Spatial Gaussian Process Regression for Bayesian Optical Flow Estimation (Appendices)

Andy Wang & **Luca Orquiza**
awang1@jh.edu lorquiz1@jh.edu

Appendix A

Here we prove $\tilde{\boldsymbol{F}}|\boldsymbol{F} \sim \mathcal{N}\left(\boldsymbol{F}, \boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}\right)$ and $\boldsymbol{F} \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{K}\right)$ implies $\tilde{\boldsymbol{F}} \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{K}\right)$. For simplicity of derivations, we assume $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}} = \sigma_{\boldsymbol{\eta}}^2 \boldsymbol{I}$:

$$\begin{aligned} p\left(\tilde{\boldsymbol{F}}\right) &= \int p\left(\tilde{\boldsymbol{F}}|\boldsymbol{F}\right) p\left(\boldsymbol{F}\right)\, d\boldsymbol{F} \\ &\propto \int \exp\left(-\frac{1}{2}\left(\tilde{\boldsymbol{F}}-\boldsymbol{F}\right)^T \frac{1}{\sigma_{\boldsymbol{\eta}}^2}\boldsymbol{I}\left(\tilde{\boldsymbol{F}}-\boldsymbol{F}\right)\right) \exp\left(-\frac{1}{2}\left(\boldsymbol{F}-\boldsymbol{\mu}\right)^T \boldsymbol{K}^{-1}\left(\boldsymbol{F}-\boldsymbol{\mu}\right)\right) d\boldsymbol{F} \\ &= \int \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma_{\boldsymbol{\eta}}^2}\boldsymbol{F}^T\boldsymbol{F} - \frac{2}{\sigma_{\boldsymbol{\eta}}^2}\tilde{\boldsymbol{F}}^T\boldsymbol{F} + \frac{1}{\sigma_{\boldsymbol{\eta}}^2}\tilde{\boldsymbol{F}}^T\tilde{\boldsymbol{F}} + \boldsymbol{F}^T\boldsymbol{K}^{-1}\boldsymbol{F} - 2\boldsymbol{\mu}^T\boldsymbol{K}^{-1}\boldsymbol{F} + \boldsymbol{\mu}^T\boldsymbol{K}^{-1}\boldsymbol{\mu}\right]\right) d\boldsymbol{F} \end{aligned}$$

We are integrating over $\boldsymbol{F}$ and briefly group all terms not involving $\boldsymbol{F}$ into a constant:

$$p\left(\tilde{\boldsymbol{F}}\right) \propto \int \exp\left(-\frac{1}{2}\left[\boldsymbol{F}^T\left(\frac{1}{\sigma_{\boldsymbol{\eta}}^2}\boldsymbol{I} + \boldsymbol{K}^{-1}\right)\boldsymbol{F} - 2\left(\frac{1}{\sigma_{\boldsymbol{\eta}}^2}\tilde{\boldsymbol{F}}^T + \boldsymbol{\mu}^T\boldsymbol{K}^{-1}\right)\boldsymbol{F} + \text{const.}\right]\right) d\boldsymbol{F}$$

Let $\boldsymbol{A} \coloneqq \frac{1}{\sigma_{\boldsymbol{\eta}}^2}\boldsymbol{I} + \boldsymbol{K}^{-1}, \boldsymbol{b}^T \coloneqq \frac{1}{\sigma_{\boldsymbol{\eta}}^2}\tilde{\boldsymbol{F}}^T - \boldsymbol{\mu}^T\boldsymbol{K}^{-1}$. We complete the square:

$$\boldsymbol{F}^T\boldsymbol{A}\boldsymbol{F} - 2\boldsymbol{b}^T\boldsymbol{F} = \left(\boldsymbol{F} - \boldsymbol{A}^{-1}\boldsymbol{b}\right)^T \boldsymbol{A}\left(\boldsymbol{F} - \boldsymbol{A}^{-1}\boldsymbol{b}\right) - \boldsymbol{b}^T\boldsymbol{A}^{-1}\boldsymbol{b}$$

