Parameter estimation for a bidimensional partially observed Ornstein-Uhlenbeck process with biological application

Benjamin Favetto, Adeline Samson

Parameter estimation for a bidimensional partially observed Ornstein-Uhlenbeck process with biological application

Running headline : Estimation for partially observed Ornstein-Uhlenbeck process

Benjamin Favetto* and Adeline Samson*

* MAP5, University Paris Descartes, France

Abstract

We consider a bidimensional Ornstein-Uhlenbeck process to describe the tissue microvascularisation in anti-cancer therapy. Data are discrete, partial and noisy observations of this stochastic differential equation (SDE). Our aim is the estimation of the SDE parameters. We use the main advantage of a one-dimensional observation to obtain an easy way to compute the exact likelihood using the Kalman filter recursion. We also propose a recursive computation of the gradient and hessian of the log-likelihood based on Kalman filtering, which allows to implement an easy numerical maximisation of the likelihood and the exact maximum likelihood estimator (MLE). Furthermore, we establish the link between the observations and an ARMA process and we deduce the asymptotic properties of the MLE. We show that this ARMA property can be generalised to a higher dimensional underlying Ornstein-Uhlenbeck diffusion. We compare this estimator with the one obtained by the well-known EM algorithm on simulated data.

Key Words: ARMA process, EM algorithm, Hidden Markov Model, Kalman filter, Maximum likelihood estimation, Ornstein-Uhlenbeck process, Partial observations

1 Introduction

Stochastic continuous-time models are a useful tool to describe biological or physiological systems based on continuous evolution (see *e.g.* Ditlevsen and De Gaetano (2005), Ditlevsen *et al.* (2005), Picchini *et al.* (2006)). The biological context of this work is the modeling of tissue microvascularisation in anti-cancer therapy. This microcirculation is usually modeled by a bidimensional deterministic differential system which describes the circulation of a contrast agent between two compartments (see Brochot *et al.* (2006) and appendix A). However, this deterministic model is unable to capture the random fluctuations observed along time. In this paper, we consider a stochastic version of this system to take into account random variations around the deterministic solution by adding a Brownian motion on each compartment. This leads to a bidimensional stochastic differential equation (SDE) defined as:

$$\left\{ \begin{array}{rcl} dP(t) & = & (\alpha a(t) - (\lambda + \beta)P(t) + kI(t))dt + \sigma_1 dW_1(t) \\ dI(t) & = & (\lambda P(t) - kI(t))dt + \sigma_2 dW_2(t) \end{array} \right. \quad (1)$$

where $P(t)$ and $I(t)$ represent contrast agent concentrations in each compartment, $a(t)$ is an input function assumed to be known, α, β, λ and k are unknown positive parameters, W_1 and W_2 are two independent Brownian motions on $\mathbb{R}$, and σ_1, σ_2 are the constant diffusion terms. We assume that $(P(0), I(0))$ is a random variable independent of (W_1, W_2). In our biological context, only the sum $S(t) = P(t) + I(t)$ can be measured. So (1) is changed into:

$$\begin{cases} dS(t) &= (\alpha a(t) - \beta S(t) + \beta I(t))\, dt + \sigma_1 dW_1(t) + \sigma_2 dW_2(t) \\ dI(t) &= (\lambda S(t) - k I(t))\, dt + \sigma_2 dW_2(t) \end{cases} \tag{2}$$

Noisy and discrete measurements $(y_i, i = 0, \ldots, n)$ of $S(t)$ are performed at times $t_0 = 0 < t_1 < \ldots < t_n = T$. The observation model is thus:

$$y_i = S(t_i) + \sigma \varepsilon_i,\ \varepsilon_i \sim \mathcal{N}(0, 1)$$

where $(\varepsilon_i)_{i=0,\ldots,n}$ are assumed to be independent and σ is the unknown standard deviation of the Gaussian noise. To evaluate the effect of the treatment on a patient, it is of importance to have a proper estimation of all unknown parameters from this data set. The aim of this paper is to investigate this problem both theoretically and numerically on simulated data.

