Composition of Movement Primitives

Andrea Pierré

June 9, 2025

Contents

1 ProMPs

1.1 Recap

From (Paraschos et al., 2013, 2018):

- q_t: joint angle over time
- $\dot{q}_t$: joint velocity over time
- $\boldsymbol{\tau} = \{q_t\}_{t=0\ldots T}$: trajectory
- $\boldsymbol{w}$: weight vector of a single trajectory $[n \times 1]$
- ϕ_t: basis function
- n: number of basis functions
- $\boldsymbol{\Phi}_t = [\phi_t, \dot{\phi}_t]$: $n \times 2$ dimensional time-dependent basis matrix
- $z(t)$: monotonically increasing phase variable
- $\boldsymbol{\epsilon}_y \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_y)$: zero-mean i.i.d. Gaussian noise

$$\boldsymbol{\Phi}_t = \begin{bmatrix} \phi_1 & \dot{\phi}_1 \\ \vdots & \vdots \\ \phi_n & \dot{\phi}_n \end{bmatrix} \tag{1}$$

$$\boldsymbol{y}_t = \begin{bmatrix} q_t \\ \dot{q}_t \end{bmatrix} = \boldsymbol{\Phi}_t^\top \boldsymbol{w} + \boldsymbol{\epsilon}_y \tag{2}$$

$$p(\boldsymbol{\tau}|\boldsymbol{w}) = \prod_t \mathcal{N}\Big(\boldsymbol{y}_t|\boldsymbol{\Phi}_t^\top \boldsymbol{w}, \boldsymbol{\Sigma}_y\Big) \tag{3}$$

$$p(\boldsymbol{\tau};\boldsymbol{\theta}) = \int p(\boldsymbol{\tau}|\boldsymbol{w}) \cdot p(\boldsymbol{w};\boldsymbol{\theta}) d\boldsymbol{w} \tag{4}$$

1.2 Coupling between joints

$$p(\boldsymbol{y}_t|\boldsymbol{w}) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{y}_{1,t} \\ \vdots \\ \boldsymbol{y}_{d,t} \end{bmatrix} \middle| \begin{bmatrix} \boldsymbol{\Phi}_t^\top & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{\Phi}_t^\top \end{bmatrix} \boldsymbol{w}, \boldsymbol{\Sigma}_y \right) = \mathcal{N}\Big(\boldsymbol{y}_t|\boldsymbol{\Psi}_t \boldsymbol{w}, \boldsymbol{\Sigma}_y\Big) \tag{5}$$

with:

- $\boldsymbol{w} = [\boldsymbol{w}_1^\top, \ldots, \boldsymbol{w}_n^\top]^\top$: combined weight vector $[n \times n]$
- $\boldsymbol{\Psi}_t$: block-diagonal basis matrix containing the basis functions and their derivatives for each dimension
- $\boldsymbol{y}_{i,t} = [q_{i,t}, \dot{q}_{i,t}]^\top$: joint angle and velocity for the i^{th} joint

1.3 Hierarchical Bayesian Model

The Hierarchical Bayesian Model used in ProMPs is illustrated in Fig. 1.

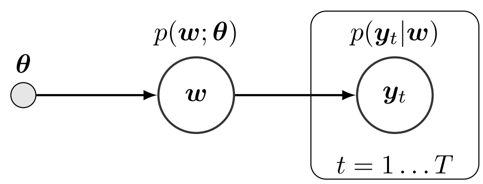

Figure 1: Hierarchical Bayesian Model used in ProMPs.

