The Bayesian geostatistical approach is described in detail by Kitanidis and Vomvoris (1983), Hoeksema and Kitanidis (1984), Kitanidis (1995), and Nowak and Cirpka (2004) among others. The mathematics are reviewed here.

The Bayesian Geostatistical Approach

In the Bayesian geostatistical approach, the posterior pdf is calculated as

$$p(\mathbf{s}|\mathbf{y}) \propto \underbrace{\exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{h}(\mathbf{s}))^T \mathbf{R}^{-1}(\mathbf{y}-\mathbf{h}(\mathbf{s}))\right)}_{\mathrm{L}(\mathbf{y}|\mathbf{s})} \underbrace{\exp\left(-\frac{1}{2}(\mathbf{s}-\mathbf{X}\beta^*)^T \mathbf{G}_{\mathbf{ss}}^{-1}(\mathbf{s}-\mathbf{X}\beta^*)\right)}_{\mathrm{p}(\mathbf{s})} \tag{3.1}$$

where $\mathbf{s}$ is the $m \times 1$ vector of parameter values at distributed spatial locations in the model, $\mathbf{X}\beta^*$ is the prior mean, $\mathbf{G}_{\mathbf{ss}}$ is the $m \times m$ prior covariance of $(\mathbf{s}-\mathbf{X}\beta^*)$, $\mathbf{h}(\mathbf{s})$ is the $n \times 1$ vector of modeled forecasts colocated with observations $(\mathbf{y})$, and $\mathbf{R}$ is the $n \times n$ epistemic uncertainty covariance, modeled as $\sigma_R^2\mathbf{W}$ where σ_R^2 represents epistemic uncertainty, and $\mathbf{W}$ is an $n \times n$ diagonal matrix containing the observation weights. In general terms, the likelihood function $(L(\mathbf{y}|\mathbf{s}))$ characterizes the misfit between model forecasts and observations whereas the prior pdf $(p(\mathbf{s}))$ defines a characteristic (such as smoothness or continuity) that is assumed to apply to the parameter field. The prior pdf also serves the role of regularization. The best estimate of $\mathbf{s}$ maximizes the posterior pdf. A computationally efficient method to find the best estimates of $\mathbf{s}$ and β $\left(\hat{\mathbf{s}} \text{ and } \hat{\beta}, \text{ respectively}\right)$ is through

$$\hat{\mathbf{s}} = \mathbf{X}\hat{\beta} + \mathbf{Q}_{\mathbf{ss}}\mathbf{H}^{\mathbf{T}}\xi \tag{3.2}$$

which is the superposition of the prior mean (first term) and an innovation term that accounts for deviations of the model outputs from the observations (second term). In this context, β in the first term is not the prior mean, but is the best estimate of the mean (mapped onto the parameter field through the $m \times p$ $\mathbf{X}$ matrix) whereas the second term is fluctuations about the estimated mean. $\mathbf{H}$ in the second term (often referred to as the $n \times m$ Jacobian, sensitivity, or susceptibility matrix) is the sensitivity of observation values to parameter values where $H_{ij} = \frac{\partial \mathbf{h}(\mathbf{s})_i}{\partial \mathbf{s}_j}$, which can be calculated by using either finite difference or adjoint-state methods. In bgaPEST, finite-difference calculations for $\mathbf{H}$ are made by using PEST whereas adjoint-state calculations depend on the specific model being used and must be made by using an external program.

The values for $\hat{\beta}$ and ξ are found by solving the $(n+p) \times (n+p)$ linear system of cokriging equations

$$\begin{bmatrix} \mathbf{Q}_{\mathbf{yy}} & \mathbf{HX} \\ \mathbf{X}^T\mathbf{H}^T & -\mathbf{Q}_{\beta\beta}^{-1} \end{bmatrix} \begin{bmatrix} \xi \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ -\mathbf{Q}_{\beta\beta}^{-1}\beta^* \end{bmatrix} \tag{3.3}$$

where $\mathbf{Q}_{\mathbf{yy}}$ is the $n \times n$ auto-covariance matrix of the observations, defined as $\mathbf{HQ}_{\mathbf{ss}}\mathbf{H}^{\mathbf{T}} + \mathbf{R}$, n is the number of observations, and p is the number of beta associations. In hydrogeologic applications, the numerical forward model is typically nonlinear. Further nonlinearity can be induced by using a logarithmic or power transformation, which is a convenient way to enforce non-negativity on parameters.

