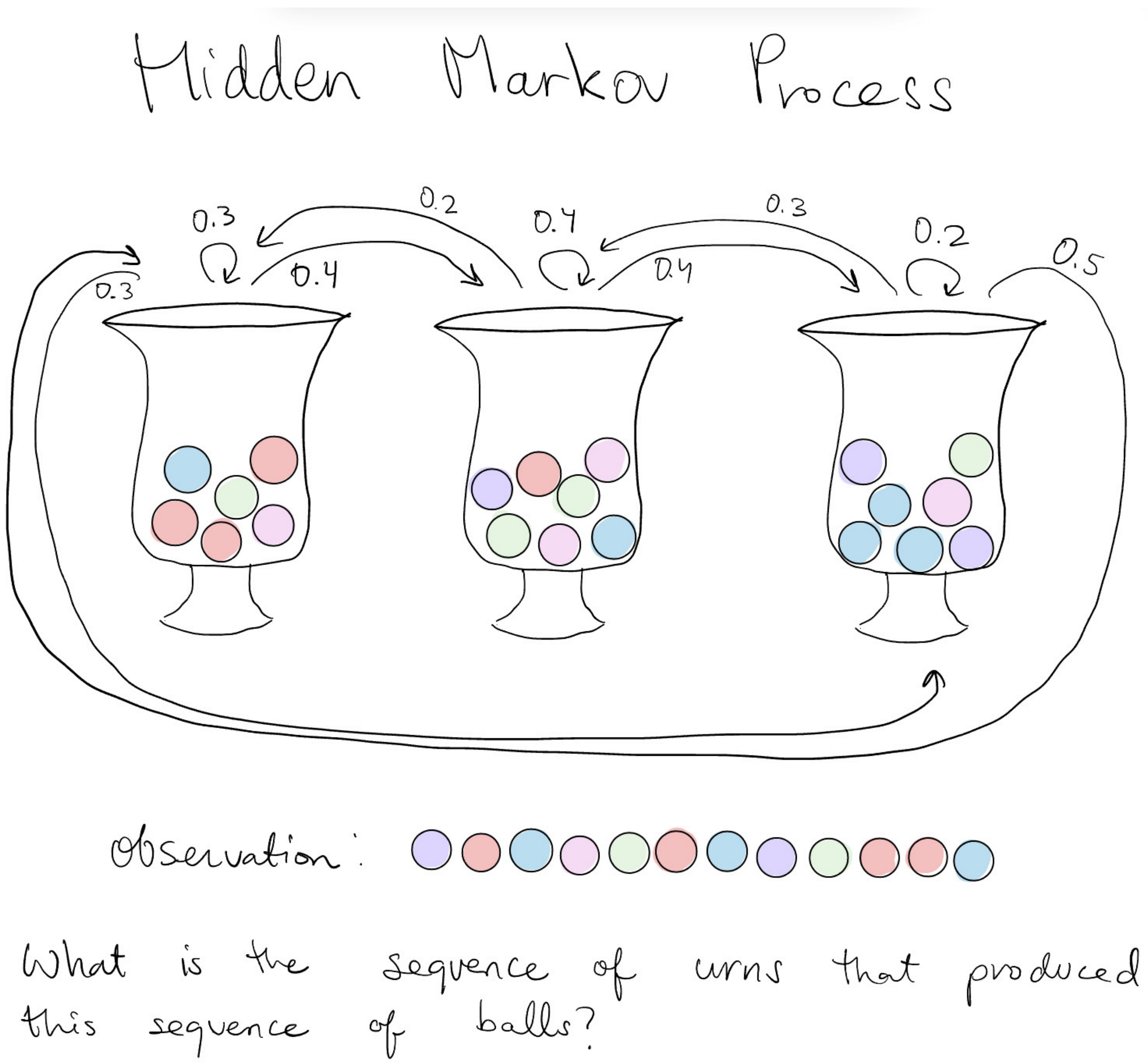
MoML Filtering Presentation Note

1 The Filtering Problem

1.1 State-Space Model



We consider a discrete-time stochastic dynamical system where a hidden *state process* is indirectly observed through noisy measurements.

1.2 State-Space Model

Let $\{X_t\}_{t\geq 0}$ denote the hidden state process and $\{Y_t\}_{t\geq 1}$ the observation process.

State evolution.

$$X_{t+1} = \tilde{a}(X_t, V_t), \tag{1}$$

where $\{V_t\}_{t\geq 0}$ is i.i.d. with distribution P_V, independent of X_0. The initial state satisfies

$$X_0 \sim p_{X_0}.$$

Observation model.

$$Y_t = \tilde{h}(X_t, W_t), \tag{2}$$

where $\{W_t\}_{t\geq 1}$ is i.i.d. with distribution P_W, independent of $\{V_t\}$ and X_0.

