

Composition of Movement Primitives

Andrea Pierré

May 24, 2025

Contents

1 ProMPs	1
1.1 Recap	1
1.2 Coupling between joints	2
1.3 Hierarchical Bayesian Model	2
1.4 Via-Points Modulation	2
1.4.1 Do we actually get the desired mean by applying the conditioning update?	3
1.4.2 Multi via-points	3
2 Composition of MPs	4
2.1 Blending	4
2.2 Stitching	4
A Hierarchical Bayesian Model proof	5
B Via-Points conditioning proof	6

1 ProMPs

1.1 Recap

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

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

$$\Phi_t = \begin{bmatrix} \phi_1 & \dot{\phi}_1 \\ \vdots & \vdots \\ \phi_n & \dot{\phi}_n \end{bmatrix} \quad (1)$$

$$\mathbf{y}_t = \begin{bmatrix} q_t \\ \dot{q}_t \end{bmatrix} = \Phi_t^\top \mathbf{w} + \epsilon_y \quad (2)$$

$$p(\tau|\mathbf{w}) = \prod_t \mathcal{N}(\mathbf{y}_t | \Phi_t^\top \mathbf{w}, \Sigma_y) \quad (3)$$

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

1.2 Coupling between joints

$$p(\mathbf{y}_t|\mathbf{w}) = \mathcal{N}\left(\begin{bmatrix} \mathbf{y}_{1,t} \\ \vdots \\ \mathbf{y}_{d,t} \end{bmatrix} \middle| \begin{bmatrix} \Phi_t^\top & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \Phi_t^\top \end{bmatrix} \mathbf{w}, \Sigma_y\right) = \mathcal{N}(\mathbf{y}_t | \Psi_t \mathbf{w}, \Sigma_y) \quad (5)$$

with:

- $\mathbf{w} = [\mathbf{w}_1^\top, \dots, \mathbf{w}_n^\top]^\top$: combined weight vector $[n \times n]$
- Ψ_t : block-diagonal basis matrix containing the basis functions and their derivatives for each dimension
- $\mathbf{y}_{i,t} = [q_{i,t}, \dot{q}_{i,t}]^\top$: 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.

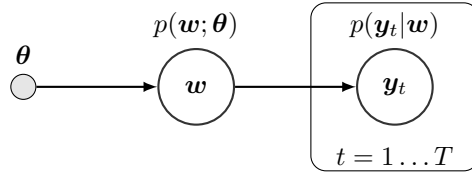


Figure 1: Hierarchical Bayesian Model used in ProMPs.

- $\theta = \{\mu_w, \Sigma_w\}$
- $p(\mathbf{w}; \theta) = \mathcal{N}(\mathbf{w} | \mu_w, \Sigma_w)$: prior over the weight vector \mathbf{w} , with parameters θ , assumed to be Gaussian

$$p(\mathbf{y}_t; \theta) = \int \mathcal{N}(\mathbf{y}_t | \Psi_t^\top \mathbf{w}, \Sigma_y) \cdot p(\mathbf{w}; \theta) d\mathbf{w} \quad (6)$$

$$= \int \mathcal{N}(\mathbf{y}_t | \Psi_t^\top \mathbf{w}, \Sigma_y) \cdot \mathcal{N}(\mathbf{w} | \mu_w, \Sigma_w) d\mathbf{w} \quad (7)$$

$$= \mathcal{N}(\mathbf{y}_t | \Psi_t^\top \mu_w, \Psi_t^\top \Sigma_w \Psi_t + \Sigma_y) \quad (8)$$

See Appendix A for the proof.

1.4 Via-Points Modulation

- $\mathbf{x}_t^* = [\mathbf{y}_t^*, \Sigma_t^*]$: desired observation
- \mathbf{y}_t^* : desired position and velocity vector at time t
- Σ_t^* : accuracy of the desired observation