Parametric inference for discretely observed general SDEs has been widely investigated. Genon-Catalot and Jacod (1993) and Kessler (1997) propose estimators based on minimization of suitable contrasts and study the asymptotic distribution of these estimators when the sampling interval tends to zero as the number of observations tends to infinity. For fixed sampling interval, Bibby and Sørensen (1995) propose martingale estimating functions. In a biological context, Ditlevsen *et al.* (2005) propose an estimation method based on simulation. Picchini *et al.* (2008) propose estimators based on the Hermite expansion of the transition densities. When combining the case of discrete, partial and noisy observations, parameter estimation is a more delicate statistical problem. In this context, it is classical to estimate the unobserved signal (filtering) (see *e.g.* Cappé *et al.* (2005)). However, our aim is the estimation of SDE parameters. In this paper, we use the main advantage of a one-dimensional observation y and the Gaussian framework of all distributions to obtain an easy way to compute the exact likelihood. For this, we solve and discretize the SDE (2). Then we use the Kalman filter recursion to compute the exact likelihood. We also obtain a recursive computation of the exact gradient and hessian of the log-likelihood

based on Kalman filtering, which allows us to implement an easy numerical maximisation of the likelihood using a gradient method and to compute the exact maximum likelihood estimator. The exact observed Fisher information matrix is also directly obtained. We develop a second approach based on the EM algorithm. This method has been first proposed by Shumway and Stoffer (1982) and Segal and Weinstein (1989).
In Section 2, we study the SDE. We detail in Section 3 the computation of the exact likelihood, the score and hessian functions. We present the EM method in Section 4. In Section 5, we establish the link between the observations and an ARMA process. This allows to deduce the asymptotic properties of the maximum likelihood estimator. Section 6 contains numerical results based on simulated data. This allows to compare the two estimation methods. Appendix A describes briefly the biological background. Appendix B, C, D and E contain some proofs and auxiliary results. In particular, the ARMA property can be generalised to a higher dimensional underlying Ornstein-Uhlenbeck diffusion.

2 Study of the stochastic differential equation

Introducing $U(t) = (S(t), I(t))'$, (2) can be written in a matrix form:

$$\left\{ \begin{array}{rcl} dU(t) & = & (F(t) + G\,U(t))dt + \Sigma dW(t), U(0) = U_0 \\ y_i & = & J\,U(t_i) + \sigma\varepsilon_i \end{array} \right.$$

where $J = (1 \;\; 0)$ and

$$F(t) = \begin{pmatrix} \alpha a(t) \\ 0 \end{pmatrix}, G = \begin{pmatrix} -\beta & \beta \\ \lambda & -k \end{pmatrix}, dW(t) = \begin{pmatrix} dW_1(t) \\ dW_2(t) \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_1 & \sigma_2 \\ 0 & \sigma_2 \end{pmatrix}$$

The process $(U(t))$ is a bidimensional Ornstein-Uhlenbeck diffusion, which can be explicitly solved. From the biological context (see Appendix A), the parameters satisfy $\beta, k, \lambda > 0$ and $\lambda < k$. This implies that G is diagonalizable with two distinct negative eigenvalues. Setting $d = (\beta - k)^2 + 4\beta\lambda \geq 0$, the eigenvalues of G are:

$$\mu_1 = \frac{-(\beta + k) - \sqrt{d}}{2} \quad \text{and } \mu_2 = \frac{-(\beta + k) + \sqrt{d}}{2}$$

The diagonal matrix D of eigenvalues and the matrix P of eigenvectors are:

$$D = \begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix},\ P = \begin{pmatrix} 1 & 1 \\ \frac{\beta-k-\sqrt{d}}{2\beta} & \frac{\beta-k+\sqrt{d}}{2\beta} \end{pmatrix} \text{ with } D = P^{-1}GP.$$

***Proposition* 1** *Let $X(t) = P^{-1}U(t)$ and $\Gamma = (\Gamma^{kj})_{1\leq k,j\leq 2} = P^{-1}\Sigma$. Then, for $t, h \geq 0$, we have:*

$$X(t+h) = e^{Dh}X(t) + B(t,t+h) + Z(t,t+h) \tag{3}$$

where

$$B(t,t+h) = e^{D(t+h)} \int_t^{t+h} e^{-Ds} P^{-1} F(s) ds \tag{4}$$

$$Z(t,t+h) = e^{D(t+h)} \int_t^{t+h} e^{-Ds} \Gamma dW_s. \tag{5}$$

Therefore, the conditional distribution of $X(t+h)$ given $X(s), s \leq t$ is

$$\mathcal{N}_2\left(e^{Dh}X(t) + B\left(t,t+h\right), R\left(t,t+h\right)\right)$$

where

$$R(t,t+h) = \left(\frac{e^{(\mu_k+\mu_{k'})h}-1}{\mu_k+\mu_{k'}} \left(\Gamma\Gamma'\right)^{kk'}\right)_{1\leq k,k'\leq 2} \tag{6}$$