1.3 Probabilistic Interpretation

It is convenient to describe the model via conditional probability kernels.

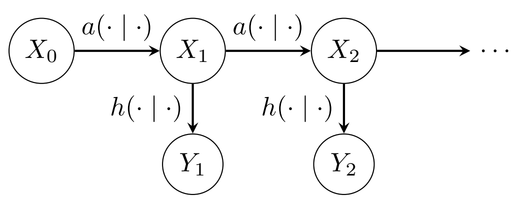
Transition kernel.

$$a(x \mid x') \;\triangleq\; p_{X_{t+1}|X_t}(x \mid x'). \tag{3}$$

Observation kernel.

$$h(y \mid x) \;\triangleq\; p_{Y_t|X_t}(y \mid x). \tag{4}$$

Initial distribution. The distribution p_{X_0} serves as the prior on X_0.



1.4 Objective of Filtering

Define the filtering distribution

$$\pi_t \;\triangleq\; p_{X_t|Y_{1:t}}(\,\cdot \mid y_1, \ldots, y_t), \tag{5}$$

where $Y_{1:t} = (Y_1, \ldots, Y_t)$. The goal is to compute π_t recursively over time.

2 Filtering Equations

Given $\pi_t = p_{X_t|Y_{1:t}}$, compute $\pi_{t+1} = p_{X_{t+1}|Y_{1:t+1}}$ via a prediction step followed by a correction step:

$$\pi_t \xrightarrow{\text{time update}} \pi_{t+1|t} \xrightarrow{\text{information update}} \pi_{t+1}.$$

2.1 Time Update (Prediction Step)

Define the predictive distribution

$$\pi_{t+1|t} \triangleq p_{X_{t+1}|Y_{1:t}}.$$

Using the law of total probability and the Markov property,

$$\pi_{t+1|t}(x) = \int a(x \mid x')\, \pi_t(x')\, dx'. \tag{6}$$

Prediction Operator

Define the prediction operator $\mathcal{A}$ by

$$(\mathcal{A}\pi)(x) \triangleq \int a(x \mid x')\, \pi(x')\, dx'. \tag{7}$$

Then

$$\pi_{t+1|t} = \mathcal{A}\pi_t.$$

2.2 Information Update (Bayes Rule)

Using conditional independence of Y_{t+1} and $Y_{1:t}$ given X_{t+1},

$$\pi_{t+1}(x) = \frac{h(y_{t+1} \mid x)\, \pi_{t+1|t}(x)}{\int h(y_{t+1} \mid x')\, \pi_{t+1|t}(x')\, dx'}. \tag{8}$$

Correction Operator

Define $\mathcal{B}_y$ by

$$(\mathcal{B}_y\pi)(x) \triangleq \frac{h(y \mid x)\, \pi(x)}{\int h(y \mid x')\, \pi(x')\, dx'}. \tag{9}$$

Then

$$\pi_{t+1} = \mathcal{B}_{y_{t+1}}\, \pi_{t+1|t}.$$

2.3 Full Filtering Recursion

$$\pi_{t+1} = \mathcal{B}_{y_{t+1}}\, \mathcal{A}\, \pi_t, \qquad \pi_0 = p_{X_0}. \tag{10}$$

Remarks

- In general, $\mathcal{A}$ and $\mathcal{B}_y$ are infinite-dimensional.
- Exact implementation of $\pi_{t+1} = \mathcal{B}_{y_{t+1}}\mathcal{A}\pi_t$ is typically infeasible except in special cases.

Linear-Gaussian Case

When the dynamics and observation models are linear and all noises are Gaussian, the recursion admits a finite-dimensional closed-form solution: the *Kalman filter*.

3 Kalman Filter

3.1 Model Description

$$X_{t+1} = A_t X_t + V_t, \tag{11}$$
$$Y_t = H_t X_t + W_t, \tag{12}$$

where $X_t \in \mathbb{R}^n$, $Y_t \in \mathbb{R}^m$, and A_t, H_t are matrices. Assume

$$X_0 \sim \mathcal{N}(\mu_0, \Sigma_0),$$
$$V_t \sim \mathcal{N}(0, Q_t),$$
$$W_t \sim \mathcal{N}(0, R_t),$$

with $\{X_0, V_0, V_1, \ldots, W_1, W_2, \ldots\}$ mutually independent and $R_t \succ 0$.

Filtering Objective

Compute $p_{X_t|Y_{1:t}}$. In the linear–Gaussian setting this distribution is Gaussian, determined by its mean and covariance.