Provided that the nonlinearities introduced are not too extreme, a solution can be obtained through successive linearizations following the quasi-linear extension (Kitanidis, 1995). The forward model, $\mathbf{h}(\mathbf{s})$ is expanded about the current best estimate of the parameters $\tilde{\mathbf{s}}$

$$\mathbf{h}(\mathbf{s}) \approx \mathbf{h}(\tilde{\mathbf{s}}) + \tilde{\mathbf{H}}(\mathbf{s}-\tilde{\mathbf{s}}) \tag{3.4}$$

where $\tilde{\mathbf{H}}$, as a function of $\tilde{\mathbf{s}}$, is evaluated at each linearization. We assign the subscript $_k$ to indicate iteration

number, and correct the measurements for the kth linearization as

$$\mathbf{y}_k^{'} = \mathbf{y} - \mathbf{h}(\tilde{\mathbf{s}}_k) + \tilde{\mathbf{H}}_k \tilde{\mathbf{s}}_k. \tag{3.5}$$

Then the cokriging equations (equation 3.3) are updated

$$\begin{bmatrix} \tilde{\mathbf{Q}}_{\mathbf{yy},k} & \tilde{\mathbf{H}}_k \mathbf{X} \\ \mathbf{X}^T \tilde{\mathbf{H}}_k^T & -\mathbf{Q}_{\beta\beta}^{-\mathbf{1}} \end{bmatrix} \begin{bmatrix} \xi_k \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \mathbf{y}_k^{'} \\ -\mathbf{Q}_{\beta\beta}^{-\mathbf{1}} \beta^* \end{bmatrix} \tag{3.6}$$

where $\tilde{\mathbf{Q}}_{\mathbf{yy},k} = \tilde{\mathbf{H}}_k \mathbf{Q}_{\mathbf{ss}} \tilde{\mathbf{H}}_k^T + \mathbf{R}$. From this set of equations, the next estimate of $\mathbf{s}$ is

$$\tilde{\mathbf{s}}_{k+1} = \mathbf{X}\hat{\beta}_k + \mathbf{Q}_{\mathbf{ss}} \tilde{\mathbf{H}}_k^T \xi_k. \tag{3.7}$$

This process can be iterated until there is minimal difference in the parameter estimates or minimal further improvement in the objective function. The objective function, which we seek to minimize, is $-\ln p(\mathbf{s}|\mathbf{y})$ which is equivalent to maximizing equation 3.1

$$\Phi_T = -\frac{1}{2} \ln p(\mathbf{s}|\mathbf{y}) = (\mathbf{s} - \mathbf{X}\beta^*)^T \mathbf{G}_{\mathbf{ss}}^{-\mathbf{1}} (\mathbf{s} - \mathbf{X}\beta^*) - \frac{1}{2} (\mathbf{y} - \mathbf{h}(\mathbf{s}))^T \mathbf{R}^{-\mathbf{1}} (\mathbf{y} - \mathbf{h}(\mathbf{s})) \tag{3.8}$$

Line Search

In some cases, numerical instability makes convergence difficult. A line search is implemented in which a linear search is performed between the most recent best estimate of the parameters ($\hat{\mathbf{s}}$) and the current linearization of the parameters ($\widetilde{\mathbf{s}}$), seeking a parameter value that minimizes an objective function.

The line search optimizes a single parameter, ρ, along a linear dimension between $\hat{\mathbf{s}}$ and $\widetilde{\mathbf{s}}$ as

$$\mathbf{s}_{opt} = \hat{\mathbf{s}}\rho + \widetilde{\mathbf{s}}(1-\rho) \tag{3.9}$$

where $\mathbf{s}_{opt}$ minimizes the objective function, Φ_T, using a Nelder-Mead simplex (see, for example, Press and others, 1992), which guarantees monotonic decrease in Φ_T over successive iterations. It is recommended to limit the number of line-search iterations to a relatively low number, because the goal of handling weak linearity is balanced against the computations required to perform the line search. The greatest advantage is likely achieved in the first few (less than five) iterations. The role of the line search is not to find a minimum value of Φ_T because the nonlinearity of the overall problem prevents it. Rather, the line search is meant to be a correction of search direction for stability.

