Coupling MCMC and EM algorithms for stochastic differential equations with hidden components

Joint works with Susanne Ditlevsen (Copenhagen Univ) and Sophie Donnet (INRA)

Adeline Leclercq Samson

Laboratoire Jean Kuntzmann, Université Joseph Fourier, Grenoble

Objectives of the talk

1. Framework of the deterministic EM algorithm
   - Gaussian mixture
   - Gaussian linear mixed model
   - Gaussian linear state-space model
2. More complex models and limits of EM
   - Non Gaussian hierarchical models
   - Hidden Markov Models
   - Multi dimensional stochastic differential equations partially observed
3. Stochastic EM algorithm coupled with
   - MCMC
   - Particle filter/Sequential Monte Carlo
   - PMCMC

Framework of the EM algorithm

- Incomplete data model
  - Observed data $(y_{0:n})$
  - Hidden components $(z_{0:n})$
  - Complete data $(y_{0:n}, z_{0:n})$ with distribution $p(y_{0:n}, z_{0:n}, \theta)$

- Likelihood of the observed data $(y_{0:n})$ may be non explicit

$$p(y_{0:n}; \theta) = \int p(y_{0:n}, z_{0:n}; \theta) dz_{0:n}$$

- EM algorithm [Dempster, Laird, Rubin, 1977], iteration m
  - *E Step*: computation of $Q_{m+1}(\theta) = \mathbb{E}\left[\log p(y_{0:n}, z_{0:n}; \theta) \,|\, y_{0:n}, \widehat{\theta}_m\right]$
  - *M Step*: update $\widehat{\theta}_{m+1} = \arg\max_{\theta} Q_{m+1}(\theta)$

- Convergence results [Wu, 1983]

Example 1: Finite mixture models

$$p(y_{0:n}, \theta) = \sum_{k=1}^{p} \pi_k p_k(y_{0:n}, \theta_k)$$

- π_k unknown mixing proportions, $\sum_{k=1}^{p} \pi_k = 1$
- θ_k unknown parameters of the kth component
- $\theta = (\theta_1, \ldots, \theta_p, \pi_1, \ldots, \pi_{p-1})$

- Examples of applications
  - Image segmentation
  - Handwriting recognition
  - Topics in a document (if the vocabulary size is not too large)

- EM algorithm [Dempster, Laird, Rubin, 1977], [McLachlan, Krishnan 1997]
  - $z_{ki} = 1$ if y_i belongs to component k
  - E step: $p(\widehat{z_{ki} = 1|y_i}, \hat{\theta}^{(m)})^{(m)} = \hat{\pi}_k^{(m)} p(y_i, \hat{\theta}_k^{(m)}) / \sum_{\ell=1}^{p} \hat{\pi}_\ell^{(m)} p(y_i, \hat{\theta}_\ell^{(m)})$
  - M step: depends on the mixture (explicit for Gaussian, Poisson mixtures)

Example 2: Regression model with random effects

$$\begin{aligned} y_{ij} &= X_{ij}\phi_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \\ \beta_i &\sim \mathcal{N}(\beta, \Omega) \end{aligned}$$

- y_{ij} jth observation of group i, X_{ij} design matrix of group i
- ϕ_i random parameters of group i (hidden components z)
- $\theta = (\beta, \Omega, \sigma^2)$

- Examples of applications
  - Groups are patients: longitudinal data
  - Groups are hierarchical structures

- EM algorithm [Pinheiro and Bates, 2001]
  - E step: $p(\phi_i | y_i; \theta)$ is Gaussian, $Q_{m+1}(\theta) = \mathbb{E}\left[\log p(y, \phi; \theta) \,|y, \widehat{\theta}_m\right]$ is explicit
  - M step: explicit

Example 3: Linear Gaussian state-space model

$$\begin{aligned} y_i &= BZ_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2) \\ Z_{i+1} &= AZ_i + \eta_i, \quad \eta_i \sim \mathcal{N}(0, \Omega) \end{aligned}$$

- y_i: output, ith observation
- Z_i hidden Markov process
- $\theta = (A, B, \Omega, \sigma^2)$

