Bayesian Filters for Affine Term Structures

Ivan Sidorov

April 16, 2025

Table of Contents

Definition (Zero-Coupon Yield Curve)

Let $P(t,\tau)$ denote the price at time t of a zero-coupon bond with tenor τ. The zero-coupon yield $y(t,\tau)$ is defined by

$$P(t,\tau) = e^{-y(t,\tau)\tau} \Rightarrow y(t,\tau) = -\frac{1}{\tau}\ln\big(P(t,\tau)\big).$$

The zero-coupon yield curve at time t is thus the mapping: $\tau \mapsto y(t,\tau)$.

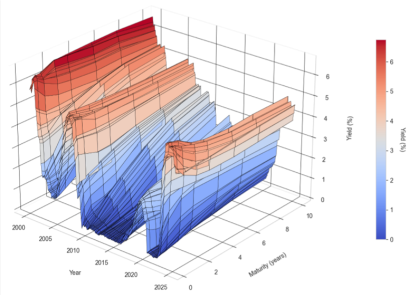

Figure: USD Zero-Coupon Bond Yield Curve Surface.

Bond Pricing. Bond price is the expectation under risk-neutral measure $\mathbb{Q}$ conditional on historical filtration $\mathcal{F}_t$ of a discount factor

$$P(t,\tau) \;=\; \mathbb{E}^{\mathbb{Q}}\Big[\exp\Big(-\int_t^{t+\tau} r_s\, ds\Big)\;\Big|\;\mathcal{F}_t\Big], \text{ where } r_t = \lim_{\tau\to 0^+} y(t,\tau)$$

Latent Factors. Instantaneous rate is assumed to be an affine function of a state vector $X_t \in \mathbb{R}^d$:

$$r_t \;=\; \alpha_0 \;+\; \alpha_1^\top X_t,$$

State Dynamics. Consider a homogeneous diffusion evolving as

$$dX_t \;=\; \mu(X_t)dt \;+\; \Sigma\big(X_t\big)\, dW_t, \quad X_0 = x,$$

where drift and diffusion squared are affine functions of latent factors

$$\mu(X_t) = A_0 + A_1^\top X_t,\; \Sigma\big(X_t\big)\,\Sigma\big(X_t\big)^\top = B_0 + B_1^\top X_t.$$

If B_0, B_1 are positive semi-definite matrices and parameters are such that the SDE has a unique strong solution, then X_t is an affine process.

Two-Factor Vasicek:

$$\begin{cases} \mathrm{d}X_t^{(1)} = k_1\left(\theta_1 - X_t^{(1)}\right)\,dt + \sigma_1\,\mathrm{d}W_t^{(1)}, \\ \mathrm{d}X_t^{(2)} = k_2\left(\theta_2 - X_t^{(2)}\right)\,dt + \sigma_2\,\mathrm{d}W_t^{(2)}, \\ d\langle W^{(1)}, W^{(2)}\rangle_t = \rho dt, \end{cases} \qquad r_t = \alpha_0 + \alpha_1 X_t^{(1)} + \alpha_2 X_t^{(2)}.$$

Independent Two-Factor CIR:

$$\begin{cases} \mathrm{d}X_t^{(1)} = k_1\left(\theta_1 - X_t^{(1)}\right)\,dt + \sigma_1\sqrt{X_t^{(1)}}\,\mathrm{d}W_t^{(1)}, \\ \mathrm{d}X_t^{(2)} = k_2\left(\theta_2 - X_t^{(2)}\right)\,dt + \sigma_2\sqrt{X_t^{(2)}}\,\mathrm{d}W_t^{(2)}, \\ d\langle W^{(1)}, W^{(2)}\rangle_t = 0, \end{cases} \qquad r_t = \alpha_0 + \alpha_1 X_t^{(1)} + \alpha_2 X_t^{(2)}.$$

Heston:

$$\begin{cases} dX_t = k_1\left(\theta_1 - X_t\right)\,dt + \sqrt{v_t}\,dW_t^{(1)}, \\ dv_t = k_2\left(\theta_2 - v_t\right)dt + \sigma\sqrt{v_t}\,dW_t^{(2)}, \\ d\langle W^{(1)}, W^{(2)}\rangle_t = \rho\,dt, \end{cases} \qquad r_t = \alpha_0 + \alpha_1 X_t + \alpha_2 v_t,$$

v_t is spanned - it can be hedged by observed variables.