This yields:

$$p\left(\tilde{\boldsymbol{F}}\right) \propto \exp\left(-\frac{1}{2}\left(\boldsymbol{F} - \boldsymbol{A}^{-1}\boldsymbol{b}\right)^T \boldsymbol{A}\left(\boldsymbol{F} - \boldsymbol{A}^{-1}\boldsymbol{b}\right)\right) \exp\left(\frac{1}{2}\left[\boldsymbol{b}^T\boldsymbol{A}^{-1}\boldsymbol{b} - \frac{1}{\sigma_{\boldsymbol{\eta}}^2}\tilde{\boldsymbol{F}}^T\tilde{\boldsymbol{F}} - \boldsymbol{\mu}^T\boldsymbol{K}^{-1}\boldsymbol{\mu}\right]\right) d\boldsymbol{F}$$

The first term integrates to the inverse of the Gaussian normalizing factor. Meanwhile, the second term does not depend on $\boldsymbol{F}$. Reincorporating the initial normalizing factor, we obtain:

$$p\left(\tilde{\boldsymbol{F}}\right) = \frac{1}{\left(2\pi\sigma_{\boldsymbol{\eta}}\right)^{2N}\sqrt{\det\left(\boldsymbol{K}\right)}} \sqrt{\left(2\pi\right)^{2N}\det\left(\boldsymbol{A}^{-1}\right)} \exp\left(\frac{1}{2}\left[\boldsymbol{b}^T\boldsymbol{A}^{-1}\boldsymbol{b} - \frac{1}{\sigma_{\boldsymbol{\eta}}^2}\tilde{\boldsymbol{F}}^T\tilde{\boldsymbol{F}} - \boldsymbol{\mu}^T\boldsymbol{K}^{-1}\boldsymbol{\mu}\right]\right)$$

We use the intermediate result:

$$\boldsymbol{A} = \boldsymbol{K}^{-1} + \frac{1}{\sigma_{\boldsymbol{\eta}}^2}\boldsymbol{I} = \boldsymbol{K}^{-1}\left(\boldsymbol{I} + \frac{1}{\sigma_{\boldsymbol{\eta}}^2}\boldsymbol{K}\right) = \frac{1}{\sigma_{\boldsymbol{\eta}}^2}\boldsymbol{K}^{-1}\left(\boldsymbol{K} + \sigma_{\boldsymbol{\eta}}^2\boldsymbol{I}\right)$$

to compute $\det\left(\boldsymbol{A}^{-1}\right)$:

$$\det\left(\boldsymbol{A}\right)^{-\frac{1}{2}} = \left(\frac{1}{\sigma_{\boldsymbol{\eta}}^{2n}}\right)^{-\frac{1}{2}} \det\left(\boldsymbol{K}^{-1}\right)^{-\frac{1}{2}} \det\left(\boldsymbol{K} + \sigma_{\boldsymbol{\eta}}^2\boldsymbol{I}\right)^{-\frac{1}{2}} = \sigma_{\boldsymbol{\eta}}^n \det\left(\boldsymbol{K}\right)^{\frac{1}{2}} \det\left(\boldsymbol{K} + \sigma_{\boldsymbol{\eta}}^2\boldsymbol{I}\right)^{-\frac{1}{2}}$$

Substituting this into the above expression yields:

$$p\left(\tilde{\boldsymbol{F}}\right) = \frac{1}{\sqrt{2\pi\det\left(\boldsymbol{K} + \sigma_{\boldsymbol{\eta}}^2\boldsymbol{I}\right)}} \exp\left(\frac{1}{2}\left[\boldsymbol{b}^T\boldsymbol{A}^{-1}\boldsymbol{b} - \frac{1}{\sigma_{\boldsymbol{\eta}}^2}\tilde{\boldsymbol{F}}^T\tilde{\boldsymbol{F}} - \boldsymbol{\mu}^T\boldsymbol{K}^{-1}\boldsymbol{\mu}\right]\right)$$

Preprint.