- Examples of applications
  - Ion channel modeling
  - Speech Recognition

- EM algorithm [Cappe, Moulines, Ryden, 2005]
  - E step: Forward-Backward Kalman filter to estimate $\mathbb{E}(Z_i | y_{0:n}, \widehat{\theta}^{(m)})$
  - M step: explicit

Limits of the deterministic EM algorithm

- Implementation
  - Initialization of EM
  - Slow convergence
- More complex models
  - Non Gaussian hierarchical models
  - Hidden Markov Models
  - Partially observed stochastic differential equations (SDE)
- E step is not explicit
- Stochastic versions of the EM algorithm coupled to MCMC/SMC are needed

Stochastic differential equations

- Extension of state-space model
  - Continuous time, continuous space

$$dX_t = f(X_t, \theta)dt + a(X_t, \sigma)dB_t$$

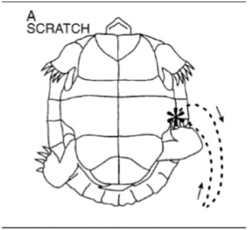
  - X_t stochastic process, B_t Brownian motion
  - f drift function, a diffusion coefficient

- Examples of hidden components
  - X_t multi-dimensional and only the first coordinate observed
  - Random parameters: $dX_{it} = f(X_{it}, \phi_i)dt + a(X_{it}, \sigma)dB_{it}$, $\phi_i \sim \mathcal{N}(\mu, \Omega)$

- Examples of applications
  - Neuron dynamics
  - Population pharmacokinetics

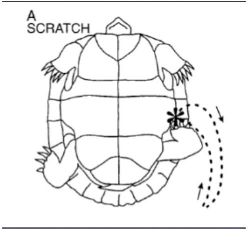
Neuronal experiment

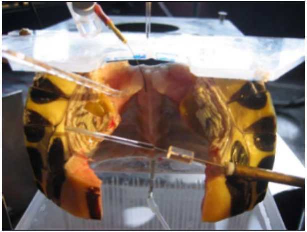
Turtle scratch reflex



Neuronal experiment

Turtle scratch reflex

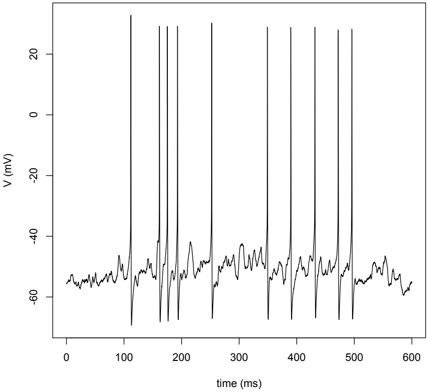




[Berg et al 2007, 2008]

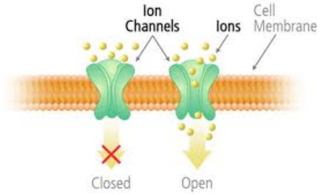
Analyzed data

- Measurements of the membrane potential: difference in voltage between interior and exterior of the neuron
  - $\Delta = 0.1$ ms, $n = 6000$
  - $t_i = i\Delta, \quad i = 1, \dots, n$
  - $V_i = V_{t_i}$



Stochastic neuron models

- [Morris, Lecar, 1981], [Ditlevsen, Greenwood, 2011]
  - V_t: Membrane potential
  - Ion channels: C_a, K, L
    - Conductances g_{C_a}, g_K, g_L
    - Equil. potent. V_{C_a}, V_K, V_L
  - U_t: Proportion of opened potassium ion channels



$$\begin{aligned} dV_t &= (-g_{C_a} m_\infty(V_t)(V_t - V_{C_a}) - g_K U_t(V_t - V_K) - g_L(V_t - V_L) + I)dt + \gamma d\tilde{B}_t \\ dU_t &= \left(\alpha_\phi(V_t)(1 - U_t) - \beta_\phi(V_t)U_t\right) dt + a_\sigma(V_t, U_t)dB_t \end{aligned}$$

