DERIVATIVES OF THE MVN LOG-LIKELIHOOD WITH RESPECT TO θ PARAMETERS

GAL AV-GAY

1. Background

The **DACE** (cite) stochastic process model uses a correlation function with unknown values for its parameters for the Multivariate Gaussian. Using maximum likelihood estimates for the correlation parameters in the correlation function, the **DACE** model has been shown to provide a robust and flexible fit for high-dimensional data (cite). This does not seem to be the case with Bayesian estimates for the correlation parameters, which result in a significant increase in error for cross-validated prediction of a response variable. In an attempt to survey the Bayesian estimate and the reasons for its misgivings, we obtain the first and second derivatives of the log-likelihood with respect to the k correlation parameters $\theta_j, j \in \{1, ..., k\}$. The Hessian matrix (Second derivative) will allow us to obtain the information, which will give us an approximation of the variance of these parameters.

2. Multivariate Gaussian

$$\frac{1}{(2\pi)^{n/2}(\sigma^2)\,|\mathbf{R}|^{1/2}} \exp\left[-\frac{(y - \mathbf{B}F)'\mathbf{R}^{-1}(y - \mathbf{B}F)}{2\sigma^2}\right]$$

$\mathbf{R}$ in the above distribution is the correlation matrix, for which the entries are defined, for example by the power-exponential correlation, for which p_j is known, and with $p_j = 2$ it becomes the gaussian correlation:

$$\mathbf{R}(x, x') = \prod_{j-1}^{k} \exp[-\theta_j |x_j - x'_j|^{p_j}]$$

There are n data points x, where each x is a k-dimensional data point, and there exists a θ_j for each dimension of the data. Therefore, every θ_j coefficient is used to determine every entry of the correlation matrix $\mathbf{R}$. The derivative of the correlation matrix at a particular point can be determined with respect to a particular θ_j as such:

$$\frac{d\mathbf{R}(x, x')}{d\theta_j} = -|x_j - x'_j|^{p_j}\mathbf{R}(x, x')$$

3. Maximum Likelihood Estimates

The maximum likelihood estimates for the regression parameters $\mathbf{B}$ and the variance σ^2 are as follows:

$$\widehat{\mathbf{B}} = (F^T R^{-1} F)^{-1} F^T R^{-1} y \tag{1}$$

$$\widehat{\sigma^2} = \frac{(y - \widehat{\mathbf{B}}F)' \mathbf{R}^{-1} (y - \widehat{\mathbf{B}}F)}{n} \tag{2}$$

4. Log Likelihood

The log likelihood, when evaluated with the maximum likelihood estimates for the regression parameters and the variance, is the following:

$$-\frac{n}{2} \ln(\widehat{\sigma^2}) - \frac{1}{2} \ln |\mathbf{R}| + const \tag{3}$$

To take the derivative of this log-likelihood with respect to the θ_j parameters, we must consider how $\widehat{\sigma^2}$ is a function of the correlation matrix $\mathbf{R}$. This results in the following expression for $\widehat{\sigma^2}$:

$$\widehat{\sigma^2} = \frac{y^T \mathbf{R}^{-1} y - y^T \mathbf{R}^{-1} F (F^T \mathbf{R}^{-1} F)^{-1} F^T \mathbf{R}^{-1} y}{n} \tag{4}$$

5. First Derivative of Log Likelihood with respect to θ_j parameters

$$\frac{d}{d\theta_j} \left[-\frac{n}{2} \ln(\widehat{\sigma^2}) - \frac{1}{2} \ln |\mathbf{R}| \right] = -\frac{n}{2\widehat{\sigma^2}} \frac{\partial \widehat{\sigma^2}}{\partial \theta_j} - \frac{1}{2} Trace[\mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_j}] \tag{5}$$

The expression in (5) was evaluated using the fact that $\frac{d}{dx} \ln |A| = Tr[A^{-1} \frac{dA}{dx}]$. The problem lies in evaluating $\frac{\partial \widehat{\sigma^2}}{\partial \theta_j}$: Using product-rule and chain-rule and considering that, $\frac{\partial R^{-1}}{\partial X} = -R^{-1} \frac{\partial R}{\partial X} R^{-1}$, we show how $n \frac{\partial \widehat{\sigma^2}}{\partial \theta_j}$ can be expressed below.

$$-y^T \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_j} \mathbf{R}^{-1} y$$

$$+2y^T \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_j} \mathbf{R}^{-1} F (F^T \mathbf{R}^{-1} F)^{-1} F^T \mathbf{R}^{-1} y$$

$$-y^T \mathbf{R}^{-1} F (F^T \mathbf{R}^{-1} F)^{-1} F^T \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_j} \mathbf{R}^{-1} F (F^T \mathbf{R}^{-1} F)^{-1} F^T \mathbf{R}^{-1} y$$

where R^{-T} is equal to R^{-1} since $\mathbf{R}$ is a symmetric Matrix.

We can then substitute in $\widehat{\mathbf{B}} = (F^T R^{-1} F)^{-1} F^T R^{-1} y$, so that our expression becomes the following:

$$-y^T \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_j} \mathbf{R}^{-1} y$$
$$+2y^T \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_j} \mathbf{R}^{-1} F \widehat{\mathbf{B}}$$
$$-\widehat{\mathbf{B}}^T F^T \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_j} \mathbf{R}^{-1} F \widehat{\mathbf{B}}$$

The above sum is rearranged as the following:

$$n \frac{\partial \widehat{\sigma^2}}{\partial \theta_j} = -(y - F\widehat{\mathbf{B}})^T \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_j} \mathbf{R}^{-1} (y - F\widehat{\mathbf{B}}) \tag{6}$$

6. Computing the First Derivative

We decompose the correlation matrix R into an upper triangular matrix multiplied by its transpose, and this allows us to solve for a new vector, $\tilde{y}$, and a new matrix $\tilde{F}$.

