Notes on adaptive filtering of continuous-time linear stochastic systems

László Gerencsér

IMTA SZTAKI, Budapest

Nonlinear PDEs, Stochastic Control and Filtering: New Methods and Applications - In honour of the 75th Birthday of Professor Nicolai Krylov

ICMS, Edinburgh, UK, 29 May - 2 June 2017
Outline

- Continuous-time linear stochastic systems (LSS)
- Recursive maximum likelihood (RML) estimation
- Convergence in $L_q$
- Stability of hybrid linear stochastic systems
- Off-line vs. on-line estimation
- Mixed, stochastic-deterministic algorithms
Continuous-time LSS, I.

Consider a linear stochastic state-space system

\[
x_t = A(\theta^*)x_t \, dt + B(\theta^*) \, du_t
\]

\[
dy_t = C(\theta^*)x_t \, dt + dv_t,
\]

with \(-\infty < t < +\infty\), where \(\theta \in D \subset \mathbb{R}^p\), where \(D\) is an open set.

The true parameter \(\theta^*\) is assumed to be unknown.

The problem: find an approximate value of the filtered state-process.
Continuous-time LSS, II.

Rewrite the above system in innovation form:

\[ x_t = A(\theta^*)x_t \, dt + K(\theta^*)dw_t \quad (3) \]

\[ dy_t = C(\theta^*)x_t \, dt + dw_t, \quad (4) \]

The auxiliary problem: estimate the true parameter \( \theta^* \) assuming that the above systems matrices are known for any \( \theta \).
Continuous-time LSS, III.

**Condition**

For all $\theta \in D$ the system is stable and inverse stable, i.e. the matrices

$$A(\theta) \quad \text{and} \quad A(\theta) - K(\theta)C(\theta)$$

are stable. Moreover, $A, K, C$ are $C^3$.

- $dw_t$ is the innovation of $dy_t$
- $K(\theta^*)$ is the Kalman–gain
The likelihood function

The problem: estimate $\theta^*$ on the basis of $\{y(s) : 0 \leq s \leq t\}$.

Fix $\theta$, and invert the system to get $d\bar{\varepsilon}_t(\theta)$:

\[
d\bar{x}_t(\theta) = A(\theta)\bar{x}_t(\theta)dt + K(\theta)(dy_t - C(\theta)\bar{x}_t(\theta))dt
\]

\[
d\bar{\varepsilon}_t(\theta) = dy_t - C(\theta)\bar{x}_t(\theta)dt.
\]

Then the (conditional) negative log-likelihood is, with $l = \dim y$,

\[
V_t(\theta) = \frac{1}{2} \int_0^t \frac{|d\bar{\varepsilon}_s(\theta)|^2 - l}{ds} ds.
\]
Continuous-time RML, I.

Stochastic regression models:


Continuous-time RML. II

RML for dynamic models:

- Gerencsér, Gyöngy, and Michaletzky,
  Proc. of the 9th IFAC World Congress, 1984.


- Levanony, Shwartz, and Zeitouni,
The likelihood equation

Differentiation w.r.t. $\theta$ gives the likelihood equation

$$\frac{\partial}{\partial \theta} V_t(\theta) = V_{\theta t}(\theta) = \int_0^t \dot{\epsilon}_{\theta s}(\theta) d\overline{\epsilon}_s(\theta) = 0,$$

(7)

where $\dot{\epsilon}_{\theta t}$ denotes the mixed derivatives $\frac{\partial}{\partial t} \overline{\epsilon}_{\theta t}$.

The solution $\bar{\theta}_t$ is the ML estimator.
The asymptotic likelihood, I.

Write

\[ d\tilde{\epsilon}_t(\theta) = (C(\theta^*)\bar{x}_t(\theta^*)dt + dw_t) - C(\theta)\bar{x}_t(\theta)dt. \]

Define

\[ \bar{z}_s(\theta) = C(\theta)\bar{x}_s(\theta). \]

Then the asymptotic negative log-likelihood function is

\[ \mathcal{W}(\theta) = \lim_{t \to \infty} \frac{1}{2} \mathbb{E} (\bar{z}_t(\theta) - \bar{z}_t(\theta^*))^T (\bar{z}_t(\theta) - \bar{z}_t(\theta^*)). \]  \hspace{1cm} (8)
The asymptotic likelihood, II.