$\tilde{B}_t, B_t$: 2 independent Brownian motions

$\alpha(v) = \frac{1}{2}\phi \cosh\left(\frac{v-V_3}{2V_4}\right)\left(1 + \tanh\left(\frac{v-V_3}{V_4}\right)\right)$, $\beta(v) = \frac{1}{2}\phi \cosh\left(\frac{v-V_3}{2V_4}\right)\left(1 - \tanh\left(\frac{v-V_3}{V_4}\right)\right)$

- Objectives: estimation of parameters $\theta = (g_{C_a}, g_K, g_L, V_{C_a}, V_K, I, \phi, \sigma, \gamma)$
  - Discrete observations $V_{0:n} = (V_0, \dots, V_n)$
  - Hidden coordinate (U_t)

Parametric estimation for multi dimensional SDE

- All coordinates observed
  - Euler contrast [Genon-Catalot, Jacod, 1993], [Kessler, 1997]
  - Martingale estimating functions [Bibby and Sorensen, 1995]
  - Stochastic EM with MCMC[Donnet, Samson, 2008]
  - Bayesian MCMC approach (Euler or Girsanov) [Donnet, Foulley, Samson, 2010], [Jensen, Papaspiliopoulos, Ditlevsen, 2014]

- Only one coordinate observed
  - EM algorithm + Kalman filter [Favetto, Samson, 2010], [Cuenod et al., 2011]
  - EM algorithm + linearization of the SDE [Huys, Paninski, 2009]
  - SAEM algorithm + particle filter [Ditlevsen, Samson, 2014]
  - Bayesian MCMC approach (Exact or Girsanov) [Fearnhead et al, 2008], [Ditlevsen, Jensen, Papaspiliopoulos, Samson]

Multi dimensional SDE

$$\begin{aligned} dV_t &= f(V_t, U_t)dt + \gamma(V_t, U_t)d\tilde{B}_t \\ dU_t &= b(V_t, U_t)dt + a_\sigma(V_t, U_t)dB_t \end{aligned}$$

- Unknown transition density
  - Approximation by Euler scheme

$$\begin{aligned} V_i &= V_{i-1} + \Delta f_\theta(V_{i-1}, U_{i-1}) + \sqrt{\Delta}\,\gamma(V_{i-1}, U_{i-1})\,\tilde{\eta}_i \\ U_i &= U_{i-1} + \Delta b_\theta(V_{i-1}, U_{i-1}) + \sqrt{\Delta}\sigma(V_{i-1}, U_{i-1})\eta_i \end{aligned}$$

    $(\tilde{\eta}_i)$, (η_i) ind. centered Gaussian variables
  - Explicit transition density of the approximate model

Only partial observations available

- V_i observed, U_i hidden
- (V_i, U_i) Markovian but not (U_i)
- Our model is a degenerate Hidden Markov Model
  - set $X_i = (V_i, U_i)$, with Markov kernel $R(X_{i-1}, dX_i) = p_\Delta(dV_i, dU_i | V_{i-1}, U_{i-1})$
  - $Y_i = X_i^{(1)}$ with transition kernel $F(X, dY) = \mathbb{1}_{\{Y = X^{(1)}\}}$

  $\rightarrow$ the kernel F is zero almost everywhere
- Our model is a general dynamic model
  - Hidden process: $U_0 \sim \mu(dU_0)$, $U_i | (U_{0:i-1}, V_{0:i-1}) \sim K(dU_i | U_{0:i-1}, V_{0:i-1})$
  - Observed process: $V_i | (U_{0:i}, V_{0:i-1}) \sim G(dV_i | U_{0:i}, V_{0:i-1})$

Expectation-Maximization (EM) algorithm

- Likelihood non explicit, even with the Euler scheme

$$p_\Delta(V_{0:n};\theta) = \int \prod_{i=1}^{n} p_\Delta(V_i, U_i | V_{i-1}, U_{i-1};\theta) dU_{0:n}$$

- Incomplete data model
  - Observed data $(V_{0:n})$
  - Complete data $(V_{0:n}, U_{0:n})$

