SUPPLEMENT TO "INFLUENCE DIAGNOSTICS FOR RIDGE REGRESSION USING THE KULLBACK-LEIBLER DIVERGENCE"

ALONSO OGUEDA AND FELIPE OSORIO

Abstract. In this supplement, we present explicit expressions to determine the influence of the ith observation on the ridge estimator and obtain ∇_{KL} and $\boldsymbol{F}_{KL}$ which define the local curvature considering several perturbation schemes.

Appendix A. Case-deletion influence measure based on the Kullback-Leibler divergence

Proof of Proposition 1. We have that $\boldsymbol{\delta} = \mathsf{E}(\widehat{\boldsymbol{\beta}}_\lambda) - \mathsf{E}(\widehat{\boldsymbol{\beta}}_\lambda(i))$ can be written as

$$\boldsymbol{\delta} = \boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{S}_\lambda^{-1}(i)\boldsymbol{X}_{(i)}^\top\boldsymbol{X}_{(i)}\boldsymbol{\beta},$$

due to the fact that $\boldsymbol{X}_{(i)}^\top\boldsymbol{X}_{(i)} = \boldsymbol{X}^\top\boldsymbol{X} - \boldsymbol{x}_i\boldsymbol{x}_i^\top$, and

$$\boldsymbol{S}_\lambda^{-1}(i) = \boldsymbol{S}_\lambda^{-1} + \frac{\boldsymbol{S}_\lambda^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^\top\boldsymbol{S}_\lambda^{-1}}{1-h_{ii}(\lambda)},$$

it follows that

$$\begin{aligned}\boldsymbol{\delta} &= \boldsymbol{S}_\lambda^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^\top\boldsymbol{\beta} - \frac{1}{1-h_{ii}(\lambda)}\boldsymbol{S}_\lambda^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^\top\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta} + \frac{h_{ii}(\lambda)}{1-h_{ii}(\lambda)}\boldsymbol{S}_\lambda^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^\top\boldsymbol{\beta}\\ &= -\frac{1}{1-h_{ii}(\lambda)}\boldsymbol{S}_\lambda^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^\top\boldsymbol{S}_\lambda^{-1}(\boldsymbol{X}^\top\boldsymbol{X} - \boldsymbol{S}_\lambda)\boldsymbol{\beta} = \frac{\lambda}{1-h_{ii}(\lambda)}\boldsymbol{S}_\lambda^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^\top\boldsymbol{S}_\lambda^{-1}\boldsymbol{\beta}.\end{aligned}$$

By defining $q(\boldsymbol{\beta}) = \boldsymbol{x}_i^\top\boldsymbol{S}_\lambda^{-1}\boldsymbol{\beta}$, we can notice that,

$$\begin{aligned}\boldsymbol{S}_\lambda(i)\boldsymbol{\delta} &= \frac{\lambda}{1-h_{ii}(\lambda)}(\boldsymbol{S}_\lambda - \boldsymbol{x}_i\boldsymbol{x}_i^\top)\boldsymbol{S}_\lambda^{-1}\boldsymbol{x}_i q(\boldsymbol{\beta}) = \frac{\lambda}{1-h_{ii}(\lambda)}(1-h_{ii}(\lambda))\boldsymbol{x}_i q(\boldsymbol{\beta})\\ &= \lambda q(\boldsymbol{\beta})\boldsymbol{x}_i.\end{aligned}$$

Thus,

$$\begin{aligned}\boldsymbol{\delta}^\top\,\mathsf{Cov}(\widehat{\boldsymbol{\beta}}_\lambda(i))\boldsymbol{\delta} &= \frac{1}{\sigma^2}\boldsymbol{\delta}^\top\boldsymbol{S}_\lambda(i)(\boldsymbol{X}_{(i)}^\top\boldsymbol{X}_{(i)})^{-1}\boldsymbol{S}_\lambda(i)\boldsymbol{\delta}\\ &= \frac{\lambda^2 q^2(\boldsymbol{\beta})}{\sigma^2}\boldsymbol{x}_i^\top\Big\{(\boldsymbol{X}^\top\boldsymbol{X})^{-1} + \frac{(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}}{1-h_{ii}}\Big\}\boldsymbol{x}_i\\ &= \frac{\lambda^2 q^2(\boldsymbol{\beta})}{\sigma^2}\Big(h_{ii} + \frac{h_{ii}^2}{1-h_{ii}}\Big) = \frac{\lambda^2 q^2(\boldsymbol{\beta})}{\sigma^2}\Big(\frac{h_{ii}}{1-h_{ii}}\Big) \qquad \text{(A.1)}\end{aligned}$$