- $\boldsymbol{\theta} = \{\boldsymbol{\mu}_w, \boldsymbol{\Sigma}_w\}$
- $p(\boldsymbol{w};\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{w}|\boldsymbol{\mu}_w, \boldsymbol{\Sigma}_w)$: prior over the weight vector $\boldsymbol{w}$, with parameters $\boldsymbol{\theta}$, assumed to be Gaussian

$$\begin{aligned} p(\boldsymbol{y}_t;\boldsymbol{\theta}) &= \int \mathcal{N}\Big(\boldsymbol{y}_t|\boldsymbol{\Psi}_t^\top \boldsymbol{w}, \boldsymbol{\Sigma}_y\Big) \cdot p(\boldsymbol{w};\boldsymbol{\theta})\, d\boldsymbol{w} && (6) \\ &= \int \mathcal{N}\Big(\boldsymbol{y}_t|\boldsymbol{\Psi}_t^\top \boldsymbol{w}, \boldsymbol{\Sigma}_y\Big) \cdot \mathcal{N}\Big(\boldsymbol{w}|\boldsymbol{\mu}_{\boldsymbol{w}}, \boldsymbol{\Sigma}_{\boldsymbol{w}}\Big)\, d\boldsymbol{w} && (7) \\ &= \mathcal{N}\Big(\boldsymbol{y}_t|\boldsymbol{\Psi}_t^\top \boldsymbol{\mu}_{\boldsymbol{w}}, \boldsymbol{\Psi}_t^\top \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi}_t + \boldsymbol{\Sigma}_y\Big) && (8) \end{aligned}$$

See Appendix A for the proof.

1.4 Via-Points Modulation

- $\boldsymbol{x}_t^\star = [\boldsymbol{y}_t^\star, \boldsymbol{\Sigma}_t^\star]$: desired observation
- $\boldsymbol{y}_t^\star$: desired position and velocity vector at time t
- $\boldsymbol{\Sigma}_t^\star$: accuracy of the desired observation

Using Bayes rule:

$$p(\boldsymbol{w}|\boldsymbol{x}_t^\star) = \frac{p(\boldsymbol{x}_t^\star|\boldsymbol{w}) \cdot p(\boldsymbol{w})}{p(\boldsymbol{x}_t^\star)} \tag{9}$$

$$p(\boldsymbol{w}|\boldsymbol{x}_t^\star) \propto \mathcal{N}\left(\boldsymbol{y}_t^\star|\boldsymbol{\Psi}_t^\top \boldsymbol{w}, \boldsymbol{\Sigma}_t^\star\right) \cdot \mathcal{N}(\boldsymbol{w}|\boldsymbol{\mu}_w, \boldsymbol{\Sigma}_w) \tag{10}$$

$$\boldsymbol{\mu}_{\boldsymbol{w}}^{[new]} = \boldsymbol{\mu}_{\boldsymbol{w}} + \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi}_t \left(\boldsymbol{\Sigma}_y^\star + \boldsymbol{\Psi}_t^\top \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi}_t\right)^{-1} (\boldsymbol{y}_t^\star - \boldsymbol{\Psi}_t^\top \boldsymbol{\mu}_{\boldsymbol{w}}) \tag{11}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{w}}^{[new]} = \boldsymbol{\Sigma}_{\boldsymbol{w}} - \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi}_t \left(\boldsymbol{\Sigma}_y^\star + \boldsymbol{\Psi}_t^\top \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi}_t\right)^{-1} \boldsymbol{\Psi}_t^\top \boldsymbol{\Sigma}_{\boldsymbol{w}} \tag{12}$$

See Appendix B for the proof.

1.4.1 Do we actually get the desired mean by applying the conditioning update?

Proof that the posterior mean equals the observed mean.

$$\mathbb{E}[\boldsymbol{y}_t|\boldsymbol{x}_t^\star] = \boldsymbol{\mu}_{\boldsymbol{y}_t|\boldsymbol{x}_t^\star} = \boldsymbol{\Psi}_t^\top \boldsymbol{\mu}_{\boldsymbol{w}|\boldsymbol{x}_t^\star} = \boldsymbol{\Psi}_t^\top \boldsymbol{\mu}_{\boldsymbol{w}} + \boldsymbol{\Psi}_t^\top \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi}_t \left(\boldsymbol{\Sigma}_t^\star + \boldsymbol{\Psi}_t^\top \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi}_t\right)^{-1} (\boldsymbol{y}_t^\star - \boldsymbol{\Psi}_t^\top \boldsymbol{\mu}_{\boldsymbol{w}}) \tag{13}$$