We express $\boldsymbol{b}^T\boldsymbol{A}^{-1}\boldsymbol{b}$ in terms of $\tilde{\boldsymbol{F}},\boldsymbol{\mu},\boldsymbol{K}$. Let $\boldsymbol{K}_\sigma \coloneqq \boldsymbol{K}+\sigma_{\boldsymbol{\eta}}^2\boldsymbol{I}$:

$$\begin{aligned}
\boldsymbol{b}^T\boldsymbol{A}^{-1}\boldsymbol{b} &= \left(\frac{1}{\sigma_{\boldsymbol{\eta}}^2}\tilde{\boldsymbol{F}} - \boldsymbol{K}^{-1}\boldsymbol{\mu}\right)^T\left(\boldsymbol{I}+\frac{1}{\sigma_{\boldsymbol{\eta}}^2}\boldsymbol{K}\right)^{-1}\boldsymbol{K}\left(\frac{1}{\sigma_{\boldsymbol{\eta}}^2}\tilde{\boldsymbol{F}}+\boldsymbol{K}^{-1}\boldsymbol{\mu}\right)\\
&= \sigma_{\boldsymbol{\eta}}^2\left(\frac{1}{\sigma_{\boldsymbol{\eta}}^2}\tilde{\boldsymbol{F}}+\boldsymbol{K}^{-1}\boldsymbol{\mu}\right)^T\boldsymbol{K}_\sigma^{-1}\boldsymbol{K}\left(\frac{1}{\sigma_{\boldsymbol{\eta}}^2}\tilde{\boldsymbol{F}}+\boldsymbol{K}^{-1}\boldsymbol{\mu}\right)\\
&= \sigma_{\boldsymbol{\eta}}^2\left[\frac{1}{\sigma_{\boldsymbol{\eta}}^4}\tilde{\boldsymbol{F}}^T\boldsymbol{K}_\sigma^{-1}\boldsymbol{K}\tilde{\boldsymbol{F}}+\frac{2}{\sigma_{\boldsymbol{\eta}}^2}\tilde{\boldsymbol{F}}^T\boldsymbol{K}_\sigma^{-1}\boldsymbol{\mu}+\boldsymbol{\mu}^T\boldsymbol{K}^{-1}\boldsymbol{K}_\sigma^{-1}\boldsymbol{\mu}\right]\\
&= \frac{1}{\sigma_{\boldsymbol{\eta}}^2}\tilde{\boldsymbol{F}}^T\tilde{\boldsymbol{F}} - \tilde{\boldsymbol{F}}^T\boldsymbol{K}_\sigma^{-1}\tilde{\boldsymbol{F}}+2\tilde{\boldsymbol{F}}^T\boldsymbol{K}_\sigma^{-1}\boldsymbol{\mu}+\sigma_{\boldsymbol{\eta}}^2\boldsymbol{\mu}^T\boldsymbol{K}^{-1}\boldsymbol{K}_\sigma^{-1}\boldsymbol{\mu}
\end{aligned}$$

Using this, the previous expression becomes:

$$\begin{aligned}
p\left(\tilde{\boldsymbol{F}}\right) &\propto \exp\left(\frac{1}{2}\left[-\tilde{\boldsymbol{F}}^T\boldsymbol{K}_\sigma^{-1}\tilde{\boldsymbol{F}}+2\tilde{\boldsymbol{F}}\boldsymbol{K}_\sigma^{-1}\boldsymbol{\mu}+\boldsymbol{\mu}^T\left(\sigma_\eta^2\boldsymbol{K}^{-1}\boldsymbol{K}_\sigma^{-1}-\boldsymbol{K}^{-1}\right)\boldsymbol{\mu}\right]\right)\\
&= \exp\left(\frac{1}{2}\left[-\tilde{\boldsymbol{F}}^T\boldsymbol{K}_\sigma^{-1}\tilde{\boldsymbol{F}}+2\tilde{\boldsymbol{F}}^T\boldsymbol{K}_\sigma^{-1}\boldsymbol{\mu}+\text{const.}\right]\right)
\end{aligned}$$

We again complete the square:

$$-\tilde{\boldsymbol{F}}^T\boldsymbol{K}_\sigma^{-1}\tilde{\boldsymbol{F}}+2\tilde{\boldsymbol{F}}\boldsymbol{K}_\sigma^{-1}\boldsymbol{\mu} = -\left(\tilde{\boldsymbol{F}}-\boldsymbol{\mu}\right)^T\boldsymbol{K}_\sigma^{-1}\left(\tilde{\boldsymbol{F}}-\boldsymbol{\mu}\right)+\boldsymbol{\mu}^T\boldsymbol{K}_\sigma^{-1}\boldsymbol{\mu}$$