$$\mathbf{R} = U^T U \therefore \mathbf{R}^{-1} = U^{-1} U^{-T} \tag{7}$$

$$U^{-T} y = \tilde{y} \therefore U^T \tilde{y} = y \tag{8}$$

$$U^{-T} F = \tilde{F} \therefore U^T \tilde{F} = F \tag{9}$$

The new $\widehat{\mathbf{B}}$ expression is as follows:

$$\widehat{\mathbf{B}} = (\tilde{F}^T \tilde{F})^{-1} \tilde{F^T} \tilde{y}$$

We then apply a QR decomposition to $\tilde{F}$. We will use the notation 'T' instead of 'R' to avoid ambiguity. The QR decomposition we use will result in a square T and rectangular Q, where $Q^T Q = I$, but $Q Q^T \neq I$.

$$\tilde{F} = Q_1 T_1 \tag{10}$$

This greatly simplifies our $\widehat{\mathbf{B}}$ expression due to the T's canceling out:

$$\widehat{\mathbf{B}} = (T_1^T Q_1^T Q_1 T_1)^{-1} T_1^T Q_1^T \tilde{y} = T_1^{-1} Q_1^T \tilde{y} \tag{11}$$

When we take the Cholesky decomposition of the $\mathbf{R}^{-1}$ matrices and the QR decomposition of the $\tilde{F}$ matrices in the derivative expression in (6), we obtain the following:

$$-(\tilde{y} - Q_1 T_1 \widehat{\mathbf{B}})^T \mathbf{U}^{-T} \frac{\partial \mathbf{R}}{\partial \theta_j} \mathbf{U}^{-1} (\tilde{y} - Q_1 T_1 \widehat{\mathbf{B}}) \tag{12}$$

Plugging in our expression for $\widehat{\mathbf{B}}$ (11) to our derivative expression (12) results in the following simplification:

$$-(\tilde{y} - Q_1 Q_1^T \tilde{y})^T \mathbf{U}^{-T} \frac{\partial \mathbf{R}}{\partial \theta_j} \mathbf{U}^{-1} (\tilde{y} - Q_1 Q_1^T \tilde{y}) \tag{13}$$

The next step is to back-solve for the $\tilde{U} = \mathbf{U}^{-1}(\tilde{y} - Q_1 Q_1^T \tilde{y})$ vector.

$$\tilde{U}^T \frac{\partial \mathbf{R}}{\partial \theta_j} \tilde{U} \tag{14}$$

Finally we can compute our full derivative,

$$\frac{d}{d\theta_j} \left[-\frac{n}{2} \ln(\widehat{\sigma^2}) - \frac{1}{2} \ln |\mathbf{R}| \right] = -\frac{1}{2\widehat{\sigma^2}} \tilde{U}^T \frac{\partial \mathbf{R}}{\partial \theta_j} \tilde{U} - \frac{1}{2} Trace[\mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_j}] \tag{15}$$

The results of this first derivative calculation correspond to finite difference approximations of the derivative, as shown in the figures below:

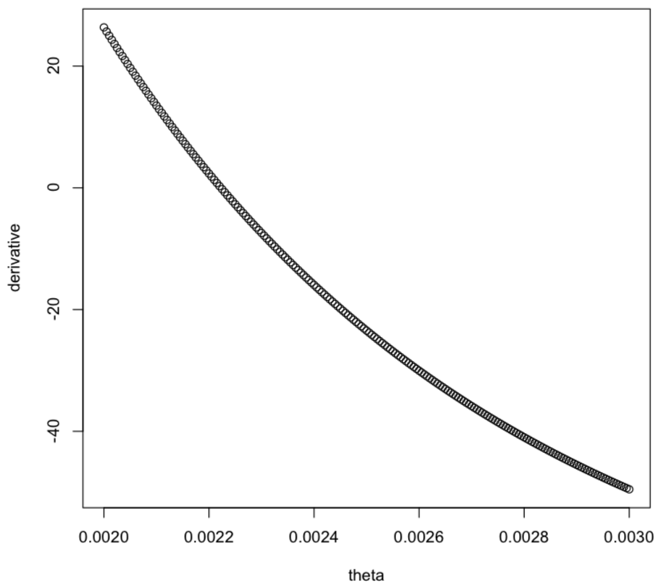


FIGURE 1. Derivative at various theta

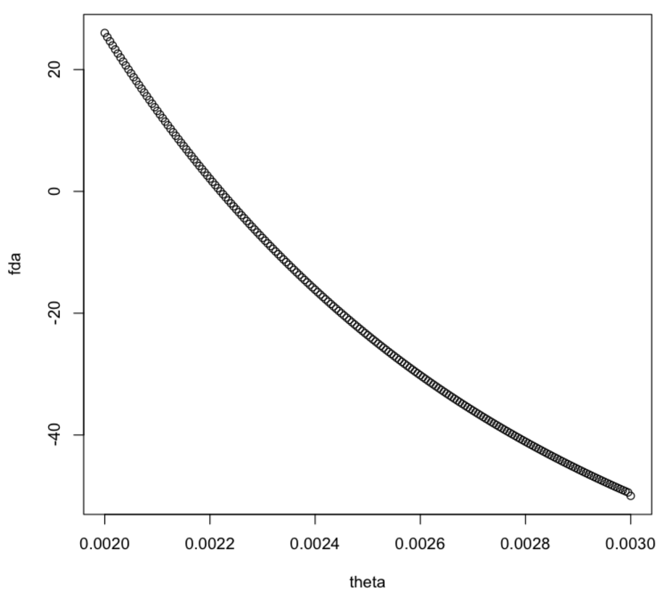


FIGURE 2. Finite Difference approximation of derivative at various theta

7. Second Derivative of the Log Likelihood

To obtain the Hessian of the log-likelihood with respect to the k θ parameters, we differentiate with respect to all possible pairs of θ parameters. The Hessian matrix $H(k,k)$ is a symmetrical matrix that contains at each point $H(i,j)$ the resulting second derivative with respect to each pair θ_j and θ_i, where $i,j \in \{0,...,k\}$. From the first derivative of the log-likelihood (5) we derive the following expression for our second derivative with respect to θ_j and θ_i:

$$\frac{\partial^2 LogLikelihood}{\partial\theta_j\partial\theta_i} = \tag{16}$$

$$\left(\frac{n}{2(\widehat{\sigma}^2)^2}\cdot\frac{\partial\widehat{\sigma}^2}{\partial\theta_i}\cdot\frac{\partial\widehat{\sigma}^2}{\partial\theta_j}\right)-\left(\frac{n}{2\widehat{\sigma}^2}\cdot\frac{\partial^2\widehat{\sigma}^2}{\partial\theta_j\partial\theta_i}\right)-\frac{1}{2}Trace\left[\mathbf{R}^{-1}\frac{\partial^2\mathbf{R}}{\partial\theta_j\partial\theta_i}-\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\theta_i}\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\theta_j}\right]$$

The first derivative terms and the $\widehat{\sigma}^2$ term are given (6),(4). The terms inside the Trace function can be solved for using forward and backwards substitution, this will be done in the next section. This leaves one last term, the second derivative of $\widehat{\sigma}^2$ with respect to θ_j and θ_i. Using the product rule five times on (6) we express $n\frac{\partial^2\widehat{\sigma}^2}{\partial\theta_j\partial\theta_i}$ below.

$$n\frac{\partial^2\widehat{\sigma}^2}{\partial\theta_j\partial\theta_i} = \tag{17}$$

$$-2\left[\frac{\partial}{\partial\theta_i}(y-F\widehat{B})^T\right]\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\theta_j}\mathbf{R}^{-1}(y-F\widehat{B})$$
$$+2(y-F\widehat{B})^T\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\theta_i}\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\theta_j}\mathbf{R}^{-1}(y-F\widehat{B})$$
$$-(y-F\widehat{B})^T\mathbf{R}^{-1}\frac{\partial^2\mathbf{R}}{\partial\theta_j\partial\theta_i}\mathbf{R}^{-1}(y-F\widehat{B})$$

Where the vector $\frac{\partial}{\partial\theta_i}(y-F\widehat{B})$ is expressed below, using the definition of $\widehat{B}$ in (1).

$$-F(F^T\mathbf{R}^{-1}F)^{-1}F^T\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\theta_i}\mathbf{R}^{-1}F(F^T\mathbf{R}^{-1}F)^{-1}F^T\mathbf{R}^{-1}y \tag{18}$$
$$+F(F^T\mathbf{R}^{-1}F)^{-1}F^T\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\theta_i}\mathbf{R}^{-1}y$$

Since

$$\widehat{\mathbf{B}} = (F^TR^{-1}F)^{-1}F^TR^{-1}y$$

,

We can factor the derivative (18) into the following form:

$$(y - F\widehat{B})^T \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_i} \mathbf{R}^{-1} F (F^T \mathbf{R}^{-1} F)^{-1} F^T$$

therefore,

$$n \frac{\partial^2 \widehat{\sigma}^2}{\partial \theta_j \partial \theta_i} = \tag{19}$$

$$-2(y - F\widehat{B})^T \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_i} \mathbf{R}^{-1} F (F^T \mathbf{R}^{-1} F)^{-1} F^T \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_j} \mathbf{R}^{-1} (y - F\widehat{B})$$
$$+2(y - F\widehat{B})^T \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_i} \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_j} \mathbf{R}^{-1} (y - F\widehat{B})$$
$$-(y - F\widehat{B})^T \mathbf{R}^{-1} \frac{\partial^2 \mathbf{R}}{\partial \theta_j \partial \theta_i} \mathbf{R}^{-1} (y - F\widehat{B})$$

We find that we can further factor this overall expression into the following:

$$n \frac{\partial^2 \widehat{\sigma}^2}{\partial \theta_j \partial \theta_i} = \tag{20}$$

$$(y - F\widehat{B})^T \mathbf{R}^{-1} \left[2 \frac{\partial \mathbf{R}}{\partial \theta_i} \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_j} - 2 \frac{\partial \mathbf{R}}{\partial \theta_i} \mathbf{R}^{-1} F (F^T \mathbf{R}^{-1} F)^{-1} F^T \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_j} - \frac{\partial^2 \mathbf{R}}{\partial \theta_j \partial \theta_i} \right] \mathbf{R}^{-1} (y - F\widehat{B})$$

$$= (y - F\widehat{B})^T \mathbf{R}^{-1} \left[2 \frac{\partial \mathbf{R}}{\partial \theta_i} \mathbf{R}^{-1} (\mathbf{I} - F (F^T \mathbf{R}^{-1} F)^{-1} F^T \mathbf{R}^{-1}) \frac{\partial \mathbf{R}}{\partial \theta_j} - \frac{\partial^2 \mathbf{R}}{\partial \theta_j \partial \theta_i} \right] \mathbf{R}^{-1} (y - F\widehat{B})$$

It is evident that the above expressions are symmetrical in terms of interchanging θ_i with θ_j, therefore satisfying this property of a second derivative.