Bond Pricing.

Given the short rate is an affine function of a time-homogeneous diffusion X_t, the bond price is

$$\begin{aligned}
P(t,\tau) &= \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_t^{t+\tau} r(X_s)\,ds\right)\Big|\mathcal{F}_t\right] && \text{(by definition)}\\
&= \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_t^{t+\tau} r(X_s)\,ds\right)\Big|X_t = x\right] && \text{(Markov property)}\\
&= \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_0^{\tau} r(X_s)\,ds\right)\Big|X_0 = x\right] && \text{(time-homogeneity)}\\
&= F(\tau,x) && \text{(under assumptions)}
\end{aligned}$$

Also we assume exponentially affine pricing function:

$$F(\tau,x) = \exp\left(A(\tau) + B(\tau)^\top x\right),$$

where $A(\tau) \in \mathbb{R}$, $B(\tau) \in \mathbb{R}^d$ are the functions to be determined.

Feynman-Kac Theorem. Let X_t be an Ito process that admits a unique strong solution and $r(X_t), \phi(X_\tau)$ - continuous functions, then

$$u(\tau, x) = \mathbb{E}^{\mathbb{Q}} \left[\exp\left(-\int_0^\tau r(X_s)\, ds \right) \cdot \phi(X_\tau) \middle| X_0 = x \right]$$

solves the PDE:

$$-\frac{\partial u}{\partial \tau} + \mu(x)^\top \nabla u(\tau, x) + \frac{1}{2} \mathrm{Tr}\left[\Sigma(x)\Sigma(x)^\top \nabla^2 u \right] - r(x)\, u = 0, \quad u(0, x) = \phi(x)$$

Bond Pricing. For zero-coupon bonds, the payoff is $\phi(x) \equiv 1$, so $u(\tau, x)$ is the bond pricing function.

PDE for Affine Term Structure.
Derivatives used in PDE:

$$\frac{\partial P}{\partial \tau} = \left(A'(\tau) + B'(\tau)^\top x\right) P, \quad \nabla P = B(\tau)\, P, \quad \nabla^2 P = B(\tau) B(\tau)^\top\, P$$

Substituting into the PDE:

$$-\left(A' + B'^\top x\right) P + \mu(x)^\top B\, P + \tfrac{1}{2}\mathrm{Tr}\left[\Sigma(x)\Sigma(x)^\top BB^\top\right] P - r(x)\, P = 0$$

Dividing through by $P(\tau, x)$ and using affine forms:

$$\mu(x) = A_0 + A_1^\top x, \quad \Sigma(x)\Sigma(x)^\top = B_0 + B_1^\top x, \quad r(x) = \alpha_0 + \alpha_1^\top x$$

yields:

$$-A' - B'^\top x + B^\top A_0 + B^\top A_1^\top x + \tfrac{1}{2}\mathrm{Tr}(B_0 BB^\top) + \tfrac{1}{2}\mathrm{Tr}(B_1^\top x\, BB^\top) - \alpha_0 - \alpha_1^\top x = 0$$

Deriving Riccati Equations.
Group terms by powers of x:

$$\begin{cases} -A'(\tau) + B(\tau)^\top A_0 + \frac{1}{2}\mathrm{Tr}\left(B_0 B B^\top\right) - \alpha_0 = 0 & \text{(constant)} \\ -B'(\tau)^\top + B(\tau)^\top A_1^\top + \frac{1}{2}\mathrm{Tr}\left(B_1^\top B B^\top\right) - \alpha_1^\top = 0 & \text{(linear in } x\text{)} \end{cases}$$

These equations define the Riccati system for $A(\tau)$ and $B(\tau)$:

$$\begin{cases} B'(\tau) & = -\alpha_1 + A_1 B(\tau) + \frac{1}{2}\mathrm{Tr}\left(B_1^\top B(\tau) B(\tau)^\top\right) \\ A'(\tau) & = -\alpha_0 + A_0^\top B(\tau) + \frac{1}{2}\mathrm{Tr}\left(B_0 B(\tau) B(\tau)^\top\right) \end{cases}$$

with initial conditions: $P(0,x) = 1 \Leftrightarrow \begin{cases} A(0) = 0 \\ B(0) = 0 \end{cases}$

Example: Vasicek Model.