If $a(t) \equiv 0$, $(X(t))$ has a Gaussian stationary distribution with a null mean and a covariance matrix equal to

$$V = \left(\frac{1}{-(\mu_k+\mu_{k'})} \left(\Gamma\Gamma'\right)^{kk'}\right)_{1\leq k,k'\leq 2}$$

If $\lim_{t\to\infty} a(t) = c \geq 0$, $(X(t))$ converges in distribution towards a Gaussian distribution with a mean equal to $(\frac{c}{|\mu_1|}, \frac{c}{|\mu_2|})'$ and the covariance matrix V.

Proof. See Appendix B.

3 Parameter estimation by maximum likelihood

Our aim is to estimate the unknown parameters $\alpha, \beta, \lambda, k, \sigma_1, \sigma_2$ and σ from observations $y_{0:n} = (y_0, \ldots, y_n)$. As the law of $((X(t)), \varepsilon_i, i = 0, \ldots, n)$ is Gaus-

sian, the likelihood of $y_{0:n}$ can be explicitly evaluated. However, the direct maximization of this likelihood requires the inversion of a matrix of dimension $2(n+1) \times 2(n+1)$ (the covariance matrix of $(X(t_i))$). This inversion can be numerically instable. In this section, we present an alternative method for the computation of the exact likelihood based on Kalman filtering, which does not require any matrix inversion. This is due to the fact that data are one-dimensional. Moreover, it is worth stressing that we need not come back to the initial process $(U(t))$ for computing the likelihood. Indeed, as $(U(t))$ is not observed, we can use either $(U(t))$ or any other transformation of $(U(t))$ even involving unknown parameters. As $(X(t))$ is simpler, we consider the following transformed model:

$$\left\{ \begin{array}{rcl} dX(t) & = & (DX(t) + P^{-1}F(t))dt + \Gamma dW_t,\ X(0) = P^{-1}U_0 = X_0 \\ y_i & = & JPX(t_i) + \sigma\varepsilon_i \end{array} \right. \tag{7}$$

Given the particular form of our vector $J = (1\ \ 0)$ and the fact that the eigenvectors can be chosen up to a proportionnality constant, we have

$$H = JP = (1\ \ 1).$$

It is especially interesting for further computations of the gradient and hessian of the likelihood that H does not depend of any unknown parameter. From model (7) and (3)-(6), we deduce the following discrete-time evolution system where $X_i = X(t_i)$:

$$\left\{ \begin{array}{rcl} X_i & = & A_i X_{i-1} + B_i + \eta_i,\ \ \eta_i \sim \mathcal{N}(0, R_i) \\ y_i & = & HX_i + \sigma\varepsilon_i \end{array} \right. \tag{8}$$

where $A_i = \exp(D(t_i - t_{i-1}))$, $B_i = B(t_{i-1}, t_i)$, $R_i = R(t_{i-1}, t_i)$.

3.1 Computation of the exact likelihood

This discrete model is a hidden Markov model (HMM) (Cappé *et al.*, 2005): (X_i) is a hidden Markov chain on $\mathbb{R}^2$ and, conditionally on (X_i), observations (y_i) are independent. Genon-Catalot and Laredo (2006) study the maximum likelihood estimator for general HMM. They specialize the exact likelihood in the case where the unobserved Markov chain is a Gaussian one-dimensional AR(1) process. We generalize this computation to the case where the unobserved Markov chain is a bidimensional AR(1) process. Let ϕ denote the vector of

unknown parameters and $y_{0:i} = (y_0, \ldots, y_i)$ the vector of observations until time t_i. By recursive conditioning, it is sufficient to compute the distribution of y_i given $y_{0:i-1}$:

$$L(\phi, y_{0:n}) = p(y_0; \phi) \prod_{i=1}^{n} p(y_i | y_{0:i-1}; \phi).$$

But the conditional law of y_i given $y_{0:i-1}$ can be evaluated by

$$p(y_i | y_{0:i-1}; \phi) = \int p(y_i | X_i; \phi) p(X_i | y_{0:i-1}; \phi) dX_i$$

Then, as the innovation noise η_i of the hidden Markov chain, and the observation noise ε_i are Gaussian variables, by elementary computations on Gaussian laws, we are able to get the law of y_i given $y_{0:i-1}$ if we know the mean and covariance of the Gaussian conditional law of X_i given $y_{0:i-1}$. This conditional distribution can be exactly computed using Kalman recursions as described below.