We set the observed covariance $\boldsymbol{\Sigma}_t^\star$ to 0 so as to have perfect accuracy around our observed position.

$$\begin{aligned} \boldsymbol{\Psi}_t^\top \boldsymbol{\mu}_{\boldsymbol{w}|\boldsymbol{x}_t^\star} &= \boldsymbol{\Psi}_t^\top \boldsymbol{\mu}_{\boldsymbol{w}} + \cancel{\boldsymbol{\Psi}_t^\top \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi}_t} \left(\cancel{\boldsymbol{\Psi}_t^\top \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi}_t}\right)^{-1} (\boldsymbol{y}_t^\star - \boldsymbol{\Psi}_t^\top \boldsymbol{\mu}_{\boldsymbol{w}}) && (14) \\ &= \cancel{\boldsymbol{\Psi}_t^\top \boldsymbol{\mu}_{\boldsymbol{w}}} + \boldsymbol{y}_t^\star - \cancel{\boldsymbol{\Psi}_t^\top \boldsymbol{\mu}_{\boldsymbol{w}}} && (15) \\ &= \boldsymbol{y}_t^\star && (16) \end{aligned}$$

□

1.4.2 Multi via-points

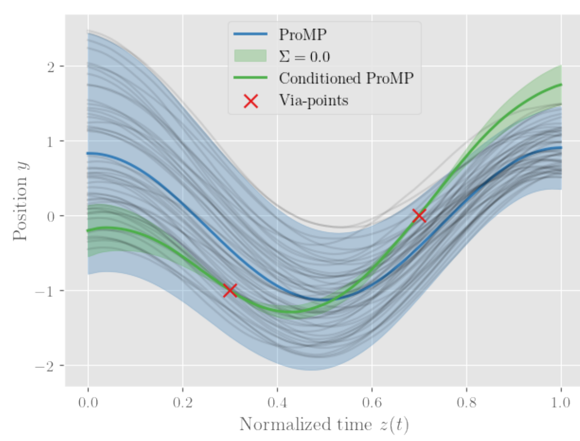

Figure 2: Example of ProMP with two via-points.

1. For the first via-point conditioning update with the observed via-point $\boldsymbol{x}_{t_1}^\star = [\boldsymbol{y}_{t_1}^\star, \boldsymbol{\Sigma}_{t_1}^\star]$, we can directly apply Eq. (11) and (12), with $\boldsymbol{\Psi}_{t_1}$ the observation matrix at time t_1:

$$\boldsymbol{\mu}_{\boldsymbol{w}|\boldsymbol{x}_{t_1}^\star} = \boldsymbol{\mu}_{\boldsymbol{w}} + \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi}_{t_1} \left(\boldsymbol{\Sigma}_{t_1}^\star + \boldsymbol{\Psi}_{t_1}^\top \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi}_{t_1}\right)^{-1} (\boldsymbol{y}_{t_1}^\star - \boldsymbol{\Psi}_{t_1}^\top \boldsymbol{\mu}_{\boldsymbol{w}}) \tag{17}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{w}|\boldsymbol{x}_{t_1}^\star} = \boldsymbol{\Sigma}_{\boldsymbol{w}} - \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi}_{t_1} \left(\boldsymbol{\Sigma}_{t_1}^\star + \boldsymbol{\Psi}_{t_1}^\top \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi}_{t_1}\right)^{-1} \boldsymbol{\Psi}_{t_1}^\top \boldsymbol{\Sigma}_{\boldsymbol{w}} \tag{18}$$

2. For the second via-point conditioning update with the observed via-point $\boldsymbol{x}^\star_{t_2} = [\boldsymbol{y}^\star_{t_2}, \boldsymbol{\Sigma}^\star_{t_2}]$, the prior is the posterior from the first via-point, *i.e.*, $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1}}, \boldsymbol{\Sigma}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1}})$, the likelihood is $\boldsymbol{y}^\star_{t_2} \sim \mathcal{N}(\boldsymbol{\Psi}_{t_2}^\top \boldsymbol{w}, \boldsymbol{\Sigma}^\star_{t_2})$, with $\boldsymbol{\Psi}_{t_2}$ the observation matrix at time t_2, and the posterior update becomes:

$$\boldsymbol{\mu}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1},\boldsymbol{x}^\star_{t_2}} = \boldsymbol{\mu}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1}} + \boldsymbol{\Sigma}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1}} \boldsymbol{\Psi}_{t_2} \Big(\boldsymbol{\Sigma}^\star_{t_2} + \boldsymbol{\Psi}_{t_2}^\top \boldsymbol{\Sigma}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1}} \boldsymbol{\Psi}_{t_2}\Big)^{-1} (\boldsymbol{y}^\star_{t_2} - \boldsymbol{\Psi}_{t_2}^\top \boldsymbol{\mu}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1}}) \tag{19}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1},\boldsymbol{x}^\star_{t_2}} = \boldsymbol{\Sigma}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1}} - \boldsymbol{\Sigma}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1}} \boldsymbol{\Psi}_{t_2} \Big(\boldsymbol{\Sigma}^\star_{t_2} + \boldsymbol{\Psi}_{t_2}^\top \boldsymbol{\Sigma}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1}} \boldsymbol{\Psi}_{t_2}\Big)^{-1} \boldsymbol{\Psi}_{t_2}^\top \boldsymbol{\Sigma}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1}} \tag{20}$$

3. For the k^{th} via-point conditioning update with the observed via-point $\boldsymbol{x}^\star_{t_k} = [\boldsymbol{y}^\star_{t_k}, \boldsymbol{\Sigma}^\star_{t_k}]$, the prior is the posterior after conditioning on the previous $k-1$ via-points, *i.e.*, $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1},\ldots,\boldsymbol{x}^\star_{t_{k-1}}}, \boldsymbol{\Sigma}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1},\ldots,\boldsymbol{x}^\star_{t_{k-1}}})$, the likelihood is $\boldsymbol{y}^\star_{t_k} \sim \mathcal{N}(\boldsymbol{\Psi}_{t_k}^\top \boldsymbol{w}, \boldsymbol{\Sigma}^\star_{t_k})$, with $\boldsymbol{\Psi}_{t_k}$ the observation matrix at time t_k, and the posterior update becomes:

$$\begin{aligned}\boldsymbol{\mu}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1},\ldots,\boldsymbol{x}^\star_{t_k}} &= \boldsymbol{\mu}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1},\ldots,\boldsymbol{x}^\star_{t_{k-1}}} \\ &\quad + \boldsymbol{\Sigma}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1},\ldots,\boldsymbol{x}^\star_{t_{k-1}}} \boldsymbol{\Psi}_{t_k} \Big(\boldsymbol{\Sigma}^\star_{t_k} + \boldsymbol{\Psi}_{t_k}^\top \boldsymbol{\Sigma}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1},\ldots,\boldsymbol{x}^\star_{t_{k-1}}} \boldsymbol{\Psi}_{t_k}\Big)^{-1} (\boldsymbol{y}^\star_{t_k} - \boldsymbol{\Psi}_{t_k}^\top \boldsymbol{\mu}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1},\ldots,\boldsymbol{x}^\star_{t_{k-1}}})\end{aligned} \tag{21}$$

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1},\ldots,\boldsymbol{x}^\star_{t_k}} &= \boldsymbol{\Sigma}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1},\ldots,\boldsymbol{x}^\star_{t_{k-1}}} \\ &\quad - \boldsymbol{\Sigma}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1},\ldots,\boldsymbol{x}^\star_{t_{k-1}}} \boldsymbol{\Psi}_{t_k} \Big(\boldsymbol{\Sigma}^\star_{t_k} + \boldsymbol{\Psi}_{t_k}^\top \boldsymbol{\Sigma}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1},\ldots,\boldsymbol{x}^\star_{t_{k-1}}} \boldsymbol{\Psi}_{t_k}\Big)^{-1} \boldsymbol{\Psi}_{t_k}^\top \boldsymbol{\Sigma}_{\boldsymbol{w}|\boldsymbol{x}^\star_{t_1},\ldots,\boldsymbol{x}^\star_{t_{k-1}}}\end{aligned} \tag{22}$$