3.2 Minimum Mean-Square Error (MMSE) Estimation

Let $X \in \mathbb{R}^n$ and $Y \in \mathbb{R}^m$. Over a class $\mathcal{V}$, the MSE-optimal estimator is

$$\hat{X} = \arg\min_{Z \in \mathcal{V}} \mathbb{E}\left[\|X - Z\|^2\right]. \tag{13}$$

Orthogonality Principle

There exists a unique $Z^\star \in \mathcal{V}$ such that

$$\|X - Z^\star\|^2 \leq \|X - Z\|^2, \quad \forall Z \in \mathcal{V},$$

iff

$$X - Z^\star \perp \mathcal{V}, \tag{14}$$

equivalently,

$$\mathbb{E}\Big[(X - Z^\star)^\top Z\Big] = 0, \qquad \forall Z \in \mathcal{V}. \tag{15}$$

MMSE Estimation

Let $\mathcal{L}^2$ be the space of square-integrable $\mathbb{R}^n$-valued random vectors,

$$\mathcal{L}^2 \triangleq \left\{ Z : \Omega \to \mathbb{R}^n \,\middle|\, \mathbb{E}[\|Z\|^2] < \infty \right\},$$

with inner product

$$\langle Z_1, Z_2 \rangle_{\mathcal{L}^2} \triangleq \mathbb{E}[Z_1^\top Z_2].$$

Then

$$\mathbb{E}[\|X - Z\|^2] = \|X - Z\|^2_{\mathcal{L}^2}.$$

Conditional Expectation as an Orthogonal Projection

Let $\mathcal{V} \subset \mathcal{L}^2$ be closed. Then
$$Z^\star = \arg\min_{Z \in \mathcal{V}} \mathbb{E}[\|X - Z\|^2]$$
is the orthogonal projection of X onto $\mathcal{V}$.

Case 1: Constant estimators. If $\mathcal{V}$ is the space of constants, then $Z^\star = \mathbb{E}[X]$.

Case 2: Affine functions of Y. If $\mathcal{V}$ is the space of affine functions of Y, then
$$Z^\star = \mathbb{E}[X] + K\big(Y - \mathbb{E}[Y]\big) \;\triangleq\; \widehat{\mathbb{E}}[X \mid Y], \tag{16}$$
where
$$K = \mathrm{Cov}(X, Y)\ \mathrm{Cov}(Y, Y)^{-1}. \tag{17}$$
If $\mathrm{Cov}(Y, Y)$ is not invertible, any K satisfying $K\ \mathrm{Cov}(Y, Y) = \mathrm{Cov}(X, Y)$ is admissible.

Case 3: All measurable functions of Y. If $\mathcal{V}$ is the space of all square-integrable functions of Y, then $Z^\star = \mathbb{E}[X \mid Y]$.

Optimality of Conditional Expectation

For any estimator Z that is a function of Y,
$$\mathbb{E}\big[\|X - \mathbb{E}[X]\|^2\big] \geq \mathbb{E}\Big[\|X - \widehat{\mathbb{E}}[X \mid Y]\|^2\Big] \geq \mathbb{E}\big[\|X - \mathbb{E}[X \mid Y]\|^2\big]\,. \tag{18}$$

Jointly Gaussian Case

If (X, Y) are jointly Gaussian, then
$$\mathbb{E}[X \mid Y] = \widehat{\mathbb{E}}[X \mid Y], \qquad X \mid Y \sim \mathcal{N}(\mathbb{E}[X \mid Y],\, \mathrm{Cov}(e))\,,$$
where $e \triangleq X - \mathbb{E}[X \mid Y]$ and
$$\mathrm{Cov}\big(e, \mathbb{E}[X \mid Y]\big) = 0, \qquad \mathrm{Cov}(X) = \mathrm{Cov}(e) + \mathrm{Cov}\big(\mathbb{E}[X \mid Y]\big).$$

Best Affine MMSE Estimator

Assume $\mathbb{E}[X \mid Y] = KY + b$. Enforcing $\mathbb{E}[e] = 0$ gives $b = \mathbb{E}[X] - K\mathbb{E}[Y]$, and enforcing $\mathrm{Cov}(e, Y) = 0$ gives
$$\mathrm{Cov}(X, Y) - K\ \mathrm{Cov}(Y) = 0.$$
If $\mathrm{Cov}(Y)$ is invertible,
$$K = \mathrm{Cov}(X, Y)\ \mathrm{Cov}(Y)^{-1}.$$

3.3 Kalman Filter Setup

Consider

$$
\begin{aligned}
X_{t+1} &= A_t X_t + V_t, \\
Y_t &= H_t X_t + W_t,
\end{aligned}
\tag{19}
$$