- EM algorithm [Dempster, Laird, Rubin, 1977], iteration m
  - *E Step*: computation of $Q_{m+1}(\theta) = \mathbb{E}_\Delta \left[\log p_\Delta(V_{0:n}, U_{0:n};\theta) \,|\, V_{0:n}, \widehat{\theta}_m \right]$
  - *M Step*: update $\widehat{\theta}_{m+1} = \arg\max_\theta Q_{m+1}(\theta)$

- Convergence results [Wu, 1983]

Stochastic Approximation algorithm

[Delyon, Lavielle, Moulines, 1999]

- SAEM algorithm
  - *E Step*
    - *S Step*: simulation of $U_{0:n}^{(m)}$ under $p_\Delta(U_{0:n}|V_{0:n}; \widehat{\theta}_m)$
    - *SA Step*: stochastic approximation of Q_{m+1}

$$Q_{m+1}(\theta) = (1-\alpha_m)Q_m(\theta) + \alpha_m \log p_\Delta(V_{0:n}, U_{0:n}^{(m)}; \theta)$$

  - *M Step*: $\widehat{\theta}_{m+1} = \arg\max_\theta Q_{m+1}(\theta)$

- Convergence results
  - Complete likelihood in exponential family with regular statistics
  - Independence of samples $(U_{0:n}^{(m)})$
  - $\sum_m \alpha_m = \infty$, $\sum_m \alpha_m^2 < \infty$

$$\widehat{\theta}_m \xrightarrow[m\to\infty]{a.s.} \text{\textit{(local) max of likelihood}}\ p_\Delta(V_{0:n}; \theta)$$

Simulation step

- Simulation under $p_\Delta(U_{0:n}|V_{0:n};\theta)$ not explicit
- Filtering problem
  - Computation of $\pi_{\Delta,n,\theta}f = \mathbb{E}_\Delta\left(f(U_n)|V_{0:n};\theta\right)$
  - Kalman filter when SDE is linear and Gaussian
- Why the HMM point of view is ill-posed here
  - $X_i = (V_i, U_i)$ the hidden chain and $Y_i = X_i^{(1)}$
  - filtering problem $\pi_{\Delta,n,\theta}f$: ratio of two quantities where $F(X_0; Y_1)$ appears
  - But $F(X_{n-1}; Y_n) = \mathbb{1}_{\{Y_n = X_{n-1}^{(1)}\}}$!
- Particle filter/Sequential Monte Carlo SMC
  - [Del Moral et al, 2001; Doucet et al, 2001; Chopin, 2004; ...]
  - Principle
    - Simulation of K particles $U_{0:n}^k$ and computation of weights $w_n(U_{0:n}^k)$
    - Empirical measure $\Psi_{n;\theta}^K = \sum_{k=1}^K w_n(U_{0:n}^k)\mathbf{1}_{U_{0:n}^k}$

Particle filter for partially observed SDE

- References
  - Autonomous SDE [Del Moral, Jacod, 2001]

$$\begin{aligned} dV_t &= f(V_t, U_t)dt + \gamma(V_t, U_t)d\tilde{B}_t \\ dU_t &= b(U_t)dt + \sigma(U_t)dB_t \end{aligned}$$

  - Gradient drift [Fearnhead et al, 2008]

$$\begin{aligned} dV_t &= f(V_t, U_t)dt + d\tilde{B}_t \\ dU_t &= b(V_t, U_t)dt + dB_t \end{aligned}$$

- Iterative particle filter for non autonomous hidden process [Ditlevsen, Samson, 2014]
  - *At time* $i = 0$: $\forall\, k = 1, \ldots, K$
    1. simulation of U_0^k
    2. calculation of weights $w_0(U_0^k) = p_\Delta(V_0 | U_0^k; \theta)$
  - *At time* $i = 1, \ldots, n$: $\forall\, k = 1, \ldots, K$
    1. resampling particles U'^k_{i-1}
    2. simulation of $U_i^k \sim q(\cdot | V_i, U'^k_{i-1}; \theta)$
    3. calculation of weights

$$w_i\left(U_{0:i}^k\right) = \frac{p_\Delta(V_{0:i}, U_{0:i}^k; \theta)}{p_\Delta(V_{0:i-1}, U'^k_{0:i-1}; \theta) q(U_i^k | V_i, U'^k_{i-1}; \theta)}$$