Alternative Batch Formulation Instead of iterative updates, we could condition on all via-points simultaneously by stacking the observations:

$$\boldsymbol{y}^\star = \begin{bmatrix}\boldsymbol{y}^\star_{t_1} \\ \vdots \\ \boldsymbol{y}^\star_{t_k}\end{bmatrix}, \quad \boldsymbol{\Psi} = \begin{bmatrix}\boldsymbol{\Psi}_{t_1} \\ \vdots \\ \boldsymbol{\Psi}_{t_k}\end{bmatrix}, \quad \boldsymbol{\Sigma}^\star = \mathrm{diag}(\boldsymbol{\Sigma}^\star_{t_1}, \ldots, \boldsymbol{\Sigma}^\star_{t_k}) \tag{23}$$

$$\boldsymbol{\mu}_{\boldsymbol{w}|\{\boldsymbol{x}^\star_{t_k}\}_{k=1}^K} = \boldsymbol{\mu}_{\boldsymbol{w}} + \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi} \Big(\boldsymbol{\Sigma}^\star + \boldsymbol{\Psi}^\top \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi}\Big)^{-1} (\boldsymbol{y}^\star - \boldsymbol{\Psi}^\top \boldsymbol{\mu}_{\boldsymbol{w}}) \tag{24}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{w}|\{\boldsymbol{x}^\star_{t_k}\}_{k=1}^K} = \boldsymbol{\Sigma}_{\boldsymbol{w}} - \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi} \Big(\boldsymbol{\Sigma}^\star + \boldsymbol{\Psi}^\top \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi}\Big)^{-1} \boldsymbol{\Psi}^\top \boldsymbol{\Sigma}_{\boldsymbol{w}} \tag{25}$$

2 Gaussian mixture modeling (GMM)/Gaussian mixture regression (GMR) recap

2.1 Gaussian mixture modeling (GMM)

- π_k: mixture weights
- $\boldsymbol{\theta} := \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \pi_k : k = 1, \ldots, K\}$: collection of all parameters of the model
- r_{nk}: responsibility of the k^{th} mixture component for the n^{th} data point
- N: number of data points
- $N_k := \sum_{n=1}^N r_{nk}$: total responsibility of the k^{th} mixture component for the entire dataset

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \sum_{k=1}^K \pi_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{26}$$

$$0 \le \pi_k \le 1, \quad \sum_{k=1}^K \pi_k = 1 \tag{27}$$

$$r_{nk} := \frac{\pi_k \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \tag{28}$$

Update of the GMM means:

$$\boldsymbol{\mu}_k^{new} = \frac{\sum_{n=1}^{N} r_{nk}\boldsymbol{x}_n}{\sum_{n=1}^{N} r_{nk}} \tag{29}$$

Update of the GMM covariances:

$$\boldsymbol{\Sigma}_k^{new} = \frac{1}{N_k}\sum_{n=1}^{N} r_{nk}(\boldsymbol{x}_n - \boldsymbol{\mu}_k)(\boldsymbol{x}_n - \boldsymbol{\mu}_k)^\top \tag{30}$$

Update of the GMM mixture weights:

$$\pi_k^{new} = \frac{N_k}{N}, \quad k = 1, \ldots, K \tag{31}$$

Algorithm 1: Expectation Maximization (EM) algorithm

1. Initialize $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$, π_k
2. *E-step*: Evaluate responsibilities r_{nk} for every data point $\boldsymbol{x}_n$ using current parameters $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$, π_k:

$$r_{nk} = \frac{\pi_k \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \tag{32}$$

3. *M-step*: Re-estimate parameters $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$, π_k using the current responsibilities r_{nk} (from E-step):

$$\boldsymbol{\mu}_k = \frac{1}{N_k}\sum_{n=1}^{N} r_{nk}\boldsymbol{x}_n \tag{33}$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k}\sum_{n=1}^{N} r_{nk}(\boldsymbol{x}_n - \boldsymbol{\mu}_k)(\boldsymbol{x}_n - \boldsymbol{\mu}_k)^\top \tag{34}$$

$$\pi_k = \frac{N_k}{N} \tag{35}$$

3 Composition of MPs

3.1 Stitching

See Fig. 3.

