MA8702 Advanced Computer-intensive Statistics

Department of Mathematical Sciences,
NTNU, Trondheim, Norway.

Hidden Markov models & Sequential Monte Carlo methods
Part II: Continuous HMM

Henning Omre, 01.04.2021

HIDDEN MARKOV MODELS / Sequential Monte Carlo

→ Focus: temporal state n-vector

$\underset{n\times 1}{s_t}$; $t \in T_s : \{0, 1, \ldots, t, \ldots, T, T+1\}$ $<$ categorial / continuous

T ← current time

Define:

$s = [s_0, \ldots, s_t, \ldots, s_T, s_{T+1}] \rightarrow s_{i:j} = [s_i, \ldots, s_j]$ $i \le j$

→ Available: temporal observation m-vector

d_t ; $t \in T_d : \{0, 1, \ldots, t, \ldots, T\}$

T ← current time

$d = [d_0, \ldots, d_t, \ldots, d_T] \rightarrow d_{i:j} = [d_i, \ldots, d_j]$ $i \le j$

→ Assess s from $d \rightarrow [s|d]$.

$d_0 \quad d_{t-1} \quad d_t \quad d_{t+1} \quad d_T$

$s_0 \quad s_{t-1} \quad s_t \quad s_{t+1} \quad s_T \quad s_{T+1}$

→ Bayesian framework:

$$p(\mathbf{s}|\mathbf{d};\theta_{\ell p}) = \left[p(\mathbf{d};\theta_{\ell p})\right]^{-1} p(\mathbf{d}|\mathbf{s};\theta_{\ell})\, p(\mathbf{s};\theta_{p})$$

posterior

likelihood

prior

$$p(\mathbf{d};\theta_{\ell p}) = \int p(\mathbf{d}|\mathbf{s};\theta_{\ell})\, p(\mathbf{s};\theta_{p})\, d\mathbf{s}$$

→ Distinguish:

Smoothing $p(\mathbf{s}_{0:T}|\mathbf{d}_{0:T})$ – $p(\mathbf{s}_0|\mathbf{d}_{0:T})$ inversion

$p(\mathbf{s}_t|\mathbf{d}_{0:T})$ marg smoothing

Filtering $p(\mathbf{s}_T|\mathbf{d}_{0:T})$ – $p(\mathbf{s}_{T+1}|\mathbf{d}_{0:T})$ forecasting

$\mathbf{d}_0$ $\quad$ $\mathbf{d}_{t-1}$ $\quad$ $\mathbf{d}_t$ $\quad$ $\mathbf{d}_{t+1}$ $\quad$ $\mathbf{d}_T$

$\mathbf{s}_0$ $\quad$ $\mathbf{s}_{t-1}$ $\quad$ $\mathbf{s}_t$ $\quad$ $\mathbf{s}_{t+1}$ $\quad$ $\mathbf{s}_T$ $\quad$ $\mathbf{s}_{T+1}$

Standard Model Assumptions

Cappe et al [2005]

Likelihood model:

$$[\mathbf{d}|\mathbf{s}] \Rightarrow p(\mathbf{d}|\mathbf{s}) = \prod_t p(\mathbf{d}_t|\mathbf{s}) = \prod_t p(\mathbf{d}_t|\mathbf{s}_t)$$

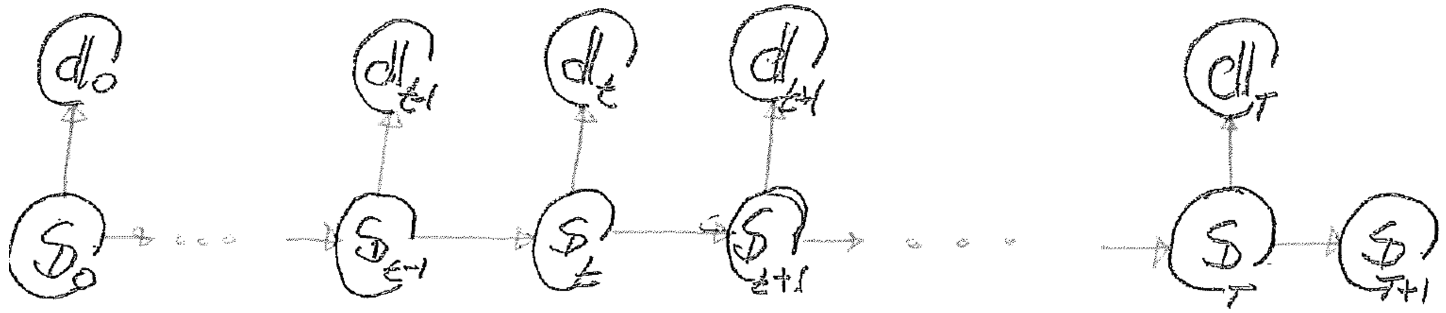
cond. indep   single site response

Prior model:

$$\mathbf{s} \Rightarrow p(\mathbf{s}) = p(\mathbf{s}_0)\prod_t p(\mathbf{s}_t|\mathbf{s}_{t-1:0})$$

$$= p(\mathbf{s}_0)\prod_t p(\mathbf{s}_t|\mathbf{s}_{t-1})$$

Markov assump



Moja et al [2019]

Posterior model:

$$[s|d] \Rightarrow p(s|d) = [\underbrace{p(d)}_{\text{const}}]^{-1}\, p(d|s)\, p(s)$$

$$= [p(d)]^{-1}\, p(d_0|s_0)\, p(s_0) \prod_t p(d_t|s_t)\, p(s_t|s_{t-1})$$

$$= p(s_0|d_{0:T}) \prod_t p(s_t|s_{t-1}, d_{t:T})$$

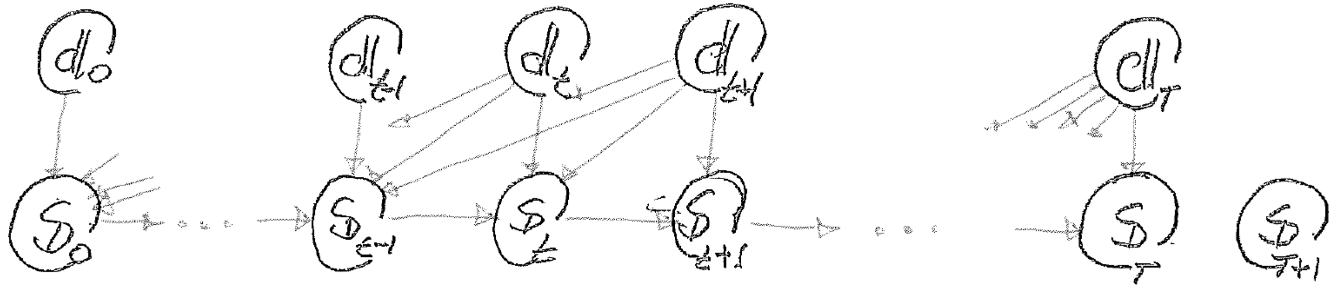
non-stat Markov mod

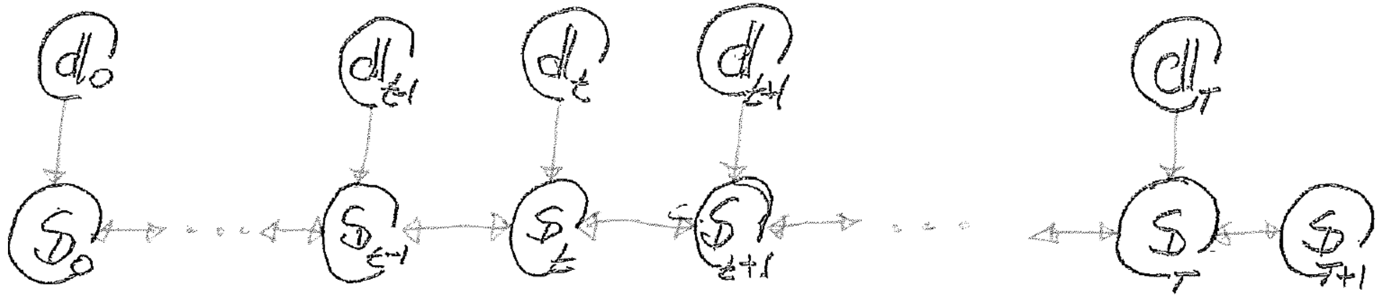
NOTE:

$$[s_t|s_{-t}, d] \Rightarrow p(s_t|s_{-t}, d)$$

$$= p(s_t|s_{t+1}, s_{t-1}, d_t)$$

Markov mod





Hidden Markov model:

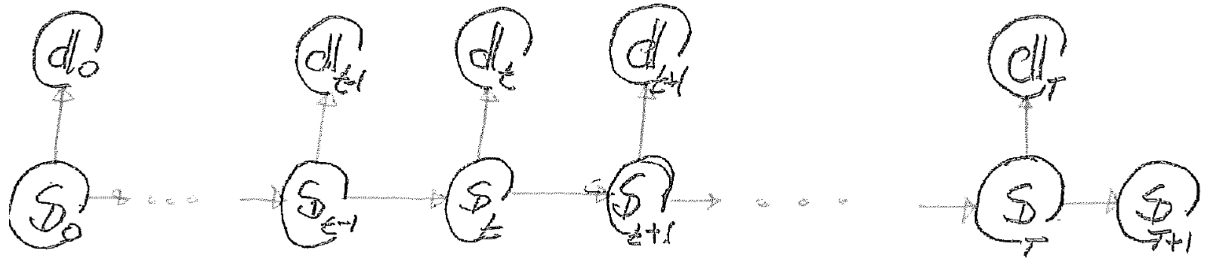
Smoothing:

$$p(s_{0:T} | d_{0:T}) = p(s|d) = p(s_0 | d_{0:T}) \prod_t p(s_t | s_{t-1}, d_{t:T})$$

← Posterior model

Marginal Smoothing

$$p(s_t | d_{0:T}) \text{ -- marginalization } p(s|d)$$



NOTE:

Forward Recursions

← can be made in real time!

Filtering:

$$p(s_T | d_{0:T})$$

Recursion PF:

$$p(s_t | d_{0:t-1})$$

$$p(s_t | d_{0:t}) = \left[p(d_t | d_{0:t-1})\right]^{-1} p(d_t | s_t) \int p(s_t | s_{t-1}) \, p(s_{t-1} | d_{0:t-1}) \, ds_{t-1}$$

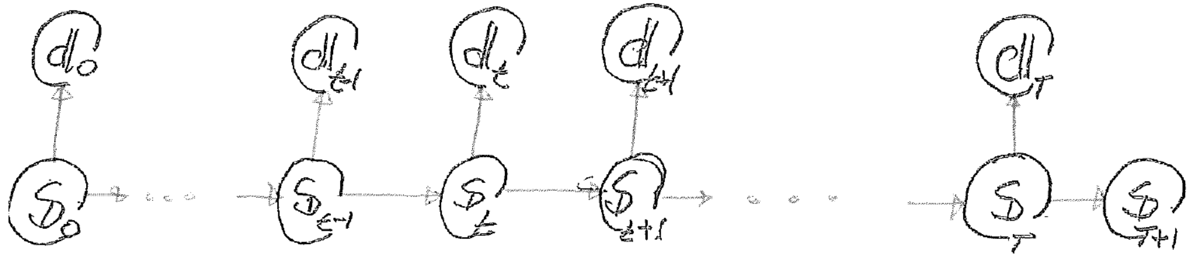
conditioning * forwarding

Recursion EnKF

$$p(s_{t+1} | d_{0:t}) = \int p(s_{t+1} | s_t) \left[p(d_t | d_{0:t-1})\right]^{-1} p(d_t | s_t) \, p(s_t | d_{0:t-1}) \, ds_t$$

forwarding * conditioning

Hidden Markov model:



Special Case – Trad Seq. Monte Carlo

Likelihood model

$$p(d|s) = \prod_t p(d_t|s_t)$$

Prior model:

$$p(s) = p(s_0)\prod_t p(s_t|s_{t-1}) = p(s_0)\prod_t \delta(s_t|s_{t-1})$$

Dirac

hence

$$[s_t|s_{t-1}] = s_{t-1} = s_0$$

Posterior model:

$$p(s_0|d) = p(s_0)\times\prod_t p(d_t|s_0)$$

NOTE:

Sequential conditioning

smoothing

filtering

identical!

simplest

Hidden Markov model

Viterbi [1967]

Posterior model:

$$p(\mathbf{s}|\mathbf{d}) = p(s_0|\mathbf{d}_{0:T}) \prod_t p(s_t|s_{t-1}, \mathbf{d}_{t:T})$$

Usually interested in:

Simulations:

$\mathbf{s}^i \leadsto p(\mathbf{s}|\mathbf{d})$; $i = 1, \dots, n_s$ – Sequential Alg

Predictions:

$\hat{\mathbf{s}}_E = E\{\mathbf{s}|\mathbf{d}\}$ – use $E_t\{s_t\} = E_{t-1}\{E_t\{s_t|s_{t-1}\}\}$ – Seq. Alg

$\hat{\mathbf{s}}_{MAP} = MAP\{\mathbf{s}|\mathbf{d}\}$ – use Viterbi Alg

$\hat{\mathbf{s}}_{MMAP} = MMAP\{\mathbf{s}|\mathbf{d}\}$ – Sequential Alg

Prediction Precision:

$\sigma_E^2 = Diag\{Var\{\mathbf{s}|\mathbf{d}\}\}$

– use $Var_t\{s_t\} = E_{t-1}\{Var_t\{s_t|s_{t-1}\}\}$

$+ Var_{t-1}\{E_t\{s_t|s_{t-1}\}\}$ – Seq. Alg

$\underset{n\times 1}{\mathbf{p}_s} = Marg\{p(\mathbf{s}|\mathbf{d})\}$ – Sequential Alg

Hidden Markov model

Model parameter inference:

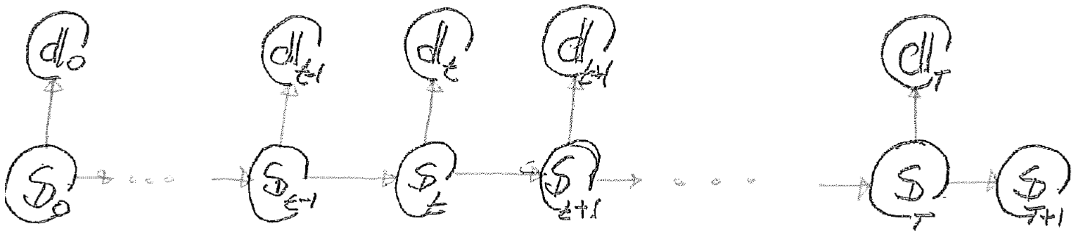
Criterion - maximum (marginal) likelihood

$$\hat{\theta}_{lp} = \underset{\theta_{lp}}{\operatorname{argmax}} \{ p(d; \theta_l, \theta_p) \}$$

NOTE:

$$p(d;\theta_{lp}) = \int_0 \cdots \int_T p(d|s;\theta_l)\, p(s;\theta_p)\, ds_T \cdots ds_0$$

$$= \int_0 \cdots \int_T p(d_0|s_0;\cdot)\, p(s_0;\cdot) \prod_t p(d_t|s_t;\cdot)\, p(s_t|s_{t-1};\cdot)\, ds_T \cdots ds_0$$



$$= \int_0 p(d_0|s_0)\, p(s_0) \int_1 p(d_1|s_1)\, p(s_1|s_0) \int \cdots$$

$$\boxed{p(d_{t:T}|s_{t-1})}$$

$$\times \int_T p(d_T|s_T)\, p(s_T|s_{T-1})\, ds_T \cdots ds_0$$

Sequential integration!

