

Glossary of important terms for GW1876

April 22, 2018

Notes on mathematics symbology

It would be really nice if a totally consistent set of symbols was used in the mathematics of parameter estimation and uncertainty analysis. But....no such luck! So many researchers in different fields over a long time have contributed to the work used in this class. As a result, this glossary is an attempt to highlight some general symbology and clarify some terms.

First of all, though, one thing that is consistent (mostly!) is the general linear algebra notation used throughout the class.

scalar values Lowercase, non-bold font indicates a scalar (single) value: x, y, z

vectors Lowercase, bold font indicates a vector of values: $\mathbf{x}, \mathbf{y}, \mathbf{z}$

matrices Uppercase, bold font indicates a matrix of values: $\mathbf{X}, \mathbf{Q}, \mathbf{J}$ A matrix with $\langle \cdot \rangle^T$ indicates a matrix transpose. A matrix with $\langle \cdot \rangle^{-1}$ indicates a matrix inverse.

matrix multiplication Then, matrix multiplication (with either other matrices or vectors) is expressed simply by adjacent matrices: $\mathbf{X}\mathbf{y}, \mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X}$

Glossary of terms and equations

Parameters Variable input values for models, typically representing system properties and forcings. Values to be estimated in the history matching process. Typically identified as k, p , or x, θ , ($\mathbf{k}, \mathbf{p}, \boldsymbol{\theta}$ or \mathbf{x} for multiple parameters in a vector).

Observation Measured system state values. These values are used to compare with model outputs collocated in space and time. The term is often used to mean *both* field measurements and outputs from the model. When referring to a measured value, observations are typically identified by the variables y or o (\mathbf{y} or \mathbf{o} for multiple parameters in a vector)

Modeled Equivalent (aka Simulated Equivalent) A modeled value collocated in time and space with an observation. There are various ways to identify a single or multiple modeled equivalent values (and, to make things confusing, they are often *also* called “observations”!)

Single values

1. $f(x)$
2. $X(\beta)$
3. $M(p)$

Multiple values

1. $\mathbf{X}\beta$
2. $\mathbf{M}\mathbf{p}$
3. *NOBS* Number of observations/simulated equivalents in the inverse model setup
4. *NPAR* Number of adjustable input parameters in the inverse model setup

Forecasts Model outputs for which field observations are not available. Typically these values are simulated under an uncertain future condition.

Phi Objective function, defined as the weighted sum of squares of residuals. Phi (aka Φ) is typically calculated as

$$\Phi = \sum_{i=1}^n \left(\frac{y_i - f(x_i)}{w_i} \right)^2 \quad \text{or} \quad \Phi = (\mathbf{y} - \mathbf{J}\mathbf{x})^T \mathbf{Q}^{-1} (\mathbf{y} - \mathbf{J}\mathbf{x}) \quad (1)$$

When regularization is included, an additional term is added, quantifying a penalty assessed for parameter sets that violate the preferred conditions regarding the parameter values. In such a case, the value of Φ as stated above is renamed Φ_m for “measurement PHI” and the additional regularization term is named Φ_r . A scalar (γ) parameter controls the tradeoff between these two dual components of the total objective function Φ_t

$$\Phi_t = \Phi_m + \gamma\Phi_r \quad (2)$$

Residuals The difference between observation values and modeled equivalents $r_i = y_i - f(x_i)$

Sensitivity The incremental change of an observation (modeled equivalent, actually) due to an incremental change in a parameter. Typically expressed as a finite-difference approximation of a partial derivative: $\frac{\partial y_i}{\partial x}$

Jacobian Matrix A matrix of the sensitivity of all observations in an inverse model to all parameters. This is often shown as a matrix by various names \mathbf{X} , \mathbf{J} , or \mathbf{H} . Each element of the matrix is a single sensitivity value $\frac{\partial y_i}{\partial x_j}$ for $i \in NOBS$, $j \in NPAR$

Regularization A preferred condition pertaining to parameters, the deviation from which, elicits a penalty added to the objective function. This serves as a balance between the level of fit or “measurement Phi” (Φ_M) and the coherence with soft knowledge/previous conditions/prior knowledge/regularization (Φ_R). These terms can also be interpreted as the likelihood function and prior distribution in Bayes’ theorem (see below)

PHIMLIM A PEST input parameter the governs the strength with which regularization is applied to the objective function. A high value of PHIMLIM indicates a strong penalty for deviation from preferred parameter conditions while a low value of PHIMLIM indicates a weak penalty. The reason this “dial” is listed as a function of PHIM (e.g. Φ_M) is because it can then be interpreted as a limit on how well we want to fit the observation data. PHIMLIM is actually controlling the value γ shown above formally trading off Φ_m and Φ_r .

Gaussian (multivariate) The equation for Gaussian (Normal) distribution for a single variable (x) is

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}} \quad (3)$$

where μ is the mean and σ is the standard deviation The equation for a multivariate Gaussian for a vector of k variables (\mathbf{x}) is

$$f(\mathbf{x}|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{-\frac{1}{2} ((\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu))} \quad (4)$$

where μ is a k -length vector of mean values, Σ is the covariance matrix, and $|\Sigma|$ is the determinant of the covariance matrix. These quantities are often abbreviated as $\mathcal{N}(\mu, \sigma)$ and $\mathcal{N}(\mu, \Sigma)$ for univariate and multivariate Gaussian distributions, respectively.

