Probabilistic Modeling of Human Movements for Intention Inference

Supplementary Technical Details

Dynamics-GP Posterior Predictive Distribution

The posterior predictive distribution of the dynamics GP evaluated at \((x_{t-1}, g)\) is given by

\[
p(x_t|x_{t-1}, g) = GP(m_X(x_{t-1}, g), \Sigma_X(x_{t-1}, g)),
\]

\[
m_X(x_{t-1}, g)_a = m_{prior}(x_{t-1}, g) + k_X((x_{t-1}, g), \bar{X})K_X^{-1}(y_a - m_{prior}(x_{t-1}, g)),
\]

\[
\Sigma_X(x_{t-1}, g) = \sigma_X^2 I, \quad \sigma_X^2 = k_X((x_{t-1}, g), (x_{t-1}, g)) - k_X((x_{t-1}, g), \bar{X})K_X^{-1}k_X(\bar{X}, (x_{t-1}, g)).
\]

In Eq. (2), we defined \(y_a\) to be the set of training targets in dimension \(a\), where \(a = 1, \ldots, d\). Moreover, \(\bar{X}\) are the training inputs \((x_{ij}, g_j)\), \(i = 1, \ldots, N, j = 1, \ldots, J\), and \(K_X\) is the kernel matrix with entries \(K_X(i, j) = k_X([x_i, g_i], [x_j, g_j])\), \(i, j = 1, \ldots, M\).

Cross-Covariance

**Proposition 1** The cross-covariance \(\Sigma_{x_{t-1}, x_t | T} = \text{cov}_{x_{t-1}, x_t} | x_t, Z\) is given as

\[
\Sigma_{x_{t-1}, x_t | T} = J_{t-1} \Sigma_{x_{t-1}, x_t | T}^{-1}
\]

for \(t = 1, \ldots, T\), where \(J_{t-1}\) is defined as

\[
J_{t-1} = \Sigma_{x_{t-1}, x_t | t-1} \Sigma_{x_{t-1}, x_t | t-1}^{-1}.
\]

**Proof 1** We start the proof by writing out the definition of the desired cross-covariance

\[
\Sigma_{x_{t-1}, x_t | T} = \text{cov}_{x_{t-1}, x_t} | x_t, Z = E_{x_{t-1}, x_t} | x_t, \Sigma_{x_{t-1}, x_t} | x_t, Z - E_{x_{t-1}} | x_t, Z E_{x_t} | x_t, Z^\top
\]

\[
= \int \int x_{t-1}p(x_{t-1} | x_t, Z)x_t^\top p(x_t | Z) dx_{t-1} dx_t - \mu_{t-1}^x | t \mu_t^x | t^\top.
\]

Here, we plugged in the means \(\mu_{t-1}^x | T, \mu_t^x | T\) of the marginal smoothing distributions \(p(x_{t-1} | Z)\) and \(p(x_t | Z)\), respectively, and factored \(p(x_{t-1}, x_t | Z) = p(x_{t-1} | x_t, Z)p(x_t | Z)\). The inner integral in Eq. (7) determines the mean of the conditional distribution \(p(x_{t-1} | x_t, Z)\), which is given
by
\[ \mathbb{E}_{x_{t-1}}[x_{t-1}|x_t, Z] = \mu_{t-1|T}^x + J_{t-1}(x_t - \mu_{t|T}^x). \quad (8) \]

The matrix \( J_{t-1} \) is defined in Eq. (5).

**Remark 1** In Eq. (8), we conditioned on all measurements, not only on \( z_1, \ldots, z_{t-1} \), and computed an updated state distribution \( p(x_t|Z) \).

Using Eq. (8) for the inner integral in Eq. (7), the desired cross-covariance is
\[ \Sigma_{t-1,t|T}^x = \int \mathbb{E}_{x_{t-1}}[x_{t-1}|x_t, Z|x_t^T p(x_t|Z) dx_t - \mu_{t-1|T}^x (\mu_{t|T}^x)^T \]
\[ = \int (\mu_{t-1|T}^x + J_{t-1}(x_t - \mu_{t|T}^x))x_t^T p(x_t|z_{1:T}) dx_t - \mu_{t-1|T}^x (\mu_{t|T}^x)^T. \quad (10) \]

The smoothing results for the marginals, yield \( \mathbb{E}_{x_t}[x_t|Z] = \mu_{t|T}^x \) and \( \text{cov}_{x_t}[x_t|Z] \), respectively. After factorizing Eq. (10), we subsequently plug these moments in and obtain
\[ \Sigma_{t-1,t|T}^x = \mu_{t-1|T}^x (\mu_{t|T}^x)^T + J_{t-1}(\Sigma_{t|T}^x + \mu_{t|T}^x (\mu_{t|T}^x)^T - \mu_{t-1|T}^x (\mu_{t|T}^x)^T) - \mu_{t-1|T}^x (\mu_{t|T}^x)^T \]
\[ = J_{t-1} \Sigma_{t|T}, \quad (11) \]
which concludes the proof of Proposition 1.

**Predictions with a Linear Kernel**

In the following, we briefly go through the computations to compute the mean and the variance of the predictive distribution when we map a Gaussian input distribution through a GP with a linear kernel.

**Predictive Mean**

\[ \mathbb{E}[h(x_t)] = \mathbb{E}_{x_t} \left[ \mathbb{E}_h[h(x_t)|x_t] \right] = \mathbb{E}_{x_t} \left[ m(x_t) | x_t \right], \quad (13) \]

where the inner (conditional) expectation corresponds to the GP posterior mean function evaluated at \( x_t \).