This yields:

$$p\left(\tilde{\boldsymbol{F}}\right) \propto \exp\left(-\frac{1}{2}\left(\tilde{\boldsymbol{F}}-\boldsymbol{\mu}\right)^T\boldsymbol{K}_\sigma^{-1}\left(\tilde{\boldsymbol{F}}-\boldsymbol{\mu}\right)\right)\exp\left(\frac{1}{2}\boldsymbol{\mu}^T\left[\boldsymbol{K}_\sigma^{-1}+\sigma_\eta^2\boldsymbol{K}^{-1}\boldsymbol{K}_\sigma^{-1}-\boldsymbol{K}^{-1}\right]\boldsymbol{\mu}\right)$$

The middle term in the second exponent is:

$$\boldsymbol{K}_\sigma^{-1}+\sigma_\eta^2\boldsymbol{K}^{-1}\boldsymbol{K}_\sigma^{-1}-\boldsymbol{K}^{-1} = \left(\boldsymbol{I}+\sigma_{\boldsymbol{\eta}}^2\boldsymbol{K}^{-1}\right)\boldsymbol{K}_\sigma^{-1}-\boldsymbol{K}^{-1} = \boldsymbol{K}^{-1}\boldsymbol{K}_\sigma\boldsymbol{K}_\sigma^{-1}-\boldsymbol{K}^{-1} = \boldsymbol{O}$$

Thus we achieve the final expression for the density:

$$p\left(\tilde{\boldsymbol{F}}\right) = \frac{1}{\sqrt{2\pi\det\left(\boldsymbol{K}+\boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}\right)}}\exp\left(-\frac{1}{2}\left(\tilde{\boldsymbol{F}}-\boldsymbol{\mu}\right)^T\left(\boldsymbol{K}+\boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}\right)^{-1}\left(\tilde{\boldsymbol{F}}-\boldsymbol{\mu}\right)\right)$$

and subsequently the target distribution:

$$\tilde{\boldsymbol{F}} \sim \mathcal{N}\left(\boldsymbol{\mu},\boldsymbol{K}+\boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}\right)$$

Appendix B

Here we prove that if $\boldsymbol{X}_1, \boldsymbol{X}_2$ are multivariate normal, i.e.:

$$\begin{bmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\right)$$

then the conditional distribution of $\boldsymbol{X}_1|\boldsymbol{X}_2$ is:

$$\boldsymbol{X}_1|\boldsymbol{X}_2 \sim \mathcal{N}\left(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\boldsymbol{X}_2 - \boldsymbol{\mu}_2\right), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)$$

Conditioning on $\boldsymbol{X}_2$ means $\boldsymbol{X}_2$ is fixed. Thus:

$$p_{\boldsymbol{X}_1|\boldsymbol{X}_2}(\boldsymbol{x}_1|\boldsymbol{x}_2) = \frac{p_{\boldsymbol{X}_1,\boldsymbol{X}_2}(\boldsymbol{x}_1,\boldsymbol{x}_2)}{p_{\boldsymbol{X}_2}(\boldsymbol{x}_2)} \propto p_{\boldsymbol{X}_1,\boldsymbol{X}_2}(\boldsymbol{x}_1,\boldsymbol{x}_2)$$

Let $\tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}^{-1}$. This density is proportional to:

$$\exp\left(-\frac{1}{2}\begin{bmatrix}\boldsymbol{x}_1 - \boldsymbol{\mu}_1\\ \boldsymbol{x}_2 - \boldsymbol{\mu}_2\end{bmatrix}^T \begin{bmatrix}\tilde{\boldsymbol{\Sigma}}_{11} & \tilde{\boldsymbol{\Sigma}}_{12}\\ \tilde{\boldsymbol{\Sigma}}_{21} & \tilde{\boldsymbol{\Sigma}}_{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_1 - \boldsymbol{\mu}_1\\ \boldsymbol{x}_2 - \boldsymbol{\mu}_2\end{bmatrix}\right)$$

Notice that $\tilde{\boldsymbol{\Sigma}}_{ij}$ equals $\left(\boldsymbol{\Sigma}^{-1}\right)_{ij}$ and not $(\boldsymbol{\Sigma}_{ij})^{-1}$. Let Q be the exponent. We expand:

$$Q \propto (\boldsymbol{x}_1 - \boldsymbol{\mu}_1)^T \tilde{\boldsymbol{\Sigma}}_{11}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1) + 2(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)^T \tilde{\boldsymbol{\Sigma}}_{12}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2) + (\boldsymbol{x}_2 - \boldsymbol{\mu}_2)^T \tilde{\boldsymbol{\Sigma}}_{22}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)$$