Define

$$
\hat{X}_{t|t} \triangleq \mathbb{E}[X_t \mid Y_{1:t}], \qquad P_{t|t} \triangleq \mathrm{Cov}\Big(X_t - \hat{X}_{t|t}\Big)\,. \tag{20}
$$

$$
\hat{X}_{t+1|t} \triangleq \mathbb{E}[X_{t+1} \mid Y_{1:t}], \qquad P_{t+1|t} \triangleq \mathrm{Cov}\Big(X_{t+1} - \hat{X}_{t+1|t}\Big)\,. \tag{21}
$$

Kalman Filter Recursion

$$
\Big(\hat{X}_{t|t},\, P_{t|t}\Big) \xrightarrow{\text{time update}} \Big(\hat{X}_{t+1|t},\, P_{t+1|t}\Big) \xrightarrow{\text{information update}} \Big(\hat{X}_{t+1|t+1},\, P_{t+1|t+1}\Big)
$$

Assume that X_0, $\{V_t\}$, and $\{W_t\}$ are mutually uncorrelated. Then,

$$
\hat{X}_{t+1|t} = A_t \hat{X}_{t|t}, \tag{22}
$$

$$
P_{t+1|t} = A_t P_{t|t} A_t^\top + Q_t. \tag{23}
$$

With the observation at time $t+1$,

$$
K_{t+1} \triangleq P_{t+1|t} H_{t+1}^\top \Big(H_{t+1} P_{t+1|t} H_{t+1}^\top + R_{t+1}\Big)^{-1}, \tag{24}
$$

$$
\hat{X}_{t+1|t+1} = \hat{X}_{t+1|t} + K_{t+1}\Big(Y_{t+1} - H_{t+1}\hat{X}_{t+1|t}\Big), \tag{25}
$$

$$
P_{t+1|t+1} = P_{t+1|t} - K_{t+1}\Big(H_{t+1} P_{t+1|t} H_{t+1}^\top + R_{t+1}\Big) K_{t+1}^\top. \tag{26}
$$

Kalman Filter Theorem

Theorem (Kalman Filter). Consider the state-space model in (19) and assume that X_0, $\{V_t\}$, and $\{W_t\}$ are mutually uncorrelated

1. The best linear estimator of $\mathbb{E}[X_t \mid Y_{1:t}]$ is given by the Kalman recursion $(\hat{X}_{t|t}, P_{t|t})$.

2. If, in addition, X_0, $\{V_t\}$, and $\{W_t\}$ are jointly Gaussian, then

$$
X_t \mid Y_{1:t} \sim \mathcal{N}\Big(\hat{X}_{t|t},\, P_{t|t}\Big)\,,
$$

   and the Kalman filter yields the MMSE-optimal estimator.

3.4 Extended Kalman Filter

For nonlinear models, the *extended Kalman filter* (EKF) applies the Kalman updates to a local linearization.

Nonlinear State-Space Model

Consider

$$X_{t+1} = f_t(X_t) + V_t, \tag{27}$$
$$Y_{t+1} = g_{t+1}(X_{t+1}) + W_{t+1}, \tag{28}$$

with

$$V_t \sim \mathcal{N}(0, Q_t), \qquad W_{t+1} \sim \mathcal{N}(0, R_{t+1}),$$

mutually independent across time and independent of X_0.

Linearization

Let $(\hat{X}_{t|t}, P_{t|t})$ approximate $p_{X_t|Y_{1:t}}$ as Gaussian. Define the Jacobians

$$A_t \triangleq \left.\frac{\partial f_t(x)}{\partial x}\right|_{x=\hat{X}_{t|t}} \in \mathbb{R}^{n\times n}, \tag{29}$$
$$H_{t+1} \triangleq \left.\frac{\partial g_{t+1}(x)}{\partial x}\right|_{x=\hat{X}_{t+1|t}} \in \mathbb{R}^{m\times n}. \tag{30}$$

EKF Recursion

Time Update (Prediction Step)

$$\hat{X}_{t+1|t} = f_t(\hat{X}_{t|t}), \tag{31}$$
$$P_{t+1|t} = A_t P_{t|t} A_t^\top + Q_t. \tag{32}$$

Information Update (Measurement Update)

$$\hat{Y}_{t+1|t} \triangleq g_{t+1}(\hat{X}_{t+1|t}), \tag{33}$$
$$\nu_{t+1} \triangleq Y_{t+1} - \hat{Y}_{t+1|t}, \tag{34}$$
$$S_{t+1} \triangleq H_{t+1} P_{t+1|t} H_{t+1}^\top + R_{t+1}, \tag{35}$$
$$K_{t+1} \triangleq P_{t+1|t} H_{t+1}^\top S_{t+1}^{-1}, \tag{36}$$
$$\hat{X}_{t+1|t+1} = \hat{X}_{t+1|t} + K_{t+1}\nu_{t+1}, \tag{37}$$
$$P_{t+1|t+1} = P_{t+1|t} - K_{t+1} S_{t+1} K_{t+1}^\top. \tag{38}$$

