from the mean (instead of their absolute values). Thus you have a measure of the variation between the points. Modifying eq. (1) accordingly,

(12)
$$\hat{p}' = \sum_{i}^{n} (\alpha_i)\,(\hat{p}_i - bg_i)$$

and then to get the final estimated field, the deviation estimate is added to the background at that point:

(13)
$$\hat{p}(x) = \hat{p}'(x) + bg(x)$$

Thus, in order to calculate an interpolated field, the user must supply an input data field composed of n $\hat{p}_i(x_i)$ points, an error variance at each of those points, σ_{jj}^2, a background field, bg which is defined at all input data locations as well as all output (interpolated) locations, a correlation length, c, and a function with which to calculate the correlation values in R and S.

(10) $$A = \alpha_i$$

$$q^2 S = E p p_j$$

$$q^2 R = E p_i p_j$$

R is the correlation matrix between all pairs of data points; S is the correlation vector between the missing point, x, and each data points; A is the vector of weighting functions, and q is the standard deviation of the measured field (a normalization factor).

The correlation function for an isotropic model is generally chosen to be of this form:

(11) $$R(d_j) = \exp\left(-\left(\frac{d_j}{c}\right)^2\right)$$

where d is the separation distance between each pair of the n known points (at x_j). c is the correlation length, a parameter that must be chosen for the field. The function used to describe the correlation of S is of the same form, except that the distance d, in that case, is taken between each of the known measured points (for each j) and the missing point at x.

Equation (9) is solved for A to yield the series of weighting functions for each point which are then to be plugged back into equation (1) to give the estimated value, the new, interpolated field, at each points, x
For the case of a non-zero mean of the field p, the background value (or mean) must be first subtracted from the measured values to yield an estimated field based on the deviation of points

Simplifying and using equation (3) (which gives 0 for the value of the third, fourth and sixth terms), gives,

(8) $$\alpha_i\left[Ep_ip_j + \sigma_{jj}^{2}\right] = Epp_j$$

α_i **are the weighting factors for each data point i. How each point is weighted can be seen, from this equation, to depend on its relation to the missing point, its relation to other data points, and its associated error. Thus, data points which are far from the missing point, x, receive a proportionally smaller weight, similarly if they are associated with a large measurement error. Points which are clustered together (relative to the determined length scale) are also discounted to avoid biasing.**

But, since this is an equation where a sum is performed over all i for each j, it is analogous to a matrix formulation:

$$\left[\begin{bmatrix} Ep_0p_0 & Ep_1p_0 & \dots \\ Ep_0p_1 & Ep_1p_1 & \dots \\ \dots & \dots & \dots \end{bmatrix} + \begin{bmatrix} \sigma^2{}_{00} & 0 & \dots \\ 0 & \sigma^2{}_{11} & \dots \\ \dots & \dots & \dots \end{bmatrix}\right]\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \dots \end{bmatrix} = \begin{bmatrix} Epp_0 & Epp_1 & \dots \end{bmatrix}$$

(9) $$\left[R + \frac{\sigma^2}{q^2}I\right]A = S$$

where I is the identity matrix, S and A are vectors, and R is a matrix, defined by:

the value, σ_{jj}^{2}.

This method of estimating unknown points in a field is called optimal interpolation because the weighting functions are chosen so that the variance of the difference between the estimated field and the"true" field is "optimized".

$$\varepsilon = (\hat{p} - p)$$

Defining the variance of the difference field, and inserting equation (1) gives:

(4) $$var\varepsilon = E\left(\varepsilon^{2}\right) = E(\hat{p} - p)^{2} = E\left(\sum_{i}^{n} \alpha_{i}\hat{p}_{i} - p\right)^{2}$$

The derivative with respect to all weighting functions of the variance of this error, ε, is set equal to zero to minimize this difference:

(5) $$\frac{\partial}{\partial\alpha_{j}}(var\varepsilon) = 0$$

Taking the derivative with respect to the weighting function, for any j, inserting equation (2),and switching the order of the two linear operators, E() and $\frac{\partial}{\partial\alpha_{j}}$(), gives:

(6) $$2E\left(\left(\left[\sum_{i}^{n} \alpha_{i}(p_{i} + \delta_{i})\right] - p\right) \times (p_{j} + \delta_{j})\right) = 0$$

where this is a sum over all i. Distributing,

(7) $$2E[\alpha_{i}(p_{i}p_{j} + \delta_{i}\delta_{j} + p_{i}\delta_{j} + p_{j}\delta_{i}) - pp_{j} - p\delta_{j}] = 0$$

The goal is to estimate an unknown "true" field, p(x), at any point x, using a series of n weighting functions, α_i, and a known series of n measurements, $\hat{p}_i(x_i)$, taken at the surrounding points, x_i (for a field with mean = zero):

(1)

$$\hat{p} = \sum_{i}^{n} (\alpha_i) \hat{p}_i$$

where $\hat{p}$ is the estimated (or interpolated) field at a particular point, x. Each of these measurements is assumed to have some associated error, so that they deviate from the "true" field as well:

(2)

$$\hat{p}_i = p_i + \delta_i$$

and these measurement errors are assumed to be uncorrelated, so that the cross-correlation variance is zero:

(3)

$$E(\delta_i \delta_j) = 0$$

$$E\left(\delta^2{}_j\right) = \sigma_{jj}{}^2$$

and,the error is also assumed to be uncorrelated with the measurement itself, p_i, as well as the true field, p:

$$E p_i \delta_j = 0$$

where E is the expectation value operator (the mean). And the variance of measurement error at each point,x_j is defined to be