$dX_t = \mu(X_t)dt + \Sigma(X_t)dW_t$, where $\mu(x) = \kappa(\theta - x)$, $\Sigma(x) = \sigma$, $r(X_t) = X_t$

Bond price is assumed to have the form: $P(\tau, X_t) = \exp\left(A(\tau) + B(\tau)X_t\right)$
Matching coefficients in the PDE gives:

$$\begin{cases} B'(\tau) = -1 - \kappa B(\tau), & B(0) = 0 \\ A'(\tau) = \kappa\theta B(\tau) + \frac{1}{2}\sigma^2 B(\tau)^2, & A(0) = 0 \end{cases}$$

Step 1: Solve the linear ODE:

$$B(\tau) = Ce^{-\kappa\tau} - \frac{1}{\kappa} \Rightarrow B(0) = 0 \Rightarrow C = \frac{1}{\kappa} \Rightarrow B(\tau) = \frac{1 - e^{-\kappa\tau}}{\kappa}$$

Step 2: Using $B(\tau)$ integrate $A'(\tau)$: $A(\tau) = \int_0^\tau \left[\kappa\theta B(s) + \frac{1}{2}\sigma^2 B(s)^2\right] ds$
Carrying out the integration gives:

$$A(\tau) = \left(\theta - \frac{\sigma^2}{2\kappa^2}\right)(B(\tau) - \tau) - \frac{\sigma^2}{4\kappa} B(\tau)^2$$

Risk-neutral measure $\mathbb{Q}$.

- Bond prices and yields are derived under the risk-neutral measure $\mathbb{Q}$
- Models ensure no-arbitrage pricing: $P(t, \tau) = \mathbb{E}^{\mathbb{Q}}[\dots]$

Historical measure $\mathbb{P}$.

- To model historical yield data, we work under $\mathbb{P}$
- We assume the risk-premia is affine in state:

$$\lambda(X_t) = \lambda_0 + \lambda_1^\top X_t$$

- By Girsanov's theorem we change the measure:

$$dW_t^{\mathbb{P}} = dW_t^{\mathbb{Q}} + \lambda(X_t)dt$$

- The drift of X_t changes:

$$\mu^{\mathbb{P}}(X_t) = \mu^{\mathbb{Q}}(X_t) + \Sigma(X_t) \cdot \lambda(X_t)$$

- The same latent process is interpreted statistically, not for pricing

Discrete-time setting.

- In discrete time, no Girsanov theorem is needed
- We only require the domains to be equal for equivalence of measures to switch from $\mathbb{P}$ to $\mathbb{Q}$
- Measure change is done via the Radon–Nikodym derivative:

$$\mathbb{E}^{\mathbb{Q}}[Z] = \mathbb{E}^{\mathbb{P}}\left[Z \cdot \frac{d\mathbb{Q}}{d\mathbb{P}}\right]$$

Final Model.
We arrive to the following system, that further will be calibrated and filtered on historical data.

$$\begin{cases} dX_t \;=\; \mu^{\mathbb{P}}(X_t)dt \;+\; \Sigma^{\mathbb{P}}(X_t)\, dW_t, \\ y(t,\tau) = -\dfrac{A^{\mathbb{Q}}(\tau)}{\tau} \;-\; \dfrac{B^{\mathbb{Q}}(\tau)^{\top}}{\tau} X_t + v_t, \end{cases}$$

where

- X_t is the affine latent state precess,
- $y(t,\tau)$ is a set of observed zero-coupon yields for different τ,
- $A^{\mathbb{Q}}(\tau), B^{\mathbb{Q}}(\tau)$ defined by the latent dynamics parameters in risk-neutral measure,
- v_t is the measurement noise.

