

Bayesian Scientific Computing

Day 5

Daniela Calvetti, Erkki Somersalo

Case Western Reserve University
Department of Mathematics, Applied Mathematics and Statistics

Lyngby, December 2019

Dynamic Inverse Problems

- The unknown of primary interest is a time-dependent random variable (stochastic process)

$$X_t, \quad t \geq 0.$$

- Observations at discrete time instances. Assuming additive noise,

$$B_j = G(t_j, X_{t_j}) + \varepsilon_j, \quad t_1 < t_2 < \dots$$

- The prior information may include an evolution model, e.g., an ordinary or stochastic differential equation.

Inverse problem: Estimate X_t through the posterior model

$$\pi_t(x_t | \mathcal{B}_t), \quad \mathcal{B}_t = \{b_1, \dots, b_j \mid t_j \leq t < t_{j+1}\}.$$

Discrete time model

Discrete time stochastic process

$$X_0, X_1, \dots, \quad X_0 \sim \pi_0(x_0).$$

with stochastic evolution model

$$X_{t+1} = F_t(X_t) + V_{t+1}, \quad V_t = \text{innovation},$$

Observation model

$$B_t = G(X_t) + W_t.$$

Posterior density

$$\pi(X_t | \mathcal{B}_t), \quad \mathcal{B}_t = \{b_1, b_2, \dots, b_t\}.$$

Introductory Example

Description of the model:

- A hunter walks in a forest with speed $v \leq v_{\max}$
- At $t = 0$, the hunter is at $x = (0, 0)$.
- The dog of the hunter is at the distance $d \leq d_{\max}$ from the hunter
- The dog is observed at discrete times $t_1 < t_2 < \dots t_m$.

Problem: *Estimate the trajectory of the hunter.*

Introductory Example

Particle approach: For each time step t_j , generate a sample of possible positions of the hunter,

$$\mathcal{S}_j = \{x_j^1, x_j^2, \dots, x_j^N\}.$$

Modeling step by step: At $t = t_0 = 0$, we know *with certainty* that the hunter is at the origin. Therefore,

$$\mathcal{S}_0 = \{x_0^1, x_0^2, \dots, x_0^N\},$$

where

$$x_0^k = (0, 0), \quad 1 \leq k \leq N.$$

Assign to each *particle* the same probability,

$$w_0^k = \frac{1}{N}, \quad 1 \leq k \leq N.$$

Introductory Example

At $t = t_1$, before observing the dog, where is the hunter?

Propagate each particle:

$$\hat{x}_1^k = x_0^k + w_1^k, \quad 1 \leq k \leq N,$$

where w_1^k is a random step. It must satisfy

$$\|w_1^k\| \leq v_{\max}(t_1 - t_0) = \gamma_1.$$

The direction of the step is arbitrary, so we choose a model

$$W_1 = \gamma_1 S \begin{bmatrix} \cos \Theta \\ \sin \Theta \end{bmatrix},$$

where

$$S \sim \text{Uniform}([0, 1]), \quad \Theta \sim \text{Uniform}([0, 2\pi]).$$

Introductory Example

Predictive sample, based on no data:

$$\widehat{\mathcal{S}}_1 = \{\widehat{x}_1^1, \widehat{x}_j^2, \dots, \widehat{x}_1^N\}.$$

Now the first observation of the dog arrives. Denote it by $b_1 \in \mathbb{R}^2$.

Likelihood (prob. distribution of the position of the dog, assuming that we know where the hunter is at time $t = 1$):

$$\pi_{B_1|x_1}(b_1 | x_1) \propto \chi_{D(x_1, d_{\max})}(b_1) = \begin{cases} 1, & \text{if } \|x_t - b_1\| \leq d_{\max}, \\ 0 & \text{otherwise.} \end{cases}$$

Above,

$$D(x_1, d_{\max}) = \{b \in \mathbb{R}^2 \mid \|b - x_1\| \leq d_{\max}\}.$$



Observing the dog.

Introductory Example

To each \hat{x}_1^k , assign a weight:

$$w_1^k = \pi_{B_1|X_1}(b_1 | \hat{x}_1^k), \quad w_1^k \leftarrow \frac{w_1^k}{\sum_{\ell=1}^N w_1^\ell}.$$

The weight w_1^k expresses the likelihood of b_1 to take place if \hat{x}_1^k is the position of the hunter.

With the current likelihood model,

$$\pi_{B_1|X_1}(b_1 | \hat{x}_1^k) = \begin{cases} 1, & \text{if } \|\hat{x}_1^k - b_1\| \leq d_{\max}, \\ 0 & \text{otherwise.} \end{cases}$$

Introductory Example

Resampling: From the sample

$$\{(\hat{x}_1^1, w_1^1), (\hat{x}_1^2, w_1^2), \dots, (\hat{x}_1^N, w_1^N)\},$$

draw with replacement N new particles. This is the new sample:

$$\mathcal{S}_1 = \{x_1^1, x_1^2, \dots, x_1^N\},$$

Observe: Typically, some particles \hat{x}_1^k are repeated several times, while some other particles do not appear at all.

Introductory Example

Estimating the hunter's position at $t = t_1$: Set

$$\bar{x}_1 = \frac{1}{N} \sum_{k=1}^N x_1^k.$$



Now we can repeat the process again, to find $\bar{x}_2, \bar{x}_3, \dots$



Spotting the hunter (or dog-owner).

Introductory Example

The likelihood we used was harsh in the sense that it gave no chance to some particles.

Softer likelihood: Replace the densities with Gaussian approximations.

At $t = t_1$, before observing the dog, where is the hunter?

Propagate each particle:

$$\hat{x}_1^k = x_0^k + w_1^k, \quad 1 \leq k \leq N,$$

where w_1^k is a realization of W_1 ,

$$W_1 \sim \mathcal{N}(0, \gamma_1^2 I_2).$$

Introductory Example

The first observation arrives: Use a Gaussian likelihood model,

$$\pi_{B_1|x_1}(b_1 | x_1) \propto \exp\left(-\frac{1}{2d_0^2} \|b_1 - x_1\|^2\right).$$

Assign likelihoods to proposed particles:

$$w_1^k = \exp\left(-\frac{1}{2d_0^2} \|b_1 - \hat{x}_1^k\|^2\right), \quad w_1^k \leftarrow \frac{w_1^k}{\sum_{\ell=1}^N w_1^\ell}.$$

