Optimizing the ln likelihood function in George

author: Karen Y. Ng

Kernel used:

$$k(r) = \Sigma(r^2) + c\delta_{ij}$$

where $\Sigma(r^2)$ is ExpSquaredKernel or its derivatives. The second term is the WhiteKernel in George

Parametrization used in George

See george_examples/basic_properties_of_george.ipynb

$$k(r^2) = \lambda^{-1} \exp\left(-\frac{r^2}{2l^2}\right)$$

Parametrization for better visualization...

It is hard to visualize the likelihood surface in the original scale, so we perform a log transformation

$$a = \log_{10}(\lambda^{-1})$$

$$b = \log_{10}(l^2)$$

To have:

$$k(r^2) = 10^a \exp\left(-\frac{r^2}{2 \times 10^b}\right)$$

How new parametrization relates to ρ

$$b = \log_{10}(l^2) \tag{1}$$

$$=\log_{10}(1/\beta)\tag{2}$$

$$= -\log_{10}\left(-\frac{1}{4}\ln\rho\right) \tag{3}$$

Note on transformation of variables

Jacobian needed to preserve the area of integrated PDF

$$f_y(\vec{y}) = f_x(\vec{x}) |\det(J)|$$

Only when the transformed variable is the one that we integrate with respect to.

Let

$$\vec{x} = \begin{pmatrix} \lambda^{-1} \\ l^2 \end{pmatrix} = \begin{pmatrix} B^a \\ B^b \end{pmatrix} \tag{4}$$

where l^2 is the characteristic length.

$$f_{y}(\vec{y}) = f_{y}(a,b) = f_{x}(\lambda^{-1}, l^{2}) \left| \det \begin{pmatrix} \frac{\partial \lambda^{-1}}{\partial a} & \frac{\partial \lambda^{-1}}{\partial b} \\ \frac{\partial l^{2}}{\partial a} & \frac{\partial l^{2}}{\partial b} \end{pmatrix} \right|$$

$$= f_{x}(\lambda^{-1}, l^{2}) \left| \det \begin{pmatrix} \frac{\partial \exp(a \ln B)}{\partial a} & \frac{\partial \exp(a \ln B)}{\partial b} \\ \frac{\partial \exp(b \ln B)}{\partial a} & \frac{\partial \exp(b \ln B)}{\partial b} \end{pmatrix} \right|$$

$$= f_{x}(B^{a}, B^{b})(B^{a} + B^{b}) \ln B$$

$$(5)$$

$$= f_x(\lambda^{-1}, l^2) \left| \det \left(\begin{array}{c} \frac{\partial \exp(a \ln B)}{\partial a} & \frac{\partial \exp(a \ln B)}{\partial b} \\ \frac{\partial \exp(b \ln B)}{\partial a} & \frac{\partial \exp(b \ln B)}{\partial b} \end{array} \right) \right|$$
 (6)

$$= f_x(B^a, B^b)(B^a + B^b) \ln B \tag{7}$$

When we implement the log likelihood in the MCMC we need

$$\ln L_y(a,b) = \ln L_x(B^a, B^b) + \ln(B^a + B^b) + \ln(\ln B)$$
(8)

we can ignore the last term as it is a constant w.r.t. change in a and b.

Optimizing the GP Inlikelihood