NOTE:

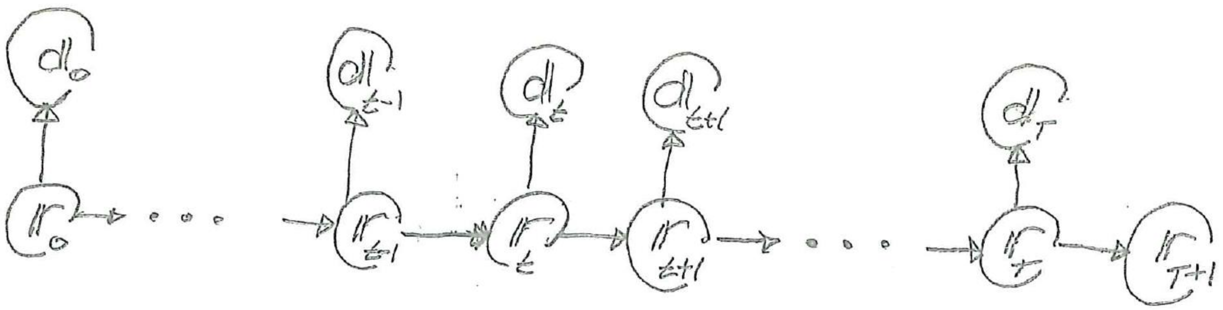
Must calculate $p(d;\theta_l,\theta_p)$ for each value of $[\theta_l, \theta_p]$.

Hidden Markov model

→ Continuous variables

$$\underset{n\times 1}{\mathbb{S}_t} \longrightarrow \underset{n\times 1}{\mathbf{r}_t} \in \mathbb{R}^n$$

$$\mathbf{r} = [\mathbf{r}_0, \dots, \mathbf{r}_t, \dots, \mathbf{r}_T, \mathbf{r}_{T+1}]$$



→ Likelihood model

RV — not specified

$$[\mathbf{d}_t | \mathbf{r}_t] = \psi_t(\mathbf{r}_t, \epsilon_{d|r}) \Rightarrow p(\mathbf{d}_t | \mathbf{r}_t)$$

Gauss-linear

- Kalman: $= \underset{m\times n}{\mathbb{H}_t}\, \underset{n\times 1}{\mathbf{r}_t} + \underset{m\times 1}{\epsilon_{d|r}} \Rightarrow \varphi_m(\mathbf{d}_t; \mathbb{H}_t \mathbf{r}_t, \Sigma_{d|r})$

→ Prior model

RV — not specified

$$\mathbf{r}_0 = \omega_0(\epsilon_{r|r}) \Rightarrow p(\mathbf{r}_0)$$

Gauss

- Kalman: $\Rightarrow \varphi_n(\mathbf{r}_0; \mu_{0|\cdot}, \Sigma_{0|\cdot})$

RV — not specified

$$[\mathbf{r}_t | \mathbf{r}_{t-1}] = \omega_t(\mathbf{r}_{t-1}, \epsilon_{r|r}) \Rightarrow p(\mathbf{r}_t | \mathbf{r}_{t-1})$$

Gauss-linear

- Kalman: $= \underset{n\times n}{\mathbb{A}_t}\, \underset{n\times 1}{\mathbf{r}_{t-1}} + \underset{n\times 1}{\epsilon_{r|r}} \Rightarrow \varphi_n(\mathbf{r}; \mathbb{A}_t \mathbf{r}_{t-1}, \Sigma_{r|r})$

→ NOTE:

Gauss — Not linear

$$[\mathbf{d}_t | \mathbf{r}_t] = \psi_t(\mathbf{r}_t) + \epsilon_{d|r} \Rightarrow p(\mathbf{d}_t | \mathbf{r}_t) = \varphi_m(\mathbf{d}; \psi_t(\mathbf{r}_t), \Sigma_{d|r})$$

Kalman model

→ Likelihood model ← Gauss-lin

$$[d_t | r_t] \to p(d_t | r_t) = \varphi_m(d; H_t r_t, \Sigma_{d|r})$$

→ Prior model

$$r_0 \to p(r_0) = \varphi_n(r; \mu_{0|\cdot}, \Sigma_{0|\cdot})$$ ← Gauss

$$[r_t | r_{t-1}] \to p(r_t | r_{t-1}) = \varphi_n(r; A_t r_{t-1}, \Sigma_{r|r})$$ ← Gauss-lin

→ Kalman Smoother:

$$p(r_{0:T}, d_{0:T}) \longrightarrow p(r_{0:T} | d_{0:T})$$ ← Gauss

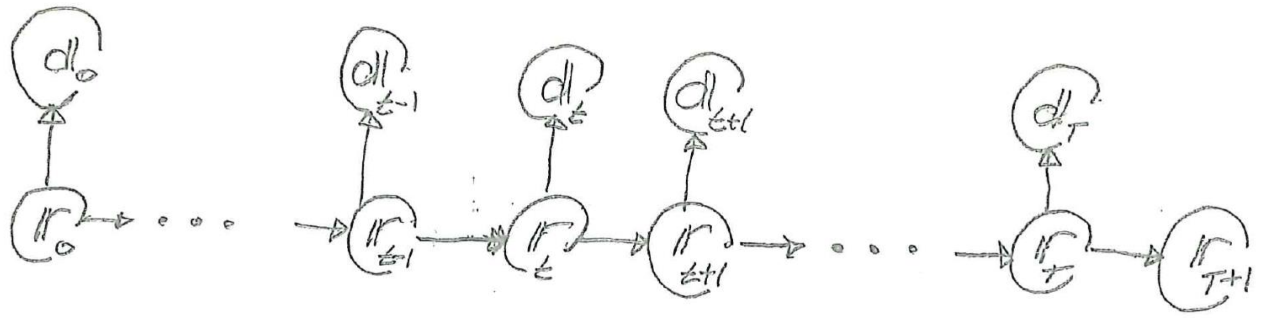
Gauss

↓ marg - easy!!!

$$p(r_t | d_{0:T})$$ ← Gauss

Alternative approach:

Recursive reverse algorithm



→ Kalman Joint recursion:

$$p(r_{0:t}, d_{0:t})$$

$$= p(d_t | r_{0:t}, d_{0:t-1})\, p(r_t | r_{0:t-1}, d_{0:t-1})\, p(r_{0:t-1}, d_{0:t-1})$$

↓ $p(d_t | r_t)$ ↓ $p(r_t | r_{t-1})$

Gauss-lin Gauss

Gauss

Gauss-lin Gauss

Gauss

Initiate:

$$p(r_0, d_0) = p(d_0 | r_0)\, p(r_0)$$ ← Gauss

Gauss-lin Gauss

Kalman model

Kalman Smoother

$$[\mathbf{r}_{0:T+1} | \mathbf{d}_{0:T}] \rightsquigarrow p(\mathbf{r}_{0:T+1} | \mathbf{d}_{0:T}) = \varphi_{n(T+2)}(\mathbf{r}; \mu_{|T}, \Sigma_{|T})$$

Interested in:

- Simulation

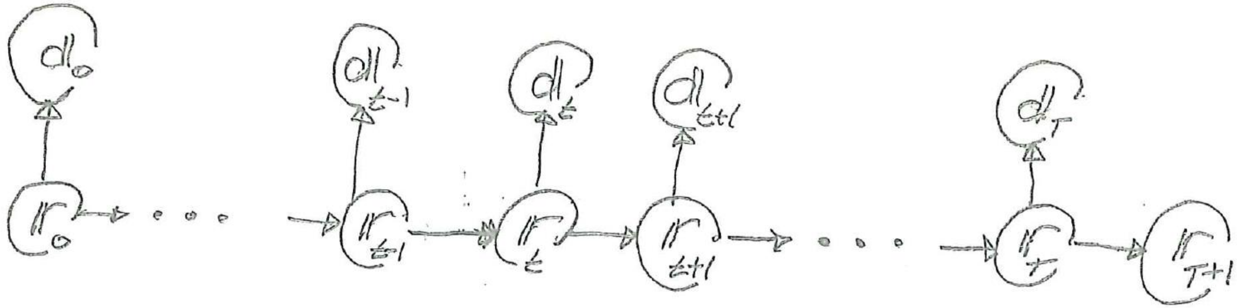
$$\mathbf{r}^i \rightsquigarrow p(\mathbf{r}|\mathbf{d}) = \varphi_{n(T+2)}(\mathbf{r}; \mu_{|T}, \Sigma_{|T}) \; ; i = 1, \ldots, n_s$$

- Prediction

$$\hat{\mathbf{r}}_E = E\{\mathbf{r}|\mathbf{d}\} = \mu_{|T} \quad \leftarrow \quad \text{same as } \hat{\mathbf{r}}_{MAP} \,\&\, \hat{\mathbf{r}}_{MMAP}$$

- Prediction Precision

$$\sigma_E^2 = \mathrm{Diag}\{\Sigma_{|T}\} \quad \leftarrow \quad \text{Gauss Pred intervals}$$



Kalman model

Kalman filter recursion

$$p(r_{t+1}|d_{0:t}) = \int \underbrace{p(r_{t+1}|r_t)}_{\text{Gauss-lin}} \underbrace{[p(d_t|d_{0:t-1})]^{-1} \underbrace{\underbrace{p(d_t|r_t)}_{\text{Gauss-lin}} \underbrace{p(r_t|d_{0:t-1})}_{\text{Gauss}}}_{\text{Gauss}}}_{\text{Gauss}} \, dr_t$$

Initiate:

$$p(r_0) \leftarrow \text{Gauss}$$

Kalman filter algorithm:

$$[r_{T+1}|d_{0:T}] \Rightarrow p(r_{T+1}|d_{0:T}) = \varphi_n(r; \mu_{T+1|T}, \Sigma_{T+1|T})$$

$$f(\boldsymbol{r}_t|\boldsymbol{d}_{0:t-1}) = \varphi_n(\boldsymbol{r}_t; \boldsymbol{\mu}^r_{t|t-1}, \boldsymbol{\Sigma}^r_{t|t-1}),$$

$$f(\boldsymbol{r}_t|\boldsymbol{d}_{0:t}) = \varphi_n(\boldsymbol{r}_t; \boldsymbol{\mu}^r_{t|t}, \boldsymbol{\Sigma}^r_{t|t}),$$

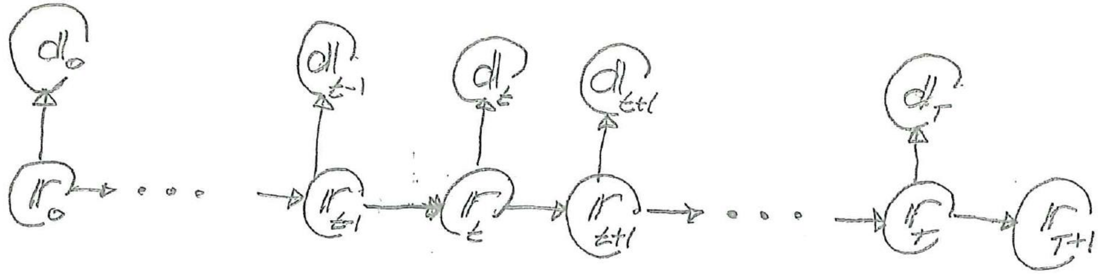
$$\boldsymbol{\mu}^r_{t|t} = \boldsymbol{\mu}^r_{t|t-1} + \boldsymbol{\Sigma}^r_{t|t-1}\boldsymbol{H}^T_t(\boldsymbol{H}_t\boldsymbol{\Sigma}^r_{t|t-1}\boldsymbol{H}^T_t + \boldsymbol{\Sigma}^d_t)^{-1}(\boldsymbol{d}_t - \boldsymbol{H}\boldsymbol{\mu}^r_{t|t-1}),$$

$$\boldsymbol{\Sigma}^r_{t|t} = \boldsymbol{\Sigma}^r_{t|t-1} - \boldsymbol{\Sigma}^r_{t|t-1}\boldsymbol{H}^T_t(\boldsymbol{H}_t\boldsymbol{\Sigma}^r_{t|t-1}\boldsymbol{H}^T_t + \boldsymbol{\Sigma}^d_t)^{-1}\boldsymbol{H}_t\boldsymbol{\Sigma}^r_{t|t-1},$$

$$f(\boldsymbol{r}_{t+1}|\boldsymbol{d}_{0:t}) = \varphi_n(\boldsymbol{r}_{t+1}; \boldsymbol{\mu}^r_{t+1|t}, \boldsymbol{\Sigma}^r_{t+1|t}),$$

$$\boldsymbol{\mu}^r_{t+1|t} = \boldsymbol{A}_t\boldsymbol{\mu}^r_{t|t},$$

$$\boldsymbol{\Sigma}^r_{t+1|t} = \boldsymbol{A}_t\boldsymbol{\Sigma}^r_{t|t}\boldsymbol{A}^T_t + \boldsymbol{\Sigma}^r_t.$$



Kalman Model extensions

- non-Gauss-linear cases

→ Extended Kalman Filter - EKF

$$[\mathbb{d}_t | \mathbb{r}_t] = \psi(\mathbb{r}_t) + \epsilon_{d|r} \longrightarrow \psi(\mu_t) + \overset{\hat{H}_t}{\psi'(\mu_t)}[\mathbb{r}_t - \mu_t] + \epsilon_{d|r}$$

$$[\mathbb{r}_t | \mathbb{r}_{t-1}] = \omega(\mathbb{r}_{t-1}) + \epsilon_{r|r} \longrightarrow \omega(\mu_{t-1}) + \underset{\hat{A}_t}{\omega'(\mu_{t-1})}[\mathbb{r}_{t-1} - \mu_{t-1}] + \epsilon_{r|r}$$

Use KF Alg on linearized model.

→ Unscented Kalman Filter - UKF

$p(\mathbb{r}_{t+1} | \mathbb{d}_{0:T})$ approx by 'design points'

- conditioning by 'numerical integration'.

→ Selection Kalman Filter - SKF

• Likelihood model:

$$[d_t | r_t] = H r_t + \epsilon_{d|r} \to p(d_t | r_t) = \varphi_m(d_t; H r_t, \Sigma_{d|r})$$ - Gauss - linear

• Prior model:

$$[r_{t+1} | r_t] = A r_t + \epsilon_{r|r} \to p(r_{t+1} | r_t) = \varphi_n(r; A r_t, \Sigma_{r|r})$$ - Gauss - linear

$$[r_0] \to p(r_0)$$ - Selection Gauss — multi-modal, skewed, peaked

• Define:

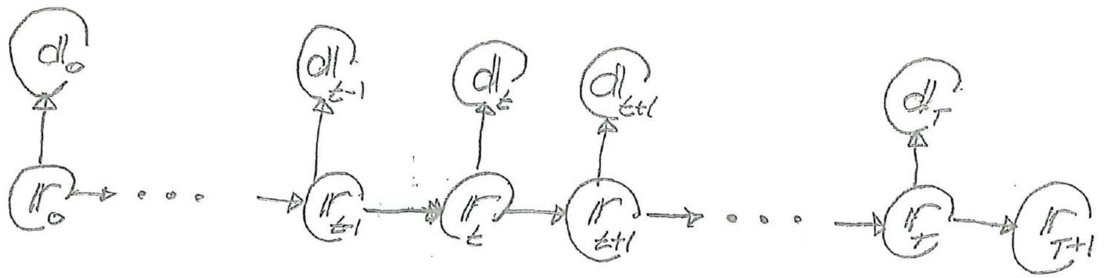
Auxiliary variables $\begin{bmatrix}\tilde{r}\\ \tilde{\nu}\end{bmatrix}_{2n\times 1} \to \varphi_{2n}\left(\begin{bmatrix}\tilde{r}\\ \tilde{\nu}\end{bmatrix}; \begin{bmatrix}\mu_{\tilde{r}}\\ \mu_{\tilde{\nu}}\end{bmatrix}, \begin{bmatrix}\Sigma_{\tilde{r}} & \Gamma_{\tilde{r}\tilde{\nu}}\\ \Gamma_{\tilde{\nu}\tilde{r}} & \Sigma_{\tilde{\nu}}\end{bmatrix}\right)$

