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

$$\begin{aligned}
b &= \log_{10}(l^2) && (1)\\
&= \log_{10}(1/\beta) && (2)\\
&= -\log_{10}\left(-\frac{1}{4}\ln\rho\right) && (3)
\end{aligned}$$

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

$$\vec{x} = \left(\begin{array}{c} \lambda^{-1} \\ l^2 \end{array} \right) = \left(\begin{array}{c} B^a \\ B^b \end{array} \right) \tag{4}$$

where l^2 is the characteristic length.

$$\begin{aligned}
f_y(\vec{y}) = f_y(a,b) &= f_x(\lambda^{-1}, l^2) \left| \det \left(\begin{array}{cc} \frac{\partial \lambda^{-1}}{\partial a} & \frac{\partial \lambda^{-1}}{\partial b} \\ \frac{\partial l^2}{\partial a} & \frac{\partial l^2}{\partial b} \end{array}) \right) \right| && (5) \\
&= f_x(\lambda^{-1}, l^2) \left| \det \left(\begin{array}{cc} \frac{\partial e^{a \ln B}}{\partial a} & \frac{\partial e^{a \ln B}}{\partial b} \\ \frac{\partial e^{b \ln B}}{\partial a} & \frac{\partial e^{b \ln B}}{\partial b} \end{array}) \right) \right| && (6) \\
&= f_x(B^a, B^b)(B^a + B^b) \ln B && (7)
\end{aligned}$$

When we implement the log likelihood in the MCMC we need

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

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

Optimizing the GP lnlikelihood