For the calculation of $\operatorname{tr}\mathsf{Cov}(\widehat{\boldsymbol{\beta}})\,\mathsf{Cov}^{-1}(\widehat{\boldsymbol{\beta}}_\lambda(i))$ it is required to obtain the trace of the $\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{S}_\lambda(i)(\boldsymbol{X}_{(i)}^\top\boldsymbol{X}_{(i)})^{-1}\boldsymbol{S}_\lambda(i)\boldsymbol{S}_\lambda^{-1}$ matrix. After a simple but tedious algebra, we obtain

$$\begin{aligned}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{S}_\lambda(i)(\boldsymbol{X}_{(i)}^\top\boldsymbol{X}_{(i)})^{-1}\boldsymbol{S}_\lambda(i)\boldsymbol{S}_\lambda^{-1} &= \boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}(\boldsymbol{S}_\lambda - \boldsymbol{x}_i\boldsymbol{x}_i^\top)\Big\{(\boldsymbol{X}^\top\boldsymbol{X})^{-1} \\ &\quad + \frac{1}{1-h_{ii}}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\Big\}(\boldsymbol{S}_\lambda - \boldsymbol{x}_i\boldsymbol{x}_i^\top)\boldsymbol{S}_\lambda^{-1} \\ &= \boldsymbol{I}_p + \frac{1}{1-h_{ii}}\boldsymbol{x}_i\boldsymbol{x}_i^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1} - \boldsymbol{x}_i\boldsymbol{x}_i^\top\boldsymbol{S}_\lambda^{-1} - \frac{h_{ii}}{1-h_{ii}}\boldsymbol{x}_i\boldsymbol{x}_i^\top\boldsymbol{S}_\lambda^{-1} \\ &\quad - \boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1} - \frac{h_{ii}}{1-h_{ii}}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1} \\ &\quad + h_{ii}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^\top\boldsymbol{S}_\lambda^{-1} + \frac{h_{ii}^2}{1-h_{ii}}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^\top\boldsymbol{S}_\lambda^{-1}.\end{aligned}$$

Therefore,

$$\begin{aligned}\operatorname{tr}\mathsf{Cov}(\widehat{\boldsymbol{\beta}}_\lambda)\,\mathsf{Cov}^{-1}(\widehat{\boldsymbol{\beta}}_\lambda(i)) &= p + \frac{h_{ii}}{1-h_{ii}} - 2h_{ii}(\lambda) - \frac{2h_{ii}h_{ii}(\lambda)}{1-h_{ii}} \\ &\quad + \frac{h_{ii}}{1-h_{ii}}\operatorname{tr}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^\top\boldsymbol{S}_\lambda^{-1} \\ &= p + \frac{1}{1-h_{ii}}\big(h_{ii} - 2h_{ii}(\lambda) + h_{ii}\operatorname{tr}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^\top\boldsymbol{S}_\lambda^{-1}\big).\end{aligned}$$

Using the SVD decomposition of $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top$ and noting that $\boldsymbol{x}_i = \boldsymbol{V}\boldsymbol{D}\boldsymbol{u}_i$, for $i = 1,\dots,n$, leads to,

$$\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^\top\boldsymbol{S}_\lambda^{-1} = \boldsymbol{V}\boldsymbol{D}^2(\boldsymbol{D}^2+\lambda\boldsymbol{I}_p)^{-1}\boldsymbol{D}\boldsymbol{u}_i\boldsymbol{u}_i^\top\boldsymbol{D}(\boldsymbol{D}^2+\lambda\boldsymbol{I}_p)^{-1}\boldsymbol{V}^\top,$$

hence

$$\operatorname{tr}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^\top\boldsymbol{S}_\lambda^{-1} = \operatorname{tr}\boldsymbol{\Delta}\boldsymbol{u}_i\boldsymbol{u}_i^\top\boldsymbol{\Delta} = \boldsymbol{u}_i^\top\boldsymbol{\Delta}^2\boldsymbol{u}_i,$$

with $\boldsymbol{\Delta} = (\boldsymbol{D}^2+\lambda\boldsymbol{I})^{-1}\boldsymbol{D}^2$. Consequently,

$$\operatorname{tr}\mathsf{Cov}(\widehat{\boldsymbol{\beta}}_\lambda)\,\mathsf{Cov}^{-1}(\widehat{\boldsymbol{\beta}}_\lambda(i)) = p + \frac{1}{1-h_{ii}}\big\{(1+\boldsymbol{u}_i^\top\boldsymbol{\Delta}^2\boldsymbol{u}_i)h_{ii} - 2h_{ii}(\lambda)\big\}. \tag{A.2}$$

In addition, as

$$\begin{aligned}|\boldsymbol{X}_{(i)}^\top\boldsymbol{X}_{(i)}| &= |\boldsymbol{X}^\top\boldsymbol{X}(\boldsymbol{I} - (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^\top)| = (1-h_{ii})|\boldsymbol{X}^\top\boldsymbol{X}|, \\ |\boldsymbol{S}_\lambda(i)| &= |\boldsymbol{S}_\lambda(\boldsymbol{I} - \boldsymbol{S}_\lambda^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^\top)| = (1-h_{ii}(\lambda))|\boldsymbol{S}_\lambda|,\end{aligned}$$

leads to write the ratio of determinants as

$$\begin{aligned}\frac{|\mathsf{Cov}(\widehat{\boldsymbol{\beta}}_\lambda)|}{|\mathsf{Cov}(\widehat{\boldsymbol{\beta}}_\lambda(i))|} &= \frac{\sigma^{2p}|\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}|}{\sigma^{2p}|\boldsymbol{S}_\lambda^{-1}(i)\boldsymbol{X}_{(i)}^\top\boldsymbol{X}_{(i)}\boldsymbol{S}_\lambda^{-1}(i)|} = \Big(\frac{|\boldsymbol{S}_\lambda(i)|}{|\boldsymbol{S}_\lambda|}\Big)^2\frac{|\boldsymbol{X}^\top\boldsymbol{X}|}{|\boldsymbol{X}_{(i)}^\top\boldsymbol{X}_{(i)}|} \\ &= \Big\{\frac{(1-h_{ii}(\lambda))|\boldsymbol{S}_\lambda|}{|\boldsymbol{S}_\lambda|}\Big\}^2\frac{|\boldsymbol{X}^\top\boldsymbol{X}|}{(1-h_{ii})|\boldsymbol{X}^\top\boldsymbol{X}|} = \frac{(1-h_{ii}(\lambda))^2}{1-h_{ii}}\end{aligned} \tag{A.3}$$