8. Computing the Second Derivative

The vector $(y - F\widehat{B})$ above can be computed using the expression for $\widehat{B}$ in (11):

$$(y - FT^{-1} Q^T \tilde{y}) \tag{21}$$

To compute the second derivative of the variance, we decompose the inverse of the correlation matrix using a Cholesky decomposition,

$$\mathbf{R}^{-1} = U^{-1} U^{-T}$$

then forward-solve

$$\tilde{F} = U^{-T} F$$

$\tilde{F}$ undergoes a QR-decomposition

$$\tilde{F} = Q_(n, m) T_(m, m)$$

we now rewrite our expression for the second derivative of the variance with respect to θ_j and θ_i as such:

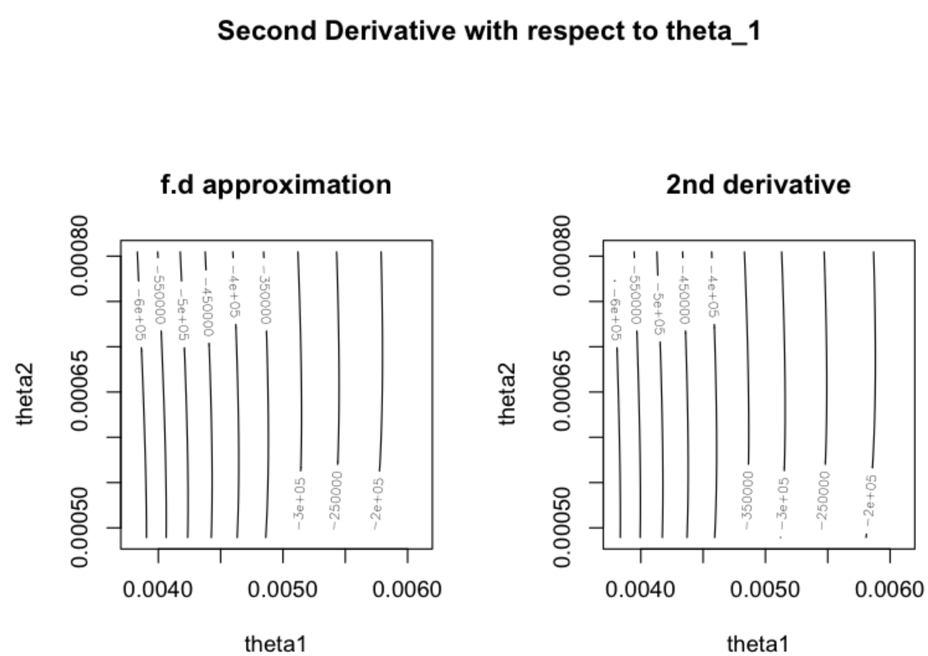
$$n\frac{\partial^2\widehat{\sigma}^2}{\partial\theta_j\partial\theta_i} = \tag{22}$$

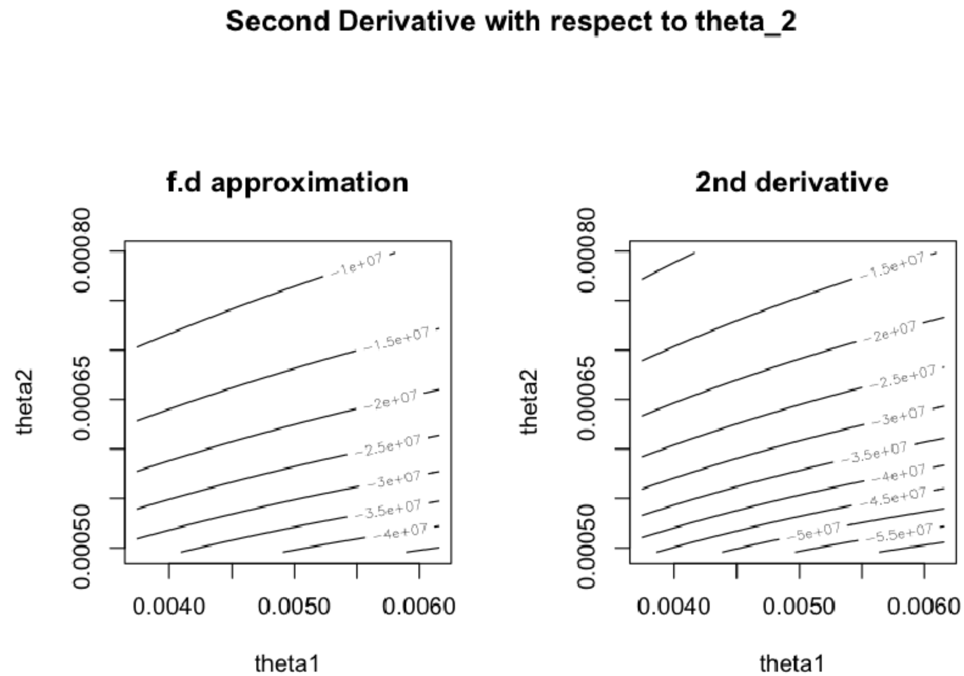
$$(y - FT^{-1}Q^T\tilde{y})^T U^{-1}U^{-T}\left[2\frac{\partial\mathbf{R}}{\partial\theta_i}\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\theta_j} - 2\frac{\partial\mathbf{R}}{\partial\theta_i}U^{-1}QQ^TU^{-T}\frac{\partial\mathbf{R}}{\partial\theta_j} - \frac{\partial^2\mathbf{R}}{\partial\theta_j\partial\theta_i}\right]U^{-1}U^{-T}(y - FT^{-1}Q^T\tilde{y})$$