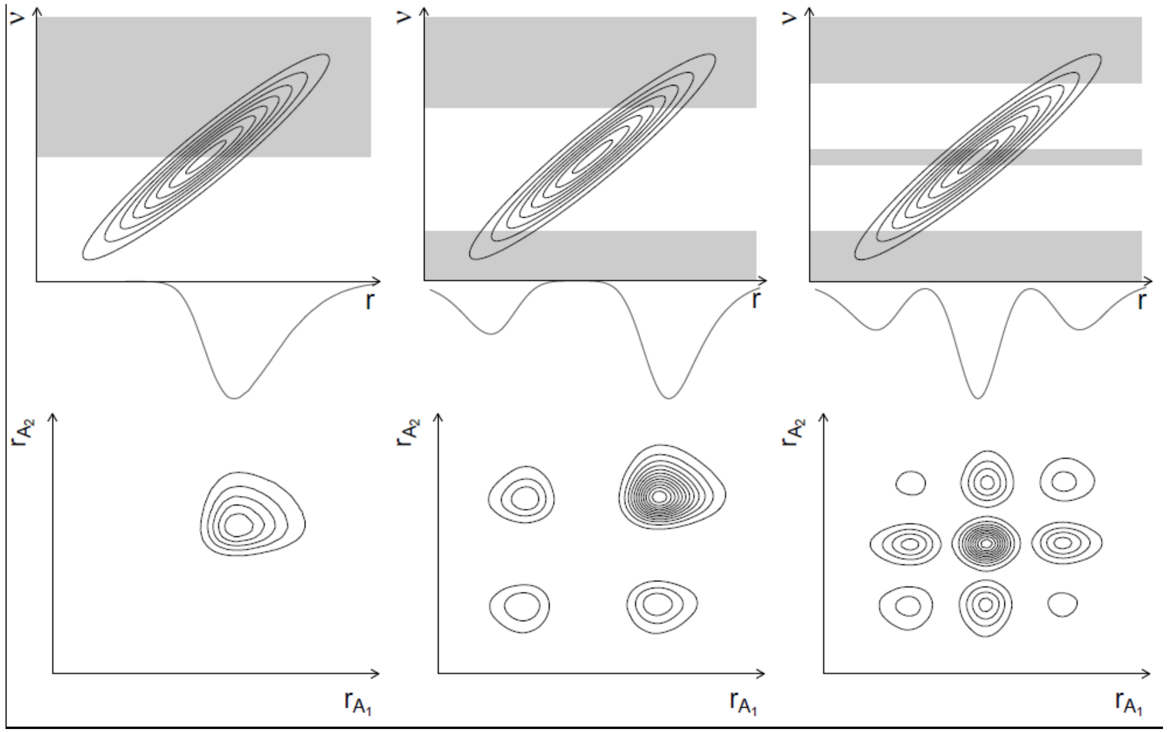
Selection set: $A \subset \mathbb{R}^n$

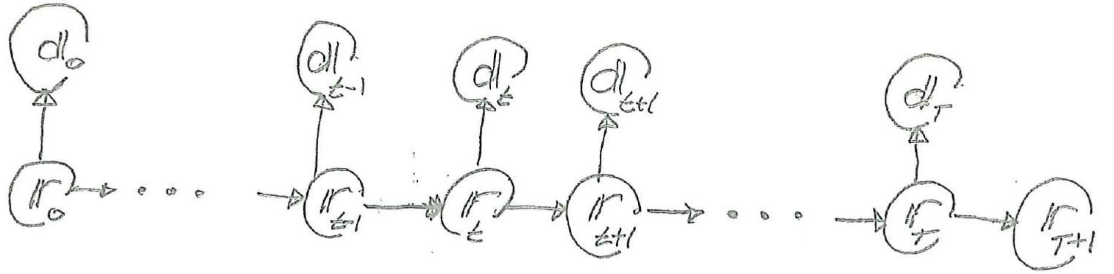
• Then

$$r_0 = [\tilde{r} | \tilde{\nu} \in A] \to p(r_0) = p(\tilde{r} | \tilde{\nu} \in A)$$

$$= [\Phi_n(A; \mu_{\tilde{\nu}}, \Sigma_{\tilde{\nu}})]^{-1} \times \Phi_n(A; \mu_{\tilde{\nu}|\tilde{r}}, \Sigma_{\tilde{\nu}|\tilde{r}}) \times \varphi_n(\tilde{r}; \mu_{\tilde{r}}, \Sigma_{\tilde{r}})$$







→ Selection Kalman Filter - SKF

- Likelihood model:

$$[d_t | r_t] = H r_t + \epsilon_{d|r} \rightarrow p(d_t | r_t) = \varphi_m(d_t; H r_t, \Sigma_{d|r})$$ - Gauss-linear

- Prior model:

$$[r_{t+1} | r_t] = A r_t + \epsilon_{r|r} \rightarrow p(r_{t+1} | r_t) = \varphi_n(r; A r_t, \Sigma_{r|r})$$ - Gauss-linear

$[r_0] \rightarrow p(r_0)$ - Selection Gauss — multi-modal, skewed, peaked

- Define:

Auxiliary variables $\underset{2n\times 1}{\begin{bmatrix}\tilde{r}\\ \tilde{\nu}\end{bmatrix}} \rightarrow \varphi_{2n}\left(\begin{bmatrix}\tilde{r}\\ \tilde{\nu}\end{bmatrix}; \begin{bmatrix}\mu_{\tilde{r}}\\ \mu_{\tilde{\nu}}\end{bmatrix}, \begin{bmatrix}\Sigma_{\tilde{r}} & \Gamma_{\tilde{r}\tilde{\nu}}\\ \Gamma_{\tilde{\nu}\tilde{r}} & \Sigma_{\tilde{\nu}}\end{bmatrix}\right)$

Selection set: $A \subset \mathbb{R}^n$

- Then

$$r_0 = [\tilde{r} | \tilde{\nu} \in A] \rightarrow p(r_0) = p(\tilde{r} | \tilde{\nu} \in A)$$

$$= [\Phi_n(A; \mu_{\tilde{\nu}}, \Sigma_{\tilde{\nu}})]^{-1} \times \Phi_n(A; \mu_{\tilde{\nu}|\tilde{r}}, \Sigma_{\tilde{\nu}|\tilde{r}}) \times \varphi_n(\tilde{r}; \mu_{\tilde{r}}, \Sigma_{\tilde{r}})$$

- KF recursion:

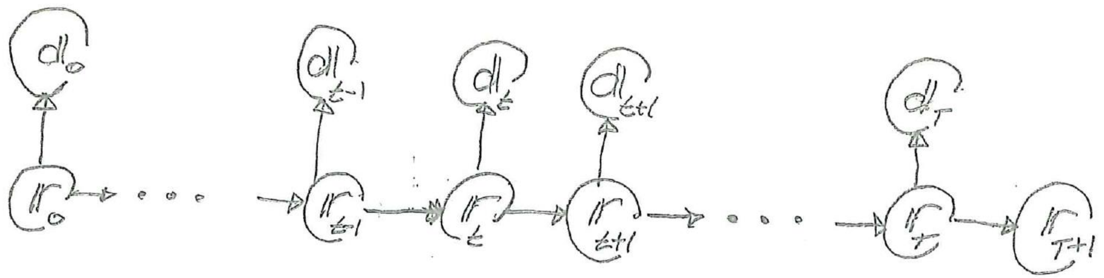
$$p(r_{t+1} | d_{0:T}) = \int \underbrace{\underbrace{p(r_{t+1} | r_t)}_{\text{Gauss-lin}} [p(d_t | d_{0:t-1})]^{-1} \underbrace{\underbrace{p(d_t | r_t)}_{\text{Gauss-lin}}\, \underbrace{p(r_t | d_{0:t-1})}_{\text{Sel-Gauss}}}_{\text{Sel-Gauss}}}_{\text{Sel-Gauss}} dr_t$$

Initiate:

$p(r_0) \leftarrow$ Sel-Gauss

→ NOTE: Selection Gauss prior model conjugate wrt Gauss-linear likelihood models!

→ Selection Kalman Filter analytically tractable!



Kalman model

→ Model parameter inference

$$\hat{\Theta}_{lp} = \underset{\Theta_{lp}}{\operatorname{argmax}} \{ p(d; \Theta_l, \Theta_p) \} \text{ - ML estimator!}$$

$[\mu_{0|\cdot}, \Sigma_{0|\cdot}, \sigma^2_{d|r}, \sigma^2_{r|r}]$

$$= \int_T \cdots \int_0 p(r_{0:T}, d_{0:T}; \Theta_{lp})\, dr_0 \cdots dr_T$$

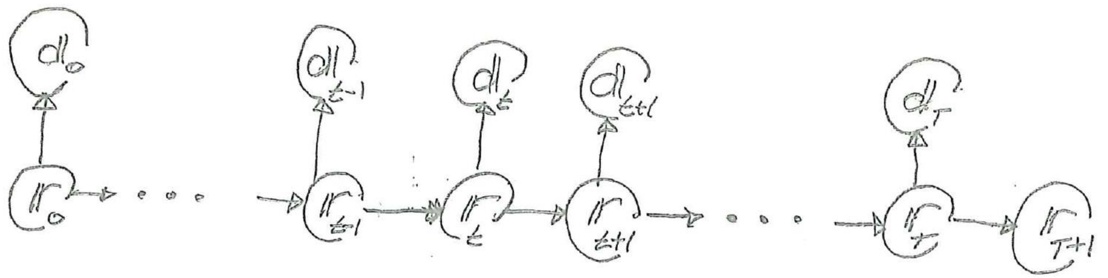
Gauss - marg - easy

HIDDEN MARKOV MODEL / ~~DATA~~ DATA ASSIMILATION

Focus on: $\mathbf{r} = [\mathbf{r}_0, \quad , \mathbf{r}_t, \dots, \mathbf{r}_T, \mathbf{r}_{T+1}]$

Observations: $\mathbf{d} = [\mathbf{d}_0, \dots, \mathbf{d}_t, \dots, \mathbf{d}_T]$

Solution $[\mathbf{r}|\mathbf{d}]$



Likelihood model:

$$p(\mathbf{d}|\mathbf{r}) = \prod_t p(\mathbf{d}_t|\mathbf{r}_t)$$

$$[\mathbf{d}_t|\mathbf{r}_t] = \psi_t(\mathbf{r}_t; \epsilon_{d|r}) \rightarrow \psi_t(\mathbf{r}_t) + \epsilon_{d|r}$$

Prior model:

$$p(\mathbf{r}) = p(\mathbf{r}_0) \prod_t p(\mathbf{r}_t|\mathbf{r}_{t-1})$$

$$\mathbf{r}_0 = \omega_i(\epsilon_{r|r})$$

$$[\mathbf{r}_{t+1}|\mathbf{r}_t] = \omega_t(\mathbf{r}_t; \epsilon_{r|r}) \rightarrow \omega_t(\mathbf{r}_t) + \epsilon_{r|r}$$

NOTE: Posterior model

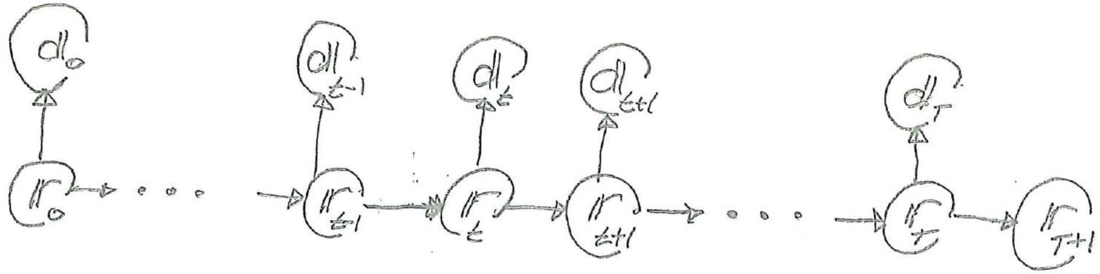
$$[\mathbf{r}|\mathbf{d}] \Leftrightarrow p(\mathbf{r}|\mathbf{d}) = p(\mathbf{r}_0|\mathbf{d}_{0:t}) \prod_t p(\mathbf{r}_t|\mathbf{r}_{t-1}, \mathbf{d}_{0:t})$$

More complicated – Sequential Monte Carlo

- Particle Algorithms – Filter/Smoother
- Ensemble Kalman Algorithms – Filter/Smoother
- [Approximate Bayesian Computation – ABC]

PARTICLE ALGORITHMS

→ The Particle Filter – $[\mathbb{r}_T|\mathbb{d}_{0:T}] \Rightarrow p(\mathbb{r}_T|\mathbb{d}_{0:T})$



Based on Filter recursion:

$$p(\mathbb{r}_t|\mathbb{d}_{0:t}) = [\underbrace{p(\mathbb{d}_t|\mathbb{d}_{0:t-1})}_{\text{const}}]^{-1} \underbrace{p(\mathbb{d}_t|\mathbb{r}_t)}_{\text{conditioning}} \underbrace{\int p(\mathbb{r}_t|\mathbb{r}_{t-1})\, p(\mathbb{r}_{t-1}|\mathbb{d}_{0:t-1})\, d\mathbb{r}_{t-1}}_{\text{forwarding}}$$

Recursions:

Set of particles with weights at time $t-1$:

$\Pi_{t-1}: \{[\mathbb{r}^i_{t-1}, w^i_{t-1}];\ i=1,\dots,n_\pi\}$ with $w^i_\cdot \geq 0;\ \sum_i w^i_\cdot = 1$

representing $p(\mathbb{r}_{t-1}|\mathbb{d}_{0:t-1})$

Perform Forwarding & Conditioning to time t:

$\Pi_t: \{[\mathbb{r}^i_t, w^i_t];\ i=1,\dots,n_\pi\}$ with $w^i_\cdot \geq 0;\ \sum_i w^i_\cdot = 1$

representing $p(\mathbb{r}_t|\mathbb{d}_{0:t})$

Sekvensiell Monte Carlo
(eller Partikkelfiltre)

Geir Storvik

Dept. of Mathematics, University of Oslo
SFI[2], Statistics for inovation
Norwegian Computing Center
SFF-CEES, Ecology and Evolution synthesis

Jevnaker 16-18. juni 2009

→ 'Brute-force' Particle Filter Algorithm

Initiate:

Set $n_\pi \in \mathbb{N}_+$

Generate $\mathbf{r}_0^i = \omega_0(\epsilon^i_{r|r})$; $i=1,\dots,n_\pi$

Set $w_0^i = \frac{1}{n_\pi}$; $i=1,\dots,n_\pi$

$\pi_0 = \{(\mathbf{r}_0^i, w_0^i); i=1,\dots,n_\pi\}$

Do for $t=1,\dots,T$

Forwarding:

$\mathbf{r}_t^i = \omega_t(\mathbf{r}_{t-1}^i, \epsilon^i_{r|r})$; $i=1,\dots,n_\pi$

Conditioning:

$\tilde{w}_t^i = w_{t-1}^i \times p(\mathbf{d}_t|\mathbf{r}_t^i)$; $i=1,\dots,n_\pi$

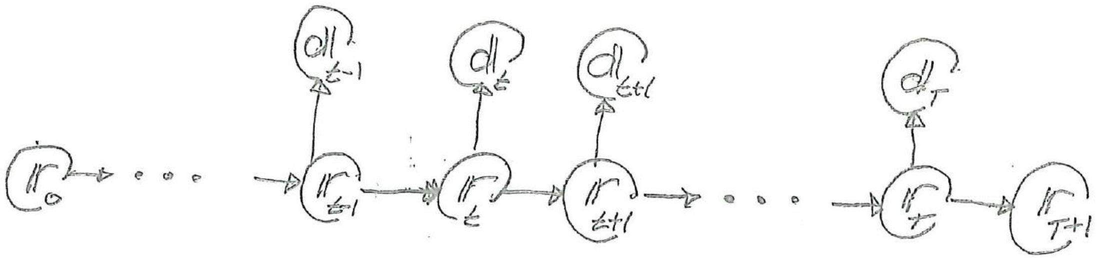
$w_t^i = \left[\sum_i \tilde{w}_t^i\right]^{-1} \tilde{w}_t^i$; $i=1,\dots,n_\pi$

$\pi_t = \{(\mathbf{r}_t^i, w_t^i); i=1,\dots,n_\pi\}$

End for

$\pi_T : \{(\mathbf{r}_T^i, w_T^i); i=1,\dots,n_\pi\}$

represents $p(\mathbf{r}_T|\mathbf{d}_{0:T})$



Non-parametric estimates:

$$\hat{P}_{n_\pi}(\mathbf{r}_T|\mathbf{d}_{0:T}) = \sum_i w_T^i I(\mathbf{r}_T^i \le \mathbf{r}_T)$$

$$\hat{p}_{n_\pi}(\mathbf{r}_T|\mathbf{d}_{0:T}) = \sum_i w_T^i \delta(\mathbf{r}_T, \mathbf{r}_T^i)$$