Thus, by collecting Equations (A.1), (A.2) and (A.3), allows to verify the proposition. □

Appendix B. Curvature derivation

In this appendix we derive the first and second differential $\mathrm{d}_\omega KL(\boldsymbol{\omega})$ and $\mathrm{d}_\omega^2 KL(\boldsymbol{\omega})$ for the each perturbation scheme defined in Section 3.2 of the manuscript. The vector ∇_{KL} and matrix $\boldsymbol{F}_{KL}$ are efficiently obtained using the differentiation method described in Magnus and Neudecker (2019) (see also Liu et al., 2024).

Proof of Proposition 2. Assume the influence function

$$\begin{aligned} KL(\boldsymbol{\omega}) = & \frac{1}{2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{\delta}(\boldsymbol{\omega}) - \frac{p}{2} \\ & + \frac{1}{2}\operatorname{tr}\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top \\ & + \log|\boldsymbol{S}_\lambda| - \frac{1}{2}\log|\boldsymbol{X}^\top\boldsymbol{X}| - \log|\boldsymbol{S}_\lambda(\boldsymbol{\omega})| + \frac{1}{2}\log|\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X}|, \end{aligned}$$

where $\boldsymbol{\delta}(\boldsymbol{\omega}) = (\boldsymbol{M}_\lambda - \boldsymbol{M}_\lambda(\boldsymbol{\omega}))\boldsymbol{\beta}$, with $\boldsymbol{M}_\lambda = \boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\boldsymbol{X}$ and $\boldsymbol{M}_\lambda(\boldsymbol{\omega}) = \boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X}$. Consider the following differentials,

$$\mathrm{d}_\omega(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1} = -(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}, \tag{B.1}$$

$$\mathrm{d}_\omega\log|\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X}| = \operatorname{tr}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X}, \tag{B.2}$$

$$\mathrm{d}_\omega\,\boldsymbol{S}_\lambda(\boldsymbol{\omega}) = \boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}, \tag{B.3}$$

$$\mathrm{d}_\omega\log|\boldsymbol{S}_\lambda(\boldsymbol{\omega})| = \operatorname{tr}\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}, \tag{B.4}$$

$$\mathrm{d}_\omega\,\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega}) = -\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega}), \tag{B.5}$$

and additionally

$$\begin{aligned} \mathrm{d}_\omega\,\boldsymbol{M}_\lambda(\boldsymbol{\omega}) = & -\{\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X} \\ & - \boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}\}. \end{aligned} \tag{B.6}$$

Using Equations (B.1) through (B.6), and after some algebra, it follows that

$$\begin{aligned} \mathrm{d}_\omega\,KL(\boldsymbol{\omega}) = & -\frac{1}{\sigma^2}\operatorname{tr}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{M}_\lambda - \boldsymbol{M}_\lambda(\boldsymbol{\omega}))\boldsymbol{\beta}\boldsymbol{\beta}^\top\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W} \\ & + \frac{1}{\sigma^2}\operatorname{tr}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{M}_\lambda - \boldsymbol{M}_\lambda(\boldsymbol{\omega}))\boldsymbol{\beta}\boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W} \\ & + \operatorname{tr}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W} \\ & - \frac{1}{2}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W} \\ & + \frac{1}{2}\operatorname{tr}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W} - \operatorname{tr}\boldsymbol{X}^\top\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}. \end{aligned}$$

Evaluating at $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, we have

$$\boldsymbol{M}_\lambda - \boldsymbol{M}(\boldsymbol{\omega}_0) = \boldsymbol{0}, \quad \boldsymbol{S}_\lambda(\boldsymbol{\omega}) = \boldsymbol{S}_\lambda, \quad \boldsymbol{X}^\top\boldsymbol{W}_0\boldsymbol{X} = \boldsymbol{X}^\top\boldsymbol{X},$$

from which it follows that

$$\begin{aligned} \mathrm{d}_\omega\,KL(\boldsymbol{\omega})\big|_{\omega=\omega_0} = & \operatorname{tr}\boldsymbol{X}^\top\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}\,\mathrm{d}\,\boldsymbol{W} - \frac{1}{2}\operatorname{tr}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W} \\ & + \frac{1}{2}\operatorname{tr}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W} - \operatorname{tr}\boldsymbol{X}^\top\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}\,\mathrm{d}\,\boldsymbol{W} = 0. \end{aligned}$$

Using the first identification theorem given in Magnus and Neudecker (2019), we obtain that $\partial KL(\boldsymbol{\omega})/\partial\boldsymbol{\omega}\big|_{\omega=\omega_0} = \boldsymbol{0}$.