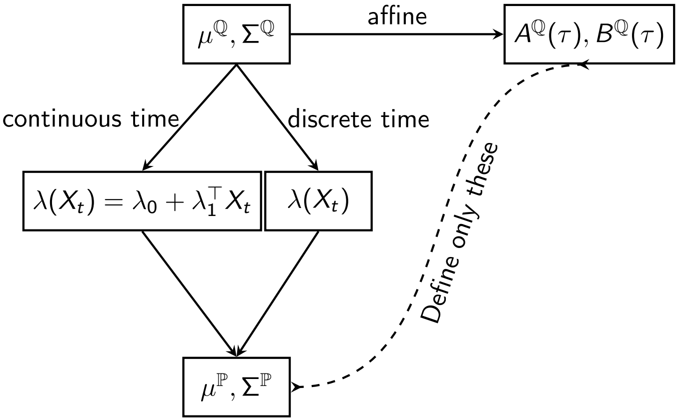

Figure: Structure of model parameterization.

Definition (Filtering)

Let X_t be a latent process and Y_t the observable process. Define the observation filtration as $\mathcal{G}_t = \sigma(Y_s,\ s \leq t)$.
The goal is to compute the filtering distribution: $\pi_t(\cdot) = \mathbb{P}(X_t \in \cdot \mid \mathcal{G}_t)$, or estimate $\widehat{X}_t = \arg\min_{Z \in L^2(\Omega, \mathcal{G}_t, \mathbb{P})} \mathbb{E}[|X_t - Z|^2] = \mathbb{E}[X_t \mid \mathcal{G}_t]$.

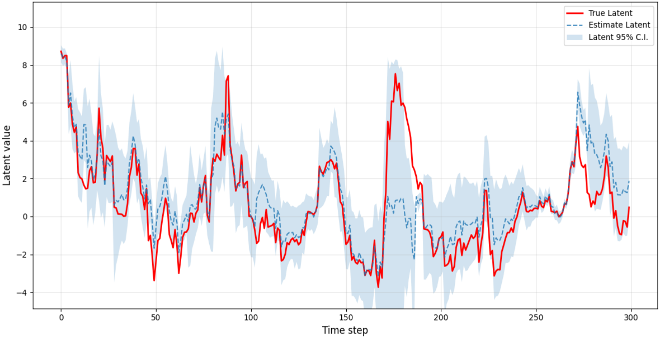

Figure: Latent state filtering.

Definition (State-Space Model (SSM))

A state-space model is a probabilistic model for two stochastic processes $(X_t)_{t \geq 0}$ and $(Y_t)_{t \geq 1}$, where:

- (X_t) is a Markov chain with initial distribution and transition dynamics: $p(x_0)$, $p(x_t \mid x_{t-1})$,
- (Y_t) are observations conditionally independent given the states: $p(y_t \mid x_{0:t}, y_{1:t-1}) = p(y_t \mid x_t)$.

Probabilistic model: $p(x_{0:T}, y_{1:T}) = p(x_0) \prod_{t=1}^{T} p(x_t \mid x_{t-1})\, p(y_t \mid x_t)$

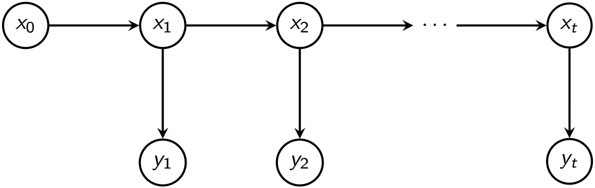

Figure: Structure of a general state-space model.

Linear, time-varying Gaussian SSM: Kalman Filter (Optimal).

$$\begin{cases} x_{t+1} = A_t\, x_t + G_t\, w_t, \\ y_t = H_t\, x_t + v_t, \\ w_t \sim \mathcal{N}(0, G_t),\;\; v_t \sim \mathcal{N}(0, R_t),\;\; \text{no autocorrelation} \end{cases}$$

Non-linear, time-varying Gaussian SSM: Linearized, Extended, Unscented (...) Filters.

$$\begin{cases} x_{t+1} = f_t(x_t) + w_t, \\ y_t = h_t(x_t) + v_t, \\ w_t \sim \mathcal{N}(0, G_t),\;\; v_t \sim \mathcal{N}(0, R_t),\;\; \text{no autocorrelation} \end{cases}$$

Non-linear, time-varying non-Gaussian SSM: Bayes, Particle Filters.