Asymptotics:

$$\hat{P}_{n_\pi}(\mathbf{r}_T|\mathbf{d}_{0:T}) \xrightarrow[n_\pi \to \infty]{} P(\mathbf{r}_T|\mathbf{d}_{0:T})$$

$$\hat{p}_{n_\pi}(\mathbf{r}_T|\mathbf{d}_{0:T}) \xrightarrow[n_\pi \to \infty]{} p(\mathbf{r}_T|\mathbf{d}_{0:T})$$

→ 'Brute-force' Particle Filter Algorithm

Initiate:

Set $n_\pi \in \mathbb{N}_+$

Generate $r_0^i = \omega_0(\epsilon_{r|r}^i)$; $i=1,\dots,n_\pi$

Set $w_0^i = \frac{1}{n_\pi}$; $i=1,\dots,n_\pi$

$\pi_0 = \{(r_0^i, w_0^i); i=1,\dots,n_\pi\}$

Do for $t=1,\dots,T$

Forwarding:

$r_t^i = \omega_t(r_{t-1}^i, \epsilon_{r|r}^i)$; $i=1,\dots,n_\pi$

Conditioning:

$\tilde{w}_t^i = w_{t-1}^i \times p(d_t \mid r_t^i)$; $i=1,\dots,n_\pi$

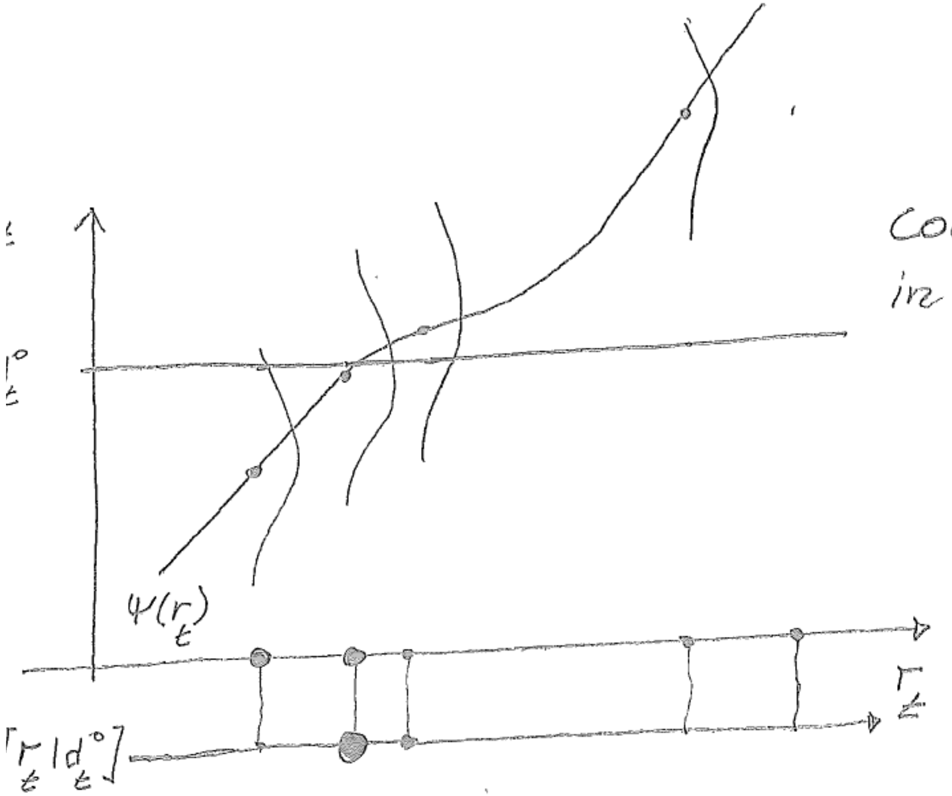
$w_t^i = \left[\sum_i \tilde{w}_t^i\right]^{-1} \tilde{w}_t^i$; $i=1,\dots,n_\pi$

$\pi_t = \{(r_t^i, w_t^i); i=1,\dots,n_\pi\}$

End for

$\pi_T : \{(r_T^i, w_T^i); i=1,\dots,n_\pi\}$

represents $p(r_T \mid d_{0:T})$



Conditioning in PF

Particle Filter Algorithms

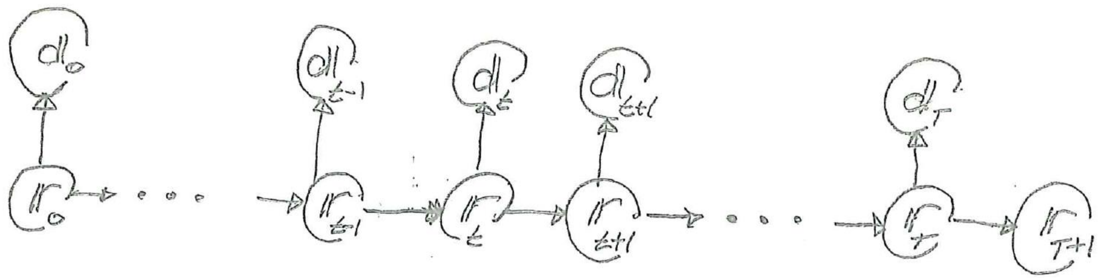
→ Challenge:

(A) Set of particles π_t 'collapses':

$w_t^j \approx 1.0$ and $w_t^i \approx 0$; $i = 1, \ldots, n_\pi$, $i \neq j$

(B): One particle fits $d_{0:T}$ much better than others

But objective fit need not be good since $\sum_i w_t^i = 1.0$



→ Mitigation strategies:

(A) - Force Forwarding move closer to d_t

Importance sampling:

$$r_t^i \rightsquigarrow q(r \mid r_{t-1}^i, d_t) \;;\; i = 1, \ldots, n_\pi$$

$$\tilde{w}_t^i = w_{t-1}^i \times \frac{p(r_t^i \mid r_{t-1}^i)}{q(r_t^i \mid r_{t-1}^i, d_t)} \times p(d_t \mid r_t^i) \;;\; i = 1, \ldots, n_\pi$$

Best : $q(r \mid r_{t-1}^i, d_t) = p(r \mid r_{t-1}, d_t)$ - not available

(B) - Balance Conditioned weights

Resampling importance sampling:

$$r_t^k \rightsquigarrow \left[r_t^j \text{ with prob } w_t^j \;;\; j = 1, \ldots, n_\pi \right] ; \; k = 1, \ldots, n_\pi$$

$$\pi_t : \left\{ \left(r_t^i, w_t^i = \frac{1}{n_\pi} \right) ; \; i = 1, \ldots, n_\pi \right\}$$

→ 'Brute-force' Particle Filter Algorithm

Initiate:

Set $n_\pi \in \mathbb{N}_+$

Generate $r_0^i = \omega_0(\epsilon^i_{r|r})$; $i=1,\dots,n_\pi$

Set $w_0^i = \frac{1}{n_\pi}$; $i=1,\dots,n_\pi$

$\Pi_0 = \{(r_0^i, w_0^i);\ i=1,\dots,n_\pi\}$

Do for $t=1,\dots,T$

Forwarding:

$r_t^i = \omega_t(r_{t-1}^i, \epsilon^i_{r|r})$; $i=1,\dots,n_\pi$

Conditioning:

$\tilde{w}_t^i = w_{t-1}^i \times p(d_t|r_t^i)$; $i=1,\dots,n_\pi$

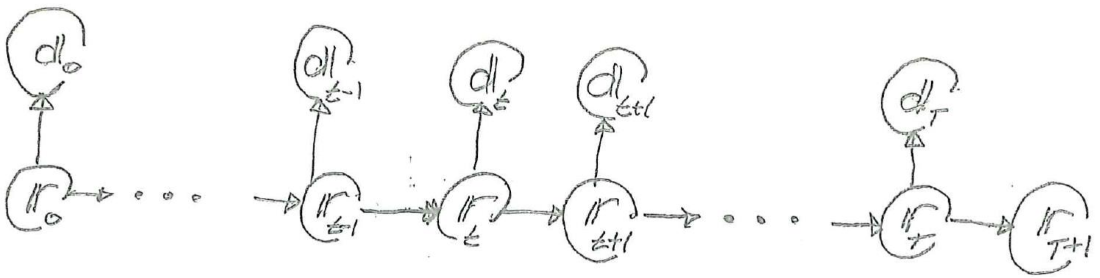
$w_t^i = \left[\sum_i \tilde{w}_t^i\right]^{-1} \tilde{w}_t^i$; $i=1,\dots,n_\pi$

$\Pi_t = \{(r_t^i, w_t^i);\ i=1,\dots,n_\pi\}$

End for

$\Pi_T : \{(r_T^i, w_T^i);\ i=1,\dots,n_\pi\}$

represents $p(r_T|d_{0:T})$

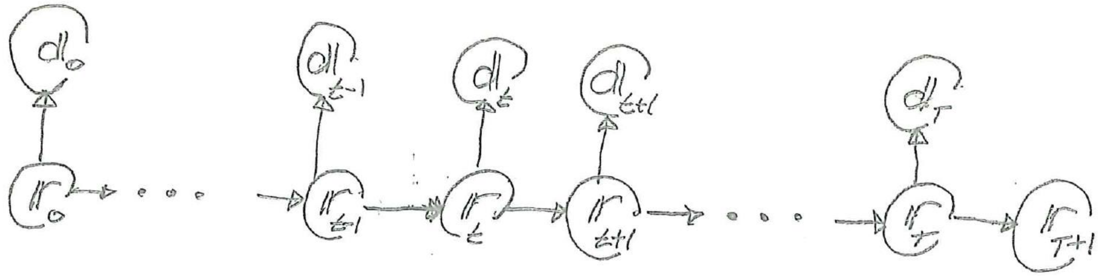


$r_t^i \sim q(r|r_{t-1}^i, d_t)$; $i=1,\dots,n_\pi$

$\tilde{w}_t^i = w_{t-1}^i \times \frac{p(r_t^i|r_{t-1}^i)}{q(r_t^i|r_{t-1}^i, d_t)} \times p(d_t|r_t^i)$; $i=1,\dots,n_\pi$

$r_t^k \sim [r_t^j \text{ with prob } w_t^j;\ j=1,\dots,n_\pi]$; $k=1,\dots,n_\pi$

$\Pi_t : \{(r_t^i, w_t^i = \frac{1}{n_\pi});\ i=1,\dots,n_\pi\}$



Particle Algorithms

→ The Particle Smoother – $[\mathbf{r}_{0:T} | \mathbf{d}_{0:T}] \approx p(\mathbf{r}_{0:T} | \mathbf{d}_{0:T})$

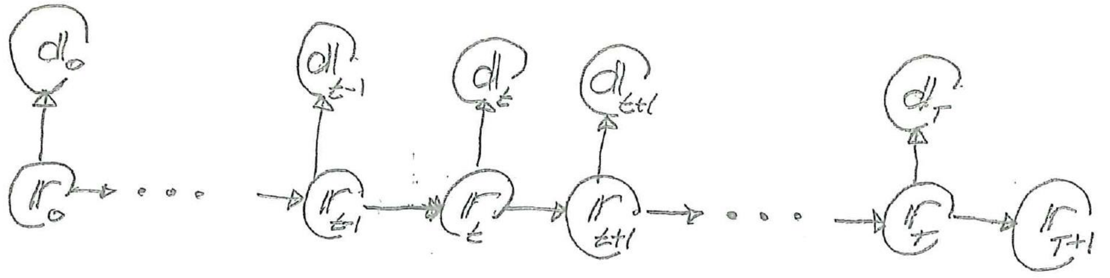
Activate Particle Filter:

$$\pi^s_T : \left\{ [\mathbf{r}^i_0, \ldots, \mathbf{r}^i_T, w^i_T];\ i = 1, \ldots, n_e \right\}$$

represents $p(\mathbf{r}_{0:T} | \mathbf{d}_{0:T})$

Challenge:

Severe particle 'collapse' for $t \ll T$.



$\longrightarrow$ Model parameter inference:

ML-estimator:

$$\hat{\Theta}_{lp} = \underset{\Theta_{lp}}{\arg\max} \{ p (dl ; \Theta_l , \Theta_p) \}$$

$$\llcorner \int_0 \cdots \int_T p(dl | r ; \Theta_e) p(r ; \Theta_p) dr_T \cdots dr_0$$

$$\rightarrow \underset{\Theta_{lp}}{\arg\max} \left\{ \sum_t \sum_i \tilde{w}_t^i (\Theta_{lp}) \right\}$$

$\llcorner$ must re-activate PF with varying Θ_{lp}.

Challenge:

Severely sensitive to sampling uncertainty

Example - HMM Continuous

Financial Stock Market

Focus on: Stock volatility

$x = [x_0, \dots, x_t, \dots, x_T, x_{T+1}]; \; x_t \in \mathbb{R}$

Available observations:

$z_t \in \mathbb{R}_\oplus$ - Stock price

$y_t = \log(z_t / z_{t-1}); \; y_t \in \mathbb{R}$ - log-ratio

$y = [y_0, \dots, y_t, \dots, y_T]$

True profile x^T and observations y.

Likelihood model:

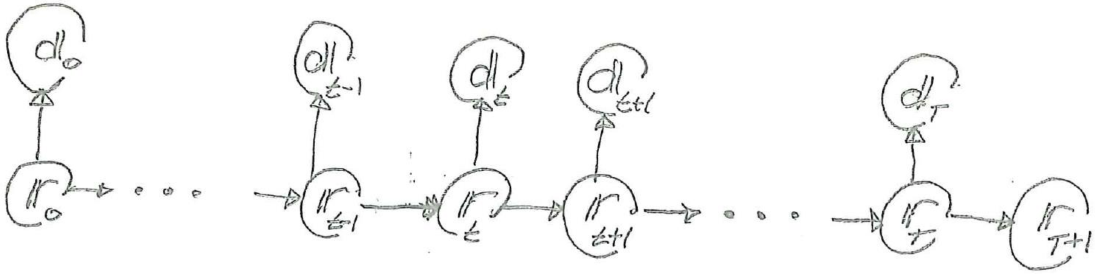
$$[y_t \mid x_t] = \beta \cdot \exp\{x_t/2\} + \epsilon_{y|x} \Rightarrow \varphi_1(y; \beta \cdot \exp\{x_t/2\}, \sigma^2_{y|x})$$

Prior model:

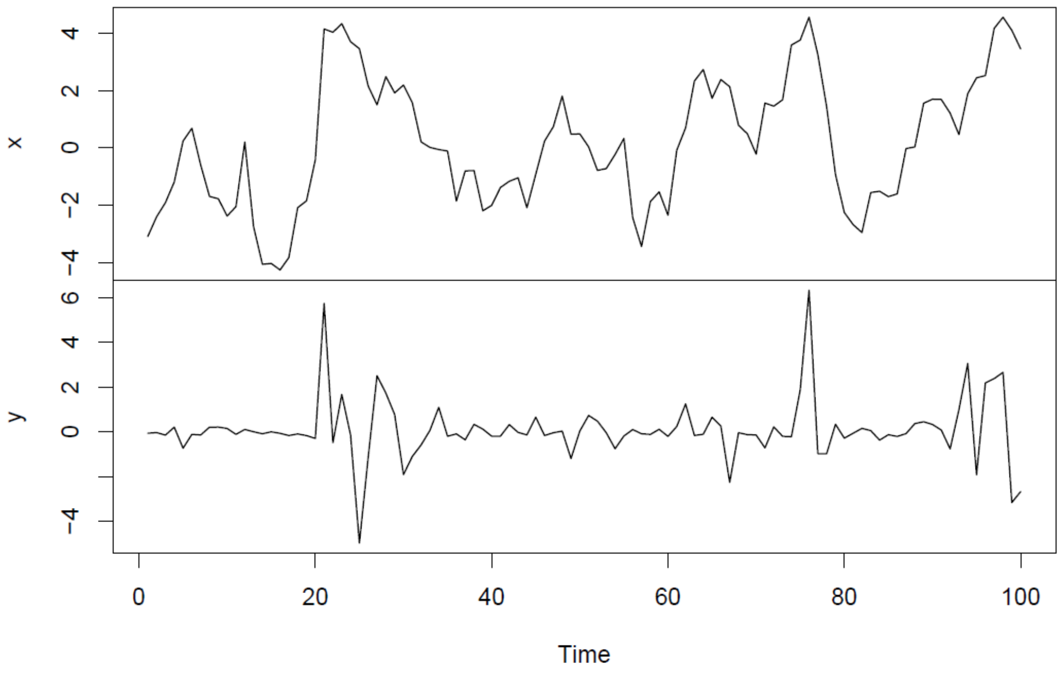
$$[x_t \mid x_{t-1}] = \alpha x_{t-1} + \epsilon_{x|x} \Rightarrow \varphi_1(x; \alpha x_{t-1}, \sigma^2_{x|x})$$