Using Bayes rule:

$$p(\mathbf{w}|\mathbf{x}_t^*) = \frac{p(\mathbf{x}_t^*|\mathbf{w}) \cdot p(\mathbf{w})}{p(\mathbf{x}_t^*)} \quad (9)$$

$$p(\mathbf{w}|\mathbf{x}_t^*) \propto \mathcal{N}(\mathbf{y}_t^*|\Psi_t^\top \mathbf{w}, \Sigma_t^*) \cdot \mathcal{N}(\mathbf{w}|\boldsymbol{\mu}_w, \Sigma_w) \quad (10)$$

$$\boldsymbol{\mu}_w^{[new]} = \boldsymbol{\mu}_w + \Sigma_w \Psi_t \left(\Sigma_y^* + \Psi_t^\top \Sigma_w \Psi_t \right)^{-1} (\mathbf{y}_t^* - \Psi_t^\top \boldsymbol{\mu}_w) \quad (11)$$

$$\Sigma_w^{[new]} = \Sigma_w - \Sigma_w \Psi_t \left(\Sigma_y^* + \Psi_t^\top \Sigma_w \Psi_t \right)^{-1} \Psi_t^\top \Sigma_w \quad (12)$$

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.

$$\mathbb{E}[\mathbf{y}_t|\mathbf{x}_t^*] = \boldsymbol{\mu}_{\mathbf{y}_t|\mathbf{x}_t^*} = \Psi_t^\top \boldsymbol{\mu}_w|\mathbf{x}_t^* = \Psi_t^\top \boldsymbol{\mu}_w + \Psi_t^\top \Sigma_w \Psi_t \left(\Sigma_y^* + \Psi_t^\top \Sigma_w \Psi_t \right)^{-1} (\mathbf{y}_t^* - \Psi_t^\top \boldsymbol{\mu}_w) \quad (13)$$

We set the observed covariance Σ_t^* to 0 so as to have perfect accuracy around our observed position.

$$\Psi_t^\top \boldsymbol{\mu}_w|\mathbf{x}_t^* = \Psi_t^\top \boldsymbol{\mu}_w + \cancel{\Psi_t^\top \Sigma_w \Psi_t} \left(\cancel{\Sigma_y^* + \Psi_t^\top \Sigma_w \Psi_t} \right)^{-1} (\mathbf{y}_t^* - \Psi_t^\top \boldsymbol{\mu}_w) \quad (14)$$

$$= \cancel{\Psi_t^\top \boldsymbol{\mu}_w} + \mathbf{y}_t^* - \cancel{\Psi_t^\top \boldsymbol{\mu}_w} \quad (15)$$

$$= \mathbf{y}_t^* \quad (16)$$

□

1.4.2 Multi via-points

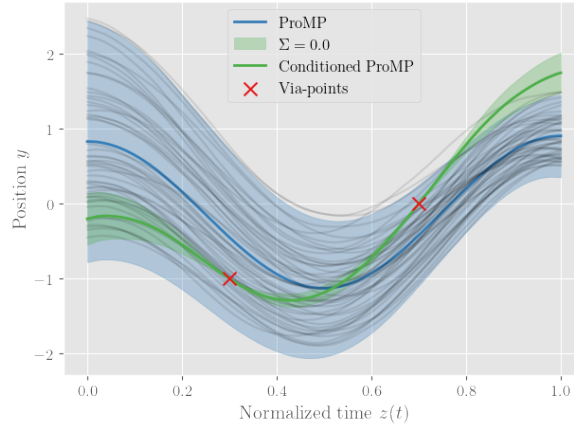


Figure 2: Example of ProMP with two via-points.

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

$$\boldsymbol{\mu}_w|\mathbf{x}_{t_1}^* = \boldsymbol{\mu}_w + \Sigma_w \Psi_{t_1} \left(\Sigma_{t_1}^* + \Psi_{t_1}^\top \Sigma_w \Psi_{t_1} \right)^{-1} (\mathbf{y}_{t_1}^* - \Psi_{t_1}^\top \boldsymbol{\mu}_w) \quad (17)$$

$$\Sigma_w|\mathbf{x}_{t_1}^* = \Sigma_w - \Sigma_w \Psi_{t_1} \left(\Sigma_{t_1}^* + \Psi_{t_1}^\top \Sigma_w \Psi_{t_1} \right)^{-1} \Psi_{t_1}^\top \Sigma_w \quad (18)$$