Particle filter convergence

Exact conditional expectation

$$\pi_{n,\theta} f = \mathbb{E}\left(f(U_{0:n}) | V_{0:n}; \theta\right)$$

Euler approximate conditional expectation

$$\pi_{\Delta,n,\theta} f = \mathbb{E}_{\Delta}\left(f(U_{0:n}) | V_{0:n}; \theta\right)$$

Particle filter approximation

$$\Psi^K_{n,\theta} f = \sum_{k=1}^{K} f(U^k_{0:n}) w_{n,\theta}(U^k_{0:n})$$

Theorem [Ditlevsen, Samson, 2014]

If $1/\Delta \geq K^{1/3}$, for any $\varepsilon > 0$, for any bounded Borel function f

$$\mathbb{P}\left(\left|\Psi^K_{n,\theta} f - \pi_{\Delta,n,\theta} f\right| \geq \varepsilon\right) \leq C_1 \exp\left(-K \frac{\varepsilon^2}{C_2 ||f||^2}\right)$$

$$\mathbb{P}\left(\left|\Psi^K_{n,\theta} f - \pi_{n,\theta} f\right| \geq \varepsilon\right) \leq C_3 \exp\left(-K \frac{\varepsilon^2}{C_4 ||f||^2}\right)$$

SAEM-SMC algorithm

- *E Step*
  - *S Step*: simulation of $U_{0:n}^{(m)}$ with particle filter approximating $p_\Delta(U_{0:n}|V_{0:n};\widehat{\theta}_m)$ with $K(m)$ particles
  - *SA Step*: stochastic approximation of Q_{m+1}

$$Q_{m+1}(\theta) = (1-\alpha_m)Q_m(\theta) + \alpha_m \log p_\Delta(V_{0:n}, U_{0:n}^{(m)}; \theta)$$

- *M Step*: $\widehat{\theta}_{m+1} = \arg\max_\theta Q_{m+1}(\theta)$

Convergence results of SAEM-SMC

SAEM assumptions

1. Incomplete data model in exponential family
2. Regularity of complete likelihood
3. $\sum_m \alpha_m = \infty$, $\sum_m \alpha_m^2 < \infty$.

SMC assumptions

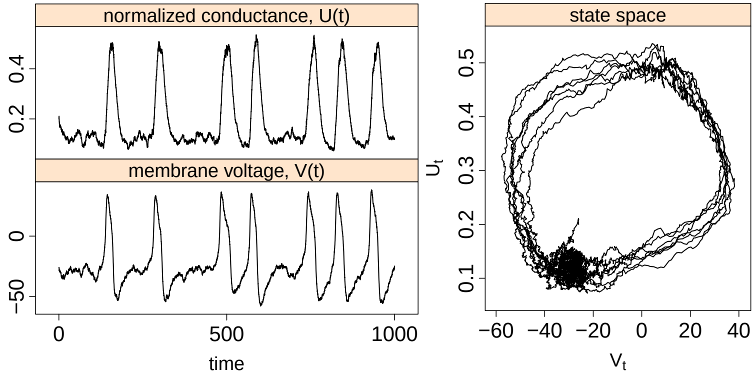
1. Number of particles $K(m) = \log(m^{1+\delta})$
2. Sufficient statistics bounded uniformly in θ.
3. Functions $p_\Delta(V_i|U_i, V_{i-1}, U_{i-1})$ bounded uniformly in θ.