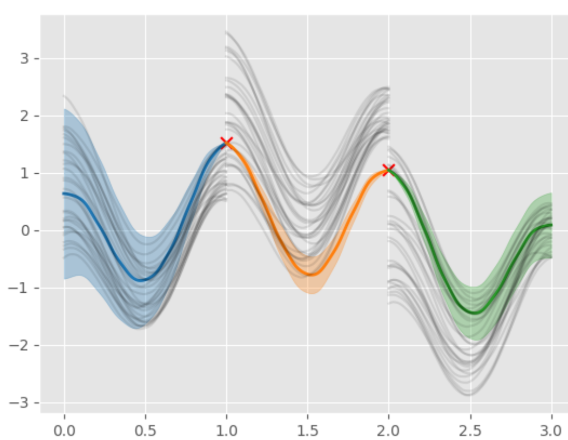

Figure 3: Stitching three ProMPs.

3.2 Piecewise Gaussian Process

References

A. Paraschos, C. Daniel, J. R. Peters, and G. Neumann, "Probabilistic Movement Primitives," in *Advances in Neural Information Processing Systems*, vol. 26. Curran Associates, Inc., 2013. [Online]. Available: https://proceedings.neurips.cc/paper/2013/hash/e53a0a2978c28872a4505bdb51db06dc-Abstract.html

A. Paraschos, C. Daniel, J. Peters, and G. Neumann, "Using probabilistic movement primitives in robotics," *Autonomous Robots*, vol. 42, no. 3, pp. 529–551, Mar. 2018. [Online]. Available: https://doi.org/10.1007/s10514-017-9648-7

M. P. Deisenroth, A. A. Faisal, and C. S. Ong, *Mathematics for machine learning.* Cambridge University Press, 2020.

C. M. Bishop and H. Bishop, *Deep Learning: Foundations and Concepts.* Springer International Publishing, 2024. [Online]. Available: https://doi.org/10.1007/978-3-031-45468-4

A Hierarchical Bayesian Model proof

Proof of Eq. (8). From (Deisenroth et al., 2020), we have the joint distribution:

$$p(\mathbf{x}_a, \mathbf{x}_b) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix} \right) \tag{36}$$

and the marginal distribution $p(\mathbf{x}_a)$ of a joint Gaussian distribution $p(\mathbf{x}_a, \mathbf{x}_b)$:

$$p(\mathbf{x}_a) = \int p(\mathbf{x}_a, \mathbf{x}_b) d\mathbf{x}_b = \mathcal{N}(\mathbf{x}_a | \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_{aa}) \tag{37}$$

Since $\boldsymbol{y}_t$ and $\boldsymbol{w}$ are jointly Gaussian, we have:

$$\begin{bmatrix} \boldsymbol{y}_t \\ \boldsymbol{w} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{y}_t} \\ \boldsymbol{\mu}_{\boldsymbol{w}} \end{bmatrix}, \begin{bmatrix} \mathrm{Cov}[\boldsymbol{y}_t, \boldsymbol{y}_t] & \mathrm{Cov}[\boldsymbol{y}_t, \boldsymbol{w}] \\ \mathrm{Cov}[\boldsymbol{w}, \boldsymbol{y}_t] & \mathrm{Cov}[\boldsymbol{w}, \boldsymbol{w}] \end{bmatrix} \right) \tag{38}$$