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

$$\boldsymbol{\mu}_{\mathbf{w}|\mathbf{x}_{t_1}^*, \mathbf{x}_{t_2}^*} = \boldsymbol{\mu}_{\mathbf{w}|\mathbf{x}_{t_1}^*} + \Sigma_{\mathbf{w}|\mathbf{x}_{t_1}^*} \Psi_{t_2} \left(\Sigma_{t_2}^* + \Psi_{t_2}^\top \Sigma_{\mathbf{w}|\mathbf{x}_{t_1}^*} \Psi_{t_2} \right)^{-1} (\mathbf{y}_{t_2}^* - \Psi_{t_2}^\top \boldsymbol{\mu}_{\mathbf{w}|\mathbf{x}_{t_1}^*}) \quad (19)$$

$$\Sigma_{\mathbf{w}|\mathbf{x}_{t_1}^*, \mathbf{x}_{t_2}^*} = \Sigma_{\mathbf{w}|\mathbf{x}_{t_1}^*} - \Sigma_{\mathbf{w}|\mathbf{x}_{t_1}^*} \Psi_{t_2} \left(\Sigma_{t_2}^* + \Psi_{t_2}^\top \Sigma_{\mathbf{w}|\mathbf{x}_{t_1}^*} \Psi_{t_2} \right)^{-1} \Psi_{t_2}^\top \Sigma_{\mathbf{w}|\mathbf{x}_{t_1}^*} \quad (20)$$

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

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{w}|\mathbf{x}_{t_1}^*, \dots, \mathbf{x}_{t_k}^*} &= \boldsymbol{\mu}_{\mathbf{w}|\mathbf{x}_{t_1}^*, \dots, \mathbf{x}_{t_{k-1}}^*} \\ &+ \Sigma_{\mathbf{w}|\mathbf{x}_{t_1}^*, \dots, \mathbf{x}_{t_{k-1}}^*} \Psi_{t_k} \left(\Sigma_{t_k}^* + \Psi_{t_k}^\top \Sigma_{\mathbf{w}|\mathbf{x}_{t_1}^*, \dots, \mathbf{x}_{t_{k-1}}^*} \Psi_{t_k} \right)^{-1} (\mathbf{y}_{t_k}^* - \Psi_{t_k}^\top \boldsymbol{\mu}_{\mathbf{w}|\mathbf{x}_{t_1}^*, \dots, \mathbf{x}_{t_{k-1}}^*}) \end{aligned} \quad (21)$$

$$\begin{aligned} \Sigma_{\mathbf{w}|\mathbf{x}_{t_1}^*, \dots, \mathbf{x}_{t_k}^*} &= \Sigma_{\mathbf{w}|\mathbf{x}_{t_1}^*, \dots, \mathbf{x}_{t_{k-1}}^*} \\ &- \Sigma_{\mathbf{w}|\mathbf{x}_{t_1}^*, \dots, \mathbf{x}_{t_{k-1}}^*} \Psi_{t_k} \left(\Sigma_{t_k}^* + \Psi_{t_k}^\top \Sigma_{\mathbf{w}|\mathbf{x}_{t_1}^*, \dots, \mathbf{x}_{t_{k-1}}^*} \Psi_{t_k} \right)^{-1} \Psi_{t_k}^\top \Sigma_{\mathbf{w}|\mathbf{x}_{t_1}^*, \dots, \mathbf{x}_{t_{k-1}}^*} \end{aligned} \quad (22)$$

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

$$\mathbf{y}^* = \begin{bmatrix} \mathbf{y}_{t_1}^* \\ \vdots \\ \mathbf{y}_{t_k}^* \end{bmatrix}, \quad \Psi = \begin{bmatrix} \Psi_{t_1} \\ \vdots \\ \Psi_{t_k} \end{bmatrix}, \quad \Sigma^* = \text{diag}(\Sigma_{t_1}^*, \dots, \Sigma_{t_k}^*) \quad (23)$$

$$\boldsymbol{\mu}_{\mathbf{w}|\{\mathbf{x}_{t_k}^*\}_{k=1}^K} = \boldsymbol{\mu}_{\mathbf{w}} + \Sigma_{\mathbf{w}} \Psi \left(\Sigma^* + \Psi^\top \Sigma_{\mathbf{w}} \Psi \right)^{-1} (\mathbf{y}^* - \Psi^\top \boldsymbol{\mu}_{\mathbf{w}}) \quad (24)$$

$$\Sigma_{\mathbf{w}|\{\mathbf{x}_{t_k}^*\}_{k=1}^K} = \Sigma_{\mathbf{w}} - \Sigma_{\mathbf{w}} \Psi \left(\Sigma^* + \Psi^\top \Sigma_{\mathbf{w}} \Psi \right)^{-1} \Psi^\top \Sigma_{\mathbf{w}} \quad (25)$$

2 Composition of MPs

2.1 Blending

2.2 Stitching

See Fig. 3.