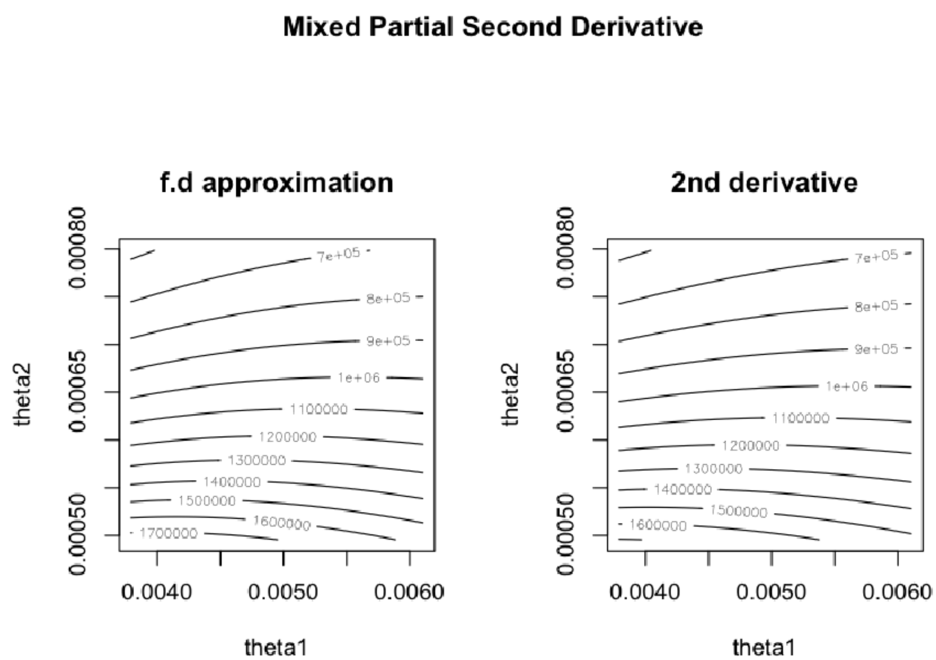
$$(y - FT^{-1}Q^T\tilde{y})^T U^{-1}U^{-T}\left[2\frac{\partial\mathbf{R}}{\partial\theta_i}U^{-1}\left[\mathbf{I} - QQ^T\right]U^{-T}\frac{\partial\mathbf{R}}{\partial\theta_j} - \frac{\partial^2\mathbf{R}}{\partial\theta_j\partial\theta_i}\right]U^{-1}U^{-T}(y - FT^{-1}Q^T\tilde{y})$$

The latter of the above expressions contains terms already evaluated for the first derivative, with the exception of $\frac{\partial^2\mathbf{R}}{\partial\theta_j\partial\theta_i}$.

Below we show how contour plots of the second derivative at various theta for a two-dimensional system correspond to finite-difference-approximations of the same second derivatives.




Second Derivative with respect to theta_2
f.d approximation
2nd derivative
theta2
theta1
0.00050
0.00065
0.00080
0.0040
0.0050
0.0060
-1e+07
-1.5e+07
-2e+07
-2.5e+07
-3e+07
-3.5e+07
-4e+07
-4.5e+07
-5e+07
-5.5e+07


Mixed Partial Second Derivative
f.d approximation
2nd derivative
theta2
theta1
0.00050
0.00065
0.00080
0.0040
0.0050
0.0060
7e+05
8e+05
9e+05
1e+06
1100000
1200000
1300000
1400000
1500000
1600000
1700000

9. Bayesian Case

When a bayesian posterior is determined for the theta parameters, our derivative calculation changes. It can be shown that the posterior distribution is a t-distribution with n-k degrees of freedom (where k is the number of regression parameters). Consequently the variance becomes the following:

$$\widehat{\sigma^2} = \frac{y^T\mathbf{R}^{-1}y - y^T\mathbf{R}^{-1}F(F^T\mathbf{R}^{-1}F)^{-1}F^T\mathbf{R}^{-1}y}{n-k} \tag{23}$$

The logarithm of the posterior for θ is then differentiated.

$$p(\theta|y) \propto \frac{\pi(\theta)}{(\widehat{\sigma}^2)^{(n-k)/2}det^{1/2}(\mathbf{R})det^{1/2}(F^T\mathbf{R}^{-1}F)} \tag{24}$$

$$ln(p(\theta|y)) = ln(\pi(\theta)) - \frac{n-k}{2}ln(\widehat{\sigma}^2) - \frac{1}{2}\ln|\mathbf{R}| - \frac{1}{2}ln\left|F^T\mathbf{R}^{-1}F\right| \tag{25}$$

The derivative with respect to θ is therefore:

$$\frac{\partial ln(p(\theta|y))}{\partial\theta} =$$

$$\frac{1}{\pi(\theta)}\frac{\partial\pi(\theta)}{\partial\theta} - \frac{n-k}{2\widehat{\sigma^2}}\tilde{U}^T\frac{\partial\mathbf{R}}{\partial\theta}\tilde{U} - \frac{1}{2}Trace[\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\theta}] - \frac{1}{2}Trace[(F^T\mathbf{R}^{-1}F)^{-1}F^T\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\theta}\mathbf{R}^{-1}F]$$

This solution depends on the choice of prior, $\pi(\theta)$.

10. Fast Bayesian Inference Case

The Fast Bayesian Inference Algorithm requires a Laplace approximation of the likelihood where θ is transformed logarithmically such that $\tau = ln(\theta)$. The formula for the variance remains unchanged, with $\mathbf{R}_\theta$ now being $\mathbf{R}_\tau$. The derivative is now taken with respect to τ rather than θ.

$$\widehat{\sigma^2}_\tau = \frac{y^T\mathbf{R}_\tau^{-1}y - y^T\mathbf{R}_\tau^{-1}F(F^T\mathbf{R}_\tau^{-1}F)^{-1}F^T\mathbf{R}_\tau^{-1}y}{n-k} \tag{26}$$

The posterior distribution is logarithmically transformed and then differentiated with respect to τ.

$$p(\tau|y) \propto \frac{\pi(\tau)}{(\widehat{\sigma}^2)^{(n-k)/2}det^{1/2}(\mathbf{R}_\tau)det^{1/2}(F^T\mathbf{R}_\tau^{-1}F)} \tag{27}$$

$$ln(p(\tau|y)) = ln(\pi(\tau)) - \frac{n-k}{2}ln(\widehat{\sigma}^2) - \frac{1}{2}\ln|\mathbf{R}_\tau| - \frac{1}{2}ln\left|F^T\mathbf{R}_\tau^{-1}F\right| \tag{28}$$