Results by Particle Filter - $n_\pi = 10\,000$

- Model parameter inference $\theta_p = (\alpha, \sigma^2_{x|x})$
- Filter model - $p(x_{T+1} \mid y_{0:T})$
  - E-prediction
  - Prediction $(1-\alpha)$ intervals



Stochastic volatility

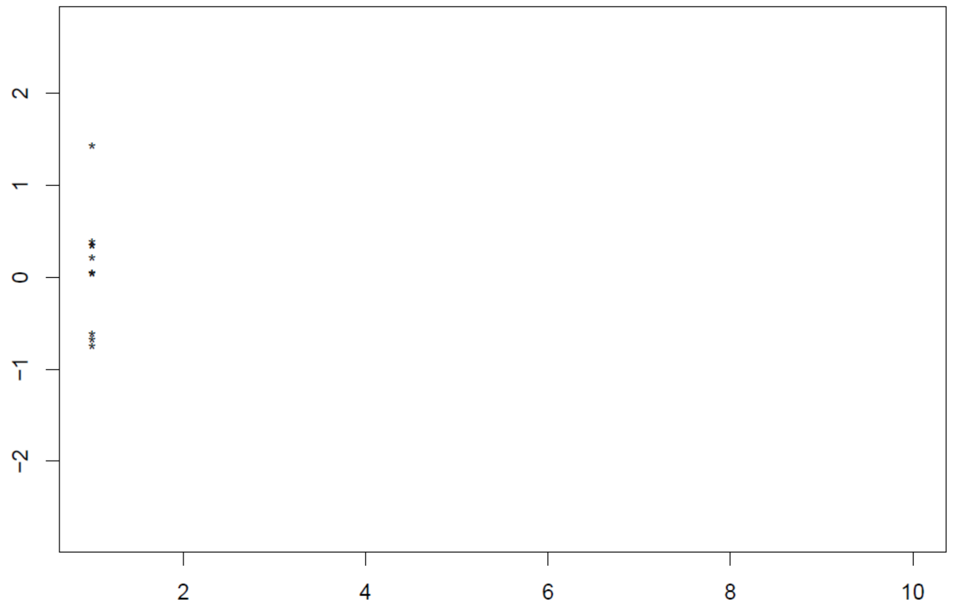


Sekvensiell Monte Carlo (eller Partikkelfiltre)

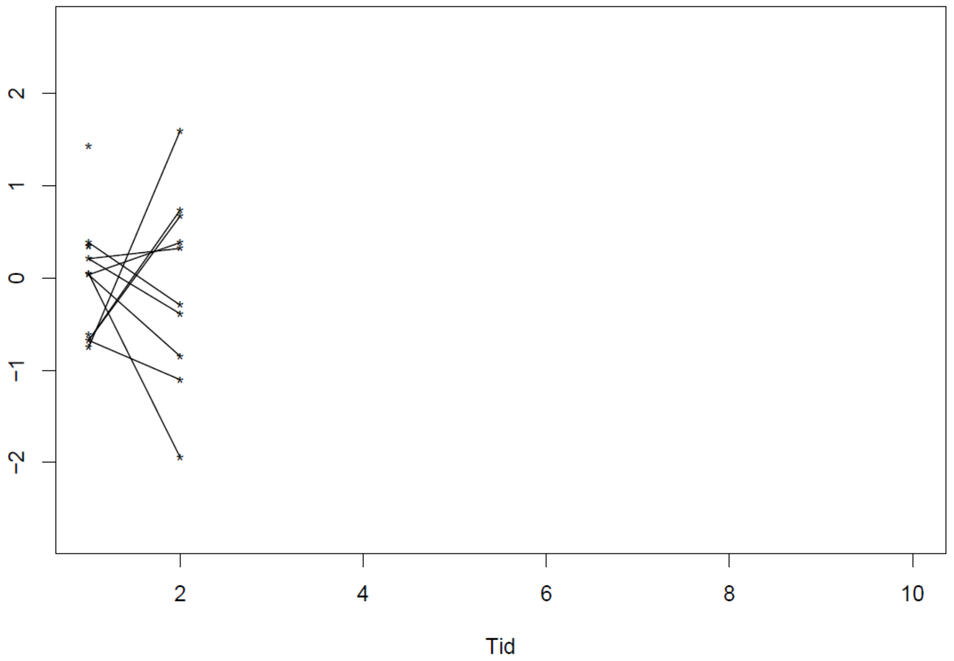
Geir Storvik

Dept. of Mathematics, University of Oslo
SFI[2], Statistics for inovation
Norwegian Computing Center
SFF-CEES, Ecology and Evolution synthesis

