7 Appendix (supplementary)

7.1 Partial Derivatives and Gradient

The expression of $\boldsymbol{g}^{(\eta)}(\boldsymbol{\theta})$ used for updating the GP hyper-parameters, $\boldsymbol{\theta}$, in (12) is obtained as:

$$\begin{aligned}\boldsymbol{g}^{(\eta)}(\boldsymbol{\theta}) &= -2\cdot \boldsymbol{y}_V^T \boldsymbol{K}_{VT}(\boldsymbol{\theta}_h)\boldsymbol{z}_T^\eta \\ &+ (\boldsymbol{z}_T^\eta)^T \boldsymbol{K}_{VT}(\boldsymbol{\theta}_h)^T \boldsymbol{K}_{VT}(\boldsymbol{\theta}_h)\boldsymbol{z}_T^\eta \\ &+ \boldsymbol{\lambda}^T \boldsymbol{K}_{TT}(\boldsymbol{\theta}_h)\boldsymbol{z}_T^\eta \\ &+ \rho(\sigma_e^2 \boldsymbol{z}_T^\eta - \boldsymbol{y}_T)^T \boldsymbol{K}_{TT}(\boldsymbol{\theta}_h)\boldsymbol{z}_T^\eta \\ &+ \frac{\rho}{2}(\boldsymbol{z}_T^\eta)^T \boldsymbol{K}_{TT}(\boldsymbol{\theta}_h)\boldsymbol{K}_{TT}(\boldsymbol{\theta}_h)\boldsymbol{z}_T^\eta. \end{aligned} \tag{17}$$

For each element of $\boldsymbol{\theta}$ (denoted as θ_i), its partial derivative is computed as:

$$\begin{aligned}\frac{\partial \boldsymbol{g}^{(\eta)}(\boldsymbol{\theta})}{\partial \theta_i} &= -2\cdot \boldsymbol{y}_V^T \frac{\partial \boldsymbol{K}_{VT}(\boldsymbol{\theta}_h)}{\partial \theta_i}\boldsymbol{z}_T^\eta \\ &+ (\boldsymbol{z}_T^\eta)^T \frac{\partial \boldsymbol{K}_{VT}^T(\boldsymbol{\theta}_h)}{\partial \theta_i}\boldsymbol{K}_{VT}(\boldsymbol{\theta}_h)\boldsymbol{z}_T^\eta \\ &+ (\boldsymbol{z}_T^\eta)^T \boldsymbol{K}_{VT}^T(\boldsymbol{\theta}_h)\frac{\partial \boldsymbol{K}_{VT}(\boldsymbol{\theta}_h)}{\partial \theta_i}\boldsymbol{z}_T^\eta \\ &+ \boldsymbol{\lambda}^T \frac{\partial \boldsymbol{K}_{TT}(\boldsymbol{\theta}_h)}{\partial \theta_i}\boldsymbol{z}_T^\eta \\ &+ \rho(\sigma_e^2 \boldsymbol{z}_T^\eta - \boldsymbol{y}_T)^T \frac{\partial \boldsymbol{K}_{TT}(\boldsymbol{\theta}_h)}{\partial \theta_i}\boldsymbol{z}_T^\eta \\ &+ \frac{\rho}{2}(\boldsymbol{z}_T^\eta)^T \boldsymbol{K}_{TT}(\boldsymbol{\theta}_h)\frac{\partial \boldsymbol{K}_{TT}(\boldsymbol{\theta}_h)}{\partial \theta_i}\boldsymbol{z}_T^\eta \\ &+ \frac{\rho}{2}(\boldsymbol{z}_T^\eta)^T \frac{\partial \boldsymbol{K}_{TT}(\boldsymbol{\theta}_h)}{\partial \theta_i}\boldsymbol{K}_{TT}(\boldsymbol{\theta}_h)\boldsymbol{z}_T^\eta. \end{aligned} \tag{18}$$

7.2 Explicit Form of Kernel Functions

The expressions for the selected kernels that we use for the synthetic data are listed below.

- **Squared Exponential (SE) Kernel**
  SE kernel is usually regarded as the default kernel for GP models, due to its great universality as well as many good properties. The length scale l in an SE kernel specifies the width of the kernel and thereby determines the smoothness of the regression function.

$$k_{se}(x,x') = \sigma^2 exp\left(-\frac{(x-x')^2}{2l^2}\right)$$

- **Locally Periodic (LP) Kernel**
  Periodicity is another important pattern that people always get interested, especially in modeling time series data. Most of the real data do not repeat themselves exactly. Therefore combining a local kernel together with a periodic kernel into a locally periodic kernel, is considered to allow the shape of the repeating patterns to evolve over time:

$$k_{lp}(x,x') = \\ \sigma^2 exp\left(-\frac{2sin^2(\pi|x-x'|/p)}{l^2}\right) exp\left(-\frac{(x-x')^2}{2l^2}\right)$$

- **Composite SE + LP Kernel**
  One good thing about using kernel function is its flexibility in combining various kernel components, which allows multiplications and/or additions over different kernels to capture different features of the data. In our experiments, we added up one SE kernel and one LP kernel to model local periodicity with trend.

$$k_{se+lp}(x,x') = \sigma^2 exp\left(-\frac{(x-x')^2}{2l_1^2}\right) + \\ \sigma^2 exp\left(-\frac{2sin^2(\pi|x-x'|/p)}{l_2^2}\right) exp\left(-\frac{(x-x')^2}{2l_2^2}\right)$$