3.1.1 Kalman filter

To ease the reading, the parameter ϕ is omitted. The Kalman filter is an iterative procedure which computes recursively the following conditional distributions

$$\begin{array}{rcll} \mathcal{L}(X_i | y_{0:i-1}) & = & \mathcal{N}_2(\hat{X}_i^-, P_i^-) & (prediction) \\ \mathcal{L}(X_i | y_{0:i}) & = & \mathcal{N}_2(\hat{X}_i, P_i) & (filter) \end{array}$$

where

$$\begin{array}{rcl} \hat{X}_i^- = \mathbb{E}(X_i | y_{0:i-1}) & \text{and} & P_i^- = \mathbb{E}((X_i - \hat{X}_i^-)(X_i - \hat{X}_i^-)') \\ \hat{X}_i = \mathbb{E}(X_i | y_{0:i}) & \text{and} & P_i = \mathbb{E}((X_i - \hat{X}_i)(X_i - \hat{X}_i)') \end{array}$$

Let us assume that the law of X_0 is Gaussian. Initial values for the algorithm are:

$$X_0 \sim \mathcal{N}(\hat{X}_0^-, P_0^-)$$

Next we have the recursive formulae obtained using (8):

$$\begin{array}{rl} \hat{X}_i^- = A_i \hat{X}_{i-1} + B_i, \quad P_i^- = A_i P_{i-1} A_i' + R_i, & i \geq 1 \\ \hat{X}_i = \hat{X}_i^- + K_i (y_i - H \hat{X}_i^-), \quad P_i = (I - K_i H) P_i^-, & i \geq 0 \end{array} \tag{9}$$

where $K_i = P_i^- H' (H P_i^- H' + \sigma^2)^{-1}$ (see *e.g.* Cappé *et al.* (2005)).

3.1.2 Computation of the exact likelihood of the observations

The conditional distribution of y_i given $y_{0:i-1}$ is Gaussian and one-dimensional. Let $m_i(\phi) = \mathbb{E}_\phi(y_i|y_{0:i-1})$ and $V_i(\phi) = Var_\phi(y_i|y_{0:i-1})$ denote its mean and variance which are given using (8) by

$$m_i(\phi) = H\hat{X}_i^-, \quad V_i(\phi) = HP_i^-H' + \sigma^2$$

where $\hat{X}_i^-$ and P_i^- depend on ϕ. The exact likelihood of $y_{0:n}$ is thus equal to

$$L(\phi, y_{0:n}) = \prod_{i=0}^{n} \frac{1}{\sqrt{2\pi V_i(\phi)}} \exp\left(-\frac{1}{2}\frac{(y_i - m_i(\phi))^2}{V_i(\phi)}\right). \tag{10}$$

Relations (9) imply that there exist two functions F_ϕ and G_ϕ such that

$$m_i(\phi) = F_\phi(m_{i-1}(\phi)),\ V_i(\phi) = G_\phi(V_{i-1}(\phi)) \tag{11}$$

These iterative relations are used to compute the derivatives of the log-likelihood.

3.2 Computation of the maximum likelihood estimator

To compute the exact maximum likelihood estimator (MLE), we use a conjugate gradient method, which relies on the explicit knowledge of the gradient and hessian of the log-likelihood. Both can be exactly computed using formula (10) and observing that the derivatives of $m_i(\phi)$ and $V_i(\phi)$ can be explicitly and recursively computed by derivating formulae (11). These computations are detailed below.