$$\begin{aligned}
\boldsymbol{\mu}_{\boldsymbol{y}_t} &= \mathbb{E}[\boldsymbol{y}_t] && (39)\\
&= \mathbb{E}[\boldsymbol{\Psi}_t^\top \boldsymbol{w} + \boldsymbol{\epsilon}_y] && (40)\\
&= \boldsymbol{\Psi}_t^\top \, \mathbb{E}[\boldsymbol{w}] + \mathbb{E}[\boldsymbol{\epsilon}_y] && (41)\\
&= \boldsymbol{\Psi}_t^\top \boldsymbol{\mu}_{\boldsymbol{w}} + 0 && (42)\\
&= \boldsymbol{\Psi}_t^\top \boldsymbol{\mu}_{\boldsymbol{w}} && (43)
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}[\boldsymbol{y}_t, \boldsymbol{y}_t] &= \mathrm{Cov}[\boldsymbol{\Psi}_t^\top \boldsymbol{w} + \boldsymbol{\epsilon}_y] && (44)\\
&= \mathrm{Cov}[\boldsymbol{\Psi}_t^\top \boldsymbol{w}] + \mathrm{Cov}[\boldsymbol{\epsilon}_y] && (45)\\
&= \boldsymbol{\Psi}_t^\top \mathrm{Cov}[\boldsymbol{w}] \boldsymbol{\Psi}_t + \boldsymbol{\Sigma}_y && (46)\\
&= \boldsymbol{\Psi}_t^\top \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi}_t + \boldsymbol{\Sigma}_y && (47)
\end{aligned}$$

$$\begin{bmatrix} \boldsymbol{y}_t \\ \boldsymbol{w} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\Psi}_t^\top \boldsymbol{\mu}_{\boldsymbol{w}} \\ \boldsymbol{\mu}_{\boldsymbol{w}} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Psi}_t^\top \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi}_t + \boldsymbol{\Sigma}_y & \boldsymbol{\Psi}_t^\top \boldsymbol{\Sigma}_{\boldsymbol{w}} \\ \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi}_t & \boldsymbol{\Sigma}_{\boldsymbol{w}} \end{bmatrix} \right) \tag{48}$$

$$p(\boldsymbol{y}_t; \boldsymbol{\theta}) = \mathcal{N}\Big(\boldsymbol{y}_t | \boldsymbol{\Psi}_t^\top \boldsymbol{\mu}_{\boldsymbol{w}}, \boldsymbol{\Psi}_t^\top \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{\Psi}_t + \boldsymbol{\Sigma}_y \Big) \tag{49}$$

□

B Via-Points conditioning proof

Proof of Eq. (11) and Eq. (12). With the joint distribution $p(\mathbf{x}_a, \mathbf{x}_b)$ in Eq. (36), and from (Bishop and Bishop, 2024), the parameters of a conditional multivariate Gaussian $p(\mathbf{x}_a|\mathbf{x}_b) = \mathcal{N}\big(\boldsymbol{\mu}_{a|b}, \boldsymbol{\Sigma}_{a|b}\big)$ are the following:

$$\boldsymbol{\mu}_{a|b} = \boldsymbol{\mu}_a + \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\mathbf{x}_b - \boldsymbol{\mu}_b) \tag{50}$$

$$\boldsymbol{\Sigma}_{a|b} = \boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba} \tag{51}$$

We want the posterior $p(\boldsymbol{w}|\boldsymbol{x}_t^\star)$, knowing the likelihood $\boldsymbol{x}_t^\star|\boldsymbol{w} \sim \mathcal{N}\Big(\boldsymbol{y}_t^\star|\boldsymbol{\Psi}_t^\top \boldsymbol{w}, \boldsymbol{\Sigma}_t^\star\Big)$, and the prior $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{w}|\boldsymbol{\mu}_w, \boldsymbol{\Sigma}_w)$.

$$\begin{bmatrix}\boldsymbol{w}\\ \boldsymbol{x}_t^\star\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\boldsymbol{\mu}_{\boldsymbol{w}}\\ \boldsymbol{\Psi}_t^\top\boldsymbol{\mu}_{\boldsymbol{w}}\end{bmatrix}, \begin{bmatrix}\mathrm{Cov}[\boldsymbol{w},\boldsymbol{w}] & \mathrm{Cov}[\boldsymbol{w},\boldsymbol{x}_t^\star]\\ \mathrm{Cov}[\boldsymbol{x}_t^\star,\boldsymbol{w}] & \mathrm{Cov}[\boldsymbol{x}_t^\star,\boldsymbol{x}_t^\star]\end{bmatrix}\right) \tag{52}$$