Remarks on the EKF

- The EKF is a first-order local approximation via linearization.
- It can be biased/unstable under strong nonlinearities or large uncertainty.
- Joseph form (optional) for numerical stability:

$$P_{t+1|t+1} = (I - K_{t+1}H_{t+1})P_{t+1|t}(I - K_{t+1}H_{t+1})^\top + K_{t+1}R_{t+1}K_{t+1}^\top.$$

3.5 Unscented Kalman Filter

The *unscented Kalman filter* (UKF) propagates means/covariances through nonlinear maps using deterministic *sigma points* (unscented transform).

Nonlinear State-Space Model

$$X_{t+1} = f_t(X_t) + V_t, \tag{39}$$
$$Y_{t+1} = g_{t+1}(X_{t+1}) + W_{t+1}, \tag{40}$$

with

$$V_t \sim \mathcal{N}(0, Q_t), \qquad W_{t+1} \sim \mathcal{N}(0, R_{t+1}),$$

and $\{X_0, V_0, V_1, \ldots, W_1, W_2, \ldots\}$ mutually independent, and

$$X_t \mid Y_{1:t} \approx \mathcal{N}(\hat{X}_{t|t}, P_{t|t}).$$

Sigma Points and Weights

Let the state dimension be n, choose $\alpha > 0$, $\kappa \in \mathbb{R}$, $\beta \geq 0$, and define

$$\lambda \triangleq \alpha^2(n+\kappa) - n, \tag{41}$$
$$\gamma \triangleq \sqrt{n+\lambda}. \tag{42}$$

Given mean μ and covariance P, define $2n+1$ sigma points:

$$\chi^{(0)} \triangleq \mu, \tag{43}$$
$$\chi^{(i)} \triangleq \mu + \gamma\,[\sqrt{P}]_i, \qquad i = 1, \ldots, n, \tag{44}$$
$$\chi^{(i+n)} \triangleq \mu - \gamma\,[\sqrt{P}]_i, \qquad i = 1, \ldots, n, \tag{45}$$

where $\sqrt{P}\,\sqrt{P}^\top = P$ and $[\sqrt{P}]_i$ is the i-th column. Weights:

$$w_m^{(0)} \triangleq \frac{\lambda}{n+\lambda}, \tag{46}$$
$$w_c^{(0)} \triangleq \frac{\lambda}{n+\lambda} + (1 - \alpha^2 + \beta), \tag{47}$$
$$w_m^{(i)} = w_c^{(i)} \triangleq \frac{1}{2(n+\lambda)}, \qquad i = 1, \ldots, 2n. \tag{48}$$

UKF Recursion

Time Update (Prediction Step)

$$\chi_t^{(i)} \triangleq \chi^{(i)}(\mu = \hat{X}_{t|t},\, P = P_{t|t}), \qquad i = 0, \ldots, 2n.$$

$$\chi_{t+1|t}^{(i)} \triangleq f_t(\chi_t^{(i)}), \qquad i = 0, \ldots, 2n, \tag{49}$$
$$\hat{X}_{t+1|t} \triangleq \sum_{i=0}^{2n} w_m^{(i)} \chi_{t+1|t}^{(i)}, \tag{50}$$
$$P_{t+1|t} \triangleq \sum_{i=0}^{2n} w_c^{(i)} (\chi_{t+1|t}^{(i)} - \hat{X}_{t+1|t})(\chi_{t+1|t}^{(i)} - \hat{X}_{t+1|t})^\top + Q_t. \tag{51}$$

Information Update (Measurement Update)

$$
\begin{aligned}
\psi_{t+1|t}^{(i)} &\triangleq g_{t+1}(\chi_{t+1|t}^{(i)}), \qquad i = 0, \ldots, 2n, && (52)\\
\hat{Y}_{t+1|t} &\triangleq \sum_{i=0}^{2n} w_m^{(i)} \psi_{t+1|t}^{(i)}, && (53)\\
S_{t+1} &\triangleq \sum_{i=0}^{2n} w_c^{(i)} (\psi_{t+1|t}^{(i)} - \hat{Y}_{t+1|t})(\psi_{t+1|t}^{(i)} - \hat{Y}_{t+1|t})^\top + R_{t+1}, && (54)\\
P_{t+1}^{xy} &\triangleq \sum_{i=0}^{2n} w_c^{(i)} (\chi_{t+1|t}^{(i)} - \hat{X}_{t+1|t})(\psi_{t+1|t}^{(i)} - \hat{Y}_{t+1|t})^\top, && (55)\\
K_{t+1} &\triangleq P_{t+1}^{xy} S_{t+1}^{-1}, && (56)\\
\nu_{t+1} &\triangleq Y_{t+1} - \hat{Y}_{t+1|t}, && (57)\\
\hat{X}_{t+1|t+1} &= \hat{X}_{t+1|t} + K_{t+1}\nu_{t+1}, && (58)\\
P_{t+1|t+1} &= P_{t+1|t} - K_{t+1} S_{t+1} K_{t+1}^\top. && (59)
\end{aligned}
$$