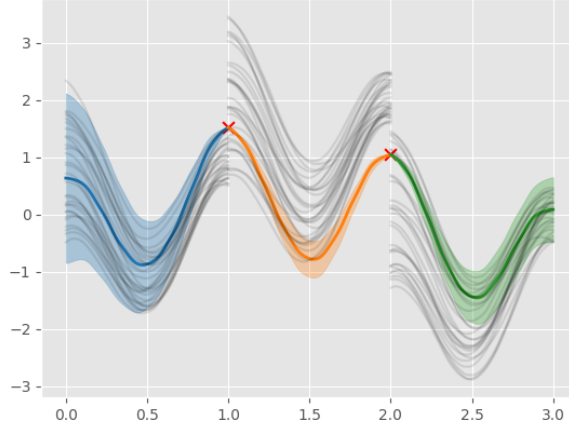


Figure 3: Stitching three ProMPs.

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:

$$p(\mathbf{x}_a, \mathbf{x}_b) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix}\right) \quad (26)$$

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

$$p(\mathbf{x}_a) = \int p(\mathbf{x}_a, \mathbf{x}_b) d\mathbf{x}_b = \mathcal{N}(\mathbf{x}_a | \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_{aa}) \quad (27)$$

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

$$\begin{bmatrix} \mathbf{y}_t \\ \mathbf{w} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_{\mathbf{y}_t} \\ \boldsymbol{\mu}_{\mathbf{w}} \end{bmatrix}, \begin{bmatrix} \text{Cov}[\mathbf{y}_t, \mathbf{y}_t] & \text{Cov}[\mathbf{y}_t, \mathbf{w}] \\ \text{Cov}[\mathbf{w}, \mathbf{y}_t] & \text{Cov}[\mathbf{w}, \mathbf{w}] \end{bmatrix}\right) \quad (28)$$

$$\boldsymbol{\mu}_{\mathbf{y}_t} = \mathbb{E}[\mathbf{y}_t] \quad (29)$$

$$= \mathbb{E}[\boldsymbol{\Psi}_t^\top \mathbf{w} + \boldsymbol{\epsilon}_y] \quad (30)$$

$$= \boldsymbol{\Psi}_t^\top \mathbb{E}[\mathbf{w}] + \mathbb{E}[\boldsymbol{\epsilon}_y] \quad (31)$$

$$= \boldsymbol{\Psi}_t^\top \boldsymbol{\mu}_{\mathbf{w}} + 0 \quad (32)$$

$$= \boldsymbol{\Psi}_t^\top \boldsymbol{\mu}_{\mathbf{w}} \quad (33)$$

$$\text{Cov}[\mathbf{y}_t, \mathbf{y}_t] = \text{Cov}[\Psi_t^\top \mathbf{w} + \epsilon_y] \quad (34)$$

$$= \text{Cov}[\Psi_t^\top \mathbf{w}] + \text{Cov}[\epsilon_y] \quad (35)$$

$$= \Psi_t^\top \text{Cov}[\mathbf{w}] \Psi_t + \Sigma_y \quad (36)$$

$$= \Psi_t^\top \Sigma_w \Psi_t + \Sigma_y \quad (37)$$

$$\begin{bmatrix} \mathbf{y}_t \\ \mathbf{w} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \Psi_t^\top \boldsymbol{\mu}_w \\ \boldsymbol{\mu}_w \end{bmatrix}, \begin{bmatrix} \Psi_t^\top \Sigma_w \Psi_t + \Sigma_y & \Psi_t^\top \Sigma_w \\ \Sigma_w \Psi_t & \Sigma_w \end{bmatrix} \right) \quad (38)$$

$$p(\mathbf{y}_t; \boldsymbol{\theta}) = \mathcal{N}(\mathbf{y}_t | \Psi_t^\top \boldsymbol{\mu}_w, \Psi_t^\top \Sigma_w \Psi_t + \Sigma_y) \quad (39)$$

