particles and each particle has an emittance, A_i $(i = 1,N)$, a total summed emittance can be defined as:

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

The rms ellipse parameters minimize I. These parameters are found most easily by multiplying Eq. (2) 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. It is found that with

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

the results are

$$\gamma = \frac{\sum y_i^2}{D} , \ \beta = \frac{\sum x_i^2}{D} , \ \alpha = -\frac{\sum x_i y_i}{D} ,$$

and

$$D = \sqrt{\sum x_i^2 \sum y_i^2 - \left(\sum x_i y_i\right)^2} \quad , \tag{5}$$

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

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

$$J = \sum A_i^2 + \lambda(\gamma\beta - \alpha^2 - 1) \quad . \tag{6}$$

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

$$\gamma\beta - \alpha^2 - 1 = 0 \quad , \tag{2}$$

$$\alpha \sum A_i y_i^2 + \gamma \sum A_i x_i y_i = 0 \quad , \tag{7}$$

and

$$\alpha \sum A_i x_i^2 + \beta \sum A_i x_i y_i = 0 \quad . \tag{8}$$

Equations (2), (3), (7), and (8) define α, β, and γ. An iteration procedure issued to solve these equations.[2] The starting values for the procedure may be obtained from Eq. (5).

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, it was assumed that the particle distribution is centered about the origin. This deficiency can be corrected by substituting $(x_i - x_o)$ for x_i and $(y_i - y_o)$ for y_i in Eq. (3) then minimizing Eq. (4b) or Eq. (6) with respect to x_o and y_o, where (x_o, y_o) is the origin of the ellipse.

*Work performed under the auspices of the U.S. Department of Energy.

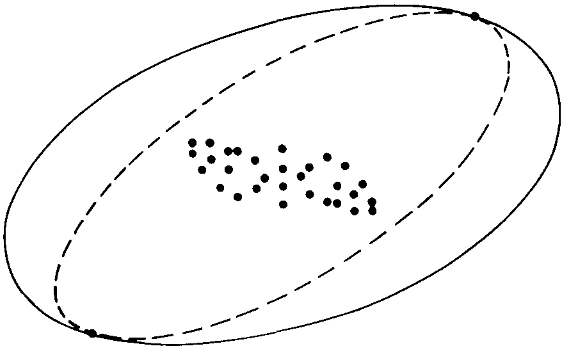

Fig. 1 Weighted and rms ellipses that encompass all particles. The inner ellipse (dashed line) resulted from the "weighted" fit.

Higher Dimensional Hyperellipsoid

The equation of a hyperellipse in an n-dimensional hyper-space is:

$$X^T \underline{A} X = 1 \quad , \tag{9}$$

where $\underline{A}$ is an n x n -dimensional symmetric matrix and X is an n-dimensional column vector. Given a set of k points (particles) represented by X_i (i = 1,k) in this hyper-space, the hyper-ellipse defining matrix $\underline{A}$ that best represents this set of points is found. The fit is by minimizing the volume enclosed by the hyper-ellipse characterized by the data. The technique[3] of minimizing the function,

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

with respect to the elements of the symmetric matrix $\underline{A}$ was not considered. This technique minimizes the distances between the data points and the surface of the hyperellipsoid defined by $\underline{A}$.

An explicit relationship between the volume of the hyperellipsoid and the shape-defining parameters of the ellipse was first obtained. The volume of an n-dimensional upright ellipsoid,

$$\sum_{i=1}^{n} \frac{X_i^2}{d_i^2} = 1 \quad , \tag{10}$$

with semiaxes of length d_i is

$$V_n = K(n) \prod_{i=1}^{n} d_i \ ; \quad K(n) = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2} + 1\right)} \quad . \tag{11}$$

Note that the sum in Eq. (10) and the product in Eq. (11) are over the n-dimensional vector space, not over the particle coordinates. Now let $\underline{U}$ be an orthogonal matrix that diagonalizes $\underline{A}$, Eq. (9), let $\underline{D}$ be the resulting diagonal matrix, ($\underline{D} = \underline{U}\underline{A}\underline{U}^T$), and let Y be the transformed vector, (Y = $\underline{U}$X). Then,

$$Y^T \underline{D} Y = 1 \quad . \tag{12}$$

Equation (12) is in the form of Eq. (10) with components of $\underline{D}$,

$$\underline{D}_{ij} = \frac{\delta_{ij}}{d_j^2} \quad , \tag{13}$$

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

$$|\underline{D}| = \left[\prod_{i=1}^{n} d_i^2 \right]^{-1} = \left(\frac{K(n)}{V_n} \right)^2 = |\underline{A}| \quad . \tag{14}$$