Now, we have

$$
\begin{aligned}
\mathrm{d}^2_\omega KL(\boldsymbol{\omega}) &= \frac{1}{\sigma^2}\boldsymbol{\beta}^\top\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{\delta}(\boldsymbol{\omega})\\
&- \frac{1}{\sigma^2}\boldsymbol{\beta}^\top\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}\boldsymbol{\delta}(\boldsymbol{\omega})\\
&+ \frac{1}{\sigma^2}\boldsymbol{\beta}^\top\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta}\\
&- \frac{1}{\sigma^2}\boldsymbol{\beta}^\top\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta}\\
&+ \frac{1}{\sigma^2}\boldsymbol{\beta}^\top\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{\delta}(\boldsymbol{\omega})\\
&- \frac{1}{\sigma^2}\boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{\delta}(\boldsymbol{\omega})\\
&- \frac{1}{\sigma^2}\boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{\delta}(\boldsymbol{\omega})\\
&+ \frac{1}{\sigma^2}\boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta}\\
&- \frac{1}{\sigma^2}\boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X}\boldsymbol{S}^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta}\\
&- \operatorname{tr}\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\\
&+ \operatorname{tr}\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\\
&+ \operatorname{tr}\boldsymbol{X}^\top\boldsymbol{S}_\lambda^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\\
&- \frac{1}{2}\operatorname{tr}\boldsymbol{X}^\top\boldsymbol{S}_\lambda^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\\
&- \frac{1}{2}\operatorname{tr}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}\\
&+ \operatorname{tr}\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}.
\end{aligned}
$$

Evaluating at $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ and $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_\lambda$, it follows

$$
\begin{aligned}
\mathrm{d}^2_\omega KL(\boldsymbol{\omega})\big|_{\omega=\omega_0,\,\theta=\widehat{\theta}_\lambda} &= -\frac{1}{\widehat{\sigma}^2_\lambda}\widehat{\boldsymbol{\beta}}_\lambda^\top\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H}\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{X}\widehat{\boldsymbol{\beta}}_\lambda\\
&+ \frac{1}{\widehat{\sigma}^2_\lambda}\widehat{\boldsymbol{\beta}}_\lambda^\top\boldsymbol{X}^\top\boldsymbol{H}(\lambda)\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H}\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H}(\lambda)\boldsymbol{X}\widehat{\boldsymbol{\beta}}_\lambda - \operatorname{tr}\boldsymbol{H}(\lambda)\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H}\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H}\\
&+ \operatorname{tr}\boldsymbol{H}(\lambda)\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H}\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H}(\lambda) - \operatorname{tr}\boldsymbol{H}(\lambda)\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H}\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H} + \operatorname{tr}\boldsymbol{H}\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H}\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H}\\
&+ \operatorname{tr}\boldsymbol{H}(\lambda)\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H}(\lambda)\,\mathrm{d}\,\boldsymbol{W} - \frac{1}{2}\operatorname{tr}\boldsymbol{H}\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H}\,\mathrm{d}\,\boldsymbol{W}.
\end{aligned}
$$

Noticing that $\widehat{\boldsymbol{Y}}_\lambda = \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\lambda = \boldsymbol{H}(\lambda)\boldsymbol{Y}$, and

$$
\boldsymbol{H}\boldsymbol{H}(\lambda) = \boldsymbol{H}(\lambda), \qquad \boldsymbol{H}(\lambda)\boldsymbol{H} = \boldsymbol{H}(\lambda),
$$

we obtain

$$
\begin{aligned}
\mathrm{d}^2_\omega KL(\boldsymbol{\omega})\big|_{\omega=\omega_0,\,\theta=\widehat{\theta}_\lambda} &= -\frac{1}{\widehat{\sigma}^2_\lambda}\operatorname{tr}\boldsymbol{H}(\lambda)\boldsymbol{C}\boldsymbol{H}(\lambda)\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H}\,\mathrm{d}\,\boldsymbol{W} + \operatorname{tr}\boldsymbol{H}^2(\lambda)\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H}\,\mathrm{d}\,\boldsymbol{W}\\
&- 2\operatorname{tr}\boldsymbol{H}(\lambda)\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H}\,\mathrm{d}\,\boldsymbol{W} + \operatorname{tr}\boldsymbol{H}(\lambda)\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H}(\lambda)\,\mathrm{d}\,\boldsymbol{W} + \frac{1}{2}\operatorname{tr}\boldsymbol{H}\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H}\,\mathrm{d}\,\boldsymbol{W},
\end{aligned}
$$

where $\boldsymbol{C} = \boldsymbol{Y}\boldsymbol{Y}^\top - \widehat{\boldsymbol{Y}}_\lambda\widehat{\boldsymbol{Y}}_\lambda^\top$.

Hence, we can write

$$\begin{aligned}\mathrm{d}^2_\omega\, KL(\boldsymbol{\omega})\big|_{\omega=\omega_0,\,\theta=\widehat{\theta}_\lambda} &= -\frac{1}{\widehat{\sigma}^2_\lambda}\operatorname{tr}\boldsymbol{H}(\lambda)\boldsymbol{C}\boldsymbol{H}(\lambda)\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H}\,\mathrm{d}\,\boldsymbol{W} + \operatorname{tr}\boldsymbol{H}(\lambda)\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H}(\lambda)\,\mathrm{d}\,\boldsymbol{W}\\ &\quad + \frac{1}{2}\operatorname{tr}\big\{\boldsymbol{H} - 4\boldsymbol{H}(\lambda) + \boldsymbol{H}^2(\lambda)\big\}\,\mathrm{d}\,\boldsymbol{W}\boldsymbol{H}\,\mathrm{d}\,\boldsymbol{W}.\end{aligned}$$