Remarks on the UKF

- The UKF does not require Jacobians; it evaluates f_t and g_{t+1} on sigma points.
- It often matches mean/covariance under nonlinear transforms more accurately than first-order linearization.
- Square-root implementations can improve numerical stability.

3.6 Ensemble Kalman Filter

The *ensemble Kalman filter* (EnKF) approximates the filtering distribution with an ensemble and uses sample statistics.

Nonlinear State-Space Model

$$
\begin{aligned}
X_{t+1} &= f_t(X_t) + V_t, && (60)\\
Y_{t+1} &= g_{t+1}(X_{t+1}) + W_{t+1}, && (61)
\end{aligned}
$$

with

$$V_t \sim \mathcal{N}(0, Q_t), \qquad W_{t+1} \sim \mathcal{N}(0, R_{t+1}),$$

all noise terms mutually independent and independent of X_0.

Ensemble Representation

Approximate $p_{X_t|Y_{1:t}}$ by $\{X_{t|t}^{(i)}\}_{i=1}^N$, with sample mean/covariance

$$
\begin{aligned}
\bar{X}_{t|t} &\triangleq \frac{1}{N} \sum_{i=1}^{N} X_{t|t}^{(i)}, && (62)\\
P_{t|t} &\triangleq \frac{1}{N-1} \sum_{i=1}^{N} (X_{t|t}^{(i)} - \bar{X}_{t|t})(X_{t|t}^{(i)} - \bar{X}_{t|t})^\top. && (63)
\end{aligned}
$$

EnKF Recursion

Time Update (Prediction Step)

$$X_{t+1|t}^{(i)} = f_t(X_{t|t}^{(i)}) + V_t^{(i)}, \qquad i = 1, \ldots, N, \tag{64}$$

where $V_t^{(i)}$ are i.i.d. $\mathcal{N}(0, Q_t)$. Compute

$$\bar{X}_{t+1|t} = \frac{1}{N} \sum_{i=1}^{N} X_{t+1|t}^{(i)}, \tag{65}$$

$$P_{t+1|t} = \frac{1}{N-1} \sum_{i=1}^{N} (X_{t+1|t}^{(i)} - \bar{X}_{t+1|t})(X_{t+1|t}^{(i)} - \bar{X}_{t+1|t})^\top. \tag{66}$$

Information Update (Measurement Update)

$$Y_{t+1|t}^{(i)} \triangleq g_{t+1}(X_{t+1|t}^{(i)}), \tag{67}$$

$$\bar{Y}_{t+1|t} \triangleq \frac{1}{N} \sum_{i=1}^{N} Y_{t+1|t}^{(i)}, \tag{68}$$

$$P_{t+1}^{xy} \triangleq \frac{1}{N-1} \sum_{i=1}^{N} (X_{t+1|t}^{(i)} - \bar{X}_{t+1|t})(Y_{t+1|t}^{(i)} - \bar{Y}_{t+1|t})^\top, \tag{69}$$

$$S_{t+1} \triangleq \frac{1}{N-1} \sum_{i=1}^{N} (Y_{t+1|t}^{(i)} - \bar{Y}_{t+1|t})(Y_{t+1|t}^{(i)} - \bar{Y}_{t+1|t})^\top + R_{t+1}, \tag{70}$$

$$K_{t+1} \triangleq P_{t+1}^{xy} S_{t+1}^{-1}. \tag{71}$$

Update each member using a perturbed observation:

$$X_{t+1|t+1}^{(i)} = X_{t+1|t}^{(i)} + K_{t+1}\Big(Y_{t+1} + W_{t+1}^{(i)} - Y_{t+1|t}^{(i)}\Big), \tag{72}$$

where $W_{t+1}^{(i)}$ are i.i.d. $\mathcal{N}(0, R_{t+1})$. Then

$$\bar{X}_{t+1|t+1} = \frac{1}{N} \sum_{i=1}^{N} X_{t+1|t+1}^{(i)}, \tag{73}$$

$$P_{t+1|t+1} = \frac{1}{N-1} \sum_{i=1}^{N} (X_{t+1|t+1}^{(i)} - \bar{X}_{t+1|t+1})(X_{t+1|t+1}^{(i)} - \bar{X}_{t+1|t+1})^\top. \tag{74}$$

Remarks on the EnKF

- The EnKF uses sample covariances from a finite ensemble.
- As $N \to \infty$, it converges (under suitable assumptions) to the Kalman filter in the linear–Gaussian case.
- It is often effective in high-dimensional systems with moderate N.