Introductory Example

Importance Resampling: From the sample

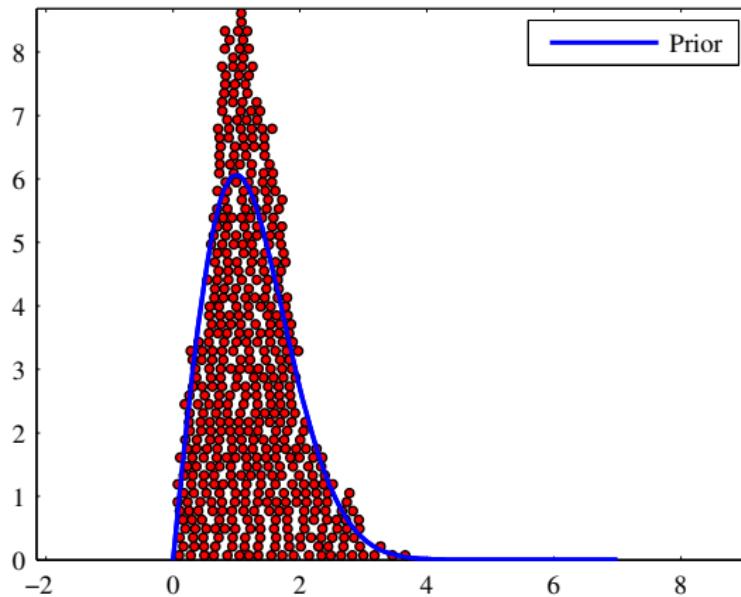
$$\{(\hat{x}_1^1, w_1^1), (\hat{x}_1^2, w_1^2), \dots, (\hat{x}_1^N, w_1^N)\},$$

draw with replacement N new particles. This is the new sample:

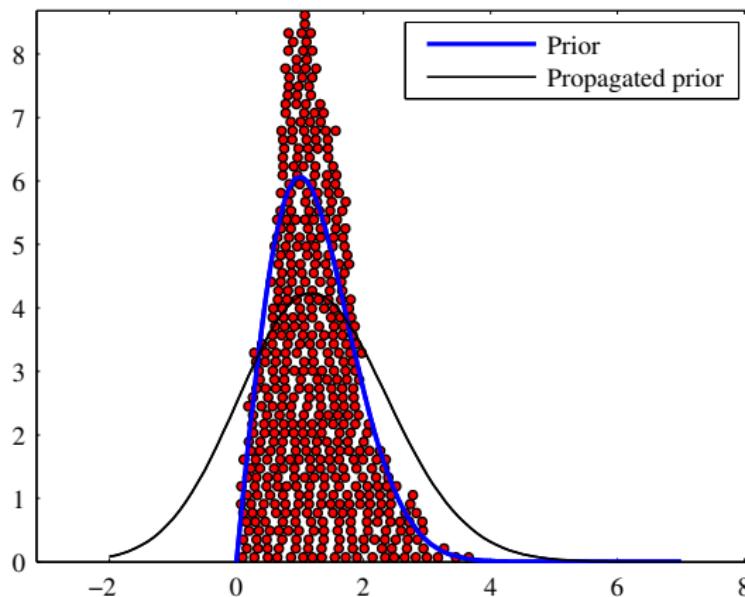
$$\mathcal{S}_1 = \{x_1^1, x_1^2, \dots, x_1^N\},$$

Repeat with $j \leftarrow j + 1$.