For the \( a \)th predictive dimension, the posterior mean is given by
\[ m(x_t) = k_Z(x_t, X)K_Z^{-1}y_a \quad (14) \]

If we now use the linear kernel
\[ k_Z(x_i, x_j) = \beta_1 x_i^T x_j + \beta_2^{-1} \delta_{ij} \quad (15) \]
and plug it into Eqs. (14) and (13), we obtain for the predictive mean

\[ E[h_a(x_t)] = \int m(x_t)p(x_t) \, dx_t = \int k_Z(x_t, x)K_Z^{-1}y_a p(x_t) \, dx_t, \tag{16} \]

where \( a = 1, \ldots, D \). We define \( \gamma_a := K_Z^{-1}y_a \). Since \( \gamma_a \) is independent of \( x_t \), we pull it out of the integral and obtain

\[ E[h_a(x_t)] = \int m(x_t)p(x_t) \, dx_t = \int k_Z(x_t, x)p(x_t) \, dx_t \gamma_a \tag{17} \]

\[ = \int \beta_1x_t^\top X^\top p(x_t) \, dx_t \gamma_a \tag{18} \]

\[ = \beta_1 \int x_t^\top p(x_t) \, dx_t X^\top \gamma_a \tag{19} \]

\[ = \beta_1(\mu_{t|t-1})^\top X^\top \gamma_a \tag{20} \]

\[ = q^\top \gamma_a. \tag{21} \]

Here, we used the fact that \( p(x_t) \) is approximated by the Gaussian \( \mathcal{N}(x_t \mid \mu_{t|t-1}, \Sigma_{t|t-1}) \). Moreover, we defined \( q \) for the linear kernel \( k_Z \) in Eq. (15) as

\[ q^\top = \int \beta_1x_t^\top X^\top p(x_t) \, dx_t \tag{22} \]

\[ = \beta_1(\mu_{t|t-1})^\top X^\top \tag{23} \]

**Predictive Variance**

In the beginning, we focus on the univariate case, i.e., we only predict a single target dimension \( (a = 1) \). Using the law of total variance, the predictive variance is given as

\[ \sigma^2_Z = E_{x_t}[\text{var}_h[h(x_t) \mid x_t]] + \text{var}_{x_t}[E_h[h(x_t) \mid x_t]], \tag{24} \]

i.e., the predictive variance is a sum of the expected (conditional) variance and the variance of the (conditional) expectation. In Eq. (24), we identify the posterior GP variance and the posterior GP mean as the conditional variance and mean, respectively. Therefore, Eq. (24) can be rewritten as

\[ \sigma^2_Z = \int k_Z(x_t, x_t)p(x_t) \, dx_t - \int k_Z(x_t, x)K_Z^{-1}k_Z(x, x_t)p(x_t) \, dx_t \\
+ \gamma_a^\top \int k_Z(x, x_t)k_Z(x_t, x)p(x_t) \, dx_t \gamma_a - (q^\top \gamma_a)^2 \tag{25} \]

\[ = \int k_Z(x_t, x_t)p(x_t) \, dx_t - \text{tr}(K_Z^{-1} \int k_Z(x, x_t)k_Z(x_t, x)p(x_t) \, dx_t) \\
+ \gamma_a^\top \int k_Z(x, x_t)k_Z(x_t, x)p(x_t) \, dx_t \gamma_a - (q^\top \gamma_a)^2. \tag{26} \]
We now define
\[ Q = \int k_Z(X, x_t)k_Z(x_t, X)p(x_t) \, dx_t, \quad (27) \]

and obtain
\[ \sigma_Z^2 = \int k_Z(x_t, x_t)p(x_t) \, dx_t - \text{tr}(K_Z^{-1}Q) + \gamma_a^\top Q\gamma_a - \gamma_a^\top qq^\top \gamma_a \quad (28) \]

For the linear kernel in Eq. (15) the \( Q \)-matrix in Eq. (27) is given as
\[ Q = \int \beta_1^2 X x_t^\top X^\top p(x_t) \, dx_t = \beta_1^2 X (\Sigma^x_{\ell|t-1} + \mu^x_{\ell|t-1}(\mu^x_{\ell|t-1})^\top) X^\top. \quad (29) \]

Hence, for the linear kernel, the predictive variance is given by
\[ \sigma_Z^2 = \beta_1(\Sigma^x_{\ell|t-1} + \mu^x_{\ell|t-1}(\mu^x_{\ell|t-1})^\top) - \text{tr}(K_Z^{-1}Q) + \gamma_a^\top (Q - qq^\top) \gamma_a \quad (30) \]

For multivariate predictions, the entries \( \Sigma_Z(a, b), a, b = 1, \ldots, D \), of the predictive covariance matrix \( \Sigma_Z \) are given by

\[ \Sigma_Z(a, b) = \begin{cases} 
\gamma_a^\top (Q - qq^\top) \gamma_a, & a \neq b \\
\beta_1(\Sigma^x_{\ell|t-1} + \mu^x_{\ell|t-1}(\mu^x_{\ell|t-1})^\top) - \text{tr}(K_Z^{-1}Q) + \gamma_a^\top (Q - qq^\top) \gamma_a, & a = b.
\end{cases} \quad (32) \]

If \( a = b \), we have to include the term \( E_{x_t}[\text{cov}_h[h_a(x_t), h_b(x_t)|x_t]] = \beta_1(\Sigma^x_{\ell|t-1} + \mu^x_{\ell|t-1}(\mu^x_{\ell|t-1})^\top) - \text{tr}(K_Z^{-1}Q), \) which equals zero for \( a \neq b \) due to the assumption that the target dimensions \( a \) and \( b \) are conditionally independent given the input.