3.2.1 New parametrization

In order to simplify the derivatives of (11), from now on, we assume that observation times are equally spaced and set

$$\Delta = t_i - t_{i-1},\ \forall i = 1, \ldots, n.$$

The parameter α is assumed to be known and for the sake of simplicity we set $a(t) \equiv 0$. Hence we have $A_i = A$, $R_i = R$, $B_i = 0$. Moreover, instead of $\phi = (\beta, \lambda, k, \sigma_1, \sigma_2, \sigma^2)$, we propose a new parametrization fitted with the discretization. We consider $\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$ where $\theta_i = e^{\mu_i \Delta}$, $i = 1, 2$ and

θ_3, θ_4 and θ_5 are explicit functions of $\mu_1, \mu_2, \sigma_1, \sigma_2$ and Δ such that

$$A = A(\theta) = \begin{pmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{pmatrix} \quad \text{and} \quad R = R(\theta) = \begin{pmatrix} \theta_3 & \theta_5 \\ \theta_5 & \theta_4 \end{pmatrix}$$

We set $\vartheta = (\theta, \sigma^2)$. Our aim is to maximize the likelihood $L(\vartheta, y_{0:n})$ with respect to ϑ. Given an estimation $\hat{\vartheta}$, $\hat{\phi}$ can be obtained by solving numerically the equation $f(\hat{\phi}) = \hat{\vartheta}$ where f is the mapping $\phi \to f(\phi) = \vartheta$. Note that, later on, we will see that only five out of the six parameters can be consistently estimated.

3.2.2 Computation of the exact gradient and hessian of the log-likelihood

Let $W_i(\vartheta) = y_i - m_i(\vartheta)$ and $l_{0:i}(\vartheta) = \log L(\vartheta, y_{0:i})$. Using (10), it comes:

$$l_{0:i}(\vartheta) = l_{0:i-1}(\vartheta) - \frac{1}{2}\log(2\pi V_i(\vartheta)) - \frac{1}{2}\frac{W_i(\vartheta)^2}{V_i(\vartheta)}. \tag{12}$$

Thus for $i = 1, \ldots, n$, $q = 1, \ldots, 6$:

$$\frac{\partial l_{0:i}}{\partial \vartheta_q}(\vartheta) = \frac{\partial l_{0:i-1}}{\partial \vartheta_q}(\vartheta) - \frac{1}{2}\frac{1}{V_i(\vartheta)}\frac{\partial V_i}{\partial \vartheta_q}(\vartheta) - \frac{W_i(\vartheta)}{V_i(\vartheta)}\frac{\partial W_i(\vartheta)}{\partial \vartheta_q} + \frac{1}{2}\frac{W_i(\vartheta)^2}{V_i(\vartheta)^2}\frac{\partial V_i(\vartheta)}{\partial \vartheta_q}, \tag{13}$$

where

$$\frac{\partial V_i(\vartheta)}{\partial \vartheta_q} = H\frac{\partial P_i^-(\vartheta)}{\partial \vartheta_q}H', 1 \le q \le 5, \ \frac{\partial V_i(\vartheta)}{\partial \sigma^2} = H\frac{\partial P_i^-(\vartheta)}{\partial \sigma^2}H' + 1$$
$$\frac{\partial W_i(\vartheta)}{\partial \vartheta_q} = -H\frac{\partial \hat{X}_i^-(\vartheta)}{\partial \vartheta_q}, 1 \le q \le 6$$

Furthermore, the derivatives of $\hat{X}_i^-(\vartheta)$ and $P_i^-(\vartheta)$ can be obtained using Kalman recursions (see appendix C). With a more cumbersome computation, second order derivatives of $l_{0:n}(\vartheta)$ can be analogously deduced from (13). For $i =$