Vectorizing, leads to

$$\begin{aligned}\mathrm{d}^2_\omega\, KL(\boldsymbol{\omega})\big|_{\omega=\omega_0,\,\theta=\widehat{\theta}_\lambda} &= -\frac{1}{\widehat{\sigma}^2_\lambda}(\mathrm{d}\operatorname{vec}\boldsymbol{W})^\top\big(\boldsymbol{H}\otimes\boldsymbol{H}(\lambda)\boldsymbol{C}\boldsymbol{H}(\lambda)\big)\,\mathrm{d}\operatorname{vec}\boldsymbol{W}\\ &\quad + \frac{1}{2}(\mathrm{d}\operatorname{vec}\boldsymbol{W})^\top\big(\boldsymbol{H}\otimes\boldsymbol{H} - 4\boldsymbol{H}(\lambda) + \boldsymbol{H}^2(\lambda)\big)\,\mathrm{d}\operatorname{vec}\boldsymbol{W}\\ &\quad + (\mathrm{d}\operatorname{vec}\boldsymbol{W})^\top\big(\boldsymbol{H}(\lambda)\otimes\boldsymbol{H}(\lambda)\big)\,\mathrm{d}\operatorname{vec}\boldsymbol{W},\end{aligned}$$

as $\boldsymbol{W} = \operatorname{diag}(\boldsymbol{\omega})$, then there exists a transition matrix $\boldsymbol{B}_n$ (see Nel, 1980) such that $\operatorname{vec}\boldsymbol{W} = \boldsymbol{B}_n\boldsymbol{\omega}$. Using the second identification theorem given in Magnus and Neudecker (2019), the result follows. □

Proof of Proposition 3. For this perturbation scheme, it is straightforward to notice that $KL(\boldsymbol{\omega})$ is given by

$$KL(\boldsymbol{\omega}) = \tfrac{1}{2}\boldsymbol{\delta}^\top\,\mathsf{Cov}^{-1}(\widehat{\boldsymbol{\beta}}_\lambda)\boldsymbol{\delta} = \tfrac{1}{2}\boldsymbol{\omega}^\top\boldsymbol{H}\boldsymbol{\omega},$$

because $\boldsymbol{\delta} = -(\boldsymbol{X}^\top\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^\top\boldsymbol{\omega}$ and $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$. Thus,

$$\mathrm{d}_\omega\, KL(\boldsymbol{\omega}) = \boldsymbol{\omega}^\top\boldsymbol{H}\,\mathrm{d}\,\boldsymbol{\omega}, \qquad \mathrm{d}^2_\omega\, KL(\boldsymbol{\omega}) = (\mathrm{d}\,\boldsymbol{\omega})^\top\boldsymbol{H}\,\mathrm{d}\,\boldsymbol{\omega}.$$

Applying the theorems of identification by Magnus and Neudecker (2019), we verify the proposition. □

Proof of Proposition 4. Consider the following influence function,

$$\begin{aligned}KL(\boldsymbol{\omega}) &= \frac{1}{2\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{\delta}(\boldsymbol{\omega}) - \frac{p}{2}\\ &\quad + \frac{1}{2}\operatorname{tr}\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\\ &\quad + \log|\boldsymbol{S}_\lambda| - \frac{1}{2}\log|\boldsymbol{X}^\top\boldsymbol{X}| - \log|\boldsymbol{S}_\lambda(\boldsymbol{\omega})| + \frac{1}{2}\log|\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega})|.\end{aligned}$$

Obtaining the first differential of $KL(\boldsymbol{\omega})$ with respect to $\boldsymbol{\omega}$ yields,

$$\begin{aligned}\mathrm{d}_\omega\, KL(\boldsymbol{\omega}) &= -\frac{a}{\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{c}_t(\mathrm{d}\,\boldsymbol{\omega})^\top\boldsymbol{X}\boldsymbol{\beta}\\ &\quad + \frac{a}{\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}\boldsymbol{\beta}\\ &\quad + \frac{a}{\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})\boldsymbol{\delta}(\boldsymbol{\omega})\\ &\quad - \frac{a}{2\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{\delta}(\boldsymbol{\omega})\\ &\quad + a\operatorname{tr}\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})\\ &\quad - \frac{a}{2}\operatorname{tr}\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\\ &\quad - 2a\boldsymbol{c}_t^\top\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top(\boldsymbol{\omega})\,\mathrm{d}\,\boldsymbol{\omega} + a\boldsymbol{c}_t^\top(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{X}^\top(\boldsymbol{\omega})\,\mathrm{d}\,\boldsymbol{\omega},\end{aligned}$$

where $\boldsymbol{Q}(\boldsymbol{\omega}) = \boldsymbol{c}_t(\mathrm{d}\,\boldsymbol{\omega})^\top\boldsymbol{X}(\boldsymbol{\omega}) + \boldsymbol{X}^\top(\boldsymbol{\omega})(\mathrm{d}\,\boldsymbol{\omega})\boldsymbol{c}_t^\top$.