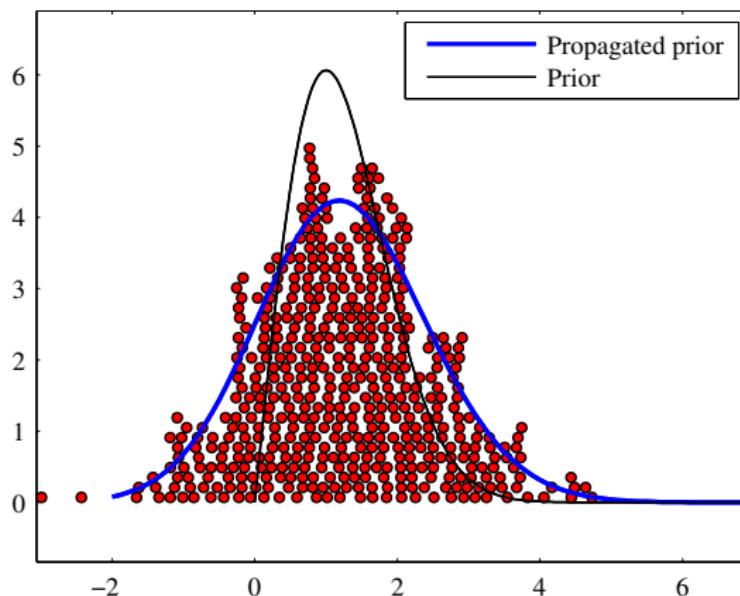
Particle visualization



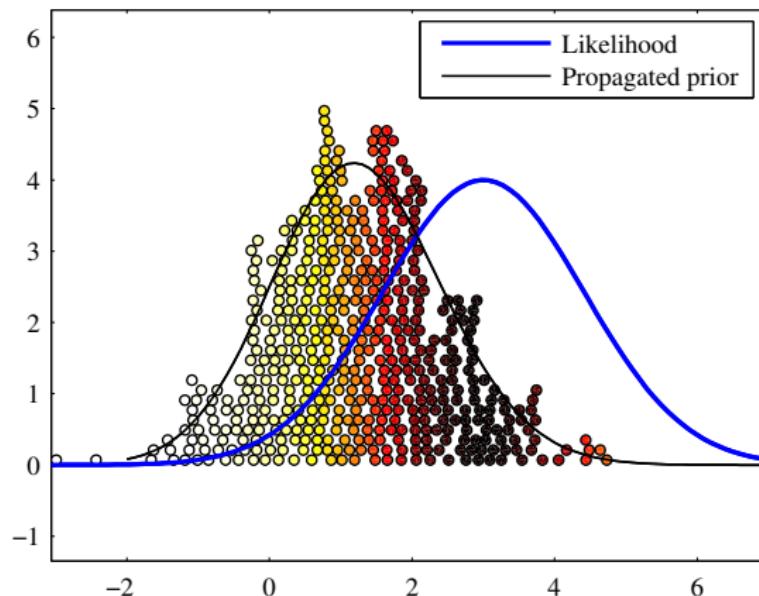
Particle visualization



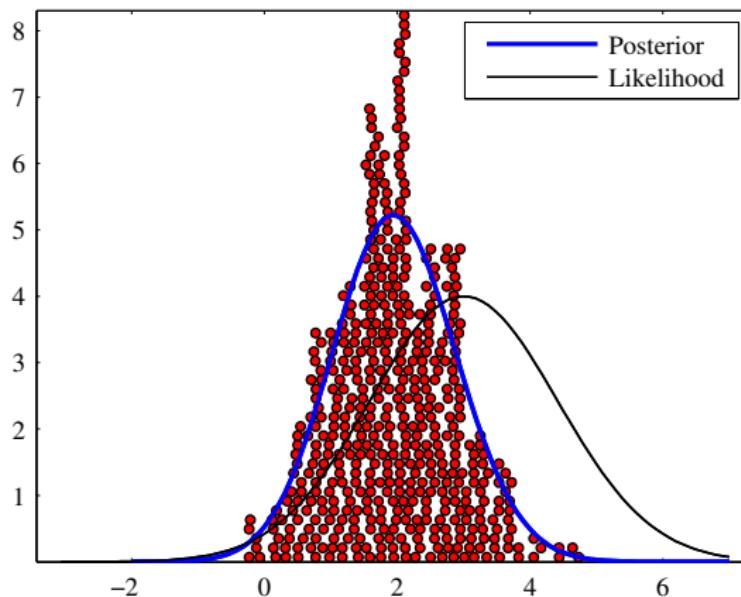
Particle visualization



Particle visualization



Particle visualization



Gaussian Model and Kalman Filter

For the Gaussian approximation, explicit formulas are available:

Recall: \hat{x}_1^k is a realization of

$$\hat{X}_1 = x_0 + W_1, \quad W_1 \sim \mathcal{N}(0, \gamma_1^2 I_2).$$

Therefore,

$$\pi_{\hat{X}_1}(\hat{x}_1) \propto \exp\left(-\frac{1}{2\gamma_1^2} \|\hat{x}_1 - x_0\|^2\right).$$

Use this as a prior when the observation comes!

Gaussian Model and Kalman Filter

Likelihood:

$$\pi_{B_1|X_1}(b_1 | x_1) \propto \exp\left(-\frac{1}{2d_0^2} \|b_1 - x_1\|^2\right).$$

Using Bayes' formula, write the posterior:

$$\pi_{X_1|B_1}(x_1 | b_1) \propto \pi_{\widehat{X}_1}(x_1) \pi_{B_1|X_1}(b_1 | x_1),$$

or explicitly,

$$\pi_{X_1|B_1}(x_1 | b_1) \propto \exp\left(-\frac{1}{2\gamma_1^2} \|x_1 - x_0\|^2 - \frac{1}{2d_0^2} \|b_1 - x_1\|^2\right).$$

Gaussian Model and Kalman Filter

The quadratic term in the exponent can be written as

$$\begin{aligned} & \frac{1}{2\gamma_1^2} \|x_1 - x_0\|^2 + \frac{1}{2d_0^2} \|b_1 - x_1\|^2 \\ &= \left(\frac{1}{2\gamma_1^2} + \frac{1}{2d_0^2} \right) \|x_1\|^2 - 2 \left(\frac{1}{2\gamma_1^2} x_0 + \frac{1}{2d_0^2} b_1 \right)^T x_1 + \dots \end{aligned}$$

Denote

$$\frac{1}{\gamma_2^2} = \left(\frac{1}{\gamma_1^2} + \frac{1}{d_0^2} \right),$$

to simplify the expression as

$$\frac{1}{2\gamma_2^2} \left(\|x_1\|^2 - 2 \left(\frac{\gamma_2^2}{\gamma_1^2} x_0 + \frac{\gamma_2^2}{d_0^2} b_1 \right)^T x_1 + \dots \right)$$

Gaussian Model and Kalman Filter

Conclusion: The posterior density is Gaussian,

$$\pi_{X_1|B_1}(x_1 | b_1) \sim \mathcal{N}(\bar{x}_1^2, \gamma_2^2 l_2),$$

where the mean is

$$\boxed{\bar{x}_1 = \frac{\gamma_2^2}{\gamma_1^2} x_0 + \frac{\gamma_2^2}{d_0^2} b_1},$$

a weighted mean of the a priori mean x_0 and the position of the dog b_1 , and the variance is the harmonic mean of the prior and likelihood variances,

$$\boxed{\frac{1}{\gamma_2^2} = \left(\frac{1}{\gamma_1^2} + \frac{1}{d_0^2} \right)}.$$

This is a particular case of *Kalman Filtering!*

Bayes filtering, basic form

Evolution–observation model:

$$X_{j+1} = F_j(X_j) + V_{j+1}, \quad j = 0, 1, 2, \dots$$

$$B_j = G_j(X_j) + E_j, \quad j = 1, 2, \dots$$

Observations, or data:

$$B_j = b_j, \quad j = 1, 2, \dots$$

We assume further that the prior probability density of X_0 is given.

Filtering algorithm

The goal is to design an algorithm along the following lines:

- Given the density of X_0 , **predict** the density of X_1 using the evolution model,
- Using the predicted density of X_1 as prior, **update** the posterior density $\pi_{X_1|B_1}(x_1 | b_1)$,
- Using the posterior density $\pi_{X_1|B_1}(x_1 | b_1)$, **predict** the density of X_2 ,
- Using the predicted density of X_2 as prior, **update** the posterior density $\pi_{X_2|B_2}(x_2 | b_2)$,
- Continue inductively.

Kalman filtering

Evolution–observation model:

$$X_{j+1} = F X_j + V_{j+1}, \quad j = 0, 1, 2, \dots$$

$$B_j = A X_j + E_j, \quad j = 1, 2, \dots$$

Assumptions about the noise processes and the initial process:

- ① Normality:

$$V_j \sim \mathcal{N}(0, \Gamma_j), \quad E_j \sim \mathcal{N}(0, \Sigma_j).$$

- ② Independence: Variables V_j, E_j , all mutually independent.

- ③ Initial density:

$$X_0 \sim \mathcal{N}(\bar{x}_0, D_0),$$

and X_0 is independent of the noise processes.

Propagation

Note: To completely specify a Gaussian density, it is enough to know the mean and the variance.

Propagation: Assume that

$$X_j \sim \mathcal{N}(\bar{x}_j, D_j).$$

① **Mean:** We have

$$X_{j+1} = F X_j + V_{j+1},$$

implying that the mean is

$$\begin{aligned}\bar{x}_{j+1} &= E\{X_{j+1}\} = F E\{X_j\} + E\{V_{j+1}\} \\ &= F \bar{x}_j.\end{aligned}$$

Hence: Propagate the mean with the propagator F .

