and it maps the $N$-vector $x$ into our parameters $\alpha$

$$\alpha = G^+ x. \quad (13.290)$$

The product $G^+ G = I_M$ is the $M \times M$ identity matrix, while

$$GG^+ = P \quad (13.291)$$

is an $N \times N$ projection operator (exercise 13.38) onto the $M \times M$ subspace for which $G^+ G = I_M$ is the identity operator. Like all projection operators, $P$ satisfies $P^2 = P$.

13.20 Karl Pearson’s Chi-Squared Statistic

The argument of the exponential (13.284) in $P(x)$ is (the negative of) Karl Pearson’s chi-squared statistic (Pearson, 1900)

$$\chi^2 = \sum_{j=1}^{N} \frac{(x_j - f(t_j; \alpha))^2}{2\sigma^2}. \quad (13.292)$$

When the function $f(t; \alpha)$ is linear (13.286) in $\alpha$, the $N$-vector $f(t_j; \alpha)$ is $f = G \alpha$. Pearson’s $\chi^2$ then is

$$\chi^2 = (x - G \alpha)^2/2\sigma^2. \quad (13.293)$$

Now (13.290) tells us that $\alpha = G^+ x$, and so in terms of the projection operator $P = GG^+$, the vector $x - G \alpha$ is

$$x - G \alpha = x - GG^+ x = (I - GG^+) x = (I - P) x. \quad (13.294)$$

So $\chi^2$ is proportional to the squared length

$$\chi^2 = \tilde{x}^2/2\sigma^2 \quad (13.295)$$

of the vector

$$\tilde{x} \equiv (I - P) x. \quad (13.296)$$

Thus if the matrix $G$ has rank $M$, and the vector $x$ has $N$ independent components, then the vector $\tilde{x}$ has only $N - M$ independent components.

**Example 13.17 (Two Position Measurements)** Suppose we measure a position twice with error $\sigma$, get $x_1$ and $x_2$, and choose $G^T = (1, 1)$. Then
the single parameter \( \alpha \) is their average \( \alpha = (x_1 + x_2)/2 \), and \( \chi^2 \) is
\[
\chi^2 = \left\{ \frac{(x_1 - x_1 + x_2)^2 + [x_2 - (x_1 + x_2)/2]^2}{2\sigma^2} \right\} \left( \frac{(x_1 - x_2)^2}{2}\right) / 2\sigma^2
\]
\[
= \left\{ \frac{(x_1 - x_2)^2 + (x_2 - x_1)^2}{2}\right\} / 2\sigma^2
\]
\[
= \left( \frac{(x_1 - x_2)^2}{2}\right) / 2\sigma^2. \tag{13.297}
\]
Thus instead of having two independent components \( x_1 \) and \( x_2 \), \( \chi^2 \) just has one \( (x_1 - x_2)/\sqrt{2} \).

We can see how this happens more generally if we use as basis vectors the \( N - M \) orthonormal vectors \( |j\rangle \) in the kernel of \( P \) (that is, the \( |j\rangle\)'s annihilated by \( P \))
\[
P|j\rangle = 0 \quad 1 \leq j \leq N - M \tag{13.298}
\]
and the \( M \) that lie in the range of the projection operator \( P \)
\[
P|k\rangle = |k\rangle \quad N - M + 1 \leq k \leq N. \tag{13.299}
\]
In terms of these basis vectors, the \( N \)-vector \( \mathbf{x} \) is
\[
\mathbf{x} = \sum_{j=1}^{N-M} x_j |j\rangle + \sum_{k=N-M+1}^{N} x_k |k\rangle \tag{13.300}
\]
and the last \( M \) components of the vector \( \tilde{\mathbf{x}} \) vanish
\[
\tilde{\mathbf{x}} = (I - P) \mathbf{x} = \sum_{j=1}^{N-M} x_j |j\rangle. \tag{13.301}
\]

