$\widehat{\boldsymbol{\alpha}}$, which maximizes the joint probability density for all the data, given by

$$\mathscr{L}(\boldsymbol{\alpha}) = \prod_i f(x_i;\boldsymbol{\alpha}) \ , \tag{17.5}$$

where $\mathscr{L}$ is called the likelihood. It is usually easier to work with $\ln\mathscr{L}$, and since both are maximized for the same set of $\boldsymbol{\alpha}$, it is sufficient to solve the *likelihood equation*

$$\frac{\partial \ln \mathscr{L}}{\partial \alpha_n} = 0 \ . \tag{17.6}$$

The solution is called the *maximum likelihood estimate* of $\boldsymbol{\alpha}$. If $\partial \ln \mathscr{L}/\partial \alpha_n$ is linear in the vicinity of the root, an efficient estimator is guaranteed. For large data samples, the central limit theorem will usually assure this condition in some significant neighborhood of zero. In this case, in the neighborhood of the maximum $\ln\mathscr{L}$ is a downward-curving paraboloid and $\mathscr{L}$ is proportional to a multivariate Gaussian.

17.3.2. *Error estimates*:

The covariance matrix V may be estimated from

$$V_{nm} = \left(E\left[-\frac{\partial^2 \ln \mathscr{L}}{\partial \alpha_n \, \partial \alpha_m}\bigg|_{\widehat{\alpha}} \right]\right)^{-1} \ . \tag{17.7}$$

If $\partial \ln \mathscr{L}/\partial \alpha_n$ is linear near $\widehat{\alpha}$, the estimator has the least possible variance and the second derivative above is constant; hence the "expectation" operation has no effect.

Geometrically, s-standard-deviation errors may be estimated as the contour $\boldsymbol{\alpha}$ satisfying

$$\ln\mathscr{L}(\boldsymbol{\alpha}) = \ln\mathscr{L}_{\rm max} - s^2/2 \ , \tag{17.8}$$

these errors may not be symmetric and may even consist of disjoint regions. All errors estimated from likelihoods should be regarded as approximate unless $d(\ln\mathscr{L})/d\alpha_m$ is linear over a wide region.

17.4. Method of least squares

The *method of least squares* can be derived from the maximum likelihood theorem. We suppose a set of N measurements at points x_i. The ith measurement y_i is assumed to be chosen from a Gaussian distribution with mean $F(x_i;\boldsymbol{\alpha})$ and variance σ_i^2. If the y_i's are independent,

$$\chi^2 = -2\ln\mathscr{L} + \text{constant} = \sum_1^N \frac{[y_i - F(x_i;\boldsymbol{\alpha})]^2}{\sigma_i^2} \ . \tag{17.10}$$

Finding the set of parameters $\boldsymbol{\alpha}$ which maximizes $\mathscr{L}$ is the same as finding the set which minimizes χ^2. If the y_i's are not independent a