Propagation

② **Covariance:** Since

$$X_{j+1} - \bar{x}_{j+1} = F(X_j - \bar{x}_j) + V_{j+1},$$

by independency,

$$\begin{aligned} E\{(X_{j+1} - \bar{x}_{j+1})(X_{j+1} - \bar{x}_{j+1})^T\} \\ &= E\{(F(X_j - \bar{x}_j) + V_{j+1})(F(X_j - \bar{x}_j) + V_{j+1})^T\} \\ &= E\{(F(X_j - \bar{x}_j)(X_j - \bar{x}_j)^T F^T\} + E\{V_{j+1} V_{j+1}^T\} \\ &= FD_j F^T + \Gamma_{j+1}. \end{aligned}$$

Hence, after propagation,

$$X_{j+1} \sim \mathcal{N}(F\bar{x}_j, FD_j F^T + \Gamma_{j+1}).$$

Correction

Updating: Consider a linear inverse problem,

$$B = AX + E,$$

where

$$X \sim \mathcal{N}(\bar{x}, D), \quad E \sim \mathcal{N}(0, \Sigma).$$

To find the posterior density $\pi_{X|B}(x | b)$ we use the formulas for the Gaussian posterior:



$$\begin{aligned}\bar{x}_{\text{post}} &= \bar{x} + DA^T(ADA^T + \Sigma)^{-1}b, \\ C_{\text{post}} &= D - DA^T(ADA^T + \Sigma)^{-1}AD.\end{aligned}$$

Kalman filtering

Putting the steps together:

- ① Initialize: $j = 0$, x_0 and D_0 given.
- ② **Prediction step:** Calculate

$$\bar{x}_{j+1} = Fx_j,$$

$$\hat{D}_{j+1} = FD_jF^T + \Gamma_{j+1}.$$

- ③ *Updating step:* Calculate

$$\bar{x}_{j+1} = \hat{x}_{j+1} + \hat{D}_{j+1}A^T(A\hat{D}_{j+1}A^T + \Sigma_{j+1})^{-1}(b_{j+1} - F\bar{x}_{j+1}),$$

$$D_{j+1} = \hat{D}_{j+1} - \hat{D}_{j+1}A^T(A\hat{D}_{j+1}A^T + \Sigma_{j+1})^{-1}A\hat{D}_{j+1}.$$

- ④ Increase j by one and repeat from 2.

Kalman filtering

In the above formulas,

$$\Delta_{j+1} = b_{j+1} - A\bar{x}_{j+1} = \text{prediction error},$$

expressing how much the propagated mean value differs from the actual data.

$$K_{j+1} = \hat{D}_{j+1} A^T (A \hat{D}_{j+1} A^T + \Sigma_{j+1})^{-1} = \text{Kalman gain matrix.}$$

With these notations, the updating formulas are written as

$$\bar{x}_{j+1} = \hat{x}_{j+1} + K_{j+1} \Delta_{j+1},$$

$$D_{j+1} = (I - K_{j+1} A) \hat{D}_{j+1}.$$

The discrete time Markov models framework

Extensions to non-linear, non-Gaussian models.

Evolution model:

$$X_{j+1} = F(X_j, \theta) + V_{j+1},$$

- F is a known propagation model
- V_{j+1} is an innovation process
- θ is a parameter: assumed known now, later to be estimated.

The observation model

$$Y_j = G(X_j) + W_j,$$

the observation noise W_j independent of X_j .

Update scheme for posterior densities given accumulated data:

$$\pi(x_j | \mathcal{B}_j) \longrightarrow \pi(x_{j+1} | \mathcal{B}_j) \longrightarrow \pi(x_{j+1} | \mathcal{B}_{j+1}).$$

Bayesian filtering

In the following, we leave out the subindices of the densities.

① Propagation step: Chapman-Kolmogorov formula

$$\begin{aligned}\pi(x_{j+1} \mid \mathcal{B}_j) &= \int \pi(x_{j+1} \mid x_j, \mathcal{B}_j) \pi(x_j \mid \mathcal{B}_j) dx_j \\ &= \int \pi(x_{j+1} \mid x_j) \pi(x_j \mid \mathcal{B}_j) dx_j,\end{aligned}$$

② Analysis step: Bayes' formula conditional on \mathcal{B}_j

$$\begin{aligned}\pi(x_{j+1} \mid \mathcal{B}_{j+1}) &= \pi(x_{j+1} \mid b_{j+1}, \mathcal{B}_j) \\ &\propto \pi(b_{j+1} \mid x_{j+1}, \mathcal{B}_j) \pi(x_{j+1} \mid \mathcal{B}_j) \\ &= \pi(b_{j+1} \mid x_{j+1}) \pi(x_{j+1} \mid \mathcal{B}_j),\end{aligned}$$

Combining:

$$\pi(x_{j+1} \mid \mathcal{B}_{j+1}) \propto \pi(b_{j+1} \mid x_{j+1}) \int \pi(x_{j+1} \mid x_j) \pi(x_j \mid \mathcal{B}_j) dx_j.$$

Bayesian filtering

Assume that the current distribution is represented in terms of a sample

$$\mathcal{S}_j = \{(x_j^1, w_j^1), (x_j^2, w_j^2), \dots, (x_j^N, w_j^N)\} \sim \pi(x_j | \mathcal{B}_j).$$

The particle version (Monte Carlo integration)

$$\pi(x_{j+1} | \mathcal{B}_{j+1}) \propto \pi(b_{j+1} | x_{j+1}) \int \pi(x_{j+1} | x_j) \pi(x_j | \mathcal{B}_j) dx_j.$$

can be written as

$$\pi(x_{j+1} | \mathcal{B}_{j+1}) \propto \pi(b_{j+1} | x_{j+1}) \sum_{n=1}^N w_j^n \pi(x_{j+1} | x_j^n).$$

Sampling Importance Resampling (SIR)

Layered sampling: For $n = 1, 2, \dots, N$,

- ① Draw a candidate particle \hat{x}_{j+1}^n from $\pi(x_{j+1} | x_j^n)$;
- ② Compute the relative likelihood $g_{j+1}^n = \pi(b_{j+1} | \hat{x}_{j+1}^n)$;
- ③ Resample with replacement from

$$\{(\hat{x}_{j+1}^1, \hat{w}_{j+1}^1), (\hat{x}_{j+1}^2, \hat{w}_{j+1}^2), \dots, (\hat{x}_{j+1}^N, \hat{w}_{j+1}^N)\}, \quad \hat{w}_{j+1}^n = \frac{g_{j+1}^n}{\sum g_{j+1}^n}.$$

Data thinning:

- Most particles \hat{x}_{j+1}^n may have vanishingly small likelihood
- Few candidate particles are sampled over and over: The new sample consists mostly of copies of few candidate particles.
- The density is poorly sampled.