Weight or Epistemic Uncertainty A value by which a residual is divided by when constructing the sum of squared residuals. In principal, $w \approx \frac{1}{\sigma}$ where σ is an approximation of the expected error between model output and collocated observation values. While the symbol σ implies a standard deviation, it is important to note that measurement error only makes up a portion of this error. Other aspects such as structural error (e.g. inadequacy inherent in all models to perfectly simulate the natural world) also contribute to this expected level of error. The reciprocal of weights are also called Epistemic Uncertainty terms.

Weight Covariance matrix (correlation matrix) In practice, this is usually a $NOBS \times NOBS$ diagonal matrix with values of weights on the diagonal representing the inverse of the observation covariance. This implies a lack of correlation among the observations. A full covariance matrix would indicate correlation among the observations which, in reality, is present but, in practice, is rarely characterized. The weight matrix is often identified as \mathbf{Q}^{-1} or Σ_ϵ^{-1}

Parameter Covariance matrix The uncertainty of parameters can be expressed as a matrix as well. This is formed also as a diagonal matrix from the bounds around parameter values (assuming that the range between the bounds indicates 4σ (e.g. 95% of a normal distribution).) In `pyemu`, some functions accept a `sigma_range` argument which can override the 4σ assumption. In many cases of our applications, parameters are spatially distributed (e.g. hydraulic conductivity fields) so a covariance matrix with off-diagonal terms can be formed to characterize not only their variance but also their correlation/covariance. We often use geostatistical variograms to characterize the covariance of parameters. The parameter covariance matrix is often identified as $C(\mathbf{p})$, Σ_θ , or \mathbf{R} .

Measurement noise/error Measurement noise is a contribution to Epistemic Uncertainty. This is the expected error of repeated measurements due to things like instrument error and also can be compounded by error of surveying a datum, location of an observation on a map, and other factors.

Structural (model) error Epistemic uncertainty is actually dominated by structural error relative to measurement noise. The structural error is the expected misfit between measured and modeled values at observation locations due to model inadequacy (including everything from model simplification due to the necessity of discretizing the domain, processes that are missing from the model, etc.)

Monte Carlo Parameter Realization A set of parameter values, often but not required to be a multi-Gaussian distribution, sampled from the mean values of specified parameter values (either starting values or, in some cases, optimal values following parameter estimation) with covariance from a set of variance values, or a covariance matrix. Can be identified as θ

Monte Carlo Observation Realization A set of observation values, often but not required to be a multi-Gaussian distribution, sampled using the mean values of measured observations and variance from the observation weight covariance matrix. Can be identified as \mathbf{d}_{obs}

Monte Carlo Ensemble A group of realizations of parameters (Θ), observations (\mathbf{D}_{obs}) and the simulated equivalent values \mathbf{D}_{sim} . Note that these three matrices are made up of column vectors representing all of the θ , \mathbf{d}_{obs} , and \mathbf{d}_{sim} vectors where \mathbf{d}_{sim}

Bayes' Theorem
$$P(\theta|\mathbf{d}) = \frac{\mathcal{L}(\mathbf{d}|\theta)P(\theta)}{P(\mathbf{d})} \dots \underbrace{P(\theta|\mathbf{d})}_{\text{posterior pdf}} \propto \underbrace{\mathcal{L}(\mathbf{d}|\theta)}_{\text{likelihood function}} \underbrace{P(\theta)}_{\text{prior pdf}}$$

where θ is a vector of parameters, and \mathbf{d} is a vector of observations. It is computationally expedient to assume that these quantities can be characterized by multivariate Gaussian distributions and, thus, characterized only by their first two moments — mean and covariance.

Posterior (multivariate distribution) The posterior distribution is the updated distribution (mean and covariance) of parameter values θ updated from their prior by an experiment (encapsulated in the likelihood function). In other words, information gleaned from observations \mathbf{d} is used to update the initial values of the parameters.

Prior (multivariate distribution) This distribution represents what we know about parameter values prior to any modeling. It is also called “soft knowledge” or “expert knowledge”. This information is more than just starting values, but also encapsulates the understanding of uncertainty (characterized through the covariance) based on direct estimation of parameters (e.g. pumping tests, geologic maps, and grain size analysis, for example). In one interpretation of the objective function, this is also where the regularization information is contained.

Likelihood (multivariate distribution) This is a function describing how much is learned from the model. It is characterized by the misfit between modeled equivalents and observations.

FOSM (aka linear uncertainty analysis) “First Order Second Moment” is a technique to use an assumption of Gaussian distributions to, analytically, calculate the covariance of model outputs considering both the prior covariance and the likelihood function. In other words, it’s an analytical calculation of the posterior covariance of parameters using Bayes’ Theorem. The equation for this calculation is the Schur Complement. The key advantage of this is that we really only need a few quantities — a Jacobian Matrix \mathbf{J} , the prior covariance of parameters Σ_θ , and the observation covariance Σ_ϵ

Schur Complement The formula used to propagate uncertainty from prior through a “notional” calibration (via the Jacobian) to the posterior update.

$$\underbrace{\bar{\Sigma}_\theta}_{\text{what we know now}} = \underbrace{\Sigma_\theta}_{\text{what we knew}} - \underbrace{\Sigma_\theta \mathbf{J}^T [\mathbf{J} \Sigma_\theta \mathbf{J}^T + \Sigma_\epsilon]^{-1} \mathbf{J} \Sigma_\theta}_{\text{what we learned}} \quad (5)$$