Evaluating at $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, and noticing that,

$$\operatorname{tr}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{Q}(\boldsymbol{\omega}_0) = 2\boldsymbol{c}_t^\top(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \,\mathrm{d}\,\boldsymbol{\omega}, \quad \operatorname{tr}\boldsymbol{S}_\lambda^{-1}\boldsymbol{Q}(\boldsymbol{\omega}_0) = 2\boldsymbol{c}_t^\top \boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top \,\mathrm{d}\,\boldsymbol{\omega},$$

leads to

$$\begin{aligned}\mathrm{d}_\omega\, KL(\boldsymbol{\omega}_0) &= 2a\boldsymbol{c}_t^\top \boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top \,\mathrm{d}\,\boldsymbol{\omega} - a\boldsymbol{c}_t^\top(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \,\mathrm{d}\,\boldsymbol{\omega} \\ &\qquad - 2a\boldsymbol{c}_t^\top \boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top \,\mathrm{d}\,\boldsymbol{\omega} + a\boldsymbol{c}_t^\top(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \,\mathrm{d}\,\boldsymbol{\omega} = 0.\end{aligned}$$

The secod differential of $KL(\boldsymbol{\omega})$ with respect to $\boldsymbol{\omega}$, is given by

$$\begin{aligned}
\mathrm{d}^2_\omega\, KL(\boldsymbol{\omega}) &= \frac{a^2}{\sigma^2}\boldsymbol{\beta}^\top \boldsymbol{X}^\top(\mathrm{d}\,\boldsymbol{\omega})\boldsymbol{c}_t^\top(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{c}_t(\mathrm{d}\,\boldsymbol{\omega})^\top \boldsymbol{X}\boldsymbol{\beta} \\
&- \frac{a^2}{\sigma^2}\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X}(\boldsymbol{\omega})\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{c}_t(\mathrm{d}\,\boldsymbol{\omega})^\top \boldsymbol{X}\boldsymbol{\beta} \\
&- \frac{a^2}{\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{c}_t(\mathrm{d}\,\boldsymbol{\omega})^\top \boldsymbol{X}\boldsymbol{\beta} \\
&+ \frac{a^2}{\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{c}_t(\mathrm{d}\,\boldsymbol{\omega})^\top \boldsymbol{X}\boldsymbol{\beta} \\
&- \frac{a^2}{\sigma^2}\boldsymbol{\beta}^\top \boldsymbol{X}^\top(\mathrm{d}\,\boldsymbol{\omega})\boldsymbol{c}_t^\top(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}\boldsymbol{\beta} \\
&+ \frac{a^2}{\sigma^2}\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X}(\boldsymbol{\omega})\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}\boldsymbol{\beta} \\
&+ \frac{a^2}{\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}\boldsymbol{\beta} \\
&- \frac{a^2}{\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}\boldsymbol{\beta} \\
&+ \frac{2a^2}{\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{c}_t(\mathrm{d}\,\boldsymbol{\omega})^\top(\mathrm{d}\,\boldsymbol{\omega})\boldsymbol{c}_t^\top \boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}\boldsymbol{\beta} \\
&+ \frac{a^2}{\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}^2(\boldsymbol{\omega})\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}\boldsymbol{\beta} \\
&+ \frac{a^2}{\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\mathrm{d}\,\boldsymbol{\omega})\boldsymbol{c}_t^\top \boldsymbol{X}\boldsymbol{\beta} \\
&- \frac{a^2}{\sigma^2}\boldsymbol{\beta}^\top \boldsymbol{X}^\top(\mathrm{d}\,\boldsymbol{\omega})\boldsymbol{c}_t^\top(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})\boldsymbol{\delta}(\boldsymbol{\omega}) \\
&+ \frac{a^2}{\sigma^2}\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X}(\boldsymbol{\omega})\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\omega)\boldsymbol{\delta}(\boldsymbol{\omega}) \\
&+ \frac{a^2}{\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})\boldsymbol{\delta}(\boldsymbol{\omega}) \\
&- \frac{a^2}{\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\omega)\boldsymbol{\delta}(\boldsymbol{\omega}) \\
&+ \frac{2a^2}{\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{c}_t(\mathrm{d}\,\boldsymbol{\omega})^\top(\mathrm{d}\,\boldsymbol{\omega})\boldsymbol{c}_t^\top \boldsymbol{\delta}(\boldsymbol{\omega}) \\
&- \frac{a^2}{\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{c}_t(\mathrm{d}\,\boldsymbol{\omega})^\top \boldsymbol{X}\boldsymbol{\beta} \\
&+ \frac{a^2}{\sigma^2}\boldsymbol{\delta}^\top \boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{Q}(\omega)\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}\boldsymbol{\beta}
\end{aligned}$$