**Example 13.18** (N position measurements) Suppose the \( N \) values of \( x_j \) are the measured values of the position \( f(t_j; \alpha) = x_j \) of some object. Then \( M = 1 \), and we choose \( G_j = g_1(t_j) = 1 \) for \( j = 1, \ldots, N \). Now \( G^T G = N \) is a \( 1 \times 1 \) matrix, the number \( N \), and the parameter \( \alpha \) is the mean \( \bar{x} \)
\[
\alpha = G^T \mathbf{x} = (G^T G)^{-1} G^T \mathbf{x} = \frac{1}{N} \sum_{j=1}^{N} x_j = \bar{x} \tag{13.302}
\]
of the \( N \) position measurements \( x_j \). So the vector \( \tilde{\mathbf{x}} \) has components \( \tilde{x}_j = x_j - \bar{x} \) and is orthogonal to \( G^T = (1, 1, \ldots, 1) \)
\[
G^T \tilde{\mathbf{x}} = \left( \sum_{j=1}^{N} x_j \right) - N \bar{x} = 0. \tag{13.303}
\]
The matrix $G^T$ has rank 1, and the vector $\tilde{x}$ has $N - 1$ independent components.

Suppose now that we have determined our $M$ parameters $\alpha$ and have a theoretical fit

$$x = f(t; \alpha) = \sum_{k=1}^{M} g_k(t) \alpha_k$$

(13.304)

which when we apply it to $N$ measurements $x_j$ gives $\chi^2$ as

$$\chi^2 = (\tilde{x})^2 / 2\sigma^2.$$  

(13.305)

How good is our fit?

A $\chi^2$ distribution with $N - M$ degrees of freedom has by (13.202) mean

$$E[\chi^2] = N - M$$

(13.306)

and variance

$$V[\chi^2] = 2(N - M).$$

(13.307)

So our $\chi^2$ should be about

$$\chi^2 \approx N - M \pm \sqrt{2(N - M)}.$$ 

(13.308)

If it lies within this range, then (13.304) is a good fit to the data. But if it exceeds $N - M + \sqrt{2(N - M)}$, then the fit isn’t so good. On the other hand, if $\chi^2$ is less than $N - M - \sqrt{2(N - M)}$, then we may have used too many parameters or overestimated $\sigma$. Indeed, by using $N$ parameters with $GG^+ = I_N$, we could get $\chi^2 = 0$ every time.

The probability that $\chi^2$ exceeds $\chi_0^2$ is the integral (13.201)

$$\Pr_n(\chi^2 > \chi_0^2) = \int_{\chi_0^2}^{\infty} P_n(\chi^2 / 2) d\chi^2 = \int_{\chi_0^2}^{\infty} \frac{1}{2\Gamma(n/2)} \left( \frac{\chi^2}{2} \right)^{n/2-1} e^{-\chi^2/2} d\chi^2$$

(13.309)

in which $n = N - M$ is the number of data points minus the number of parameters, and $\Gamma(n/2)$ is the gamma function (5.102, 4.62). So an $M$-parameter fit to $N$ data points has only a chance of $\epsilon$ of being good if its $\chi^2$ is greater than a $\chi_0^2$ for which $\Pr_{N-M}(\chi^2 > \chi_0^2) = \epsilon$. These probabilities $\Pr_{N-M}(\chi^2 > \chi_0^2)$ are plotted in Fig. 13.6 for $N - M = 2, 4, 6, 8$, and 10. In particular, the probability of a value of $\chi^2$ greater than $\chi_0^2 = 20$ respectively is 0.000045, 0.000499, 0.00277, 0.010336, and 0.029253 for $N - M = 2, 4, 6, 8$, and 10.
The Chi-Squared Test

Figure 13.6 The probabilities $\Pr_{N-M}(\chi^2 > \chi^2_0)$ are plotted from left to right for $N - M = 2, 4, 6, 8, \text{ and } 10$ degrees of freedom as functions of $\chi^2_0$.

13.21 Kolmogorov’s Test

Suppose we want to use a sequence of $N$ measurements $x_j$ to determine the probability distribution that they come from. Our empirical probability distribution is

$$P_e^{(N)}(x) = \frac{1}{N} \sum_{j=1}^{N} \delta(x - x_j). \quad (13.310)$$

Our cumulative probability for events less than $x$ then is

$$\Pr_e^{(N)}(-\infty, x) = \int_{-\infty}^{x} P_e^{(N)}(x') \, dx' = \int_{-\infty}^{x} \frac{1}{N} \sum_{j=1}^{N} \delta(x' - x_j) \, dx'. \quad (13.311)$$