$$\begin{cases} x_{t+1} = f_t(x_t, w_t), \\ y_t = h_t(x_t, v_t), \\ w_t, v_t \;\text{any distributions, no autocorrelation} \end{cases}$$

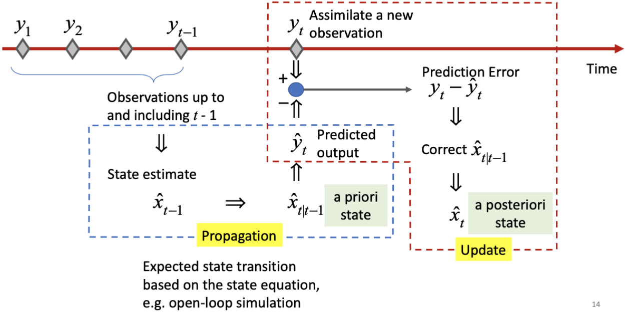

Figure: Recursive structure of the Kalman Filter

Propagation Step:

$$\begin{cases} \hat{x}_{t|t-1} = A_{t-1}\hat{x}_{t-1} \Rightarrow \hat{y}_{t|t-1} = H_t\hat{x}_{t|t-1} \\ P_{t|t-1} = A_{t-1}P_{t-1|t-1}A_{t-1}^\top + G_{t-1}Q_{t-1}G_{t-1}^\top \end{cases}$$

Update Step:

$$\begin{cases} \hat{x}_t = \hat{x}_{t|t-1} + K_t(y_t - \hat{y}_{t|t-1}) \\ P_t = (I - K_tH_t)P_{t|t-1} \end{cases}$$

Kalman Gain:

$$K_t = P_{t|t-1}H_t^\top \left(H_tP_{t|t-1}H_t^\top + R_t\right)^{-1}$$

$$\begin{tikzcd}
(\hat{x}_{t-1}, P_{t-1}) \arrow[r] \arrow[d] & (\hat{x}_{t|t-1}, P_{t|t-1}) \arrow[r] \arrow[d] & (\hat{x}_t, P_t) \arrow[d] & \textbf{Kalman Filter} \\
p(x_{t-1} \mid y_{1:t-1}) \arrow[r, "\text{Propagate}"'] & p(x_t \mid y_{1:t-1}) \arrow[r, "\text{Update}"'] & p(x_t \mid y_{1:t}) & \textbf{Bayes Filter}
\end{tikzcd}$$

at time $t-1$ — a priori estimate — a posteriori estimate

Figure: From Kalman Filter to Bayes Filter.

Propagation step: Chapman-Kolmogorov equation

$$p(x_t \mid y_{1:t-1}) = \int p(x_t \mid x_{t-1})\, p(x_{t-1} \mid y_{1:t-1})\, dx_{t-1}$$

Update step: Bayes formula

$$p(x_t \mid y_{1:t}) = \frac{p(y_t \mid x_t)\, p(x_t \mid y_{1:t-1})}{\int p(y_t \mid x_t)\, p(x_t \mid y_{1:t-1})\, dx_t}$$

1. Propagation: Proposal distribution

$$x_t^{(i)} \sim g\left(x_t \mid x_{t-1}^{(i)}, y_t\right)$$

2. Update: Sequential Importance Sampling

$$W_t^{(i)} \propto W_{t-1}^{(i)} \frac{p(y_t \mid x_t^{(i)})\, p(x_t^{(i)} \mid x_{t-1}^{(i)})}{g(x_t^{(i)} \mid x_{t-1}^{(i)}, y_t)}$$

3. Update: Sequential Importance Resampling

- Compute and compare effective sample size:

$$ESS(W_t^{1:N}) \quad \text{vs} \quad \alpha N$$

- Resample if necessary:

$$x_t^{(i)} \sim \sum_{j=1}^{N} W_t^{(j)} \delta_{x_t^{(j)}}$$

- Reset weights: $W_t^{(i)} = \frac{1}{N}$

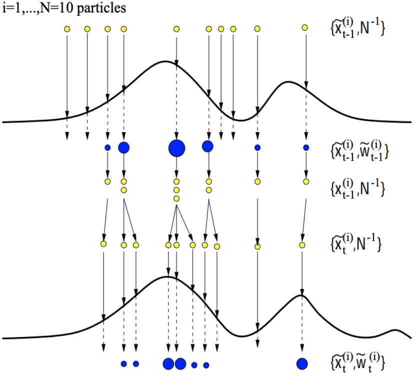

Figure: Dynamics of particles.