$$
\begin{aligned}
&+ \frac{a^2}{2\sigma^2}\boldsymbol{\beta}^\top \boldsymbol{X}^\top(\mathrm{d}\,\boldsymbol{\omega})\boldsymbol{c}_t^\top(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{\delta}(\boldsymbol{\omega})\\
&- \frac{a^2}{2\sigma^2}\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X}(\boldsymbol{\omega})\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{S}_\lambda\boldsymbol{\delta}(\boldsymbol{\omega})\\
&- \frac{a^2}{2\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{\delta}(\boldsymbol{\omega})\\
&+ \frac{a^2}{2\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{S}_\lambda\boldsymbol{\delta}(\boldsymbol{\omega})\\
&- \frac{a^2}{\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{c}_t(\mathrm{d}\,\boldsymbol{\omega})^\top(\mathrm{d}\,\boldsymbol{\omega})\boldsymbol{c}_t^\top(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{\delta}(\boldsymbol{\omega})\\
&+ \frac{a^2}{2\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{\delta}(\boldsymbol{\omega})\\
&- \frac{a^2}{2\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})\boldsymbol{\delta}(\boldsymbol{\omega})\\
&+ \frac{a^2}{2\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{c}_t(\mathrm{d}\,\boldsymbol{\omega})^\top\boldsymbol{X}\boldsymbol{\beta}\\
&- \frac{a^2}{2\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}\boldsymbol{\beta}\\
&+ a^2\,\mathrm{tr}\,\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})\\
&- a^2\,\mathrm{tr}\,\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})\\
&+ 2a^2\,\mathrm{tr}\,\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{c}_t(\mathrm{d}\,\boldsymbol{\omega})^\top(\mathrm{d}\,\boldsymbol{\omega})\boldsymbol{c}_t^\top\\
&- \frac{a^2}{2}\,\mathrm{tr}\,\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\\
&+ \frac{a^2}{2}\,\mathrm{tr}\,\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{S}_\lambda^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\\
&- a^2\,\mathrm{tr}\,\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{c}_t(\mathrm{d}\,\boldsymbol{\omega})^\top(\mathrm{d}\,\boldsymbol{\omega})\boldsymbol{c}_t^\top(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\\
&+ \frac{a^2}{2}\,\mathrm{tr}\,\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\\
&- \frac{a^2}{2}\,\mathrm{tr}\,\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})\\
&+ 2a^2\boldsymbol{c}_t^\top\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{Q}(\boldsymbol{\omega})\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top(\boldsymbol{\omega})\,\mathrm{d}\,\boldsymbol{\omega} - 2a^2\boldsymbol{c}_t^\top\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{c}_t(\mathrm{d}\,\boldsymbol{\omega})^\top\,\mathrm{d}\,\boldsymbol{\omega}\\
&- a^2\boldsymbol{c}_t^\top(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{Q}(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{X}^\top(\boldsymbol{\omega})\,\mathrm{d}\,\boldsymbol{\omega} + a^2\boldsymbol{c}_t^\top(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{c}_t(\mathrm{d}\,\boldsymbol{\omega})\,\mathrm{d}\,\boldsymbol{\omega}
\end{aligned}
$$

Consider,

$$
\boldsymbol{R} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}, \quad \boldsymbol{L} = \boldsymbol{S}_\lambda^{-1}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{S}_\lambda^{-1}, \quad \boldsymbol{m} = \boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\widehat{\boldsymbol{Y}}_\lambda,
$$

and $\boldsymbol{E}_t = \boldsymbol{c}_t\boldsymbol{c}_t^\top$. Noticing that,

$$
\begin{aligned}
(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{Q}(\boldsymbol{\omega}_0)(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{Q}(\boldsymbol{\omega}_0) &= r_{tt}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\mathrm{d}\,\boldsymbol{\omega})(\mathrm{d}\,\boldsymbol{\omega})^\top\boldsymbol{X}\\
&+ (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{c}_t(\mathrm{d}\,\boldsymbol{\omega})^\top\boldsymbol{H}(\mathrm{d}\,\boldsymbol{\omega})\boldsymbol{c}_t^\top + (\mathrm{d}\,\boldsymbol{\omega})^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{c}_t[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{c}_t(\mathrm{d}\,\boldsymbol{\omega})^\top\boldsymbol{X}\\
&+ (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\mathrm{d}\,\boldsymbol{\omega})\boldsymbol{c}_t^\top],
\end{aligned}
$$

and analogously

$$\begin{aligned}\boldsymbol{S}_\lambda^{-1}\boldsymbol{Q}(\boldsymbol{\omega}_0)(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{Q}(\boldsymbol{\omega}_0) &= r_{tt}\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top(\mathrm{d}\,\boldsymbol{\omega})(\mathrm{d}\,\boldsymbol{\omega})^\top\boldsymbol{X} + \boldsymbol{S}_\lambda^{-1}\boldsymbol{c}_t(\mathrm{d}\,\boldsymbol{\omega})^\top\boldsymbol{H}(\mathrm{d}\,\boldsymbol{\omega})\boldsymbol{c}_t^\top \\ &\quad + (\mathrm{d}\,\boldsymbol{\omega})^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{c}_t[\boldsymbol{S}_\lambda^{-1}\boldsymbol{c}_t(\mathrm{d}\,\boldsymbol{\omega})^\top\boldsymbol{X} + \boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top(\mathrm{d}\,\boldsymbol{\omega})\boldsymbol{c}_t^\top],\end{aligned}$$

and evaluating at $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_\lambda$, leads to