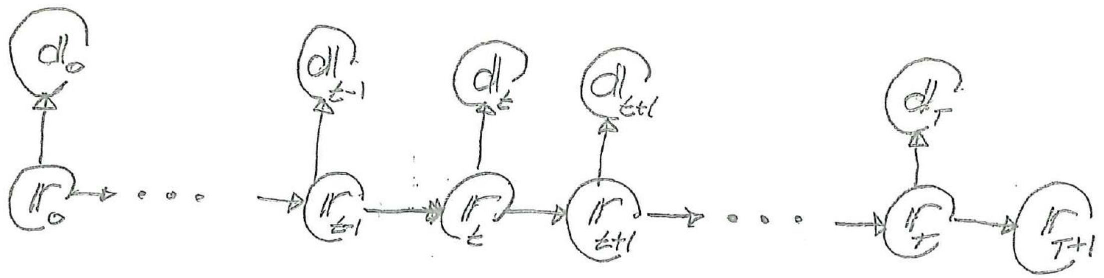
Jevnaker 16-18. juni 2009

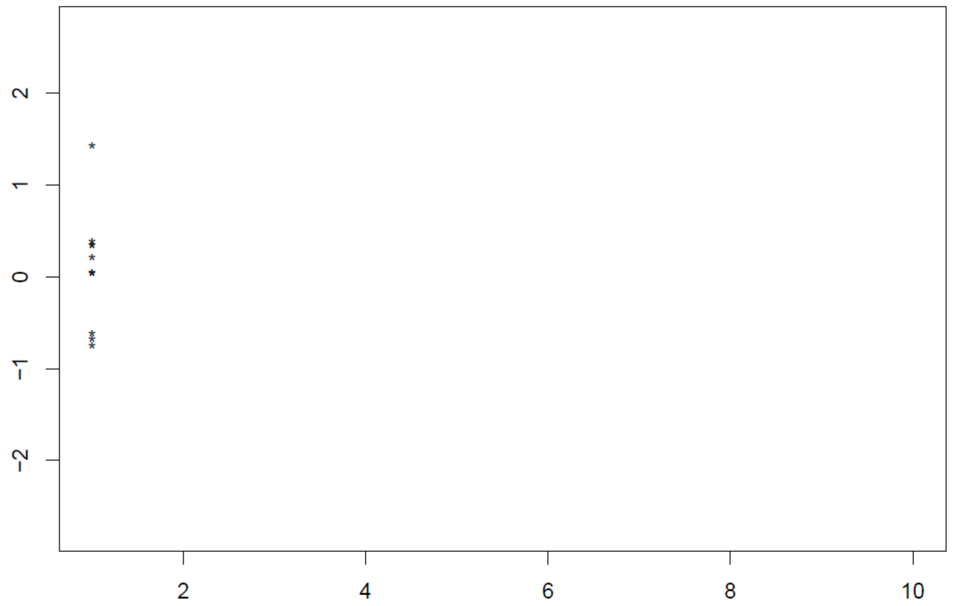


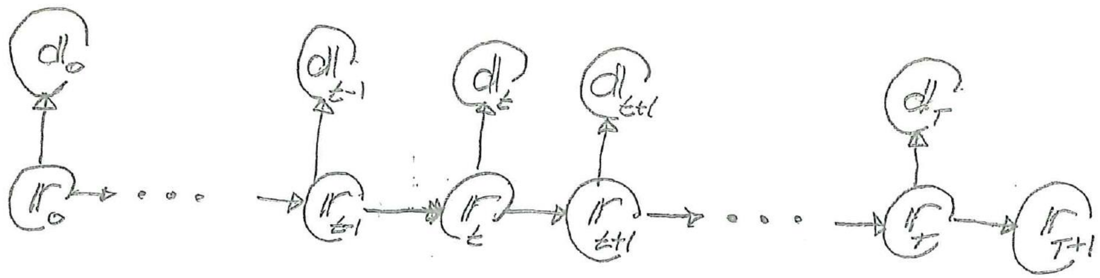
2
1
0
−1
−2
2
4
6
8
10

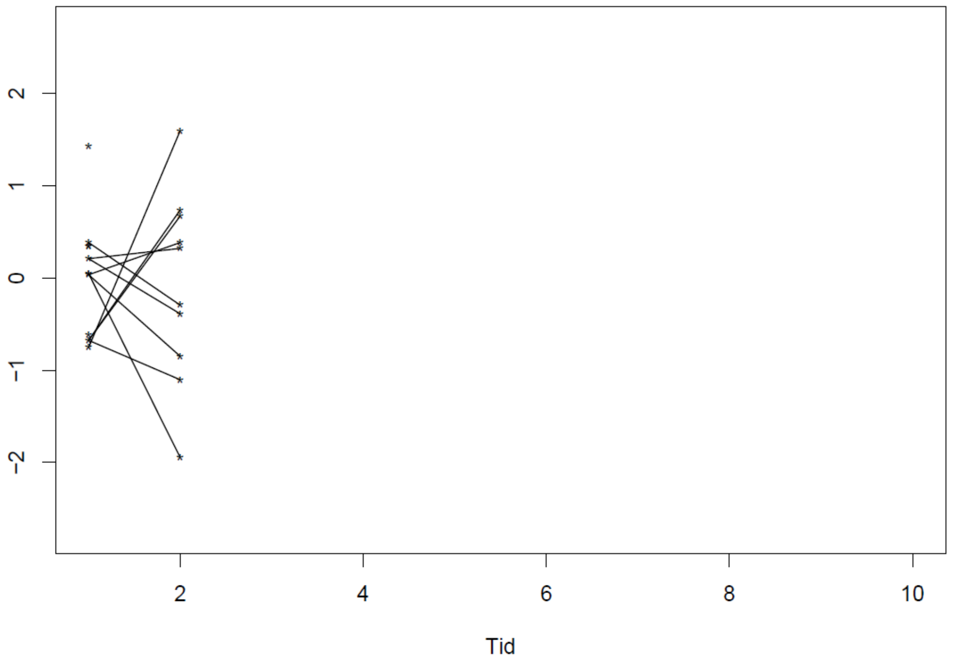


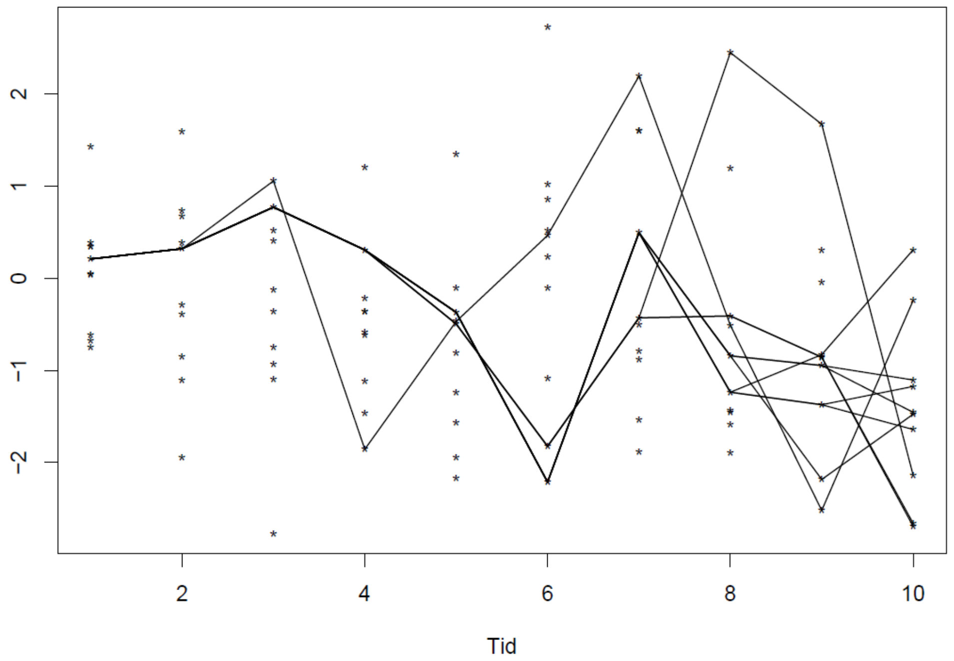
2
1
0
−1
−2
2
4
6
8
10
Tid

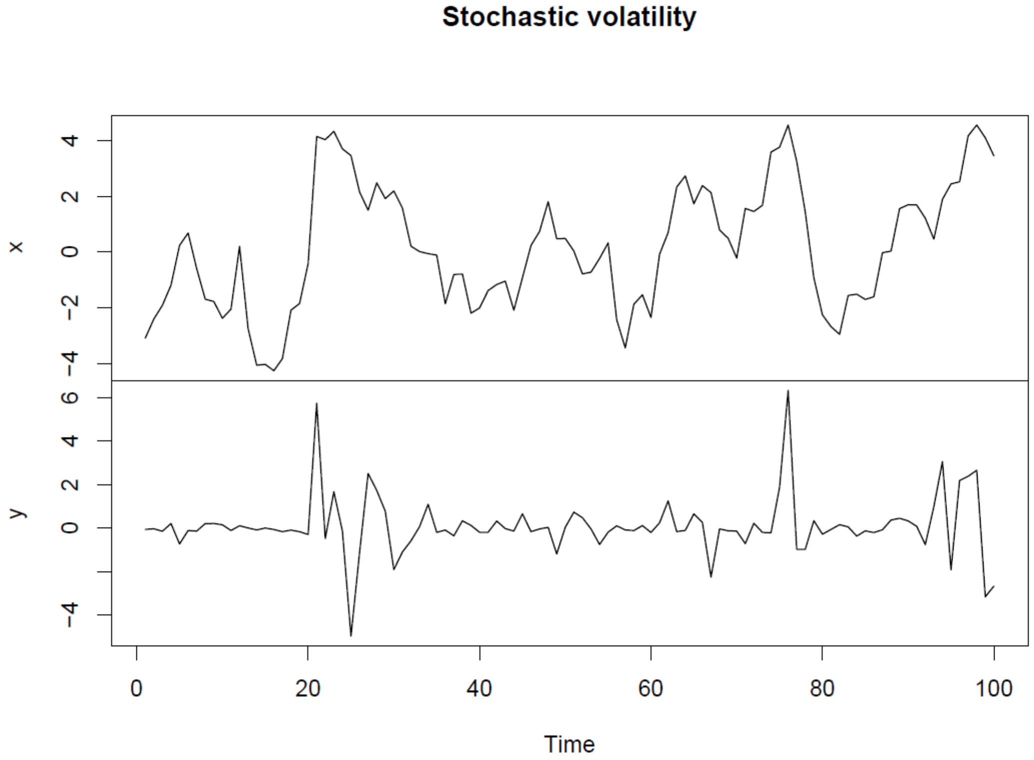


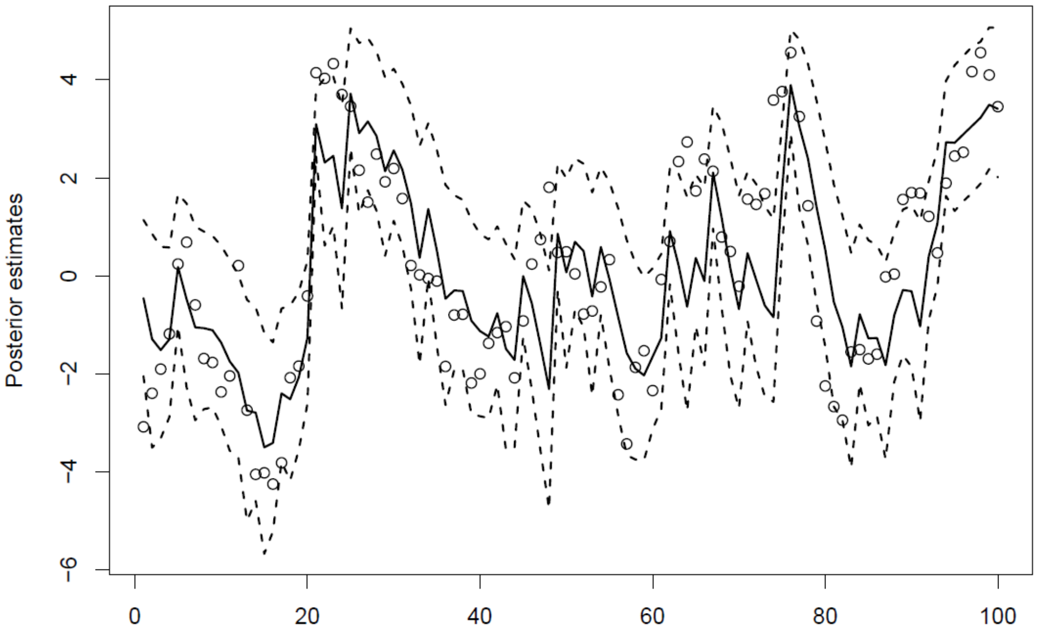

2
1
0
−1
−2
2
4
6
8
10

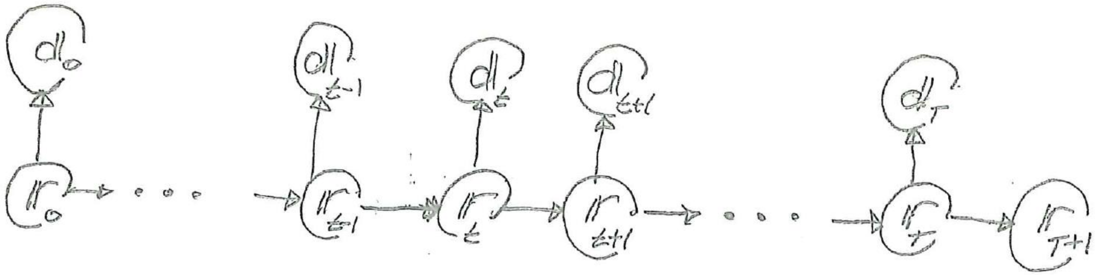

d_0
d_{t-1}
d_t
d_{t+1}
d_T
r_0
r_{t-1}
r_t
r_{t+1}
r_T
r_{T+1}

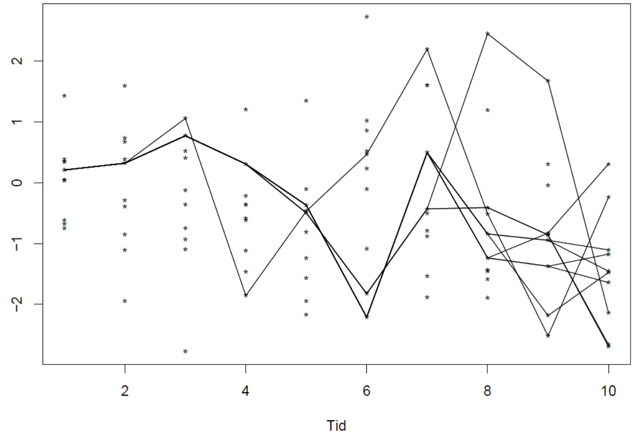

2
1
0
−1
−2
2
4
6
8
10
Tid


2
1
0
−1
−2
2
4
6
8
10
Tid


Stochastic volatility
x
y
Time
0
20
40
60
80
100

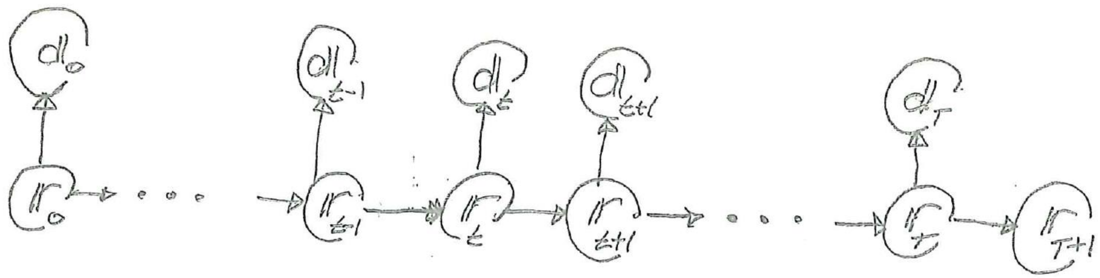

Posterior estimates
0
20
40
60
80
100


d_0
r_0
d_{t-1}
r_{t-1}
d_t
r_t
d_{t+1}
r_{t+1}
d_T
r_T
r_{T+1}


Tid
2
4
6
8
10

ENSEMBLE KALMAN ALGORITHMS

→ The Ensemble Kalman Filter $- [\mathbb{r}_{T+1} | \mathbb{d}_{0:T}] \sim p(\mathbb{r}_{T+1} | \mathbb{d}_{0:T})$



Based on Filter recursion:

$$p(\mathbb{r}_{t+1} | \mathbb{d}_{0:t}) = \int p(\mathbb{r}_{t+1} | \mathbb{r}_t) \left[p(\mathbb{d}_t | \mathbb{d}_{0:t-1}) \right]^{-1}_{\text{const}} p(\mathbb{d}_t | \mathbb{r}_t) p(\mathbb{r}_t | \mathbb{d}_{0:t-1}) \, d\mathbb{r}_t$$

< forw × cond >

Recursions:

Set of ensembles at time $t-1$:

$$\mathbb{e}^r_{t|} : \{ \mathbb{r}^{ui}_t ; i = 1, \dots, n_e \}$$

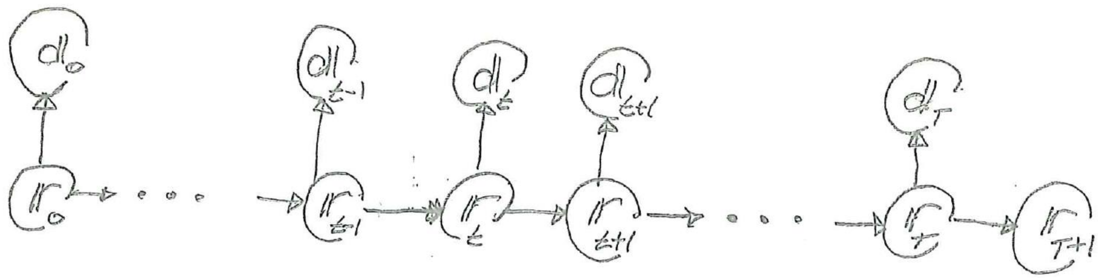
representing $p(\mathbb{r}_t | \mathbb{d}_{0:t-1})$

Perform Conditioning & Forwarding to time t:

$$\mathbb{e}^r_{t+1} : \{ \mathbb{r}^{ui}_{t+1} ; i = 1, \dots, n_e \}$$

representing $p(\mathbb{r}_{t+1} | \mathbb{d}_{0:t})$.

→ "Brute-force" Ensemble Kalman Filter Algorithm:



Initiate:

Set $n_e \in \mathbb{N}_+$

Generate $r_0^{ui} = \omega_0(\epsilon^i_{r})$; $i=1,\dots,n_e$

Generate $[d_0^i = \psi_0(r_0^{ui}, \epsilon^i_{d|r})$; $i=1,\dots,n_e$

$e_0 = \{(r_0^{ui}, d_0^i); i=1,\dots,n_e\}$

Do for $t = 0,\dots,T$

Conditioning:

Assess $\hat{\Sigma}_{rd}$ from e_t : $\hat{\Sigma}_{rd} \to \hat{K}_t = \hat{\Gamma}_{rd}[\hat{\Sigma}_d]^{-1}$

$r_t^{ci} = r_t^{ui} + \hat{K}_t[d_t - d_t^i]$; $i=1,\dots,n_e$

Forwarding:

$r_{t+1}^{ui} = \omega_t(r_t^{ci}, \epsilon^i_{r|r})$; $i=1,\dots,n_e$

$d_{t+1}^i = \psi_{t+1}(r_{t+1}^{ui}, \epsilon^i_{d|r})$; $i=1,\dots,n_e$

$e_{t+1} : \{(r_{t+1}^{ui}, d^i); i=1,\dots,n_e\}$

End for

$e^r_{T+1} : \{r^{ui}_{T+1}; i=1,\dots,n_e\}$

represent $p(r_{T+1}|d_{0:T})$

Non-parametric estimates:

$\hat{P}_{n_e}(r_{T+1}|d_{0:T}) = \frac{1}{n_e}\sum_i I(r^{ui}_{T+1} \le r_{T+1})$

$\hat{p}_{n_e}(r_{T+1}|d_{0:T}) = \frac{1}{n_e}\sum_i \delta(r_{T+1}, r^{ui}_{T+1})$

Asymptotics - $n_e \to \infty$ - NOT Very Good

→ "Brute-force" Ensemble Kalman Filter Algorithm:

Initiate:

Set $n_e \in \mathbb{N}_+$

Generate $r_0^{ui} = \omega_0(\epsilon_{r|r}^i)$; $i = 1, \ldots, n_e$

Generate $d_0^i = \psi_0(r_0^{ui}, \epsilon_{d|r}^i)$; $i = 1, \ldots, n_e$

$e_0 = \{(r_0^{ui}, d_0^i); i = 1, \ldots, n_e\}$

Do for $t = 0, \ldots, T$

Conditioning:

Assess $\hat{\Sigma}_{rd}$ from e_t : $\hat{\Sigma}_{rd} \to \hat{K}_t = \hat{\Gamma}_{rd}[\hat{\Sigma}_d]^{-1}$

$r_t^{ci} = r_t^{ui} + \hat{K}_t[d_t - d_t^i]$; $i = 1, \ldots, n_e$

Forwarding:

$r_{t+1}^{ui} = \omega_t(r_t^{ci}, \epsilon_{r|r}^i)$; $i = 1, \ldots, n_e$

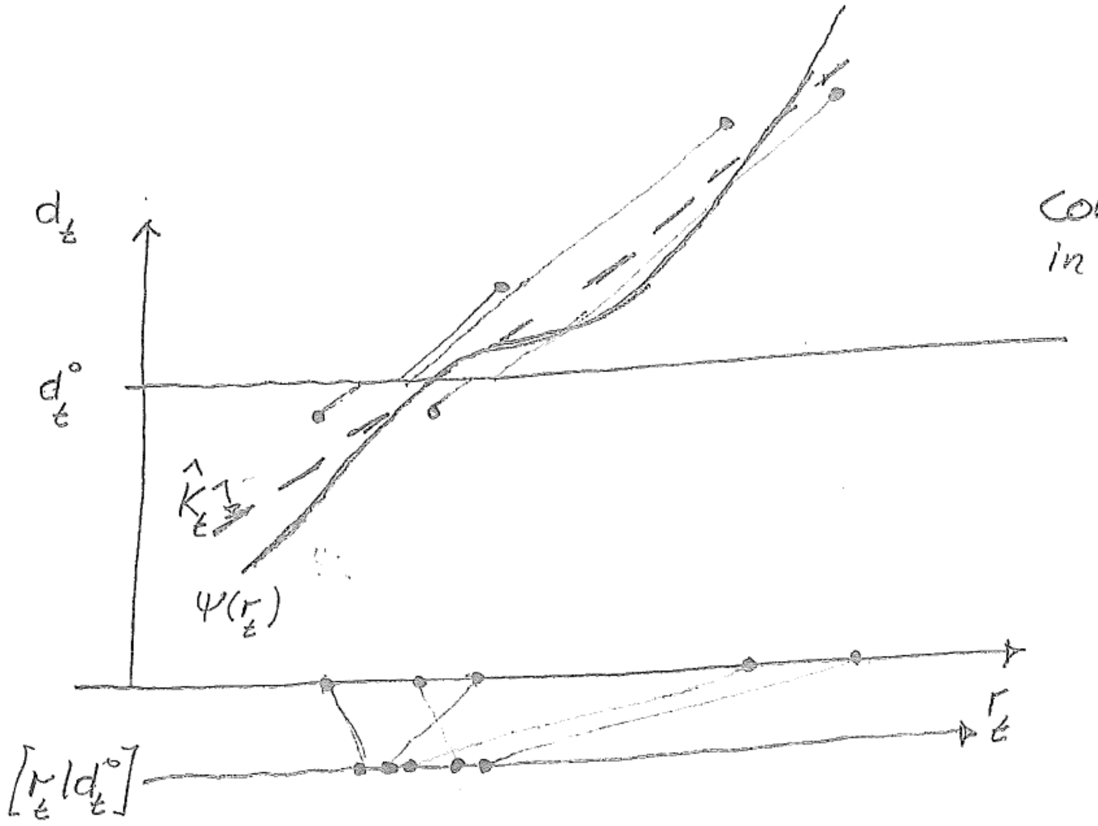
$d_{t+1}^i = \psi_{t+1}(r_{t+1}^{ui}, \epsilon_{d|r}^i)$; $i = 1, \ldots, n_e$

$e_{t+1} : \{(r_{t+1}^{ui}, d^i); i = 1, \ldots, n_e\}$

End for

$e_{T+1}^r : \{ r_{T+1}^{ui}; i = 1, \ldots, n_e \}$

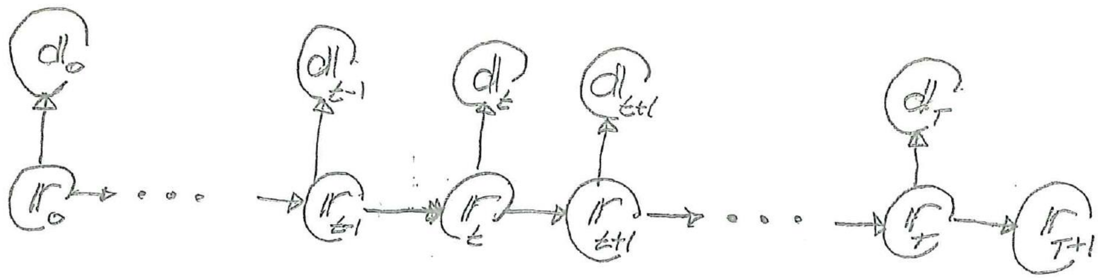
represent $p(r_{T+1} | d_{0:T})$



Conditioning in EnKF

→ Ensemble Kalman Filter Algorithms

NOTE: Conditioning is the complicated part!



Special Case:

Likelihood : $[d_t | r_t] = H r_t + \epsilon_{d|r} \Rightarrow p(d_t | r_t) = \varphi_m(d_t; H r_t, \Sigma_{d|r})$

Ensemble: $e_t : \{ r_t^{ui}; i = 1, \dots, n_e \}$ iid $\varphi_n(r; \mu_{t|t-1}, \Sigma_r)$

→ Trad EnKF – Randomized

→ Conditioning:

$$r_t^{ci} = r_t^{ui} + K_t\Big[d_t - \big(\overbrace{H r_t^{ui} + \epsilon_{d|r}}^{d_t^i}\big)\Big]$$

- $E\{r_t^{ci}\} = \mu_r + K_t[d_t - H\mu_r] = \mu_{r|d}$ ok!
- $Var\{r_t^{ci}\} = \Sigma_r + K_t[H\Sigma_r H^T + \Sigma_{d|r}]K_t^T - 2KH\Sigma_r$
  $= \Sigma_r - K_t[H\Sigma_r H^T + \Sigma_{d|r}]K_t^T = \Sigma_{r|d}$ ok!