Improvement: Auxiliary particles

Before resampling, calculate an auxiliary predictor:

$$\bar{x}_{j+1}^n = F(x_j^n)$$


We write

$$\pi(x_{j+1} \mid D_{j+1}) \propto \sum_{n=1}^N w_j^n \underbrace{\pi(b_{j+1} \mid \bar{x}_{j+1}^n)}_{=g_{j+1}^n} \frac{\pi(b_{j+1} \mid x_{j+1})}{\pi(b_{j+1} \mid \bar{x}_{j+1}^n)} \pi(x_{j+1} \mid x_j^n),$$

The quantity g_{j+1}^n is a *predictor* of how well the auxiliary particle would explain the data.

Survival of the Fittest (SOF)

Given the initial probability density $\pi_0(x_0)$,

- ① *Initialization:* Draw the particle ensemble from $\pi_0(x_0)$:

$$\mathcal{S}_0 = \{(x_0^1, w_0^1), (x_0^2, w_0^2), \dots, (x_0^N, w_0^N)\},$$

$$w_0^1 = w_0^2 = \dots = w_0^N = \frac{1}{N}.$$

Set $j = 0$.

- ② *Propagation:* Compute the predictor:

$$\bar{x}_{j+1}^n = F(x_j^n), \quad 1 \leq n \leq N.$$

Survival of the Fittest (SOF)

- ③ *Survival of the fittest:* For each n :

- (a) Compute the fitness weights

$$g_{j+1}^n = w_j^n \pi(b_{j+1} \mid \bar{x}_{j+1}^n), \quad g_{j+1}^n \leftarrow \frac{g_{j+1}^n}{\sum_n g_{j+1}^n};$$

- (b) Draw indices with replacement $\ell_n \in \{1, 2, \dots, N\}$ using probabilities

$$\Pr\{\ell_n = k\} = g_{j+1}^k;$$

- (c) Reshuffle

$$x_j^n \leftarrow x_j^{\ell_n}, \quad \bar{x}_{j+1}^n \leftarrow \bar{x}_{j+1}^{\ell_n}, \quad 1 \leq n \leq N.$$

Survival of the Fittest (SOF)



- ④ *Innovation:* For each n : Proliferate

$$x_{j+1}^n = \bar{x}_{j+1}^n + v_{j+1}^n.$$

- ⑤ *Weight updating:* For each n , compute

$$w_{j+1}^n = \frac{\pi(b_{j+1} | x_{j+1}^n)}{\pi(b_{j+1} | \bar{x}_{j+1}^n)}, \quad w_{j+1}^n \leftarrow \frac{w_{j+1}^n}{\sum_n w_{j+1}^n}.$$

- ⑥ If $j < T$, increase $j \leftarrow j + 1$ and repeat.

Estimating parameters: Sequential Monte Carlo

For the discrete time model, the propagation (and possibly the likelihood) depend on the unknown θ ,

$$x_{j+1} = F(x_j, \theta).$$

Monte Carlo integral for posterior update:

$$\begin{aligned} \pi(x_{j+1}, \theta | \mathcal{B}_{j+1}) &\propto \pi(b_{j+1} | x_{j+1}, \theta) \\ &\times \int \pi(x_{j+1} | x_j, \theta) \pi(x_j | \theta, \mathcal{B}_j) \pi(\theta | \mathcal{B}_j) dx_j, \end{aligned}$$

Sample update:

$$\mathcal{S}_j \rightarrow \mathcal{S}_{j+1}, \quad \mathcal{S}_j = \{(x_j^n, \theta_j^n, w_j^n)\}_{n=1}^N.$$

where \mathcal{S}_j is drawn from $\pi(x_j, \theta | \mathcal{B}_j)$.

Denote

$$\bar{\theta}_j = \sum_{n=1}^N w_j^n \theta_j^n, \quad C_j = \sum_{n=1}^N w_j^n (\theta_j^n - \bar{\theta}_j) (\theta_j^n - \bar{\theta}_j)^T.$$

Approximate the marginal probability density $\pi(\theta | \mathcal{B}_j)$ of θ by a **Gaussian mixture model**,

$$\pi(\theta | \mathcal{B}_j) \approx \sum_{n=1}^N w_j^n \mathcal{N}(\theta | \bar{\theta}_j^n, s^2 C_j),$$

for which we define the auxiliary particle by



$$\bar{\theta}_j^n = a \theta_j^n + (1 - a) \bar{\theta}_j,$$

where a is a shrinkage factor, $0 < a < 1$ and $a^2 + s^2 = 1$ to avoid artificial diffusion.

Approximate

$$\pi(x_{j+1}, \theta | \mathcal{B}_{j+1}) \propto \sum_{n=1}^N w_j^n \pi(b_{j+1} | x_{j+1}, \theta) \pi(x_{j+1} | x_j^n, \theta) \mathcal{N}(\theta | \bar{\theta}_j^n, s^2 C_j),$$

which we write as

$$\begin{aligned} \pi(x_{j+1}, \theta | \mathcal{B}_{j+1}) &\propto \sum_{n=1}^N \underbrace{w_j^n \pi(b_{j+1} | \bar{x}_{j+1}^n, \bar{\theta}_j^n)}_{=g_{j+1}^n} \\ &\quad \times \frac{\pi(y_{j+1} | x_{j+1}, \theta)}{\pi(b_{j+1} | \bar{x}_{j+1}^n, \bar{\theta}_j^n)} \pi(x_{j+1} | x_j^n, \theta) \mathcal{N}(\theta | \bar{\theta}_j^n, s^2 C_j), \end{aligned}$$



where the coefficient g_{j+1}^n is the fitness of the predictor

$$(\bar{x}_{j+1}^n, \bar{\theta}_j^n) = (F(x_{j+1}^n, \bar{\theta}_j^n), \bar{\theta}_j^n).$$

Propagation and innovation

The problem we are addressing assumes

$$\frac{dx}{dt} = f(t, x, \theta), \quad x(0) = x_0,$$

while the discrete propagation is written as

$$x_{j+1} = F(x_j, \theta) + v_{j+1}.$$

Questions:

- ① How do we propagate?
- ② What is the innovation?

Stiffness and synchronization

For systems which are inherently stiff, we use a good stiff solver:

$$x_{j+1} = F^{\text{exact}}(x_j, \theta) = F(x_j, \theta) + \text{approximation error},$$

where the approximation error is due to numerical integration.