$1, \ldots, n$, $q, r = 1, \ldots, 6$,

$$\begin{aligned}\frac{\partial^2 l_{0:i}}{\partial\vartheta_r\partial\vartheta_q}(\vartheta) &= \frac{\partial^2 l_{0:i-1}}{\partial\vartheta_r\partial\vartheta_q}(\vartheta) - \frac{1}{2}\frac{1}{V_i(\vartheta)}\frac{\partial^2 V_i}{\partial\vartheta_r\partial\vartheta_q}(\vartheta) + \frac{1}{2}\frac{1}{V_i^2(\vartheta)}\frac{\partial V_i}{\partial\vartheta_r}(\vartheta)\frac{\partial V_i}{\partial\vartheta_q}(\vartheta) \\ &\quad -\frac{1}{2}\left(2\frac{W_i(\vartheta)}{V_i(\vartheta)}\frac{\partial^2 W_i(\vartheta)}{\partial\vartheta_r\partial\vartheta_q} - \frac{W_i(\vartheta)^2}{V_i(\vartheta)^2}\frac{\partial^2 V_i(\vartheta)}{\partial\vartheta_r\partial\vartheta_q}\right) \\ &\quad -\left(\frac{\partial W_i(\vartheta)}{\partial\vartheta_r}\frac{\partial W_j(\vartheta)}{\partial\vartheta_q}\frac{1}{V_i(\vartheta)} - \frac{W_i(\vartheta)}{V_i(\vartheta)^2}\frac{\partial W_i(\vartheta)}{\partial\vartheta_r}\frac{\partial V_i(\vartheta)}{\partial\vartheta_q}\right) \\ &\quad +\left(\frac{\partial W_i(\vartheta)}{\partial\vartheta_q}\frac{\partial V_i(\vartheta)}{\partial\vartheta_r}\frac{W_i(\vartheta)}{V_i(\vartheta)^2} - \frac{W_i(\vartheta)^2}{V_i(\vartheta)^4}\frac{\partial V_i(\vartheta)}{\partial\vartheta_r}V_i(\vartheta)\frac{\partial V_i(\vartheta)}{\partial\vartheta_q}\right)\end{aligned} \tag{14}$$

where (see appendix C for details)

$$\frac{\partial^2 V_i(\vartheta)}{\partial\vartheta_r\partial\vartheta_q} = H\frac{\partial^2 P_i^-(\vartheta)}{\partial\vartheta_r\partial\vartheta_q}H' \text{ and } \frac{\partial^2 W_i(\vartheta)}{\partial\vartheta_r\partial\vartheta_q} = -H\frac{\partial^2 \hat{X}_i^-(\vartheta)}{\partial\vartheta_r\partial\vartheta_q}.$$

Hence, we obtain an explicit expression of $\left(-\frac{\partial^2 l_{0:n}}{\partial\vartheta_r\partial\vartheta_q}(\vartheta)\right)_{1\leq q,r\leq 6}$.

3.2.3 Maximisation of the exact likelihood

To compute the maximum likelihood estimator, the conjugate gradient algorithm is applied to minimize $\tilde{l}_{0:n}(\vartheta) = -l_{0:n}(\vartheta)$ (see Stoer and Bulirsch (1993)). Let $\nabla\tilde{l}$ denote the gradient of $\tilde{l}_{0:n}$ and $Hess\,\tilde{l}$ its hessian evaluated by (13)-(14). Starting with an arbitrary initial vector ϑ_0, we set as descent direction $u_0 = \vartheta_0$. At iteration k, given ϑ_k and u_k, the parameter and descent direction are updated by

$$\vartheta_{k+1} = \vartheta_k - \frac{\langle u_k, \nabla\tilde{l}(\vartheta_k)\rangle}{\langle u_k, Hess\,\tilde{l}(\vartheta_k)u_k\rangle}u_k,\ u_{k+1} = -\nabla\tilde{l}(\vartheta_{k+1}) + \frac{\|\nabla\tilde{l}(\vartheta_{k+1})\|}{\|\nabla\tilde{l}(\vartheta_k)\|}u_k.$$

Classical stopping conditions are used. The sequence $(\vartheta_k)_k$ converges towards the maximum of the likelihood $l_{0:n}(\vartheta)$.

4 Parameter estimation by Expectation Maximization algorithm

An alternative method to estimate $\vartheta = (\theta, \sigma^2)$ is the Expectation Maximization (EM) algorithm, proposed by Dempster *et al.* (1977), see also Shumway and Stoffer (1982) and Segal and Weinstein (1989). The EM algorithm is a classical

approach to estimate parameters of models with non-observed or incomplete data. In our case, the non-observed data are the (X_i)'s, the complete data are the (y_i, X_i)'s. The principle is to maximize

$$\vartheta \to Q(\vartheta|\vartheta_*) = \mathbb{E}(\log p(y_{0:n}, X_{0:n}; \vartheta)|y_{0:n}; \vartheta_*)$$

with $X_{0:n} = (X_0, \ldots, X_n)$. This is often easier than the maximization of the observed data log-likelihood since the log-likelihood of the complete data is generally simpler. Moreover, according to Wu (1983), as our model is an exponential family, the EM estimate sequence $(\vartheta_k)_k$ converges towards a (local) maximum of the data likelihood. The EM algorithm uses two steps: the Expectation step (E-step) and the Maximization step (M-step). Starting with an initial value (ϑ_0), the k-th iteration is

- E-step: evaluation of $Q_k(\vartheta) = Q(\vartheta \,|\, \vartheta_k)$
- M-step: update of ϑ_k by $\vartheta_{k+1} = \arg\max Q_k(\vartheta)$.