→ Square-root EnKF – Deterministic

– Conditioning

$$r_t^{ci} = \mu_r + K_t[d_t - H\mu_r] + \underset{n\times n}{B}[r_t^{ui} - \mu_r]$$

- $E\{r_t^{ci}\} = \mu_r + K_t[d_t - H\mu_r] = \mu_{r|d}$ ok
- $Var\{r_t^{ci}\} = B\Sigma_r B^T = \Sigma_r - K_t[H\Sigma_r H^T + \Sigma_{d|r}]K_t^T = \Sigma_{r|d}$ ok

$$\underset{n\times n}{B} = \Big[\Sigma_r - K_t\underset{n\times n}{[H\Sigma_r H^T + \Sigma_{d|r}]}K_t^T\Big]^{1/2} \underset{n\times n}{\Sigma_r^{-1/2}}$$

Ensemble Kalman Filter Algorithms

Challenges:

(A) Instability in assessing $\mathbb{K}_t$ from $\mathbb{e}_t$ - n_e small

$$\underset{n\times m}{\hat{\mathbb{K}}_t} = \underset{n\times m}{\hat{\Gamma}_{rd}}\Big[\underset{m\times m}{\hat{\Sigma}_d}\Big]^{-1} \quad \text{based on } \mathbb{e}_t$$

(B) Set of ensembles $\mathbb{e}_t$ 'collapses' - identical members

$$\hat{\Sigma}_r = \frac{1}{n_e}\sum_i [r_t^i - \bar{r}_t][r_t^i - \bar{r}_t]^T \xrightarrow[t \text{ increase}]{} 0 \cdot \Sigma_r$$

(C) Drift towards Gaussianity - linear conditioning

$$\hat{p}_{n_e}(r_{t+1} | d_{0:t}) \xrightarrow[t \text{ increase}]{} \varphi_n(\cdot\, ; \cdot, \cdot)$$

Mitigation strategies:

(A) Two actions:

- Localization estimators - $\hat{K}_L$
- Shrinkage estimators - $\hat{K}_S$

(B) Two actions:

- Ensemble inflation - $\mathbb{e}_I$
- Resampling ensemble - $\mathbb{e}_R$

(C) Several actions

- Gaussian anamorphosis - GA-EnKF
- Selection - S-EnKF
- Multiple data assimilation - MDA-EnKF
- Iterative - I-EnKF

Ensemble Kalman Filter Algorithms

NOTE: Conditioning is the complicated part!

$$\mathbb{e}_t : \{(\mathbf{r}_t^{ui}, \mathbf{d}_t^{i}) ; i = 1, \dots, n_e\} \qquad \Sigma_{rd} = \begin{bmatrix} \Sigma_r & \Gamma_{rd} \\ \Gamma_{dr} & \Sigma_d \end{bmatrix}$$

Assess Σ_{rd} from $\mathbb{e}_t$: $\hat{\Sigma}_{rd} \to \hat{K}_t = \hat{\Gamma}_{rd} \hat{\Sigma}_d^{-1}$

$$\mathbf{r}_t^{ci} = \mathbf{r}_t^{ui} + \hat{K}_t [\mathbf{d}_t - \mathbf{d}_t^{i}] \; ; \; i = 1, \dots, n_e$$

NOTE:

Consider:

Likelihood model:

$$[\mathbf{d}_t | \mathbf{r}_t] = H \mathbf{r}_t + \epsilon_{d|r} \Rightarrow p(\mathbf{d}_t | \mathbf{r}_t) = \varphi_m(\mathbf{d} ; \underbrace{H \mathbf{r}_t , \Sigma_{d|r}}_{\text{known}})$$

Gauss-linear

Prior model:

$$[\mathbf{r}_t | \mathbf{r}_{t-1}] = \omega_t(\mathbf{r}_{t-1}, \epsilon_{r|r}) \Rightarrow p(\mathbf{r}_t | \mathbf{r}_{t-1})$$

$$\mathbf{r}_0 = \omega_0(\epsilon_{r|r}) \Rightarrow p(\mathbf{r}_0)$$

General

Conditioning:

$$\mathbb{e}_t : \{(\mathbf{r}_t^{ui}, \mathbf{d}_t^{i}) ; i = 1, \dots, n_e\}$$

Assess Σ_r from $\mathbb{e}_t^r$: $\underset{n \times n}{\hat{\Sigma}_r} \to \hat{K}_t = \hat{\Sigma}_r H^T [H \hat{\Sigma}_r H^T + \Sigma_{d|r}]^{-1}$

$$\mathbf{r}_t^{ci} = \mathbf{r}_t^{ui} + \hat{K}_t [\mathbf{d}_t - \mathbf{d}_t^{i}] \; ; \; i = 1, \dots, n_e$$

Ensemble Kalman Filter Algorithms

NOTE: Conditioning is the complicated part!

$\mathbb{e}_t : \{(\mathbb{r}_t^{ui}, d_t^i);\ i=1,\dots,n_e\}$ $\qquad \hat{\Sigma}_{rd} = \begin{bmatrix} \Sigma_{rr} & \Gamma_{rd} \\ \Gamma_{dr} & \Sigma_d \end{bmatrix}$

Assess $\hat{\Sigma}_{rd}$ from $\mathbb{e}_t$: $\hat{\Sigma}_{rd} \to \hat{\mathbb{K}}_t = \hat{\Gamma}_{rd}\, \hat{\Sigma}_d^{-1}$

$\mathbb{r}_t^{ci} = \mathbb{r}_t^{ui} + \hat{\mathbb{K}}_t [d_t - d_t^i]$; $i=1,\dots,n_e$

(A) Assessing $\mathbb{K}_t$ from $\mathbb{e}_t$

Trad EnKF:

$$\underset{(n+m)\times(n+m)}{\hat{\Sigma}_{rd}} = \frac{1}{n_e} \sum_i \underset{(n+m)\times 1}{\begin{bmatrix} \mathbb{r}^{ui} - \bar{\mathbb{r}}^u \\ d^i - \bar{d} \end{bmatrix}} \underset{1\times(n+m)}{\begin{bmatrix} \mathbb{r}^{ui} - \bar{\mathbb{r}}^u \\ d^i - \bar{d} \end{bmatrix}^T}$$

Often $n_e < n+m$ - rank problems

Often $n_e \approx n+m$ - unstable estimates

→ Alt A1: Localization estimator

Set $\tilde{\Sigma}_{rd} = \Omega \circ \hat{\Sigma}_{rd}$ - tapering

user-specified - pos.def - $(n+m)\times(n+m)$ matrix

'Shrinkage' of $\hat{\Sigma}_{rd}$ towards Ω!

→ Alt A2: Shrinkage estimator

Assume: $\mathbb{e}_t$ iid $p\left(\begin{bmatrix}\mathbb{r}\\ d\end{bmatrix} \middle| \mu_{rd}, \Sigma_{rd}\right) = \varphi_{n+m}\left(\begin{bmatrix}\mathbb{r}\\ d\end{bmatrix}; \mu_{rd}, \Sigma_{rd}\right)$

RV

Assign conjugate priors:

$p(\mu_{rd} \mid \Sigma_{rd}) = \varphi_{n+m}(\mu; \cdot, \cdot)$

$p(\Sigma_{rd}) = W_{n+m}^{-1}(\Sigma; \Sigma_p, \nu_p)$] user-specified

inverse-Wishard!

Posterior:

$p(\Sigma_{rd} \mid \mathbb{e}_t) = W_{n+m}^{-1}(\Sigma; \Sigma_{pp}, \nu_p + n_e)$

$\Sigma_{pp} = \Sigma_p + (n_e - 1)\hat{\Sigma}_{rd}$

Set $\tilde{\Sigma}_{rd} = E\{\Sigma_{rd} \mid \mathbb{e}_t\}$

$= [\nu_p + n_e - (n+m) - 1]^{-1}\, \Sigma_{pp}$

Shrinkage of $\hat{\Sigma}_{rd}$ towards $\frac{1}{\nu_p}\Sigma_p$

Ensemble Kalman Filter Algorithms

NOTE: Conditioning is the complicated part!

$\mathbb{E}_t : \{(\mathbb{r}_t^{ui}, d_t^i) ; i=1,\ldots,n_e\}$ $\quad \Sigma_{rd} = \begin{bmatrix} \Sigma_r & \Gamma_{rd} \\ \Gamma_{dr} & \Sigma_d \end{bmatrix}$

Assess Σ_{rd} from $\mathbb{E}_t$: $\hat{\Sigma}_{rd} \rightarrow \hat{K}_t = \hat{\Gamma}_{rd} \hat{\Sigma}_d^{-1}$

$\mathbb{r}_t^{ci} = \mathbb{r}_t^{ui} + \hat{K}_t [d_t - d_t^i]$; $i=1,\ldots,n_e$

(B) Ensemble $\mathbb{E}_t$ 'collapse'

Trad EnKF: $\underset{n \times n}{\hat{\Sigma}_r^t} = \frac{1}{n_e} \sum_i [\mathbb{r}_t^{ui} - \bar{\mathbb{r}}_t^u][\mathbb{r}_t^{ui} - \bar{\mathbb{r}}_t^u]^T \xrightarrow[t \text{ incr}]{} 0 \Sigma_r$

Reason: Ensemble coupling - use same $\hat{K}_t$ all i

→ Alt B1: Ensemble inflation

Set:

$\tilde{\mathbb{r}}_t^{ci} = \mathbb{r}_t^{ui} + \hat{K}_t [d_t - d_t^i]$; $i=1,\ldots,n_e$

$\mathbb{r}_t^{ci} = \bar{\tilde{\mathbb{r}}}_t^c + \alpha [\tilde{\mathbb{r}}_t^{ci} - \bar{\tilde{\mathbb{r}}}_t^c]$; $i=1,\ldots,n_e$

user-specified - factor $\alpha > 1.0$

Increases variability in $\mathbb{E}_t$.

→ Alt B2: Resampling ensemble

Set:

For $i=1,\ldots,n_e$

$\mathbb{E}_t^i \leftarrow$ Resample Uniformly $\mathbb{E}_t$

Assess Σ_{rd} from $\mathbb{E}_t^i$: $\hat{\Sigma}_{rd}^i \rightarrow \hat{K}_t^i = \hat{\Gamma}_{rd}^i [\hat{\Sigma}_d^i]^{-1}$

$\mathbb{r}_t^{ci} = \mathbb{r}_t^{ui} + \hat{K}_t^i [d_t - d_t^i]$

end for

Decouple ensembles in $\mathbb{E}_t$ by using different $\hat{K}_t^i$

Ensemble Kalman Filter Algorithms

NOTE: Conditioning is the complicated part!

$e_t: \{(r_t^{ui}, d_t^{i}); i=1,\dots,n_e\}$ $\qquad \Sigma_{rd} = \begin{bmatrix} \Sigma_r & \Gamma_{rd} \\ \Gamma_{dr} & \Sigma_d \end{bmatrix}$

Assess Σ_{rd} from e_t: $\hat\Sigma_{rd} \to \hat K_t = \hat\Gamma_{rd}\hat\Sigma_d^{-1}$

$r_t^{ci} = r_t^{ui} + \hat K_t[d_t - d_t^{i}]; \; i=1,\dots,n_e$

(G) Drift towards Gaussianity

Trad EnKF:

$$e^r_{T+1}: \{r^{ui}_{T+1}; i=1,\dots,n_e\} \to \hat p(r_{T+1}|d_{0:T}) \xrightarrow[T\ \text{incr}]{} \varphi_n(\cdot;\cdot,\cdot)$$

Subsequent linear conditioning causes 'regressions towards mean'!

→ Alt G1: Gaussian anamorphosis - GA-EnKF

Estimate From $e_t^r \to \begin{bmatrix}\hat P_{r_1}(r)\\ \vdots \\ \hat P_{r_n}(r)\end{bmatrix}$

Define $\tilde e_t^r = \left\{ \begin{bmatrix} \Phi_1^{-1}(\hat P_{r_1}(r_{t1}^{ui});0,1) \\ \vdots \\ \Phi_1^{-1}(\hat P_{r_n}(r_{tn}^{ui});0,1) \end{bmatrix}; i=1,\dots,n_e \right\}$

univar approx Gauss

Same procedure $e_t^d \to \tilde e_t^d$ and obs $d_t \to \tilde d_t$.

Assess Σ_{rd} from $\{\tilde e_t^r, \tilde e_t^d\}$: $\tilde\Sigma_{rd}$

Trad. EnKF conditioning on $\tilde d_t$.

Back-Transform $\tilde r_t^{ci} \to r_t^{ci}$!

Perform linear conditioning in Univar-Gauss space - back-transform

Ensemble Kalman Filter Algorithms

NOTE: Conditioning is the complicated part!

$\mathbb{e}_t : \{(\mathbb{r}_t^{ui}, \mathbb{d}_t^i); i = 1, \ldots, n_e\}$ $\quad \Sigma_{rd} = \begin{bmatrix} \Sigma_r & \Gamma_{rd} \\ \Gamma_{dr} & \Sigma_d \end{bmatrix}$

Assess Σ_{rd} from $\mathbb{e}_t$: $\hat{\Sigma}_{rd} \to \hat{K}_t = \hat{\Gamma}_{rd} \hat{\Sigma}_d^{-1}$

$\mathbb{r}_t^{ci} = \mathbb{r}_t^{ui} + \hat{K}_t[\mathbb{d}_t - \mathbb{d}_t^i]$; $i = 1, \ldots, n_e$

(G) Drift towards Gaussianity

Trad EnKF:

$$\mathbb{e}_{T+1}^r : \{\mathbb{r}_{T+1}^{ui}; i = 1, \ldots, n_e\} \to \hat{p}(\mathbb{r}_{T+1} | \mathbb{d}_{0:T}) \xrightarrow[T \text{ incr}]{} \varphi_n(\cdot; \cdot, \cdot)$$

Subsequent linear conditioning causes 'regressions towards mean'!

→ Alt G2: Selection - S-EnKF

Define:

Axiliary space $\underset{2n\times 1}{\begin{bmatrix} \tilde{\mathbb{r}} \\ \tilde{\nu} \end{bmatrix}} \sim \varphi_{2n}\left(\begin{bmatrix} \tilde{\mathbb{r}} \\ \tilde{\nu} \end{bmatrix}; \begin{bmatrix} \mu_{\tilde{r}} \\ \mu_{\tilde{\nu}} \end{bmatrix}, \begin{bmatrix} \Sigma_{\tilde{r}} & \Gamma_{\tilde{r}\tilde{\nu}} \\ \Gamma_{\tilde{\nu}\tilde{r}} & \Sigma_{\tilde{\nu}} \end{bmatrix}\right)$

Selection set $A \subset \mathbb{R}^n$

Variable of interest:

$$\underset{n\times 1}{\mathbb{r}} = [\tilde{\mathbb{r}} | \tilde{\nu} \in A] \sim p(\mathbb{r}) = p(\tilde{\mathbb{r}} | \tilde{\nu} \in A)$$

$$= [\Phi_n(A; \mu_{\tilde{\nu}}, \Sigma_{\tilde{\nu}})]^{-1} \times \Phi_n(A; \mu_{\tilde{\nu}|\tilde{r}}, \Sigma_{\tilde{\nu}|\tilde{r}}) \times \varphi_n(\tilde{\mathbb{r}}; \mu_{\tilde{r}}, \Sigma_{\tilde{r}})$$

Let:

$$p(\mathbb{d}, \tilde{\nu} | \tilde{\mathbb{r}}) = p(\mathbb{d} | \tilde{\mathbb{r}}) p(\nu | \tilde{\mathbb{r}}) \Rightarrow p(\mathbb{d} | \mathbb{r}) = p(\mathbb{d} | \tilde{\mathbb{r}})$$

cond. indep

Trad EnKF on $\begin{bmatrix} \tilde{\mathbb{r}} \\ \tilde{\nu} \end{bmatrix} \in \mathbb{R}^{2n}$ condition on $\mathbb{d}_t$

Then

$$\mathbb{e}_{T+1}^{\tilde{r}\tilde{\nu}} = \{(\tilde{\mathbb{r}}_{T+1}^i, \tilde{\nu}_{T+1}^i); i = 1, \ldots, n_e\} \begin{cases} \to \hat{p}(\tilde{\mathbb{r}}_{T+1}, \tilde{\nu}_{T+1} | \mathbb{d}) \\ \to \varphi_{2n}\left(\begin{bmatrix} \tilde{\mathbb{r}} \\ \tilde{\nu} \end{bmatrix}; \hat{\mu}_{T+1|T}^{r\nu}, \hat{\Sigma}_{T+1|T}^{r\nu}\right) \end{cases}$$

Variable of interest

$$[\mathbb{r}_{T+1} | \mathbb{d}] = [\tilde{\mathbb{r}}_{T+1} | \tilde{\nu}_{T+1} \in A, \mathbb{d}] \sim p(\mathbb{r}_{T+1} | \mathbb{d}) = p(\tilde{\mathbb{r}}_{T+1} | \tilde{\nu}_{T+1} \in A, \mathbb{d})$$

$$= \cdots\cdots$$

Perform linear conditioning in auxiliary space - back-project.