We have for $\theta = \theta^*$

$$\left. \frac{\partial}{\partial \theta} W(\theta) \right|_{\theta = \theta^*} = 0. \quad (9)$$

For the Hessian of $W(\theta)$ at $\theta = \theta^*$ we have

$$\left. \frac{\partial^2}{\partial \theta_i \partial \theta_j} W(\theta) \right|_{\theta = \theta^*} = \lim_{s \to \infty} \mathbb{E} \dot{\varepsilon}_{\theta_i,s}^T(\theta) \dot{\varepsilon}_{\theta_j,s}(\theta) \right|_{\theta = \theta^*}.$$
Local identifiability

Set

\[ R^* = \left. \frac{\partial^2 W(\theta)}{\partial \theta^2} \right|_{\theta=\theta^*} = \left. W_{\theta\theta}(\theta) \right|_{\theta=\theta^*}. \]

Condition

The model class is locally identifiable:

\[ R^* > 0, \]

i.e. \( R^* \) is positive definite.
Approximations via time varying dynamics

Approximate $\tilde{x}_t(\theta_t)$ and $\tilde{e}(t, \theta_t)$ using the TV LSS

\begin{align*}
  dx_t &= A(\theta_t)x_t dt + K(\theta_t)d\varepsilon_t \tag{10} \\
  d\varepsilon_t &= dy_t - C(\theta_t)x_t dt, \tag{11}
\end{align*}

where $\theta_t$ will be an on-line estimator to be defined below.

Approximation of $\tilde{x}_{\theta,t}(\theta_t)$ is defined similarly.
Recursive maximum-likelihood (RML)

Let $R_t$ be an auxiliary $p \times p$ process approximating $R^*$.

The RML dynamics:

\begin{align}
\mathrm{d} \theta_t &= -\frac{1}{t} R_t^{-1} \dot{\epsilon}_\theta t \mathrm{d} \epsilon_t \\
\dot{R}_t &= \frac{1}{t} \left( \dot{\epsilon}_\theta t \dot{\epsilon}_\theta^T - R_t \right).
\end{align}
Resetting $\theta_t$

Let $D_{0\theta} \subset D_\theta$ be a compact domain s.t. $\theta^* \in \text{int } D_{0\theta}$ and $\theta_0 \in \text{int } D_{0\theta}$.

A key technical device: enforce $\theta_t \in D_{0\theta}$.

**Resetting:** whenever $\theta_t$ hits $\partial D_{0\theta}$ reset it as

$$\theta_t := \theta_0.$$

Similarly for $R_t$. 
Controlling the rate of $\theta_t$

Ensure slow rate of change of $\theta_t$ by enforcing

$$\frac{1}{t} \left( |x_t|^2 + \|x_{\theta,t}\|^2 \right) \leq \delta. \quad (14)$$

Reset the states $x_t, x_{\theta,t}$ to 0 when the l.h.s equals the threshold $\delta$. Get

$$0 < \tau_i < \tau_{i+1},$$

with $\tau_i = \infty$ allowed. For any finite $\tau_i$ we have $\theta_{\tau_i} = \theta_0$.

$\theta_t$ is slowly varying in a stochastic sense in $[\tau_i, \tau_{i+1})$. 
RML with resetting

Let $N_t$ denote the counting process

$$N_t = \#\{i : \tau_i \leq t\}.$$

The RML method with resetting. A stochastic Newton method:

$$d\theta_t = -\frac{1}{T_0 + t} R_t^{-1} \dot{\theta}_t d\varepsilon_t + (\theta_0 - \theta_{t-})dN_t \quad (15)$$

$$dR_t = \frac{1}{T_0 + t} \left( \dot{\varepsilon}_t \dot{\varepsilon}_t^T - R_t \right) dt + (R_0 - R_{t-}) dN_t. \quad (16)$$