4 Particle Methods

4.1 Why Particle Methods

The above integrals are rarely computable analytically. Particle methods replace continuous distributions by discrete approximations consisting of finitely many samples. This turns integrals into finite sums.

The posterior is approximated by

$$\pi_t^N(dx) = \sum_{i=1}^{N} w_t^{(i)} \, \delta_{x_t^{(i)}}(dx),$$

where particles $\{x_t^{(i)}\}$ represent possible states and weights $\{w_t^{(i)}\}$ represent their plausibility.

Expectations are approximated as

$$\int \varphi(x)\pi_t(dx) \approx \sum_{i=1}^{N} w_t^{(i)} \varphi(x_t^{(i)}).$$

4.2 Importance Sampling: Main Idea

Particle filtering is built on importance sampling. Suppose we want to compute expectations under a target density $\pi(x)$, but we can only sample from another density $q(x)$, called the *proposal*.

We correct for this mismatch using weights:

$$w(x) \propto \frac{\pi(x)}{q(x)}.$$

Thus, the proposal determines where samples are drawn, while weights adjust their contribution.

Good proposals place samples in regions where the target density is large.

4.3 Sequential Importance Sampling

In filtering, the target changes over time. Particles are propagated forward and reweighted sequentially.

We conceptually consider the joint posterior over state trajectories

$$\pi(x_{0:t} \mid y_{1:t}),$$

and approximate it by importance sampling. This leads to the recursive weight update

$$\tilde{w}_t^{(i)} = w_{t-1}^{(i)} \frac{a(x_t^{(i)} \mid x_{t-1}^{(i)}) h(y_t \mid x_t^{(i)})}{q_t(x_t^{(i)} \mid x_{t-1}^{(i)}, y_t)}.$$

Each step:

- Sample a new state from a proposal distribution.
- Multiply the old weight by a likelihood ratio.
- Normalize.

4.4 The SIR Filter

The simplest choice is to use the model dynamics as the proposal:

$$q_t(x_t \mid x_{t-1}, y_t) = a(x_t \mid x_{t-1}).$$

Then weights simplify to

$$\tilde{w}_t^{(i)} = w_{t-1}^{(i)} a(y_t \mid x_t^{(i)}).$$

Interpretation:

- Particles follow the physics.
- Observations only influence the weights.

This is known as the Sequential Importance Resampling (SIR).

4.5 Weight Degeneracy

Over time, most weights become extremely small. Only a few particles carry significant mass. This phenomenon is called *degeneracy*.

A diagnostic quantity is the effective sample size

$$N_{\text{eff}} = \frac{1}{\sum_i (w_t^{(i)})^2}.$$

Small N_{eff} indicates poor particle diversity.

4.6 Resampling

Resampling removes particles with tiny weights and replicates particles with large weights. After resampling, all particles have equal weight.

Conceptually:

- Kill unlikely particles.
- Clone likely particles.

Resampling introduces additional randomness but prevents catastrophic collapse.

4.7 Interpretation and Limitations

Particle filters provide a universal nonlinear filtering framework, it is exact as the number of particles grow. However, they suffer from:

- Curse of dimensionality,
- Loss of particle diversity after resampling,
- High variance if proposals are poor.
- Require explicit knowledge on the likelihood distribution.

These limitations motivate advanced proposals, transport-based particle flows, and deterministic approximations.

5 Optimal Transport Filter

5.1 Introduction

Coupling-based methods use a transport map for the conditioning step. The posterior is approximated with a uniformly weighted distribution of particles $\frac{1}{N}\sum_{i=1}^{N}\delta_{x_t^{(i)}}$, where each particle $x_t^{(i)}$ is updated in the following way.

$$\begin{aligned}
&\text{(propagation)} \qquad x_{t|t-1}^{(i)} \sim a(\cdot \mid x_{t-1}^{(i)}) \\
&\text{(conditioning)} \qquad x_t^{(i)} = T_t(x_{t|t-1}^{(i)}\,,\, y_t)
\end{aligned}$$

Here $T_t(\cdot\,,\, y_t)$ represents a transport map from the prior distribution $\pi_{t|t-1}(\cdot) := \mathbb{P}(x_t \in \cdot \mid y_{1:t-1})$ to the posterior distribution $\pi_t(\cdot) := \mathbb{P}(x_t \in \cdot \mid y_{1:t})$.