Bootstrap Particle Filter:

State dynamic is chosen as a proposal

$$g(x_t \mid x_{t-1}) = p(x_t \mid x_{t-1}) \Longrightarrow W_t^{(i)} \propto W_{t-1}^{(i)} p(y_t \mid x_t^{(i)}).$$

Guided Particle Filter:

Proposal depends on current observation, focuses sampling on high-likelihood regions $g^*(x_t \mid x_{t-1}, y_t) = p(x_t \mid x_{t-1}, y_t)$.

Auxiliary Particle Filter:

Predictive auxiliary weights

$$\eta_t^*(x_{t-1}^{(i)}) = \mathbb{E}\left[p(y_t \mid x_t) \mid x_{t-1}^{(i)}\right] = \int p(y_t \mid x_t)\, p(x_t \mid x_{t-1}^{(i)})\, dx_t$$

Pre-selection step: Resample particles from $\{x_{t-1}^{(i)}\}$ with probability proportional to $W_{t-1}^{(i)} \eta_t(x_{t-1}^{(i)})$, focusing on particles that are expected to generate high-likelihood observations.

Computational Complexity:

- Per step: $\mathcal{O}(N) \Rightarrow$ Over T steps: $\mathcal{O}(NT)$.
- Parallelizable across particles.

Filtering Estimates:

- Empirical approximation of $p(x_{1:t} \mid y_{1:t}) \approx \sum_{i=1}^{N} w_t^{(i)} x_{1:t}^{(i)}$ is unbiased.
- For test functions f, the Monte Carlo error:

$$\left| \mathbb{E}[f(x_{1:t}) \mid y_{1:t}] - \sum_{i=1}^{N} w_t^{(i)} f(x_{1:t}^{(i)}) \right| = \mathcal{O}\left(\frac{1}{\sqrt{N}} \right)$$

Likelihood Estimation:

- The particle filter implicitly approximates the marginal likelihood:

$$p(y_{1:T}) = \int p(x_{1:T}, y_{1:T})\, dx_{1:T} \approx \frac{1}{N} \sum_{i=1}^{N} W_T^{(i)}.$$

- This estimator is unbiased.

Problem

We aim to simultaneously infer both: the latent state trajectory $(x_{0:T})$ and unknown model parameters θ, given a sequence of observations $y_{1:T}$.

$$p(\theta, x_{0:T}, y_{1:T}) = \underbrace{p(y_{1:T} \mid x_{0:T}, \theta)}_{\text{data distribution}} \underbrace{p(x_{0:T} \mid \theta) p(\theta)}_{\text{prior}} =$$

$$= \underbrace{\prod_{t=1}^{T} \underbrace{p(y_t \mid x_t, \theta)}_{\text{observation}}}_{\text{data distribution}} \underbrace{\prod_{t=1}^{T} \underbrace{p(x_t \mid x_{t-1}, \theta)}_{\text{dynamics}} \underbrace{p(x_0 \mid \theta)}_{\text{initial state}} \underbrace{p(\theta)}_{\text{prior on parameters}}}_{\text{prior}}$$

State-of-the-art methods:

1. Particle Markov Chain Monte Carlo (PMCMC)
2. Nested Partcle Filter (NPF)
3. Sequantial Monte Carlo Squared (SMC^2)

Algorithm: Particle Metropolis–Hastings

1. **Initialize** $(m = 1)$: set $\theta[1]$ and run a particle filter to compute $\hat{Z}[1]$.
2. **For** $m = 2$ to M, iterate:
   1. Sample $\theta' \sim q(\theta \mid \theta[m])$.
   2. Sample $\hat{Z}' \sim \psi(\hat{Z} \mid \theta')$ (run a particle filter and compute likelihood estimate)
   3. With probability

$$\alpha = \min\left(1, \frac{\hat{Z}'}{\hat{Z}[m]} \cdot \frac{p(\theta')}{p(\theta[m])} \cdot \frac{q(\theta[m] \mid \theta')}{q(\theta' \mid \theta[m])}\right)$$

set $\{\theta[m+1], \hat{Z}[m+1]\} \leftarrow \{\theta', \hat{Z}'\}$ (accept candidate sample) and with probability $1 - \alpha$ set $\{\theta[m+1], \hat{Z}[m+1]\} \leftarrow \{\theta[m], \hat{Z}[m]\}$ (reject candidate sample).