If the stiffness of the systems varies a lot with the parameter values prescribing a fixed accuracy may be a problem because

- The time for the particles propagation may vary widely;
- The slowest particle determines the propagation speed
- We cannot take full advantage of parallel and vectorized computing environment.

Prescribe time, not accuracy

Thus, to improve the performance of the algorithm we

- Propagate each particle with fixed propagation time.
- Estimate the numerical accuracy for each particle
- Set the j th particle innovation variance proportional to the integration error.



This yields the innovation covariance matrix

$$V_{j+1} \sim \mathcal{N}(0, \Gamma_{j+1}),$$

where for $1 \leq i \leq d$,

$$\Gamma_{j+1} = \text{diag}(\gamma) + \varepsilon I, \quad \gamma_i = \tau^2(u_{j+1} - \hat{u}_j)^2,$$



with $\tau > 1$.

SOF with error estimate innovation

Given the initial probability density $\pi_0(x_0)$,

- ① *Initialization:* Draw the particle ensemble from $\pi_0(x_0)$:

$$\mathcal{S}_0 = \{(x_0^1, w_0^1), (x_0^2, w_0^2), \dots, (x_0^N, w_0^N)\},$$

$$w_0^1 = w_0^2 = \dots = w_0^N = \frac{1}{N}.$$

Set $j = 0$.

- ② *Propagation:* Compute the predictor using a numerical ODE solver:

$$\bar{x}_{j+1}^n = \Psi(x_j^n, h), \quad 1 \leq n \leq N.$$



- ③ *Survival of the fittest:* For each n :

- (a) Compute the fitness weights

$$g_{j+1}^n = w_j^n \pi(b_{j+1} | \bar{x}_{j+1}^n), \quad g_{j+1}^n \leftarrow \frac{g_{j+1}^n}{\sum_n g_{j+1}^n};$$

- (b) Draw indices with replacement $\ell_n \in \{1, 2, \dots, N\}$ using probabilities

$$\mathbb{P}\{\ell_n = k\} = g_{j+1}^k;$$

- (c) Reshuffle

$$x_j^n \leftarrow x_j^{\ell_n}, \quad \bar{x}_{j+1}^n \leftarrow \bar{x}_{j+1}^{\ell_n}, \quad 1 \leq n \leq N.$$

④ *Innovation:* For each n :

- (a) Using error estimate, estimate $\Gamma_{j+1}^n = \Gamma_{j+1}(x_j^n)$;
- (b) Draw $v_{j+1}^n \sim \mathcal{N}(0, \Gamma_{j+1}^n)$;
- (c) Proliferate

$$x_{j+1}^n = \bar{x}_{j+1}^n + v_{j+1}^n.$$

⑤ *Weight updating:* For each n , compute

$$w_{j+1}^n = \frac{\pi(b_{j+1} | x_{j+1}^n)}{\pi(b_{j+1} | \bar{x}_{j+1}^n)}, \quad w_{j+1}^n \leftarrow \frac{w_{j+1}^n}{\sum_n w_{j+1}^n}.$$

⑥ If $j < T$, increase $j \leftarrow j + 1$ and repeat from (ii).

The parameter estimation SMC can be also carried out concurrently.

Propagation by particle methods

- In the Gaussian linear case, no particles were needed.
- The particle filtering, on the other hand, everything is expressed in terms of particles. Large amounts of particles needed.
- **Question:** Can we find an approximation, having “the best of both worlds”?

Example: Additive Gaussian innovation:

$$X_{t+1} = F_t(X_t) + V_{t+1}, \quad V_{t+1} \sim \mathcal{N}(0, \Gamma_{t+1}),$$

$$\hat{x}_{t+1}^j = F_t(x_t^j) + v_{t+1}^j,$$

where v_{t+1}^j is drawn from the Gaussian $\mathcal{N}(0, \Gamma_{t+1})$.

Analysis step: Assimilate new data

Problem: Given the predictive sample

$$\widehat{\mathcal{S}}_{t+1} = \{\widehat{x}_{t+1}^1, \widehat{x}_{t+1}^2, \dots, \widehat{x}_{t+1}^n\},$$

representing the prior density $\pi(x_{t+1} | \mathcal{B}_t)$, find an efficient way to sample from the posterior density

$$\pi(x_{t+1} | \mathcal{B}_{t+1}) \propto \underbrace{\pi(b_{t+1} | x_{t+1})}_{\text{likelihood}} \underbrace{\pi(x_{t+1} | \mathcal{B}_t)}_{\text{prior}},$$

Ideas from linear Gaussian inverse problems

Linear inverse problem:

$$b = Ax + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma), \quad x \sim \mathcal{N}(x_0, D),$$

Posterior density

$$\pi(x | b) \propto \exp\left(-\frac{1}{2}\mathcal{E}(x, b)\right),$$

where

$$\mathcal{E}(x, b) = \|Ax - b\|_{\Sigma}^2 + \|x - x_0\|_D^2,$$

with the notation $\|z\|_M^2 = z^T M^{-1} z$.

Ideas from linear Gaussian inverse problems

Sampling from the posterior by optimization:

- For $1 \leq j \leq n$, draw two mutually independent realizations of Gaussian random variables,

$$\eta^j \sim (0, \Sigma), \quad \nu^j \sim \mathcal{N}(0, D),$$

and perturb the data and prior mean by defining



$$b^j = b + \eta^j, \quad x_0^j = x_0 + \nu^j.$$

- Define

$$x^j = \operatorname{argmin} \left\{ \|Ax - b^j\|_{\Sigma}^2 + \|x - x_0^j\|_D^2 \right\}. \quad (1)$$

Towards EnKF

Theorem

The randomized optimization algorithm defined above produces a sample

$$\mathcal{S} = \{x^1, x^2, \dots, x^n\}$$

that is distributed according to the posterior distribution



$$\pi_{\text{post}} \propto \exp\left(-\frac{1}{2}\mathcal{E}(x, b)\right).$$

Towards EnKF

Proof (outline). Symmetric decompositions of the precision matrices,

$$\Sigma^{-1} = S^T S, \quad D^{-1} = L^T L.$$

Write the Gibbs energy as

$$\mathcal{E}(x, b) = \|SAx - Sb\|^2 + \|Lx - Lx_0\|^2 = \left\| \begin{bmatrix} SA \\ L \end{bmatrix} x - \begin{bmatrix} Sb \\ Lx_0 \end{bmatrix} \right\|^2.$$