The algorithm uses a Jeffrey's prior of $\pi(\theta) \propto \prod_{j=1}^{D} 1/\theta_j$, resulting in $\pi(\tau) \propto 1$. The derivative is therefore:

$$\frac{\partial ln(p(\tau|y))}{\partial \tau} = \tag{29}$$

$$-\frac{n-k}{2\widehat{\sigma^2}}\tilde{U}_\tau^T \frac{\partial \mathbf{R}_\tau}{\partial \tau}\tilde{U}_\tau - \frac{1}{2}Trace[\mathbf{R}_\tau^{-1}\frac{\partial \mathbf{R}_\tau}{\partial \tau}] + \frac{1}{2}Trace[(F^T\mathbf{R}_\tau^{-1}F)^{-1}F^T\mathbf{R}_\tau^{-1}\frac{\partial \mathbf{R}}{\partial \tau}\mathbf{R}_\tau^{-1}F]$$

To compute the $(F^T\mathbf{R}_\tau^{-1}F)^{-1}F^T\mathbf{R}_\tau^{-1}\frac{\partial \mathbf{R}}{\partial \tau}\mathbf{R}_\tau^{-1}F$ term, we must decompose $\mathbf{R}_\tau^{-1}$ using the Cholesky decomposition (7) and solve for $\tilde{F}$ (9), and then proceed to take the QR decomposition as in (10).

$$\begin{aligned}
&(F^T\mathbf{R}_\tau^{-1}F)^{-1}F^T\mathbf{R}_\tau^{-1}\frac{\partial \mathbf{R}}{\partial \tau}\mathbf{R}_\tau^{-1}F \\
&= (\tilde{F}^T\tilde{F})^{-1}\tilde{F}^T U^{-T}\frac{\partial \mathbf{R}}{\partial \tau}U^{-1}\tilde{F} \\
&= T^{-1}Q^T U^{-T}\frac{\partial \mathbf{R}}{\partial \tau}U^{-1}\tilde{F}
\end{aligned}$$

The above expression can be solved for using backwards and forwards substitution due to that T and U are both upper triangular matrices.

Finally, we note that the derivative of our correlation matrix $\mathbf{R}$ with respect to $\tau = \ln(\theta)$ is the following:

$$\frac{d\mathbf{R}(x,x^{'})}{d\tau_i} = -e^{\tau_i}|x_i - x_i'|^{p_i}\mathbf{R}(x,x') = -\theta_i|x_i - x_i'|^{p_i}\mathbf{R}(x,x') \tag{30}$$

$$\frac{d\mathbf{R}^2(x,x^{'})}{d\tau_i d\tau_j} = e^{\tau_j}|x_j - x_j'|^{p_j}e^{\tau_i}|x_i - x_i'|^{p_i}\mathbf{R}(x,x') = \theta_j|x_j - x_j'|^{p_j}\theta_i|x_i - x_i'|^{p_i}\mathbf{R}(x,x')$$

$$\begin{aligned}
\frac{d\mathbf{R}^2(x,x^{'})}{d\tau_i^2} &= -e^{\tau_i}|x_i - x_i'|^{p_i}\mathbf{R}(x,x') + e^{2\tau_i}|x_i - x_i'|^{2p_i}\mathbf{R}(x,x') \\
&= e^{\tau_i}|x_i - x_i'|^{p_i}\mathbf{R}(x,x')\left[e^{\tau_i}|x_i - x_i'|^{p_i} - 1\right] \\
&= \theta_i|x_i - x_i'|^{p_i}\mathbf{R}(x,x')\left[\theta_i|x_i - x_i'|^{p_i} - 1\right]
\end{aligned}$$

11. Second Derivative in the FBI case

The second derivative in the FBI case is similar to the expression for the Maximum Likelihood second derivative of the log-likelihood, with the exception of there being an additional term as well as different degrees of freedom.

$$\frac{\partial^2 ln(p(\tau|y))}{\partial\tau_i\partial\tau_j} = \tag{31}$$

$$(\frac{n-k}{2(\widehat{\sigma}^2)^2}\cdot\frac{\partial\widehat{\sigma}^2}{\partial\tau_i}\cdot\frac{\partial\widehat{\sigma}^2}{\partial\tau_j})-(\frac{n-k}{2\widehat{\sigma}^2}\cdot\frac{\partial^2\widehat{\sigma}^2}{\partial\tau_j\partial\tau_i})-\frac{1}{2}Trace[\mathbf{R}_\tau^{-1}\frac{\partial^2\mathbf{R}_\tau}{\partial\tau_j\partial\tau_i}-\mathbf{R}_\tau^{-1}\frac{\partial\mathbf{R}_\tau}{\partial\tau_i}\mathbf{R}_\tau^{-1}\frac{\partial\mathbf{R}_\tau}{\partial\tau_j}]$$
$$+\frac{\partial}{\partial\tau_j}\left[\frac{1}{2}Trace[(F^T\mathbf{R}_\tau^{-1}F)^{-1}F^T\mathbf{R}_\tau^{-1}\frac{\partial\mathbf{R}_\tau}{\partial\tau_i}\mathbf{R}_\tau^{-1}F]\right]$$

Since the trace function is a linear operator, this additional term can be expressed as such,

$$\frac{\partial}{\partial\tau_j}\left[\frac{1}{2}Trace[(F^T\mathbf{R}_\tau^{-1}F)^{-1}F^T\mathbf{R}_\tau^{-1}\frac{\partial\mathbf{R}_\tau}{\partial\tau_i}\mathbf{R}_\tau^{-1}F]\right] =$$
$$\frac{1}{2}Trace[\frac{\partial}{\partial\tau_j}\left[(F^T\mathbf{R}_\tau^{-1}F)^{-1}F^T\mathbf{R}_\tau^{-1}\frac{\partial\mathbf{R}_\tau}{\partial\tau_i}\mathbf{R}_\tau^{-1}F\right]]$$