Parameter Field Anisotropy

In distributed parameter fields, such as hydraulic conductivity in groundwater models, it is common to encounter anisotropy along an axis that may or not be aligned with the coordinate (x, y, z) axes. bgaPEST allows the definition of anisotropy in a horizontal plane at any angle from the x-axis and also in the vertical direction. The general layout of horizontal anisotropy is illustrated in figure 3.1. The angle from the x-axis (specified in degrees) is defined by `horiz_angle`, and the amount of anisotropy is defined by `horiz_ratio`. The designation `p_max` refers to the direction with maximum parameter values and `p_min` refers to the direction of minimum parameter values. The ratio is used to adjust the effective distance (and thereby the covariance values) along that principal direction. The user supplies values for `horiz_angle` and `horiz_ratio` for each beta association. If some beta associations are not meant to exhibit anisotropy, the user may simply set `horiz_ratio=1.0`. If none of the beta associations exhibit anisotropy, the entire `parameter_anisotropy`

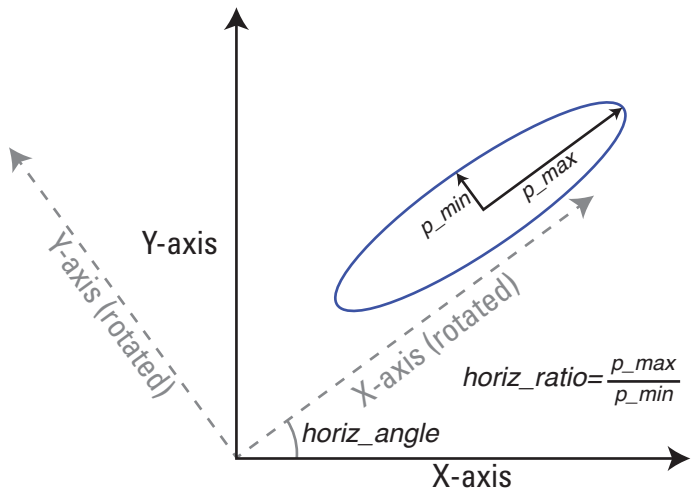

Figure 3.1. Schematic diagram of conventions defining horizontal anisotropy in bgaPEST .

block can be eliminated by setting the algorithmic control variable `par_anisotropy=0`, which means the block (if present) is ignored.

Anisotropy is introduced in the calculation of distances that are then used in the calculation of the prior covariance matrix $\mathbf{Q}_{ss}$ discussed below. Every pair of points must first be rotated into the principal direction orientation. This is accomplished by means of a rotation matrix:

$$\begin{bmatrix} x_{rot,i} \\ y_{rot,i} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} \tag{3.10}$$

where i indicates the ith point of the pair $(i = 1,2)$, θ is the angle (in degrees) specified by `horiz_angle`, x and y constitute the point coordinates in the original coordinate system, and x_{rot} and y_{rot} constitute the location projected into the coordinate system corresponding to the orientation of horizontal anisotropy. Once this projection is made, the horizontal distance is calculated as

$$distance = \sqrt{(x_{rot,1} - x_{rot,2})^2 + \texttt{horiz\_ratio} \times (y_{rot,1} - y_{rot,2})^2} \tag{3.11}$$

For three-dimensional parameter fields, a second anisotropy ratio may be specified as `vertical_ratio`. In the vertical direction, no angle is specified, so the rotation step is not required and distance is calculated as

$$distance = \sqrt{\begin{array}{c} (x_{rot,1} - x_{rot,2})^2 \\ +\texttt{horiz\_ratio} \times (y_{rot,1} - y_{rot,2})^2 \\ +\texttt{vertical\_ratio} \times (z_{rot,1} - z_{rot,2})^2 \end{array}} \tag{3.12}$$

Prior Probability Density Function

The prior pdf of $\mathbf{s}$ can be characterized as multi-Gaussian through its mean and covariance. The $(m \times 1)$ unknown parameter vector, $\mathbf{s}$, is modeled as a random process with mean

$$E[\mathbf{s}] = \mathbf{X}\beta \tag{3.13}$$