5.2 Conditioning with Transport Maps

Since the posterior π_t is approximated by an empirical distribution of particles $\{x_t^{(i)}\}_{i=1}^N$, the transport map T_t depends on the empirical distribution of the particles. We consider the problem of designing the transport map T_t in the "mean field" limit ($N = \infty$), which assumes the particles are i.i.d. according to a single mean-field distribution $\bar{\pi}_t$. Our update rules are the following.

$$\begin{aligned}
&\text{(propogation)} \qquad \bar{\pi}_{t|t-1} = \int a(\cdot \mid z)\bar{\pi}_{t-1}(z)dz \\
&\text{(conditioning)} \qquad \bar{\pi}_t = T_t(\cdot \mid y_t)_{\#}\bar{\pi}_{t|t-1}.
\end{aligned}$$

Problem: For all probability distributions π, we want to find a map T such that

$$T(\cdot \mid y)_{\#}\pi = \frac{h(y \mid \cdot)\pi(\cdot)}{\int h(y \mid z)\pi(z)dz} =: \mathcal{B}_y(\pi) \quad \forall y. \tag{75}$$

Note that the map T that satisfies (3) may not be unique, so we formulate an optimal transport problem that yields a unique solution T satisfying (3). Once we obtain a procedure for finding such a map T, we do this at every timestep t.

5.3 OT formulation

Let $x \sim \pi,\ y \sim h(\cdot \mid X)$ and let $\bar{x}$ be an independent copy of x. Let $P_{X,Y}$ denote the joint distribution of (x, y), with marginals P_X and P_Y. With this notation, the equation (3) may be written as $T(\cdot\,,\, y)_{\#}P_X = P_{X|Y}(\cdot \mid y)$ a.e., which is equivalent to

$$\big(T(\bar{x}, y),\, y\big) \sim P_{X,Y}.$$

In order to select a unique map T that satisfies the above, we formulate the **conditional Monge problem**:

$$\min_{T \in \mathcal{M}(P_X \otimes P_Y)} \mathbb{E}\big[c(T\big(\bar{x}, y\big)\,,\, \bar{x})\big] \quad \text{s.t.} \quad \big(T(\bar{x}, y),\, y\big) \sim P_{X,Y} \tag{76}$$

where $c(x, x') = \frac{1}{2}\|x - x'\|^2$ and $\mathcal{M}(P_X \otimes P_Y)$ is the set of maps that are $P_X \otimes P_Y$-measurable.

The optimization problem in (4) is viewed as the Monge problem between the independent coupling $(\bar{x}, y) \sim P_X \otimes P_Y$ and the joint distribution $(x, y) \sim P_{X,Y}$ with transport maps that are constrained to be block triangular.

We use Brenier's result to establish the existence and uniqueness of the solution to the conditional Monge problem by using the **dual Kantorovich problem**, yielding the max-min formulation

$$\max_{f \in c\text{-Concave}_x} \min_{T \in \mathcal{M}(P_X \otimes P_Y)} J(f, T) \tag{77}$$

where the objective function is

$$J(f, T) := \mathbb{E}[f(x, y) - f(T(\bar{x}, y), y) + c(T(\bar{x}, y), \bar{x})]$$

and c-Concave$_x$ is the set of functions that are c-concave in their first variable.

Assuming that π is absolutely continuous with respect to the Lebesgue measure with a convex support set $\mathcal{X}$, $\mathcal{B}_y(\pi)$ admits a density with respect to the Lebesgue measure for all y, and $\mathbb{E}[\|x\|^2] < \infty$, then there exists a unique pair $(\bar{f}, \bar{T})$, up to an additive constant for $\bar{f}$, that solves the max-min problem (5) and the map $\bar{T}(\cdot, y)$ is the optimal transport map from π to $\mathcal{B}_y(\pi)$ for *a.e.* y.

5.4 Empirical Approximation

In practice, the objective function for (5) is empirically approximated by samples according to

$$J^{(N)}(f, T) := \frac{1}{N} \sum_{i=1}^{N} [f(x^{(i)}, y^{(i)}) + \frac{1}{2} \|T(\bar{x}^{(i)}, y^{(i)}) - x^{(i)}\|^2 - f(T(\bar{x}^{(i)}, y^{(i)}), y^{(i)})]$$

and the problem becomes

$$\max_{f \in c\text{-Concave}_x} \min_{T \in \mathcal{M}(P_X \otimes P_Y)} J^{(N)}(f, T).$$

where $x^{(i)}, \bar{x}^{(i)} \overset{\text{i.i.d.}}{\sim} \bar{\pi}_{t|t-1}$, $y^{(i)} \sim h(\cdot \mid x^{(i)})$.