$$Q_t = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \eta_{v_t} & 0 \\ 0 & 0 & 0 & 0 & \eta_{w_t} \end{bmatrix}$$

Class Environment

State Update

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}, \mathbf{a}_t) + \eta_t, \eta_t \sim \mathcal{N}(\mathbf{0}, Q_t)$$

$$x_t = x_{t-1} + v_t cos\theta \cdot \Delta t$$
$$y_t = y_{t-1} + v_t sin\theta \cdot \Delta t$$
$$\theta_t = \theta_{t-1} + w_t \cdot \Delta t$$
$$v_t = G_v a_{v_{t-1}} + \eta_{v_t}$$
$$w_t = G_w a_{w_{t-1}} + \eta_{w_t}$$

$$\mathbf{a}_t = [a_{v_t}, a_{w_t}]$$

$$\mathbf{x}_t = [x_t, y_t, \theta, v_t, w_t]$$

Class Agent

Observation

$$\mathbf{z}_t = h(\mathbf{x}_k) + \xi_t, \xi_t \sim \mathcal{N}(0, R_t) \quad R_t = \begin{bmatrix} \xi_{v_t} & 0 \\ 0 & \xi_{w_t} \end{bmatrix}$$

$$\tilde{v}_t = O_v v_t + \xi_{v_t}$$
$$\tilde{w}_t = O_w w_t + \xi_{w_t}$$

$$\mathbf{z}_t = [\tilde{v}_t, \tilde{w}_t]$$

$$\mathbf{a}_t = [a_{v_t}, a_{w_t}]$$

Policy

critic

$Q(\mathbf{s}_t, \mathbf{a}_t)$

actor

$$\mathbf{a}_t = [a_{v_t}, a_{w_t}]$$

Belief Update

$$A_t = \left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}_{t-1|t-1}, \mathbf{a}_t} = \begin{bmatrix} 1 & 0 & -\hat{v}_t sin\hat{\theta} \cdot \Delta t & 0 & 0 \\ 0 & 1 & \hat{v}_t cos\hat{\theta} \cdot \Delta t & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$H_t = \left.\frac{\partial h}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}_{t-1|t-1}} = \begin{bmatrix} 0 & 0 & 0 & O_v & 0 \\ 0 & 0 & 0 & 0 & O_w \end{bmatrix}$$

Predict

$$\hat{\mathbf{x}}_{t|t-1} = f(\hat{\mathbf{x}}_{t-1|t-1}, \mathbf{a}_t)$$
$$P_{t|t-1} = A_t P_{t-1|t-1} A_t^T + Q_t$$

Update

$$\tilde{\mathbf{y}}_t = \mathbf{z}_k - h(\hat{\mathbf{x}}_{t|t-1})$$
$$S_t = H_t P_{t|t-1} H_t^T + R_k$$
$$K_t = P_{t|t-1} H_t^T S_t^{-1}$$
$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + K_t \tilde{\mathbf{y}}_t$$
$$P_{t|t} = (I - K_t H_t) P_{t|t-1}$$

$$\mathbf{s}_t = [r_t, \theta_{rel,t}, \hat{v}_t, \hat{w}_t, t, P_t]$$

simple $\mathbf{s}_t = [r_t, \theta_{rel,t}, \hat{v}_t, \hat{w}_t]$

Belief Reshape for Policy

$$r_t = \sqrt{\hat{x}_t^2 + \hat{y}_t^2}$$
$$\theta_{rel,t} = \hat{\theta}_t - atan2(-\hat{y}_t, -\hat{x})$$

$$\mathbf{b}_t = [\hat{v}_t, \hat{w}_t, \hat{\theta}_t, \hat{x}_t, \hat{y}_t, P_t]$$