Theorem [Ditlevsen, Samson, 2014]

$$\widehat{\theta}_m \xrightarrow[m\to\infty]{a.s.} \textit{(local) max of likelihood } p_\Delta(V_{0:n};\theta)$$

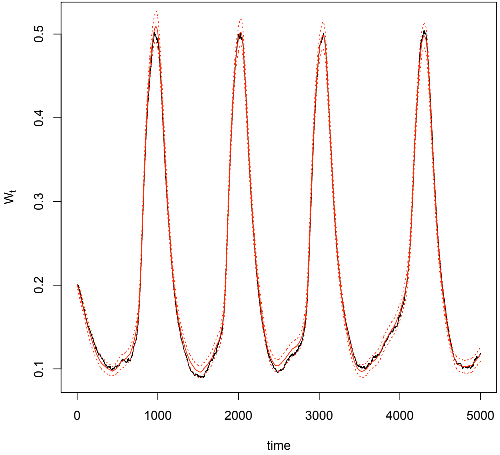
Simulations

Euler approximation of Morris-Lecar model

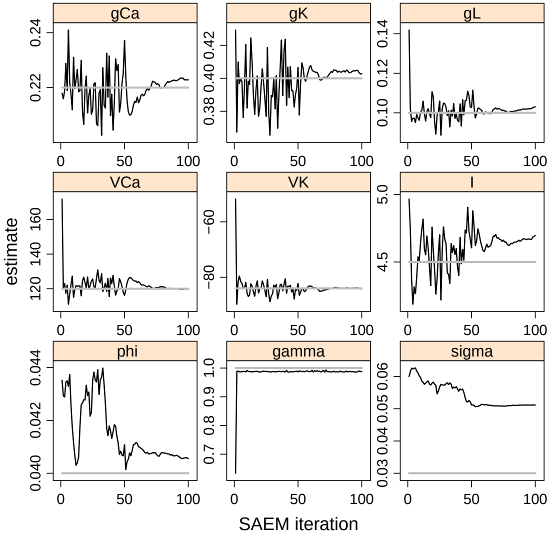


Filtering from $V_{0:n}$

Parameter θ fixed at the true value



SAEM-SMC behavior

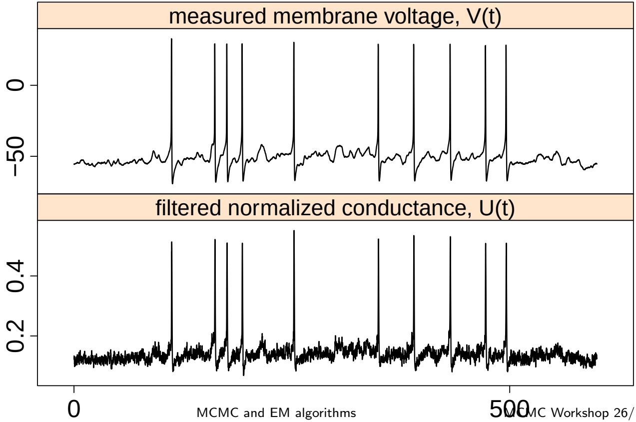


Real data results

- Estimated parameters

Parameter	g_L	g_{Ca}	g_K	σ	γ	V_K	ϕ	V_{Ca}	I
Estimate	1.29	11.56	18.63	0.09	2.69	-65.58	2.68	106.37	-65.16
SE	–	0.03	0.08	–	0.02	0.43	0.12	0.44	0.51

- Filtered path



Stochastic differential equations with random effects

$$\begin{aligned} y_{ij} &= X_{t_{ij}} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \\ dX_{it} &= f(X_{it}, \phi_i)dt + a(X_{it}, \gamma)dB_{it} \\ \phi_i &\sim \mathcal{N}(\mu, \Omega) \end{aligned}$$

- Hidden components
  - X_{it}, ϕ_i
- E-step of the EM algorithm [Donnet, Samson, 2008], [Donnet, Samson, 2014]
  - MCMC algorithm for simulating (X_{it}, ϕ_i) given $y_{i,0:n}$
    - Filtering of X_{it} given ϕ_i and $y_{i,0:n}$
    - Metropolis-Hastings algorithm simulating ϕ_i given $X_{it}$and $y_{i,0:n}$
  - Use PMCMC [Andrieu et al, 2010]

Conclusion/Perspectives

- Complex models with hidden components
- Stochastic EM algorithm coupled to
  - MCMC
  - Particle filter/SMC
  - PMCMC

- Limits/questions/perspectives
  - Influence of the number of particles
  - Computation time
  - Analysis of non "stationary" data