We expand further and ignore all constant terms (those not involving $\boldsymbol{x}_1$):

$$Q = \boldsymbol{x}_1^T\tilde{\boldsymbol{\Sigma}}_{11}\boldsymbol{x}_1 - 2\boldsymbol{x}_1^T\tilde{\boldsymbol{\Sigma}}_{11}\boldsymbol{\mu}_1 + 2\boldsymbol{x}_1^T\tilde{\boldsymbol{\Sigma}}_{12}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2) + \text{const.}$$

We complete the square by finding $\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*$:

$$(\boldsymbol{x}_1 - \boldsymbol{\mu}_*)^T \boldsymbol{\Sigma}_*^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_*) = \boldsymbol{x}_1^T\boldsymbol{\Sigma}_*^{-1}\boldsymbol{x}_1 - 2\boldsymbol{x}_1\boldsymbol{\Sigma}_*^{-1}\boldsymbol{\mu}_* = Q$$

The quadratic terms must equate:

$$\boldsymbol{x}_1^T\boldsymbol{\Sigma}_*^{-1}\boldsymbol{x}_1 = \boldsymbol{x}_1^T\tilde{\boldsymbol{\Sigma}}_{11}\boldsymbol{x}_1 \implies \boldsymbol{\Sigma}_*^{-1} = \tilde{\boldsymbol{\Sigma}}_{11} \implies \boldsymbol{\Sigma}_* = \tilde{\boldsymbol{\Sigma}}_{11}^{-1}$$

The linear terms must also equate:

$$-2\boldsymbol{x}_1\boldsymbol{\Sigma}_*^{-1}\boldsymbol{\mu}_* = -2\boldsymbol{x}_1^T\tilde{\boldsymbol{\Sigma}}_{11}\boldsymbol{\mu}_1 + 2\boldsymbol{x}_1^T\tilde{\boldsymbol{\Sigma}}_{12}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)$$

Substituting $\boldsymbol{\Sigma}_*^{-1} = \tilde{\boldsymbol{\Sigma}}_{11}$ and simplifying yields:

$$\tilde{\boldsymbol{\Sigma}}_{11}\boldsymbol{\mu}_* = \tilde{\boldsymbol{\Sigma}}_{11}\boldsymbol{\mu}_1 - \tilde{\boldsymbol{\Sigma}}_{12}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2) \implies \boldsymbol{\mu}_* = \boldsymbol{\mu}_1 - \tilde{\boldsymbol{\Sigma}}_{11}^{-1}\tilde{\boldsymbol{\Sigma}}_{12}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)$$

Thus, we have:

$$p_{\boldsymbol{X}_1|\boldsymbol{X}_2} \propto \exp\left(-\frac{1}{2}(\boldsymbol{x}_1 - \boldsymbol{\mu}_*)^T \boldsymbol{\Sigma}_*(\boldsymbol{x}_1 - \boldsymbol{\mu}_*)\right) \implies \boldsymbol{X}_1|\boldsymbol{X}_2 \sim \mathcal{N}(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*)$$

We have absorbed all constants into the final density's normalizing factor.

All that remains is to determine $\tilde{\boldsymbol{\Sigma}}_{11}, \tilde{\boldsymbol{\Sigma}}_{12}$. Zhang (2005) gives:

$$\boldsymbol{M} = \begin{bmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{C} & \boldsymbol{D}\end{bmatrix} = \begin{bmatrix}\boldsymbol{I} & \boldsymbol{B}\boldsymbol{D}^{-1}\\ \boldsymbol{O} & \boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{S} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{D}\end{bmatrix}\begin{bmatrix}\boldsymbol{I} & \boldsymbol{O}\\ \boldsymbol{D}^{-1}\boldsymbol{C} & \boldsymbol{I}\end{bmatrix}$$

where Schur complement $\boldsymbol{S} \coloneqq \boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}$. We invert:

$$\begin{aligned}\boldsymbol{M}^{-1} &= \begin{bmatrix}\boldsymbol{I} & -\boldsymbol{B}\boldsymbol{D}^{-1}\\ \boldsymbol{O} & \boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{S}^{-1} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{D}^{-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{I} & \boldsymbol{O}\\ -\boldsymbol{D}^{-1}\boldsymbol{C} & \boldsymbol{I}\end{bmatrix}\\ &= \begin{bmatrix}\boldsymbol{S}^{-1} & -\boldsymbol{S}^{-1}\boldsymbol{B}\boldsymbol{D}^{-1}\\ -\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{S}^{-1} & \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{S}^{-1}\boldsymbol{B}\boldsymbol{D}^{-1}\end{bmatrix}\end{aligned}$$