The minimizer of the Gibbs energy is the solution of the normal equations,

$$((SA)^T SA + L^T L)x = (SA)^T Sb + L^T Lx_0,$$

Towards EnKF

In terms of the covariance matrices,

$$(A^T \Sigma^{-1} A + D^{-1})x = A^T \Sigma^{-1} b + D^{-1} x_0,$$

Conditional mean estimate,

$$x_{CM} = (A^T \Sigma^{-1} A + D^{-1})^{-1} (A^T \Sigma^{-1} b + D^{-1} x_0),$$

Randomize b and x_0 ,

$$(A^T \Sigma^{-1} A + D^{-1})x = A^T \Sigma^{-1} (b + \eta) + D^{-1} (x_0 + \nu),$$

where

$$\eta \sim \mathcal{N}(0, \Sigma), \quad \nu \sim \mathcal{N}(0, D),$$

Towards EnKF

The solution x becomes a random variable, with mean

$$\mathbb{E}\{x\} = x_{CM}.$$

Furthermore,

$$(A^T \Sigma^{-1} A + D^{-1})(x - x_{CM}) = A^T \Sigma^{-1} \eta + D^{-1} \nu,$$

from which it follows that

$$\begin{aligned} & (A^T \Sigma^{-1} A + D^{-1}) \mathbb{E}\{(x - x_{CM})(x - x_{CM})^T\} (A^T \Sigma^{-1} A + D^{-1})^T \\ &= A^T \Sigma^{-1} A + D^{-1}, \end{aligned}$$

implying that

$$\text{cov}(x) = (A^T \Sigma^{-1} A + D^{-1})^{-1} = C,$$

the covariance of the posterior distribution $\pi(x | b)$. \square

EnKF algorithm

- ① Initialize: Draw by independent sampling an initial sample from the prior π_0 ,

$$\mathcal{S}_0 = \{x_0^1, x_0^2, \dots, x_0^n\},$$

and set $t = 0$.

- ② Propagate the sample:

$$\hat{x}_{t+1}^k = F(x_t^k) + v_{t+1}^k, \quad v_{t+1}^k \sim \mathcal{N}(0, \Gamma_{t+1}).$$

EnKF algorithm

- ③ Compute an estimate for the predictive covariance matrix,

$$\hat{D}_{t+1} = \frac{1}{K} \sum_{k=1}^n (\hat{x}_{t+1}^k - x_{t+1}^*) (\hat{x}_{t+1}^k - x_{t+1}^*)^\top + \alpha I_n,$$

where

$$x_{t+1}^* = \frac{1}{n} \sum_{k=1}^n \hat{x}_{t+1}^k,$$

and $\alpha > 0$ is a variance inflation parameter.

- ④ Perturb the data, generating an artificial sample of data

$$b_{t+1}^k = b_{t+1} + \eta_{t+1}^k, \quad \eta_{t+1}^k \sim \mathcal{N}(0, \Sigma_{t+1}), \quad 1 \leq k \leq n.$$

EnKF algorithm

- ⑤ For every k , solve the minimization problem

$$x_{t+1}^k = \operatorname{argmin} \left\{ \|b_{t+1}^k - G(x_{t+1})\|_{\Sigma_{t+1}}^2 + \|x_{t+1} - \hat{x}_{t+1}^k\|_{D_{t+1}}^2 \right\}.$$

- ⑥ Compute the current posterior mean and posterior covariance estimate of x_{t+1} ,

$$\bar{x}_{t+1} = \frac{1}{n} \sum_{k=1}^n x_{t+1}^k, \quad C_{t+1} = \frac{1}{n} \sum_{k=1}^n (x_{t+1}^k - \bar{x}_{t+1})(x_{t+1}^k - \bar{x}_{t+1})^\top.$$

- ⑦ if $t < T$, increase $t \leftarrow t + 1$ and continue from 2.

Observations

If the observation model is linear,

$$b_{t+1} = A_{t+1}x_{t+1} + w_{t+1},$$

the analysis step reduces to a least squares problem,

$$x_{t+1}^k = \operatorname{argmin}\left\{\|b_{t+1}^k - A_{t+1}x_{t+1}\|_{\Sigma_{t+1}}^2 + \|x_{t+1} - \hat{x}_{t+1}^k\|_{D_{t+1}}^2\right\}.$$

which has an explicit solution,

$$x_{t+1}^k = \hat{x}_{t+1}^k + K_{t+1} (b_{t+1}^k - A_{t+1} \hat{x}_{t+1}^k), \quad k = 1, 2, \dots, n$$

where the *Kalman gain* matrix K_{t+1} is

$$K_{t+1} = \hat{D}_{t+1} A_{t+1}^T \left(A_{t+1} \hat{D}_{t+1} A_{t+1}^T + \Sigma_{t+1} \right)^{-1}.$$

Extensions: Add parameters

More general parametric model:

- ① Propagation:

$$X_{t+1} = F_t(X_t, \theta) + V_{t+1}, \quad V_t = \text{innovation},$$

- ② Observation:

$$B_t = G_t(X_t, \theta) + W_t.$$

Extend the evolution model:

$$Z_{t+1} = \begin{bmatrix} X_{t+1} \\ \Theta_{t+1} \end{bmatrix} = \begin{bmatrix} F_t(X_t) \\ \Theta_t \end{bmatrix} = \mathcal{F}_t(Z_t).$$

Extensions: Add parameters

Assume that only some components of x_{t+1} are observed:

$$B_{t+1} = P X_{t+1} + W_{t+1} = \begin{bmatrix} P & 0 \end{bmatrix} Z_{t+1} + W_{t+1},$$

where P is a projection matrix.

By partitioning Γ_{t+1} as

$$\Gamma_{t+1} = \begin{bmatrix} \Gamma_{xx} & \Gamma_{x\theta} \\ \Gamma_{\theta x} & \Gamma_{\theta\theta} \end{bmatrix} \in \mathbb{R}^{(d+k) \times (d+k)},$$

the Kalman gain is

$$K_{t+1} = \begin{bmatrix} \Gamma_{xx} \\ \Gamma_{\theta x} \end{bmatrix} P^T (P \Gamma_{xx} P^T + \Sigma_{t+1})^{-1},$$

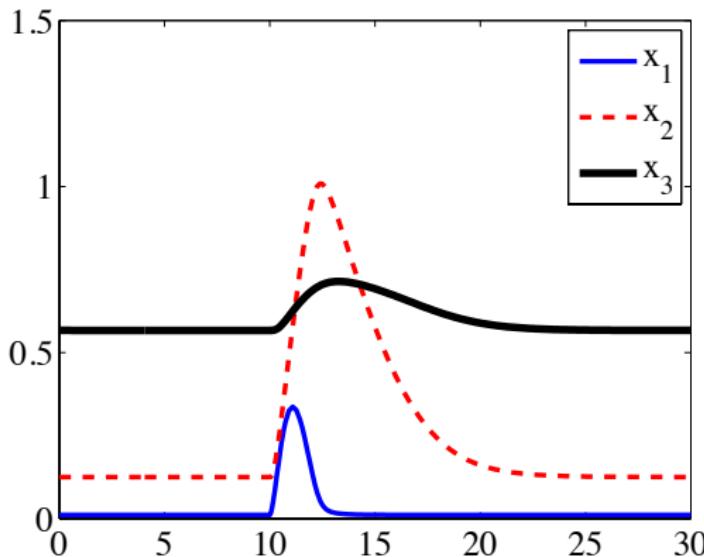
Information about θ through cross correlation.