Ensemble Kalman Filter Algorithms

NOTE: Conditioning is the complicated part!

$$\mathcal{E}_t : \{(\mathbf{r}_t^{ui}, \mathbf{d}_t^{i}) ; i = 1, \dots, n_e\} \qquad \Sigma_{rd} = \begin{bmatrix} \Sigma_r & \Gamma_{rd} \\ \Gamma_{dr} & \Sigma_d \end{bmatrix}$$

Assess Σ_{rd} from $\mathcal{E}_t$: $\hat{\Sigma}_{rd} \rightarrow \hat{K}_t = \hat{\Gamma}_{rd} \hat{\Sigma}_d^{-1}$

$$\mathbf{r}_t^{ci} = \mathbf{r}_t^{ui} + \hat{K}_t [\mathbf{d}_t - \mathbf{d}_t^{i}] ; i = 1, \dots, n_e$$

(G) Drift towards Gaussianity

Trad EnKF:

$$\mathcal{E}_{T+1}^{r} : \{\mathbf{r}_{T+1}^{ui} ; i = 1, \dots, n_e\} \rightarrow \hat{p}(\mathbf{r}_{T+1} | \mathbf{d}_{0:T}) \xrightarrow[\text{Tiner}]{} \varphi_n(\cdot ; \cdot, \cdot)$$

Subsequent linear conditioning causes 'regressions towards mean'!

→ Alt G3: Multiple data assimilation - MDA-EnKF

Likelihood model:

$$[\mathbf{d}_t | \mathbf{r}_t] = \psi(\mathbf{r}_t) + \mathcal{E}_{d|r} \Rightarrow p(\mathbf{d}_t | \mathbf{r}_t) = \varphi_m(\mathbf{d} ; \psi(\mathbf{r}_t), \Sigma_{d|r})$$

Always correct:

$$p(\mathbf{d}_t | \mathbf{r}_t) = \prod_{j=1}^{n_e} [p(\mathbf{d}_t | \mathbf{r}_t)]^{w_j} \quad ; \sum_j w_j = 1$$

Then posterior model:

$$p(\mathbf{r}_t | \mathbf{d}_t) = \text{const} \times \prod_{j=1}^{n_e} [p(\mathbf{d}_t | \mathbf{r}_t)]^{w_j} p(\mathbf{r}_t)$$

$$= \text{const} \times [p(\mathbf{d}_t | \mathbf{r}_t)]^{w_1} \times \quad [p(\mathbf{d}_t | \mathbf{r}_t)]^{w_{n_j - 1}} [p(\mathbf{d}_t | \mathbf{r}_t)]^{w_{n_j}} p(\mathbf{r}_t)$$

Sequential conditioning!

Let set $w = \{w_1, \dots, w_{n_j}\}$ - decreasing - $w_j > 0$, $\sum_j w_j = 1$

Then

$$[p(\mathbf{d}_t | \mathbf{r}_t)]^{w_j} = \text{const} \times [\exp\{-\tfrac{1}{2} [\mathbf{d}_t - \psi(\mathbf{r}_t)]^T \Sigma_{d|r}^{-1} [\mathbf{d}_t - \psi(\mathbf{r}_t)]\}]^{w_j}$$

$$= \text{const} \times \exp\{-\tfrac{1}{2} [\mathbf{d}_t - \psi(\mathbf{r}_t)]^T [\tfrac{1}{w_j} \Sigma_{d|r}]^{-1} [\mathbf{d}_t - \psi(\mathbf{r}_t)]\}$$

$$= \varphi_m(\mathbf{d} ; \psi(\mathbf{r}_t), \tfrac{1}{w_j} \Sigma_{d|r})$$

Gauss with enlarged variance

Trad EnKF with sequential decomposed conditioning on $\mathbf{d}_t$.

Perform multiple linear conditioning.

Ensemble Kalman Filter Algorithms

NOTE: Conditioning is the complicated part!

$\mathbb{e}_t : \{ (\mathbb{r}_t^{ui}, \mathbb{d}_t^i) ; i = 1, \dots, n_e \}$ $\qquad \Sigma_{rd} = \begin{bmatrix} \Sigma_r & \Gamma_{rd} \\ \Gamma_{dr} & \Sigma_d \end{bmatrix}$

Assess Σ_{rd} from $\mathbb{e}_t$: $\hat{\Sigma}_{rd} \rightarrow \hat{\mathbb{K}}_t = \hat{\Gamma}_{rd} \hat{\Sigma}_d^{-1}$

$\mathbb{r}_t^{ci} = \mathbb{r}_t^{ui} + \hat{\mathbb{K}}_t [\mathbb{d}_t - \mathbb{d}_t^i] ; \; i = 1, \dots, n_e$

(C) Drift towards Gaussianity

Trad EnKF:

$$\mathbb{e}_{T+1}^r : \{ \mathbb{r}_{T+1}^{ui} ; i = 1, \dots, n_e \} \rightarrow \hat{p}(\mathbb{r}_{T+1} | \mathbb{d}_{0:T}) \xrightarrow[T \text{ incr}]{} \varphi_n(\cdot ; \cdot, \cdot)$$

Subsequent linear conditioning causes 'regressions towards mean'!

→ Alt C4: Iterative - I-EnKF

Likelihood model:

$$[\mathbb{d}_t | \mathbb{r}_t] = \psi(\mathbb{r}_t) + \epsilon_{d|r} \Leftrightarrow p(\mathbb{d}_t | \mathbb{r}_t) = \varphi_m(\mathbb{d}; \psi(\mathbb{r}_t), \Sigma_{d|r})$$

Consider first Kalman Model: $p(\mathbb{d}_t | \mathbb{r}_t)$ - Gauss-lin

$p(\mathbb{r}_t)$ - Gauss

Then posterior

$$p(\mathbb{r}_t | \mathbb{d}_t) = \text{const} \times p(\mathbb{d}_t | \mathbb{r}_t) p(\mathbb{r}_t) \text{ - Gauss}$$

Note:

$$\mu_{r|d} = \underset{\mathbb{r}_t}{\operatorname{argmax}} \{ \text{const} \times p(\mathbb{d}_t | \mathbb{r}_t) p(\mathbb{r}_t) \}$$

$$= \underset{\mathbb{r}_t}{\operatorname{argmin}} \{ [\mathbb{d}_t - \mathbb{H}\mu_r]^T \Sigma_{d|r}^{-1} [\mathbb{d}_t - \mathbb{H}\mu_r] + [\mathbb{r}_t - \mu_r]^T \Sigma_r^{-1} [\mathbb{r}_t - \mu_r] \}$$

$$= \mu_r + \mathbb{K}[\mathbb{d}_t - \mathbb{H}\mu_r] = \mu_r + \mathbb{K}[\mathbb{d}_{t*} - \mathbb{H}\mu_r]$$

$$\mathbb{r}_t^c = \mathbb{r}_t^u + \mathbb{K}[\mathbb{d}_t - \mathbb{d}_t^i] = \mathbb{r}_t^u + \mathbb{K}[\underbrace{\mathbb{d}_t - \epsilon_{d|r}} - \mathbb{H}\mathbb{r}_t^u]$$

$\mathbb{H}\mathbb{r}_t^u + \epsilon_{d|r}$ $\qquad$ pseudo-truth?

$\mathbb{d}_t = \mathbb{d}_t^T + \epsilon_{d|r}$

Hence

$$\mathbb{r}_t^c = \underset{\mathbb{r}}{\operatorname{argmin}} \{ [\mathbb{d}_t - \epsilon_{d|r} - \mathbb{H}\mathbb{r}]^T \Sigma_{d|r}^{-1} [\mathbb{d}_t - \epsilon_{d|r} - \mathbb{H}\mathbb{r}] + [\mathbb{r} - \mathbb{r}_t^u]^T \Sigma_r^{-1} [\mathbb{r} - \mathbb{r}_t^u] \}$$

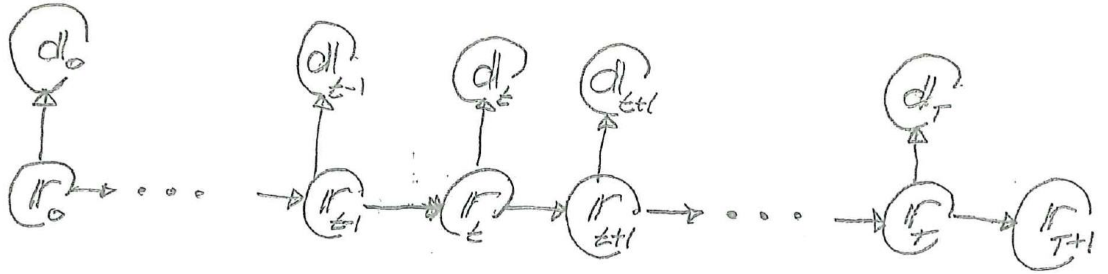
non-Lin Gauss → $\psi(\mathbb{r})$ $\qquad$ $\psi(\mathbb{r})$

Optimize iteratively - Gauss-Newton

Trad EnKF with iterative optimization conditioning on $\mathbb{d}_t$.

Perform non-linear optimization conditioning

→ Ensemble Kalman Smoother - EnKS



→ Define variable of interest:

$\mathbf{r}_t^+ = [\mathbf{r}_{0:t}]$ – NOTE: Dimension increase with t

• Likelihood model:

$$[\mathbf{d}_t | \mathbf{r}_t^+] = [\mathbf{d}_t | \mathbf{r}_{0:t}] = [\mathbf{d}_t | \mathbf{r}_t] \to p(\mathbf{d}_t | \mathbf{r}_t)$$

• Prior model:

$$[\mathbf{r}_t^+ | \mathbf{r}_{t-1}^+] = [\mathbf{r}_{t-1}^+ ; \omega(\mathbf{r}_{t-1}, \epsilon_{r|r})] \to p(\mathbf{r}_t^+ | \mathbf{r}_{t-1}^+)$$

$$\mathbf{r}_0^+ = \mathbf{r}_0 \to p(\mathbf{r}_0)$$

Trad EnKF algorithm on accumulate $\mathbf{r}_t^+$ vector

Result:

$\mathcal{E}_{T+1}^+ : \{ \mathbf{r}_{T+1}^{+ui} ; i = 1, \dots, n_e \}$ represent $p(\mathbf{r}_{0:T+1} | \mathbf{d})$

NOTE:

Dimension of Σ_{r^+d} increases fast

Must store all ensemble members at all t.

→ Model parameter inference EnKF

Bayesian framework!

Define variable of interest:

$$\begin{bmatrix} r_t \\ \theta_t \end{bmatrix} ; \ t \in T_s$$

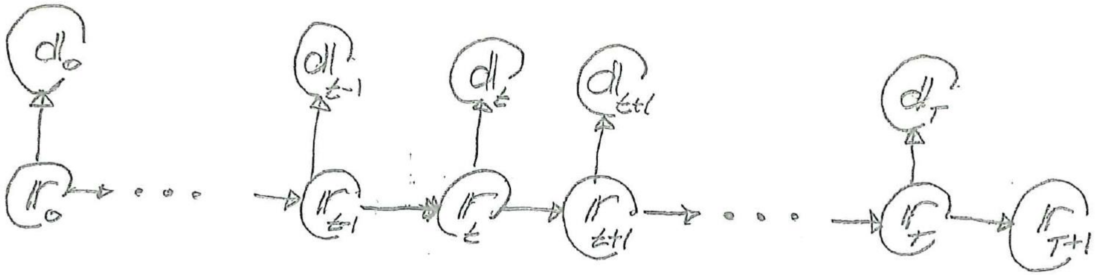
Observation likelihood:

$$\left[d_t \,\middle|\, \begin{matrix} r_t \\ \theta_t \end{matrix} \right] = \psi_t(r_t, \theta_t, \epsilon_{d|\cdot}) \to p(d_t \mid r_t ; \theta_t)$$

Prior model:

$$\left[\begin{matrix} r_{t+1} \\ \theta_{t+1} \end{matrix} \,\middle|\, \begin{matrix} r_t \\ \theta_t \end{matrix} \right] = \begin{bmatrix} \omega_t(r_t, \theta_t, \epsilon_{r|\cdot}) \\ \theta_t \end{bmatrix} \to p\left(\begin{matrix} r_{t+1} \\ \theta_{t+1} \end{matrix} \,\middle|\, r_t ; \theta_t \right)$$

$$\begin{bmatrix} r_0 \\ \theta_0 \end{bmatrix} \to p\begin{bmatrix} r_0 \\ \theta_0 \end{bmatrix}$$



Trad EnKF on extended variable $\begin{bmatrix} r_t \\ \theta_t \end{bmatrix}$

Note:

If focus is on θ only

simplest!

EnKS $\Leftrightarrow$ EnKF

smoother filter

since

$$[\theta_{t+1} \mid \theta_t] = \theta_t \to \delta(\theta ; \theta_t)$$

Dirac!

Anderson, J. (2007). An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A: Dynamic Meteorology and Oceanography*, 59:210 – 224.

Cappé, O., Moulines, E., and Ryden, T. (2005). *Inference in Hidden Markov Models (Springer Series in Statistics)*. Springer-Verlag, Berlin, Heidelberg.

Conjard, M. and Omre, H. (2020). Data assimilation in spatio-temporal models with non-Gaussian initial states—the selection ensemble Kalman model. *Applied Sciences*, 10(17):5742.

Doucet, A., de Freitas, N., and Gordon, N. (2001). *An Introduction to Sequential Monte Carlo Methods*, pages 3–14. Springer New York, New York.

Emerick, A. and Reynolds, A. (2012). History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. *Computational Geosciences*, 16:639–659.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5):101–143.

Evensen, G., 2007. Data Assimilation. The Ensemble Kalman Filter. Springer.

Evensen, G. and van Leeuwen, P. J. (2000). An Ensemble Kalman Smoother for Nonlinear Dynamics. *Monthly Weather Review*, 128(6):1852–1867.

Gaspari, G. and Cohn, S. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125:723–757.

Gordon, N. J., Salmond, D. J., and Smith, A. F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F (Radar and Signal Processing)*, 140(2):107–113.

Gu, Y. and Oliver, D. (2007). An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. *SPE Journal*, 12:438–446.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45.

Katzfuss, M., Stroud, J. R., and Wikle, C. K. (2020). Ensemble Kalman methods for high-dimensional hierarchical dynamic space-time models. *Journal of the American Statistical Association*, 115(530):866–885.

Myrseth, I. and Omre, H. (2010). Hierarchical ensemble Kalman filter. *SPE Journal*, 15(02):569–580.

Omre, H. and Rimstad, K. (2021). Bayesian spatial inversion and conjugate selection Gaussian prior models. *Journal of Uncertainty Quantification*, To appear.

Simon, E. and Bertino, L. (2009). Application of the Gaussian anamorphosis to assimilation in a 3-d coupled physical-ecosystem model of the north atlantic with the EnKF: a twin experiment. *Ocean Science*, 5(4):495–510.

Sætrom, J. and Omre, H. (2013). Uncertainty quantification in the ensemble Kalman filter. *Scandinavian Journal of Statistics*, 40(4):868–885.