Optimizing 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`

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

Transformation of variables

Jacobian needed to preserve the area of integrated PDF

$$P_y(\vec{y}) = P_x(\vec{x}) |\det(J)|$$

Transformation needed to evaluate likelihood with our parametrization

$$L_y(\lambda^{-1}, \beta) = L_x(\lambda^{-1}, 1/\beta) |\det(J)|$$

where the Jacobian is:

$$|\det(J)| = \left|\frac{\partial l^2}{\partial \beta}\right| = 1/\beta^2$$

Final ln likelihood expression

$$-\ln L_y(\lambda^{-1}, \beta) = -\ln L_x(\lambda^{-1}, 1/\beta) + 2\ln\beta$$

First term on the RHS is evaluated by George,
second term is what we need to add.

Reparametrizating the kernel by log transformation

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

As recommended by several papers / books, let

$$\lambda^{-1} = \exp a$$

$$l^2 = \exp b$$

Log transformation of variables

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

Log marginal likelihood with new parametrization

$$lnL(a, b)$$

$$= lnL(\lambda^{-1}, l^2) + ln(\lambda^{-1} l^2)$$

$$= lnL(\exp(a), \exp(b)) + ln(\exp(a)\exp(b))$$

$$= lnL(\exp(a), \exp(b)) + a + b$$

Optimizing the GP lnlikelihood