For $\boldsymbol{M} = \boldsymbol{\Sigma}$, Schur complement $\boldsymbol{\Sigma}_S \coloneqq \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$, and:

$$\tilde{\boldsymbol{\Sigma}}_{11} = \left(\boldsymbol{\Sigma}^{-1}\right)_{11} = \boldsymbol{\Sigma}_S^{-1} \qquad \tilde{\boldsymbol{\Sigma}}_{12} = \left(\boldsymbol{\Sigma}^{-1}\right)_{12} = -\boldsymbol{\Sigma}_S^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$$

From these we derive the intermediate result:

$$\tilde{\boldsymbol{\Sigma}}_{11}^{-1}\tilde{\boldsymbol{\Sigma}}_{12} = \boldsymbol{\Sigma}_S\left(-\boldsymbol{\Sigma}_S^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\right) = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$$

These results yield expressions for $\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*$ in terms of $\boldsymbol{\mu}, \boldsymbol{\Sigma}$:

$$\boldsymbol{\mu}_* = \boldsymbol{\mu}_1 - \tilde{\boldsymbol{\Sigma}}_{11}^{-1}\tilde{\boldsymbol{\Sigma}}_{12}\left(\boldsymbol{x}_2 - \boldsymbol{\mu}_2\right) = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\boldsymbol{x}_2 - \boldsymbol{\mu}_2\right)$$

$$\boldsymbol{\Sigma}_* = \tilde{\boldsymbol{\Sigma}}_{11}^{-1} = \boldsymbol{\Sigma}_S = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$

We have fully derived conditional distribution $\boldsymbol{X}_1|\boldsymbol{X}_2$. Substituting the appropriate mean and covariance parameters from the formulation of Gaussian process regression into this general expression yields the posterior distribution:

$$\boldsymbol{F}|\tilde{\boldsymbol{F}} \sim \mathcal{N}\left(\boldsymbol{\mu} + \boldsymbol{K}\left(\boldsymbol{K} + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}\right)^{-1}\left(\tilde{\boldsymbol{F}} - \boldsymbol{\mu}\right), \boldsymbol{K} - \boldsymbol{K}\left(\boldsymbol{K} + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}\right)^{-1}\boldsymbol{K}\right)$$

Appendix C

Here we show how to maximize the marginal likelihood of $\tilde{\boldsymbol{F}} \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{K} + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}\right)$:

$$p\left(\tilde{\boldsymbol{F}}\right) = \frac{1}{\sqrt{(2\pi)^{2N}\det\left(\boldsymbol{K} + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}\right)}}\exp\left(-\frac{1}{2}\left(\tilde{\boldsymbol{F}} - \boldsymbol{\mu}\right)^T\left(\boldsymbol{K} + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}\right)^{-1}\left(\tilde{\boldsymbol{F}} - \boldsymbol{\mu}\right)\right)$$

Maximizing the likelihood is equivalent to minimizing the negative log-likelihood:

$$-\log p\left(\tilde{\boldsymbol{F}}\right) = \frac{1}{2}\left(\tilde{\boldsymbol{F}} - \boldsymbol{\mu}\right)^T\left(\boldsymbol{K} + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}\right)^{-1}\left(\tilde{\boldsymbol{F}} - \boldsymbol{\mu}\right) + \frac{1}{2}\log\det\left(\boldsymbol{K} + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}\right) + N\log 2\pi$$