$$\begin{aligned}\mathrm{d}^2_\omega\, KL(\boldsymbol{\omega}_0) &= \frac{a^2 r_{tt}}{\widehat{\sigma}^2_\lambda}(\mathrm{d}\,\boldsymbol{\omega})^\top\widehat{\boldsymbol{Y}}_\lambda\widehat{\boldsymbol{Y}}_\lambda^\top\,\mathrm{d}\,\boldsymbol{\omega} - \frac{2a^2 r_{tt}}{\widehat{\sigma}^2_\lambda}(\mathrm{d}\,\boldsymbol{\omega})^\top\boldsymbol{H}(\lambda)\widehat{\boldsymbol{Y}}_\lambda\widehat{\boldsymbol{Y}}_\lambda^\top(\boldsymbol{I} - \tfrac{1}{2}\boldsymbol{H}(\lambda))\,\mathrm{d}\,\boldsymbol{\omega} \\ &\quad + a^2\Big(l_{tt} + \frac{m_t^2}{\widehat{\sigma}^2_\lambda} + r_{tt}^2\Big)(\mathrm{d}\,\boldsymbol{\omega})^\top\boldsymbol{H}\,\mathrm{d}\,\boldsymbol{\omega} - 2a^2 r_{tt}(\mathrm{d}\,\boldsymbol{\omega})^\top\boldsymbol{H}(\lambda)(\boldsymbol{I} - \tfrac{1}{2}\boldsymbol{H}(\lambda))\,\mathrm{d}\,\boldsymbol{\omega} \\ &\quad - \frac{2a^2}{\widehat{\sigma}^2_\lambda}(\mathrm{d}\,\boldsymbol{\omega})^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{E}_t\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\widehat{\boldsymbol{Y}}_\lambda\widehat{\boldsymbol{Y}}_\lambda^\top(\boldsymbol{I} - \tfrac{1}{2}\boldsymbol{H}(\lambda))\,\mathrm{d}\,\boldsymbol{\omega} \\ &\quad - 4a^2(\mathrm{d}\,\boldsymbol{\omega})^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{E}_t\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top(\boldsymbol{I} - \tfrac{1}{2}\boldsymbol{H}(\lambda))\,\mathrm{d}\,\boldsymbol{\omega} \\ &\quad + a^2(\mathrm{d}\,\boldsymbol{\omega})^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{E}_t(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{\omega} \\ &\quad + 2a^2(\mathrm{d}\,\boldsymbol{\omega})^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{E}_t\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{\omega}.\end{aligned}$$

Using the theorems of identification in Magnus and Neudecker (2019), the result follows. □

Appendix C. Penalized likelihood displacement for ridge regression

Suppose that $\boldsymbol{\beta}$ is the parameter of main interest while σ^2 is assumed as a nuisance parameter. To assess the influence of the i-th observation ($i = 1, \dots, n$), we propose to use the penalized likelihood displacement (Cook et al., 1988), which can be developed for ridge regression as,

$$LD_i(\boldsymbol{\beta}|\sigma^2) = 2\{\ell_\lambda(\widehat{\boldsymbol{\beta}}_\lambda, \widehat{\sigma}^2_\lambda) - \max_{\sigma^2}\,\ell_\lambda(\widehat{\boldsymbol{\beta}}_\lambda(i), \sigma^2)\}.$$

Because,

$$\ell_\lambda(\widehat{\boldsymbol{\beta}}_\lambda(i), \sigma^2) = -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\big(\|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\lambda(i)\|^2 + \lambda\|\widehat{\boldsymbol{\beta}}_\lambda(i)\|^2\big),$$

it is maximized at $\widehat{\sigma}^2_{\max}(i) = \big(\|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\lambda(i)\|^2 + \lambda\|\widehat{\boldsymbol{\beta}}_\lambda(i)\|^2\big)/n$. Therefore,

$$\max_{\sigma^2}\,\ell_\lambda(\widehat{\boldsymbol{\beta}}_\lambda(i), \sigma^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\widehat{\sigma}^2_{\max}(i) - \frac{n}{2}.$$

In addition,

$$\begin{aligned}\ell_\lambda(\widehat{\boldsymbol{\beta}}_\lambda, \widehat{\sigma}^2_\lambda) &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\widehat{\sigma}^2_\lambda - \frac{1}{2\widehat{\sigma}^2_\lambda}\big(\|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\lambda\|^2 + \lambda\|\widehat{\boldsymbol{\beta}}_\lambda\|^2\big) \\ &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\widehat{\sigma}^2_\lambda - \frac{n}{2}.\end{aligned}$$

This allows us to write,

$$LD_i(\boldsymbol{\beta}|\sigma^2) = n\log\Big(\frac{\widehat{\sigma}^2_{\max}}{\widehat{\sigma}^2_\lambda}\Big) = n\log\Big(\frac{\|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\lambda(i)\|^2 + \lambda\|\widehat{\boldsymbol{\beta}}_\lambda(i)\|^2}{\|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\lambda\|^2 + \lambda\|\widehat{\boldsymbol{\beta}}_\lambda\|^2}\Big).$$

References

Cook, R.D., Peña, D., Weisberg, S. (1988). The likelihood displacement: A unifying principle for influence measures. *Communications in Statistics - Theory & Methods* **17**, 623-640.

Liu, S., Trenkler, G., Kollo, T., von Rosen, D., Baksalary, O.M. (2024). Professor Heinz Neudecker and matrix differential calculus. *Statistical Papers* **65**, 2605-2639.

Magnus, J.R., Neudecker, H. (2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics (3rd Ed.)*. Wiley, Chichester.

Nel, D. (1980). On matrix differentiation in statistics. *South African Statistical Journal* **14**, 137-193.

Osorio, F. (2016). Influence diagnostics for robust P-splines using scale mixture of normal distributions. *Annals of the Institute of Statistical Mathematics* **68**, 589-619.

Shi, L., Wang, X. (1999). Local influence in ridge regression. *Computational Statistics & Data Analysis* **31**, 341-353.

Thomas, W. (1991). Influence diagnostics for the cross-validated smoothing parameter in spline smoothing. *Journal of the American Statistical Association* **86**, 693-698.

Department of Mathematical Sciences, George Mason University, USA

Current address: 4400 University Dr. Fairfax, VA 22030, USA

Orcid ID: https://orcid.org/0000-0001-5411-1484

Email address: aogueda@gmu.edu

Orcid ID: http://orcid.org/0000-0002-4675-5201

Email address: faosorios.stat@gmail.com