Note: computability is OK.
Convergence

**Theorem**

The SDE with jumps defining the RML method with resetting has a strong solution in $[0, \infty)$, we have $N_t < \infty$ a.s., and for any fixed $q > 1$

$$\sup_{s \leq t \leq qs} |\theta_t - \theta^*| = O_M(s^{-1/2})$$

in a limited sense.
All moment estimates in limited sense

Interpretation: for any $m \geq 1$ we have

$$
E^{1/m} \sup_{s \leq t \leq qs} |\theta_t - \theta^*|^m = O(s^{-1/2})
$$

if the threshold for the rate of $\theta_t$, called $\delta_0$ is sufficiently small.
OUTLINE OF PROOF
The score function

For any fixed $x_t = x$, and any $\theta$ define the instantaneous computable score:

$$H(\theta, x)dt + G(x, \theta)dw_t =: -\dot{\epsilon}\theta_s(\theta)d\bar{\epsilon}_s(\theta).$$

Then the asymptotic ML problem is: find $\theta = \theta^* \in D$ such that

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T (H(\theta, x^*_t(\theta)) dt + G(\theta, x^*_t(\theta)) dw_t) = 0. \quad (17)$$
The associated ODE

Assume $R^* = I$ is known. Define

\[ h(\theta) = -W_\theta(\theta). \]

Then the associated ODE is defined as

\[ \dot{y}_t = \frac{1}{T_0 + t} \, h(y_t). \] (18)
The ODE approximation

Following LG, SIAM J. Control and Opt. (SICON), 1992:

Fix $q > 1$, and $\sigma > 0$. Consider the solution of the ODE

$$\bar{y}_t \quad \text{with} \quad \bar{y}_\sigma = \theta_\sigma, \quad \sigma \leq t \leq q\sigma.$$

Use integral form, and define the error between stochastic and mean field:

$$d\rho_u = (H(\theta_u, x_u)du + G(\theta_u, x_u)dw_u) - h(\theta_u)du.$$
Stochastic averaging

To get a good upper bound for the tracking error

$$|\theta_t - \bar{y}_t|$$

we need to bound the integral effect of the error terms $\rho_u$:

$$r^*_\sigma = \sup_{\sigma \leq v \leq q\sigma \land \tau(\sigma)} \left| \int_{\sigma}^{\tau(\sigma)} \frac{1}{T_0 + u} d\rho_u \right|,$$

where $\tau(\sigma)$ is the first resetting point after time $\sigma$. 
Bounding the tracking error

Lemma

We have

\[ r_\sigma^* = O_M(\sigma^{-1/2}) \]

in a limited sense.

Heuristics: take a simplified problem, fix \( \theta \) and fix the weights to 1, and set:

\[
\tilde{r}_\sigma^* = \sup_{\sigma \leq v \leq q\sigma \wedge \tau(\sigma)} \left| \int_{\sigma}^{v} (H(\theta, \tilde{x}_u(\theta))du + G(\theta, x_u(\theta))dw)u - h(\theta)du \right|
\]
The Poisson equation

The classic tool for additive functionals of Markov processes. Consider

\[ \mathcal{L}(\theta) \nu(\theta, x) = H(\theta, x) - h(\theta), \]  

(19)

where \( \mathcal{L}(\theta) \) is the infinitesimal generator of \( x_t(\theta) \). Thus the l.h.s. is:

\[ \frac{\partial}{\partial x} \nu_k(\theta, x) A(\theta)x + \frac{1}{2} \text{Tr}[\frac{\partial^2}{\partial x^2} \nu_k(\theta, x) B(\theta)B^T(\theta)]. \]

Then we get with some martingale \( m_u \)

\[ \bar{r}_\sigma^* \leq \sup_{\sigma \leq \nu \leq q\sigma \wedge \tau(\sigma)} \left| \int_{\sigma}^{\nu} d\nu_u \right| + \sup_{\sigma \leq \nu \leq q\sigma \wedge \tau(\sigma)} \left| \int_{\sigma}^{\nu} d\nu_u \right| + \sup_{\sigma \leq \nu \leq q\sigma \wedge \tau(\sigma)} \left| \int_{\sigma}^{\nu} d\nu_u \right| \]
Bounding $x_t$ under state-resetting, I.