The determinant value is invariant under an orthogonal transformation. Define a new matrix $\underline{B}$ whose elements are

$$\underline{B}_{ij} = \left(\frac{V_n}{K(n)} \right)^{2/n} \underline{A}_{ij} \quad ; \tag{15}$$

then,

$$|\underline{B}| = \left[\left(\frac{V_n}{K(n)} \right)^{2/n} \right]^n |\underline{A}| = 1 \quad . \tag{16}$$

From Eq. (9), it is found that:

$$X^T \underline{B} X = \left(\frac{V_n}{K(n)} \right)^{2/n} \quad , \tag{17}$$

with

$$|\underline{B}| = 1 \quad . \tag{18}$$

Note that with n = 2, Eq. (1) is obtained.

The hyperellipse volume, determined from Eq. (17), is

$$V_n = K(n) \left[X^T \underline{B} X \right]^{n/2} \quad . \tag{19}$$

The hyperellipse volume (or some power of the volume) is minimized for k particles in an n-dimensional hyper-space. Define

$$I = \sum_{i=1}^{k} V_{n_i}^{\,m} + \lambda \left(|\underline{B}| - 1 \right) \quad , \tag{20}$$

where i is the sum over k particles, m is an exponent, n is the hyper-space dimension, and λ is a Lagrange multiplier. Let $\underline{B}$ be symmetric. Substituting Eq. (19) into Eq. (20) and absorbing the K(n) term into the Lagrange multiplier, yields

$$I = \sum_{i=1}^{k} \left(X_i^T \underline{B} X_i \right)^{nm/2} + \lambda \left(|\underline{B}| - 1 \right) \quad . \tag{21}$$

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

$$\begin{aligned} \partial I / \partial \underline{B}_{ij} &= 0 \quad (i \leqslant j = 1, \ldots, n) \quad , \\ \partial I / \partial \lambda &= 0 \quad . \end{aligned} \tag{22}$$

The restriction $(i \leq j = 1, \ldots, n)$ accounts for $\underline{B}$ being symmetric ($\underline{B}_{ij} = \underline{B}_{ji}$). Equation (22) is used to solve for $\underline{B}$. In all but the simplest cases, Eq. (22) will be nonlinear and will have to be solved by an iterative approximation scheme. If the distribution has a non-zero centroid, $(X_i - X_o)$ can be substituted for X_i in Eq. (21) which can be minimized with respect to X_o.

Given the ellipse in Eq. (17), the projected area in the two-dimensional plane defined by $\hat{\ell}$ $\hat{m}$ can be found. The gradient of Eq. (17) can be used to obtain

$$\nabla (X^T \underline{B} X) = 2 \sum_i \hat{i} \sum_j \underline{B}_{ij} X_j \quad , \tag{23}$$

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

The maximum projection of Eq. (17) occurs where the coefficients of $\hat{i}$ $(\hat{i} \neq \ell, m)$ are zero. This gives (n - 2) relations of the form

$$\sum_j B_{ij} X_j = 0 \quad (i \neq \ell, m) \quad , \tag{24}$$

which can be used to solve for X_j $(j \neq \ell, m)$ in terms of X_ℓ and X_m. Substitute these relations in Eq. (17) to get

$$\left(X_\ell , X_m \right) \underline{R} \begin{pmatrix} X_\ell \\ X_m \end{pmatrix} = \left(\frac{V_n}{K(n)} \right)^{2/n} \quad . \tag{25}$$

The area of this ellipse is

$$A = \pi \left(\frac{V_n}{K(n)} \right)^{2/n} |R|^{-1/2} \quad . \tag{26}$$

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. E. A. Wadlinger, "General Least-Squares Fitting Procedures to Minimize the Volume of a Hyperellipsoid," Los Alamos Scientific Laboratory report to be published.

3. A. Law, "A General Hyperellipsoid Least-Squares Fitting Routine," Los Alamos Scientific Laboratory, internal memorandum, June 1977.