□

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. (26), and from (Bishop and Bishop, 2024), the parameters of a conditional multivariate Gaussian $p(\mathbf{x}_a | \mathbf{x}_b) = \mathcal{N}(\boldsymbol{\mu}_{a|b}, \Sigma_{a|b})$ are the following:

$$\boldsymbol{\mu}_{a|b} = \boldsymbol{\mu}_a + \Sigma_{ab} \Sigma_{bb}^{-1} (\mathbf{x}_b - \boldsymbol{\mu}_b) \quad (40)$$

$$\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab} \Sigma_{bb}^{-1} \Sigma_{ba} \quad (41)$$

We want the posterior $p(\mathbf{w} | \mathbf{x}_t^*)$, knowing the likelihood $\mathbf{x}_t^* | \mathbf{w} \sim \mathcal{N}(\mathbf{y}_t^* | \Psi_t^\top \mathbf{w}, \Sigma_t^*)$, and the prior $\mathbf{w} \sim \mathcal{N}(\mathbf{w} | \boldsymbol{\mu}_w, \Sigma_w)$.

$$\begin{bmatrix} \mathbf{w} \\ \mathbf{x}_t^* \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\mu}_w \\ \Psi_t^\top \boldsymbol{\mu}_w \end{bmatrix}, \begin{bmatrix} \text{Cov}[\mathbf{w}, \mathbf{w}] & \text{Cov}[\mathbf{w}, \mathbf{x}_t^*] \\ \text{Cov}[\mathbf{x}_t^*, \mathbf{w}] & \text{Cov}[\mathbf{x}_t^*, \mathbf{x}_t^*] \end{bmatrix} \right) \quad (42)$$

$\text{Cov}[\mathbf{x}_t^*, \mathbf{x}_t^*]$ follows from Eq. (37).

$$\text{Cov}[\mathbf{w}, \mathbf{x}_t^*] = \text{Cov}[\mathbf{w}, \Psi_t^\top \mathbf{w} + \epsilon_y] \quad (43)$$

$$= \text{Cov}[\mathbf{w}, \Psi_t^\top \mathbf{w}] \quad (\text{Cov}[\mathbf{w}, \epsilon_y] = 0 \text{ since } \epsilon_y \text{ is independent of } \mathbf{w}) \quad (44)$$

$$= \mathbb{E}[(\mathbf{w} - \boldsymbol{\mu}_w)(\Psi_t^\top \mathbf{w} - \Psi_t^\top \boldsymbol{\mu}_w)^\top] \quad (45)$$

$$= \mathbb{E}[(\mathbf{w} - \boldsymbol{\mu}_w)(\mathbf{w} - \boldsymbol{\mu}_w)^\top \Psi_t] \quad (46)$$

$$= \text{Cov}[\mathbf{w}, \mathbf{w}] \cdot \Psi_t \quad (47)$$

$$= \Sigma_w \Psi_t \quad (48)$$

$$\begin{bmatrix} \mathbf{w} \\ \mathbf{x}_t^* \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\mu}_w \\ \Psi_t^\top \boldsymbol{\mu}_w \end{bmatrix}, \begin{bmatrix} \Sigma_w & \Sigma_w \Psi_t \\ \Psi_t^\top \Sigma_w & \Psi_t^\top \Sigma_w \Psi_t + \Sigma_t^* \end{bmatrix} \right) \quad (49)$$

Using Eq. (40) we get:

$$\boldsymbol{\mu}_{w|x_t^*} = \boldsymbol{\mu}_w + \Sigma_w \Psi_t \left(\Sigma_t^* + \Psi_t^\top \Sigma_w \Psi_t \right)^{-1} (\mathbf{y}_t^* - \Psi_t^\top \boldsymbol{\mu}_w) \quad (50)$$

Using Eq. (41) we get:

$$\Sigma_{w|x_t^*} = \Sigma_w - \Sigma_w \Psi_t \left(\Sigma_t^* + \Psi_t^\top \Sigma_w \Psi_t \right)^{-1} \Psi_t^\top \Sigma_w \quad (51)$$

□