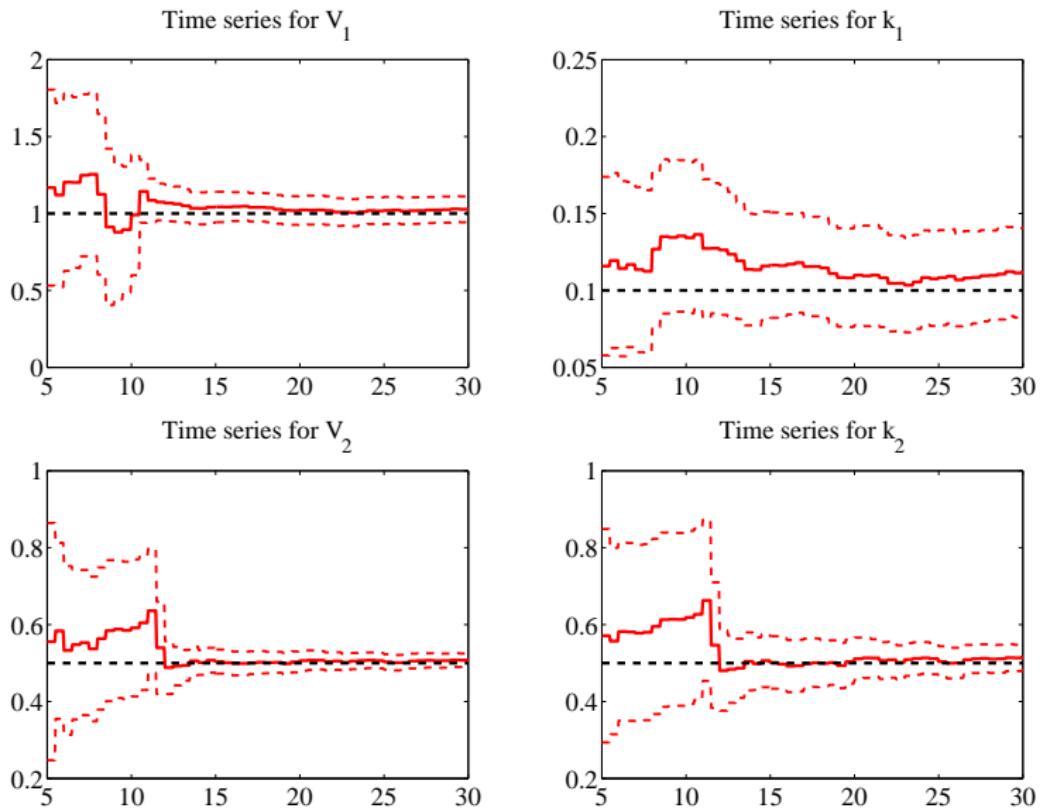
A simple example motivated by acetate metabolism

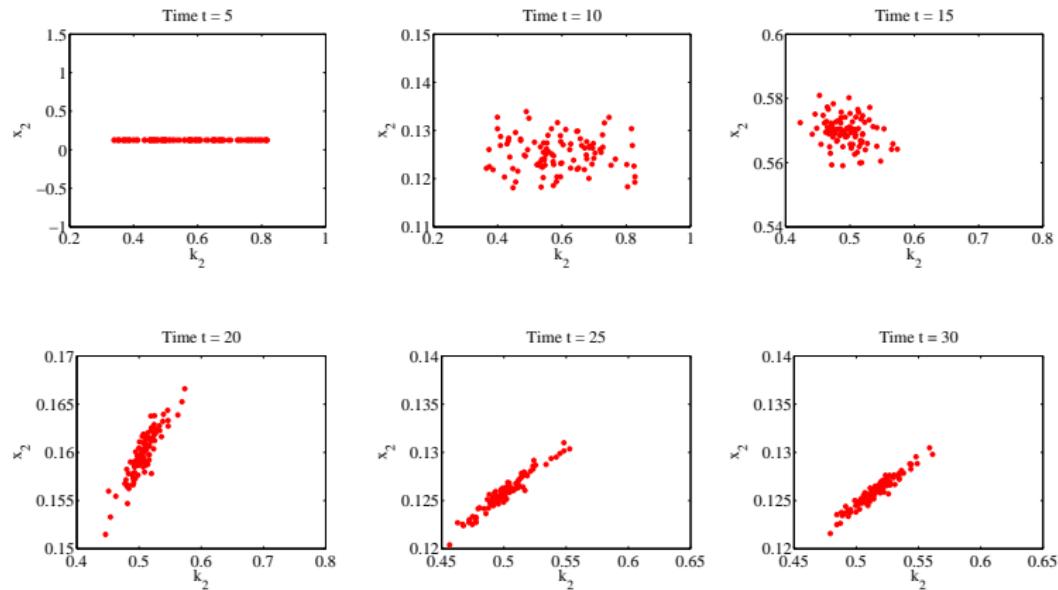
$$\begin{aligned}\frac{dx_1}{dt} &= \Phi(t) - V_1 \frac{x_1}{x_1 + k_1} \\ \frac{dx_2}{dt} &= V_1 \frac{x_1}{x_1 + k_1} - V_2 \frac{x_2}{x_2 + k_2} \\ \frac{dx_3}{dt} &= V_2 \frac{x_2}{x_2 + k_2} - \lambda(x_3 - c_0)\end{aligned}\tag{2}$$

with known parameters λ and c_0 and input function

$$\Phi(t) = A_0 + A(t - t_0) \exp(-(t - t_0)/\tau)$$







Scatter plots of the particle components k_2 versus x_2 . The component x_2 is observed, the coefficient k_2 is not, but we see that the two variables become correlated, meaning that observations of x_2 become informative for estimating k_2 .