Class Exercise 5b

2.3. Multivariate normal–normal case

Assume we are interested in an $n \times 1$ vector process $\mathbf{X}$ (e.g., u-wind components at several locations), that has prior distribution, $\mathbf{X} \sim N(\boldsymbol{\mu}, \mathbf{P})$, where for now we assume that the mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{P}$ are known. In addition, we observe the $p \times 1$ data vector $\mathbf{Y}$ and assume the following data model, $\mathbf{Y}|\mathbf{x} \sim N(\mathbf{Hx}, \mathbf{R})$, where the $p \times n$ observation matrix $\mathbf{H}$ and the observation error covariance matrix, $\mathbf{R}$, are assumed to be known.

The posterior distribution of $\mathbf{X}|\mathbf{y}$ is given by $p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$. As with the univariate case, the posterior distribution is also Gaussian:

$$\mathbf{X}|\mathbf{y} \sim N((\mathbf{H}'\mathbf{R}^{-1}\mathbf{H} + \mathbf{P}^{-1})^{-1}(\mathbf{H}'\mathbf{R}^{-1}\mathbf{y} + \mathbf{P}^{-1}\boldsymbol{\mu}), (\mathbf{H}'\mathbf{R}^{-1}\mathbf{H} + \mathbf{P}^{-1})^{-1}). \quad (5)$$

Applying some basic linear algebra, we can rewrite the posterior mean as

$$E(\mathbf{X}|\mathbf{y}) = \boldsymbol{\mu} + \mathbf{K}(\mathbf{y} - \mathbf{H}\boldsymbol{\mu}), \quad (6)$$

where $\mathbf{K} = \mathbf{PH}'(\mathbf{R} + \mathbf{HPH}')^{-1}$ is the "gain" matrix. Similarly, the posterior covariance matrix can be written as

$$\text{var}(\mathbf{X}|\mathbf{y}) = (\mathbf{I} - \mathbf{KH})\mathbf{P}. \quad (7)$$

Formulas (6) and (7) are the core of the so-called *analysis step* of DA based on linear and Gaussian assumptions. Typically, in a DA problem, $\boldsymbol{\mu}$ is the forecast from some deterministic model or long term averages based on previously collected data (e.g., climatology) and $\mathbf{P}$ is the forecast error covariance matrix. One updates the prior (forecast) mean $\boldsymbol{\mu}$ based on deviations from the observations according to the "gain" $\mathbf{K}$, which is a function of the prior and data error covariance matrices. Similarly, the prior (forecast) covariance matrix is updated according to the gain, although notably, this update is not a function of the observations, but only their location and covariance structure. This latter property relies heavily on the linear, Gaussian assumptions.

2.3.1. Relationship to kriging/optimal interpolation

We consider the relationship to Kriging (geostatistics)/Optimal Interpolation (meteorology, oceanography) by a simple example. Assume $\boxed{\mathbf{X} = [x(s_1), x(s_2), x(s_3)]'}$ at spatial locations s_i, $i = 1, 2, 3$. Also, assume we have observations at s_2 and s_3 but not s_1: $\boxed{\mathbf{y} = [y(s_2), y(s_3)]'}$ and thus $\mathbf{H}$ is defined as:

$$\mathbf{H} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Assume the prior covariance matrix that describes the (forecast) error covariance matrix is given by:

$$\boxed{\mathbf{P} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}}.$$

Note that even though we only have observations for locations 2 and 3, it is critical that we have the covariance information between all state locations of interest (e.g., 1, 2 and 3). In this case, the "gain" is given by:

$$\mathbf{K} = \mathbf{PH}'(\mathbf{R} + \mathbf{HPH}')^{-1} = \begin{pmatrix} c_{12} & c_{13} \\ c_{22} & c_{23} \\ c_{32} & c_{33} \end{pmatrix} \times \left(\mathbf{R} + \begin{pmatrix} c_{22} & c_{23} \\ c_{32} & c_{33} \end{pmatrix} \right)^{-1}.$$

For simplicity, assume $\boxed{\mathbf{R} = \sigma^2 \mathbf{I}}$ (i.e., independent measurement errors). Then, the posterior mean of $x(s_1)$ is given by:

$$E(x(s_1)|\mathbf{y}) = \mu(s_1) + w_{12}(y(s_2) - \mu(s_2)) + w_{13}(y(s_3) - \mu(s_3)),$$

where the interpolation weights, $\mathbf{w}_1 = [w_{12}, w_{13}]'$, are given by elements from the gain matrix:

$$\mathbf{w}_1' = \begin{pmatrix} c_{12} & c_{13} \end{pmatrix} \begin{pmatrix} c_{22} + \sigma^2 & c_{23} \\ c_{32} & c_{33} + \sigma^2 \end{pmatrix}^{-1}.$$

Thus, the prior mean is adjusted by a weighted combination of the anomalies (difference between observation and prior mean) at each data location.

The mean-squared prediction error (posterior variance) at $x(s_1)$ is given by

$$\mathrm{var}(x(s_1)|\mathbf{y}) = c_{11} - \begin{pmatrix} c_{12} & c_{13} \end{pmatrix} \begin{pmatrix} c_{22} + \sigma^2 & c_{23} \\ c_{32} & c_{33} + \sigma^2 \end{pmatrix}^{-1} \begin{pmatrix} c_{12} \\ c_{13} \end{pmatrix}.$$

Such spatial prediction (interpolation) is the optimal (best linear unbiased) prediction assuming the parameters, $\mathbf{R}$, $\mathbf{P}$ are known. In spatial statistics this is known as *simple kriging* [31,11] and in atmospheric/oceanographic science this is known as *optimal interpolation* [18]. See [11] for details.

Numerical example: Assume we have two observations $y_2 = 16$, $y_3 = 23$ and we are interested predicting the true process x_i at these locations and a third location x_1. Our prior mean is $\mu_i = 18$, $i = 1, 2, 3$ and our prior covariance matrix is:

$$\mathbf{P} = \begin{pmatrix} 1 & 0.61 & 0.22 \\ 0.61 & 1 & 0.37 \\ 0.22 & 0.37 & 1 \end{pmatrix}.$$

Our measurement error covariance matrix is $\mathbf{R} = 0.5\mathbf{I}$. In this case, our gain (interpolation weights) is (are):

$$\mathbf{K} = \begin{pmatrix} 0.3914 & 0.0528 \\ 0.6453 & 0.0870 \\ 0.0870 & 0.6453 \end{pmatrix}$$

and the posterior mean is:

$$E(\mathbf{X}|\mathbf{y}) = \begin{pmatrix} 17.4810 \\ 17.1442 \\ 21.0527 \end{pmatrix}$$

with posterior covariance:

$$\mathrm{var}(\mathbf{X}|\mathbf{y}) = \begin{pmatrix} 0.7508 & 0.1957 & 0.0264 \\ 0.1957 & 0.3227 & 0.0435 \\ 0.0264 & 0.0435 & 0.3227 \end{pmatrix}.$$

Class Exercise 5b (Let's do this)

1. Find the posterior mean and variance (for case when using Eq 5 and for a case when using Eq 6/7).

2. Do 1. for P=I.

see class_exercise_5b.m