Auxiliary variable $\hat{Z}$. $\mathbb{E}_{\psi(\hat{Z}|\theta)}[\hat{Z}] = p(y_{1:T} \mid \theta), \quad \hat{Z} \geq 0.$

Pseudo-marginal target distribution. $\pi(\theta \mid y_{1:T}) \propto p(y_{1:T} \mid \theta)\, p(\theta)$.

Since $p(y_{1:T} \mid \theta)$ is intractable, we define a joint target:

$$\pi(\theta \mid y_{1:T}) = \int \hat{Z} \cdot \psi(\hat{Z} \mid \theta) \cdot \frac{p(\theta)}{p(y_{1:T})} d\hat{Z} = \int \pi(\theta, \hat{Z} \mid y_{1:T}) d\hat{Z}.$$

Proposal kernel:

$$q(\theta', \hat{Z}' \mid \theta[m]) = \psi(\hat{Z}' \mid \theta') \cdot q(\theta' \mid \theta[m]).$$

Metropolis–Hastings acceptance ratio:

$$\begin{aligned}
\alpha &= \min\left(1, \frac{\pi(\theta', \hat{Z}') \cdot q(\theta[m], \hat{Z}[m] \mid \theta')}{\pi(\theta[m], \hat{Z}[m]) \cdot q(\theta', \hat{Z}' \mid \theta[m])}\right) \\
&= \min\left(1, \frac{\hat{Z}' \cdot \psi(\hat{Z}' \mid \theta') \cdot p(\theta') \cdot \psi(\hat{Z}[m] \mid \theta[m]) \cdot q(\theta[m] \mid \theta')}{\hat{Z}[m] \cdot \psi(\hat{Z}[m] \mid \theta[m]) \cdot p(\theta[m]) \cdot \psi(\hat{Z}' \mid \theta') \cdot q(\theta' \mid \theta[m])}\right) \\
&= \min\left(1, \frac{\hat{Z}'}{\hat{Z}[m]} \cdot \frac{p(\theta')}{p(\theta[m])} \cdot \frac{q(\theta[m] \mid \theta')}{q(\theta' \mid \theta[m])}\right)
\end{aligned}$$

Algorithm: Nested Particle Filter

For $i = 1, \ldots, N_{\boldsymbol{\theta}}$:

- Jittering: Draw $\bar{\boldsymbol{\theta}}^i_t \sim \kappa_{N_{\boldsymbol{\theta}}}(d\boldsymbol{\theta}|\boldsymbol{\theta}^i_{t-1})$

SMC ($N_{\boldsymbol{\theta}}$ samples) to approximate $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})$

Given $\bar{\boldsymbol{\theta}}^i_t$, for $j = 1, \ldots, N_{\boldsymbol{x}}$:

- Draw $\bar{\boldsymbol{x}}^{i,j}_t \sim p(\boldsymbol{x}_t|\bar{\boldsymbol{\theta}}^i_t, \boldsymbol{y}_{1:t-1})$
- Weights: $\tilde{u}^{i,j}_t \propto p(\boldsymbol{y}_t|\bar{\boldsymbol{x}}^{i,j}_t, \bar{\boldsymbol{\theta}}^i_t)$
- Resampling: for $m = 1, \ldots, N_{\boldsymbol{x}}$, $\tilde{\boldsymbol{x}}^{i,j}_t = \bar{\boldsymbol{x}}^{i,m}_t$ with prob. $u^{i,m}_t = \frac{\tilde{u}^{i,m}_t}{\sum_{j=1}^{N_{\boldsymbol{x}}} \tilde{u}^{i,j}_t}$

SMC ($N_{\boldsymbol{x}}$ samples) to approximate $p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \bar{\boldsymbol{\theta}}^i_t)$