Where

$$\frac{\partial}{\partial\tau_j}\left[(F^T\mathbf{R}_\tau^{-1}F)^{-1}F^T\mathbf{R}_\tau^{-1}\frac{\partial\mathbf{R}_\tau}{\partial\tau_i}\mathbf{R}_\tau^{-1}F\right] =$$

$$(F^T\mathbf{R}_\tau^{-1}F)^{-1}F^T\mathbf{R}_\tau^{-1}\frac{\partial\mathbf{R}_\tau}{\partial\tau_j}\mathbf{R}_\tau^{-1}F(F^T\mathbf{R}_\tau^{-1}F)^{-1}F^T\mathbf{R}_\tau^{-1}\frac{\partial\mathbf{R}_\tau}{\partial\tau_i}\mathbf{R}_\tau^{-1}F$$
$$-(F^T\mathbf{R}_\tau^{-1}F)^{-1}F^T\mathbf{R}_\tau^{-1}\frac{\partial\mathbf{R}_\tau}{\partial\tau_j}\mathbf{R}_\tau^{-1}\frac{\partial\mathbf{R}_\tau}{\partial\tau_i}\mathbf{R}_\tau^{-1}F$$
$$+(F^T\mathbf{R}_\tau^{-1}F)^{-1}F^T\mathbf{R}_\tau^{-1}\frac{\partial\mathbf{R}_\tau}{\partial\tau_i}\mathbf{R}_\tau^{-1}\frac{\partial\mathbf{R}_\tau}{\partial\tau_j}\mathbf{R}_\tau^{-1}F$$
$$+(F^T\mathbf{R}_\tau^{-1}F)^{-1}F^T\mathbf{R}_\tau^{-1}\frac{\partial^2\mathbf{R}_\tau}{\partial\tau_i\partial\tau_j}\mathbf{R}_\tau^{-1}F$$
$$=$$
$$(F^T\mathbf{R}_\tau^{-1}F)^{-1}F^T\mathbf{R}_\tau^{-1}\left[\frac{\partial\mathbf{R}_\tau}{\partial\tau_j}\mathbf{R}_\tau^{-1}F(F^T\mathbf{R}_\tau^{-1}F)^{-1}F^T\mathbf{R}_\tau^{-1}\frac{\partial\mathbf{R}_\tau}{\partial\tau_i}+\frac{\partial\mathbf{R}_\tau}{\partial\tau_i}\mathbf{R}_\tau^{-1}\frac{\partial\mathbf{R}_\tau}{\partial\tau_j}-\frac{\partial\mathbf{R}_\tau}{\partial\tau_j}\mathbf{R}_\tau^{-1}\frac{\partial\mathbf{R}_\tau}{\partial\tau_i}+\frac{\partial^2\mathbf{R}_\tau}{\partial\tau_i\partial\tau_j}\right]\mathbf{R}_\tau^{-1}F$$

It is useful to note that

$$(\frac{\partial\mathbf{R}_\tau}{\partial\tau_i}\mathbf{R}_\tau^{-1}\frac{\partial\mathbf{R}_\tau}{\partial\tau_j})^T = \frac{\partial\mathbf{R}_\tau}{\partial\tau_j}\mathbf{R}_\tau^{-1}\frac{\partial\mathbf{R}_\tau}{\partial\tau_i}$$

12. Computing the Second Derivative in the FBI case

Recall the expression for the second derivative of the posterior in (31). The evaluation of the first three terms in the expression is almost identical to that of the maximum likelihood case, with the exception that we now divide by $n-k$ degrees of freedom. The derivative of the correlation matrix $\mathbf{R}$ is indicated in (30). Finally, the fourth term is decomposed below so as to avoid taking the inverse directly.

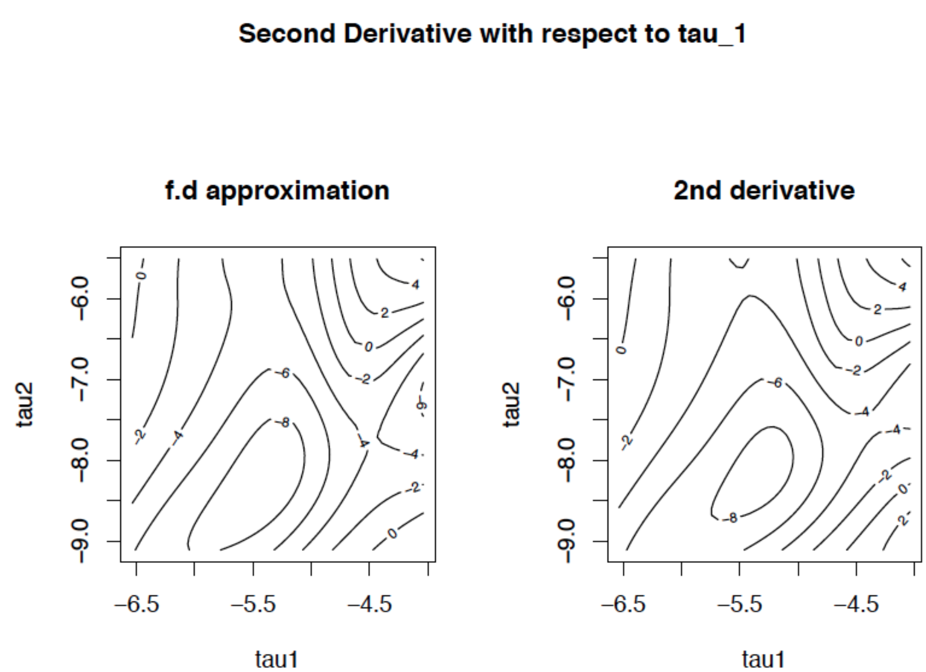
(32)