$\mathrm{Cov}[\boldsymbol{x}_t^\star, \boldsymbol{x}_t^\star]$ follows from Eq. (47).

$$\begin{aligned}
\mathrm{Cov}[\boldsymbol{w},\boldsymbol{x}_t^\star] &= \mathrm{Cov}[\boldsymbol{w}, \boldsymbol{\Psi}_t^\top\boldsymbol{w} + \boldsymbol{\epsilon}_y] && (53)\\
&= \mathrm{Cov}[\boldsymbol{w}, \boldsymbol{\Psi}_t^\top\boldsymbol{w}] \qquad (\mathrm{Cov}[\boldsymbol{w},\boldsymbol{\epsilon}_{\boldsymbol{y}}] = 0 \text{ since } \boldsymbol{\epsilon}_{\boldsymbol{y}} \text{ is independent of } \boldsymbol{w}) && (54)\\
&= \mathbb{E}[(\boldsymbol{w}-\boldsymbol{\mu}_{\boldsymbol{w}})(\boldsymbol{\Psi}_t^\top\boldsymbol{w} - \boldsymbol{\Psi}_t^\top\boldsymbol{\mu}_{\boldsymbol{w}})^\top] && (55)\\
&= \mathbb{E}[(\boldsymbol{w}-\boldsymbol{\mu}_{\boldsymbol{w}})(\boldsymbol{w}-\boldsymbol{\mu}_{\boldsymbol{w}})^\top\boldsymbol{\Psi}_t] && (56)\\
&= \mathrm{Cov}[\boldsymbol{w},\boldsymbol{w}]\cdot\boldsymbol{\Psi}_t && (57)\\
&= \boldsymbol{\Sigma}_{\boldsymbol{w}}\boldsymbol{\Psi}_t && (58)
\end{aligned}$$

$$\begin{bmatrix}\boldsymbol{w}\\ \boldsymbol{x}_t^\star\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\boldsymbol{\mu}_{\boldsymbol{w}}\\ \boldsymbol{\Psi}_t^\top\boldsymbol{\mu}_{\boldsymbol{w}}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{\boldsymbol{w}} & \boldsymbol{\Sigma}_{\boldsymbol{w}}\boldsymbol{\Psi}_t\\ \boldsymbol{\Psi}_t^\top\boldsymbol{\Sigma}_{\boldsymbol{w}} & \boldsymbol{\Psi}_t^\top\boldsymbol{\Sigma}_{\boldsymbol{w}}\boldsymbol{\Psi}_t + \boldsymbol{\Sigma}_t^\star\end{bmatrix}\right) \tag{59}$$

Using Eq. (50) we get:

$$\boldsymbol{\mu}_{\boldsymbol{w}|\boldsymbol{x}_t^\star} = \boldsymbol{\mu}_{\boldsymbol{w}} + \boldsymbol{\Sigma}_{\boldsymbol{w}}\boldsymbol{\Psi}_t\Big(\boldsymbol{\Sigma}_t^\star + \boldsymbol{\Psi}_t^\top\boldsymbol{\Sigma}_{\boldsymbol{w}}\boldsymbol{\Psi}_t\Big)^{-1}(\boldsymbol{y}_t^\star - \boldsymbol{\Psi}_t^\top\boldsymbol{\mu}_{\boldsymbol{w}}) \tag{60}$$

Using Eq. (51) we get:

$$\boldsymbol{\Sigma}_{\boldsymbol{w}|\boldsymbol{x}_t^\star} = \boldsymbol{\Sigma}_{\boldsymbol{w}} - \boldsymbol{\Sigma}_{\boldsymbol{w}}\boldsymbol{\Psi}_t\Big(\boldsymbol{\Sigma}_t^\star + \boldsymbol{\Psi}_t^\top\boldsymbol{\Sigma}_{\boldsymbol{w}}\boldsymbol{\Psi}_t\Big)^{-1}\boldsymbol{\Psi}_t^\top\boldsymbol{\Sigma}_{\boldsymbol{w}} \tag{61}$$

□