- Likelihood of $\bar{\boldsymbol{\theta}}^i_t$: $\tilde{w}^i_t = \frac{1}{N_{\boldsymbol{x}}} \sum_{j=1}^{N_{\boldsymbol{x}}} \tilde{u}^{i,j}_t$
- Resampling: for $l = 1, \ldots, N_{\boldsymbol{\theta}}$, $\{\boldsymbol{\theta}^l_t, \{\boldsymbol{x}^{l,j}_t\}_{1\le j\le N_{\boldsymbol{x}}}\} = \{\bar{\boldsymbol{\theta}}^l_t, \{\tilde{\boldsymbol{x}}^{l,j}_t\}_{1\le j\le N_{\boldsymbol{x}}}\}$ with prob. w^l_t, so that $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t}) = \frac{1}{N_{\boldsymbol{\theta}}} \sum_{i=1}^{N_{\boldsymbol{\theta}}} \delta_{\boldsymbol{\theta}^i_t}(d\boldsymbol{\theta})$

Jittering step.
To propagate $\theta_t^{(i)}$ over time, we apply a kernel:

$$\bar{\theta}_t^{(i)} \sim \kappa_{N_\theta}(d\theta \mid \theta_{t-1}^{(i)})$$

Common choice is the locally adaptive Gaussian Random Walk:

$$\bar{\theta}_t^{(i)} = \theta_{t-1}^{(i)} + \epsilon_t^{(i)}, \quad \epsilon_t^{(i)} \sim \mathcal{N}\left(0, h^2 \cdot \mathrm{Var}[\theta_{t-1}^{(1:N_\theta)}]\right)$$

SMC^2 vs. NPF:

- SMC^2: jittering is done via a particle MCMC.
- NPF: jittering is performed via a proposal (e.g. random walk).

Intuition. This step plays the same role as the propagation step for latent states. It introduces artificial dynamics on static parameters θ, making the particle approximation sequentially updatable over time.

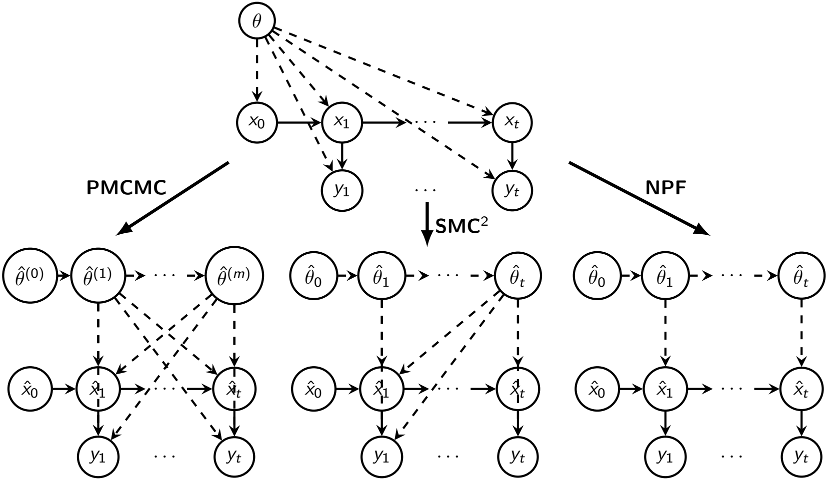

Figure: Comparison of PMCMC, SMC^2 and NPF structures.

Property	PMCMC	SMC^2	NPF
Time Complexity	$\mathcal{O}(M \cdot N_x \cdot T)$	$\mathcal{O}(N_\theta \cdot N_x \cdot T^2)$	$\mathcal{O}(N_\theta \cdot N_x \cdot T)$
Convergence Rate	$\mathcal{O}\left(\frac{1}{\sqrt{M}}\right)$	$\mathcal{O}\left(\frac{1}{\sqrt{N_x}}\right)$	$\mathcal{O}\left(\frac{1}{\sqrt{N_x}} + \frac{1}{\sqrt{N_\theta}}\right)$
Recursive	No	Partial	Yes
Parallelization	No	Across θ-particles	Across θ-particles

Table: Comparison of PMCMC, SMC^2 and NPF properties.