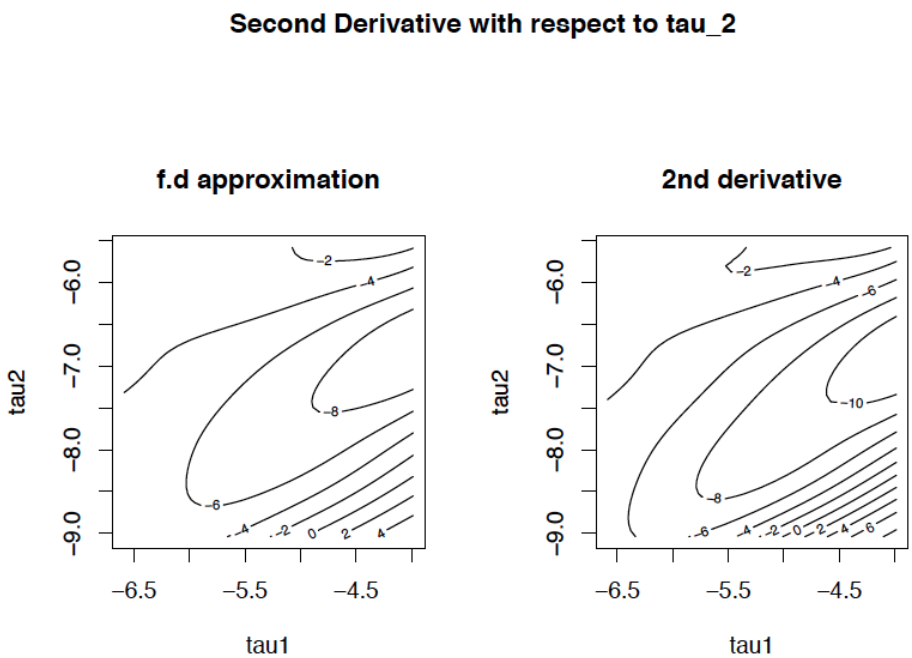
$$T^{-1}Q^TU^{-T}\left[\frac{\partial \mathbf{R}_\tau}{\partial \tau_j}U^{-1}QQ^TU^{-T}\frac{\partial \mathbf{R}_\tau}{\partial \tau_i}+\frac{\partial \mathbf{R}_\tau}{\partial \tau_i}\mathbf{R}_\tau^{-1}\frac{\partial \mathbf{R}_\tau}{\partial \tau_j}-\frac{\partial \mathbf{R}_\tau}{\partial \tau_j}\mathbf{R}_\tau^{-1}\frac{\partial \mathbf{R}_\tau}{\partial \tau_i}+\frac{\partial^2 \mathbf{R}_\tau}{\partial \tau_i \partial \tau_j}\right]U^{-1}\tilde{F}$$

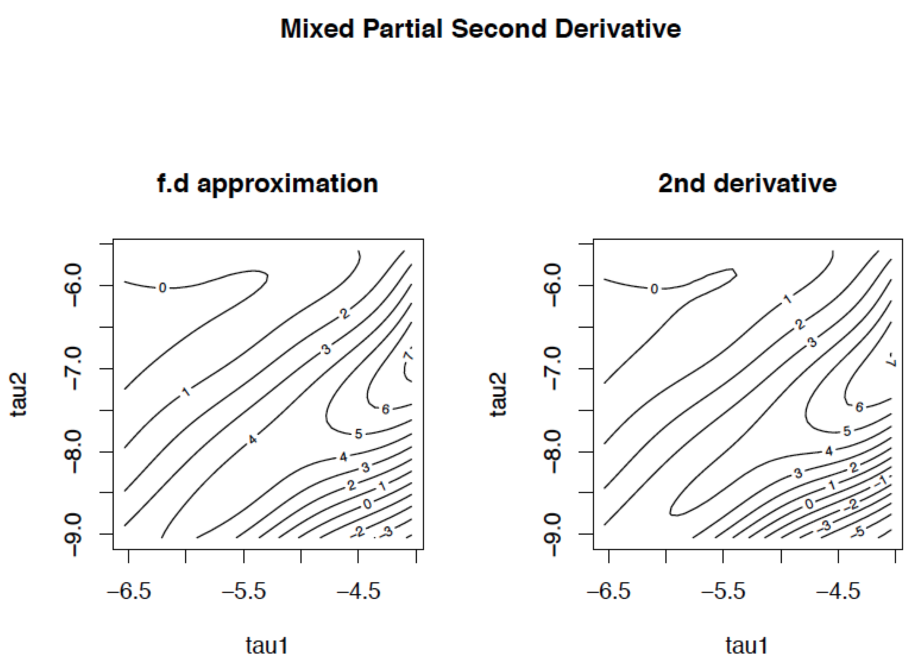
Recall that we must take half the trace of the above term and add it to our overall second derivative formula.

It has been observed in practice that the terms $\frac{\partial \mathbf{R}_\tau}{\partial \tau_i}\mathbf{R}_\tau^{-1}\frac{\partial \mathbf{R}_\tau}{\partial \tau_j}$ and $\frac{\partial \mathbf{R}_\tau}{\partial \tau_j}\mathbf{R}_\tau^{-1}\frac{\partial \mathbf{R}_\tau}{\partial \tau_i}$ will cancel each other out. However we have not yet been able to prove that these two terms are equal in all cases.

Below we show how contour plots of the second derivative at various τ values for a two-dimensional system correspond to finite-difference-approximations of the same second derivatives.




Second Derivative with respect to tau_2
f.d approximation
2nd derivative
tau2
tau1
−6.0
−7.0
−8.0
−9.0
−6.5
−5.5
−4.5


Mixed Partial Second Derivative
f.d approximation
2nd derivative
tau2
tau1
−6.0
−7.0
−8.0
−9.0
−6.5
−5.5
−4.5