To bound $\nu$ we need to bound $x_t$. Consider a simpler, time-invariant dynamics with an arbitrary $N_t$:

$$dx_t = A(\theta)x_t\,dt + B(\theta)\,dw_t + (x_0 - x_{t-})\,dN_t. \quad (20)$$

Taking into account the asymmetry in jumps, we get:

**Proposition**

For all $q \geq 1$

$$\sup_t \mathbb{E}[|x_t|^q] < \infty.$$
Bounding $x_t$ under state-resetting, II.

Extension: replace $\theta$ by a stochastically slowly varying $\theta_t$ with resetting:

**Proposition**

*Let $q \geq 0$, and assume that $E|x_0|^q < \infty$. Then for sufficiently small $\delta_0 > 0$*

$$\sup_{t \geq 0} E|x_t|^q < \infty.$$ 

A local ODE principle

Letting the initial point $\sigma$ vary we get the following key result:

**Lemma**

We have for any $q > 1$

$$\sup_{s \leq \sigma \leq t < q\sigma \wedge \tau(\sigma)} |\theta_t - \bar{y}_t| = O_M(s^{-1/2}) \text{ in a limited sense.}$$


The proof of the main theorem is completed as in SICON.
OFF-LINE vs. ON-LINE
Offline vs. on-line, I.

Idea: solve $V_{\theta t}(\theta_t) = 0$ using a stochastic implicit function theorem.


Unsettled challenge: (un)computability of composites like $\varepsilon_t(\theta_t)$. 
Let $\bar{\theta}_t$ denote the off-line estimator.

Following GL, Gyöngy and Michaletzky, 1984, we get:

$$\bar{\theta}_t - \theta_t = O_M \left( \frac{\log t}{t} \right)$$

in a limited sense.
Plan of the proof

We show that

\[ V_{\theta t}(\bar{T}_t) = 0 \quad \text{and} \quad V_{\theta t}(\theta_t) = O_M(\log t) \quad \text{i.l.s.} \quad (21) \]

Then the theorem is obtained by Taylor series expansion.

To prove the second part of (21) we show that

\[ dV_{\theta t}(\theta_t) = \frac{1}{t} b_t dt + \frac{1}{t^{1/2}} s_t dw_t \quad \text{i.l.s.} \]

For this we need the Itô-Wentzel formula.
Slow variation of $\theta_t$ relaxed

Condition (Slow variation)

The process $\theta_t \in D_{0\theta}$ satisfies

$$d\theta_t = \beta_t dt + \sigma_t dw_t + (\theta_0 - \theta_{t-}) dN_t,$$

with

$$|\beta_t| + ||\sigma_t||^2 \leq \delta \quad \text{for all } t.$$
The hybrid stochastic Desoer theorem

Consider

\[ \dot{\Phi}_t = F(\theta_t)\Phi_t, \quad \Phi_0 = I. \]

where \( F(\theta) \) is stable for all \( \theta \in D \).

**Theorem**

Assume that \( \theta_t \) is slowly varying with arbitrary \( N_t \). Then for any \( 0 \leq s \leq t \)

\[ \| \Phi_t \Phi_s^{-1} \| \leq C_s(\omega)e^{-\alpha(t-s)} \]

with some \( \alpha > 0 \), where \( C_s = O_M(1) \) in a limited sense.

DISCUSSION
Discussion

Linear stochastic systems not in innovation form:

Find the Kalman-gain matrix via a matrix Riccati diff. eqn.

We get a mixed, stochastic-deterministic method.

Convergence settled along the lines of GL, SIICON 1992.

Conclusion: adaptive filtering of LSS-s is feasible via RML with resetting
HAPPY BIRTHDAY AND
MORE HAPPINESS FOR THE YEARS TO COME!