Let $\mathcal{L} = -\log p\left(\tilde{\boldsymbol{F}}\right)$, $\boldsymbol{y} = \tilde{\boldsymbol{F}} - \boldsymbol{\mu}$, $\boldsymbol{K}_y = \boldsymbol{K} + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}$, and φ be a parameter of $\boldsymbol{m}$:

$$\frac{\partial\mathcal{L}}{\partial\varphi} = \frac{1}{2}\left[\left(\frac{\partial\boldsymbol{y}^T}{\partial\varphi}\right)\boldsymbol{K}_y^{-1}\boldsymbol{y} + \boldsymbol{y}^T\boldsymbol{K}_y\left(\frac{\partial\boldsymbol{y}}{\partial\varphi}\right)\right] = \frac{1}{2}\left(2\boldsymbol{y}^T\boldsymbol{K}_y\left(\frac{\partial\boldsymbol{y}}{\partial\varphi}\right)\right) = -\boldsymbol{y}^T\boldsymbol{K}_y\left(\frac{\partial\boldsymbol{\mu}}{\partial\varphi}\right)$$

Let θ be a parameter of $\boldsymbol{k}$. Jacobi's formula (Magnus & Neudecker, 1999) gives:

$$\frac{\partial\mathcal{L}}{\partial\theta} = \boldsymbol{y}^T\left(\frac{\partial\boldsymbol{K}_y^{-1}}{\partial\theta}\right)\boldsymbol{y} + \frac{\partial|\boldsymbol{K}_y|}{\partial\theta} = -\frac{1}{2}\boldsymbol{y}^T\boldsymbol{K}^{-1}\left(\frac{\partial\boldsymbol{K}_y}{\partial\theta}\right)\boldsymbol{K}^{-1}\boldsymbol{y} + \frac{1}{2}\mathrm{tr}\left(\boldsymbol{K}_y^{-1}\left(\frac{\partial\boldsymbol{K}_y}{\partial\theta}\right)\right)$$

Partials $\frac{\partial\boldsymbol{m}}{\partial\varphi}$ and $\frac{\partial\boldsymbol{K}_y}{\partial\theta}$ depend on the specific kernels. Our choices were:

$$\boldsymbol{m}(\cdot) = \boldsymbol{c} \qquad \boldsymbol{k}(\cdot,\cdot') = \sigma^2\exp\left(-z\right)\boldsymbol{\Sigma} \qquad z = \frac{\|\cdot - \cdot'\|^2}{2\lambda^2} \qquad \boldsymbol{\Sigma} = \begin{bmatrix}\sigma_{uu} & \sigma_{uv}\\ \sigma_{vu} & \sigma_{vv}\end{bmatrix}$$

We set $\boldsymbol{\Sigma} = \boldsymbol{I}$, but let's consider $\frac{\partial\boldsymbol{K}}{\partial\sigma_{uu}}$ (the others are analogous). We have:

$$\frac{\partial\boldsymbol{m}}{\partial\boldsymbol{c}} = \boldsymbol{1} \qquad \frac{\partial\boldsymbol{k}}{\partial\lambda} = \sigma^2\exp\left(-z\right)\Sigma\left(\frac{\|\cdot - \cdot'\|^2}{\lambda^3}\right) = \frac{\boldsymbol{k}\|\cdot - \cdot'\|^2}{\lambda^3}$$

$$\frac{\partial\boldsymbol{k}}{\partial\sigma} = 2\sigma\exp\left(-z\right)\boldsymbol{\Sigma} = \frac{2\boldsymbol{k}}{\sigma} \qquad \frac{\partial\boldsymbol{k}}{\partial\sigma_{uu}} = \sigma^2\exp\left(-z\right)\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}$$

These extend naturally to the partials below:

$$\frac{\partial\boldsymbol{\mu}}{\partial\boldsymbol{c}} = \boldsymbol{1} \qquad \frac{\partial\boldsymbol{K}}{\partial\lambda} = K\odot\frac{\|\cdot - \cdot'\|^2}{\lambda^3} \qquad \frac{\partial\boldsymbol{K}}{\partial\sigma} = \frac{2\boldsymbol{K}}{\sigma}$$

$$\frac{\partial\boldsymbol{K}}{\partial\sigma_{uu}} = \sigma^2\exp\left(-z\right)\begin{bmatrix}\boldsymbol{1}_{n\times n} & \boldsymbol{O}_{n\times n}\\ \boldsymbol{O}_{n\times n} & \boldsymbol{O}_{n\times n}\end{bmatrix}$$

These derivatives compose the gradient vector at each descent step.