In our model, function Q has an explicit expression. Recall that we have assumed that observations times are equally spaced and that $a(t) \equiv 0$. For simplicity, we also set for the initial variable $X_0 = 0$ ($\hat{X}_0^- = 0$ and $P_0^- = 0$). The complete data log-likelihood is thus equal to:

$$\log p(y_{0:n}, X_{0:n}; \vartheta) = -\frac{n+1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=0}^{n}(y_i - HX_i)^2$$
$$-\frac{1}{2}\sum_{i=1}^{n}\log(2\pi|R(\theta)|) - \frac{1}{2}\sum_{i=1}^{n}(X_i - A(\theta)X_{i-1})'R(\theta)^{-1}(X_i - A(\theta)X_{i-1}).$$

Function Q consists in taking the conditional expectation given $y_{0:n}$ under $\mathbb{P}_{\vartheta_*}$. This conditional distribution is the so-called smoothing distribution at ϑ_*. In our model, it is Gaussian and characterized by $M_{i|0:n}(\vartheta_*) = \mathbb{E}_{\vartheta_*}(X_i|y_{0:n})$ and

$$\Sigma_{i|0:n}(\vartheta_*) = Var_{\vartheta_*}(X_i|y_{0:n}),\ \Sigma_{i-1,i|0:n}(\vartheta_*) = Cov_{\vartheta_*}(X_{i-1}, X_i|y_{0:n})$$

These can be obtained through a forward-backward algorithm (see Appendix D). Thus function Q is equal to:

$$Q(\vartheta|\vartheta_*) = -\frac{n+1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=0}^{n}\left[(y_i - HM_{i|0:n}(\vartheta_*))^2 + H\Sigma_{i|0:n}(\vartheta_*)H'\right]$$
$$-\frac{n}{2}\log(2\pi|R(\theta)|) - \frac{1}{2}Tr\left\{R(\theta)^{-1}\left[C(\vartheta_*) - M(\vartheta_*)A'(\vartheta) - A(\vartheta)M'(\vartheta_*) + A(\vartheta)S(\vartheta_*)A'(\vartheta)\right]\right\}$$

where

$$\begin{aligned}
M(\vartheta_*) &= \sum_{i=1}^{n}\left(\Sigma_{i-1,i|0:n}(\vartheta_*) + M_{i|0:n}(\vartheta_*)M'_{i-1|0:n}(\vartheta_*)\right)\\
S(\vartheta_*) &= \sum_{i=1}^{n}\left(\Sigma_{i-1|0:n}(\vartheta_*) + M_{i-1|0:n}(\vartheta_*)M'_{i-1|0:n}(\vartheta_*)\right)\\
C(\vartheta_*) &= \sum_{i=1}^{n}\left(\Sigma_{i|0:n}(\vartheta_*) + M_{i|0:n}(\vartheta_*)M'_{i|0:n}(\vartheta_*)\right).
\end{aligned}$$

The matrices A, R and σ^2 are updated as

$$\begin{aligned}
A(\theta_k) &= diag(M(\vartheta_{k-1})S^{-1}(\vartheta_{k-1}))\\
R(\theta_k) &= \frac{1}{n}(C(\vartheta_{k-1}) - M(\vartheta_{k-1})S^{-1}(\vartheta_{k-1})M'(\vartheta_{k-1}))\\
\sigma_k^2 &= \frac{1}{n+1}\sum_{i=0}^{n}\left[(y_i - HM_{i|0:n}(\vartheta_{k-1}))^2 + H\Sigma_{i|0:n}(\vartheta_{k-1})H'\right]
\end{aligned}$$

5 Properties of the exact maximum likelihood estimator in the stationary case

Recall that we have assumed that $a(t) \equiv 0$. Moreover in this paragraph we assume that the initial variable X_0 has the stationary distribution $\mathcal{N}_2(0, V)$ given in Proposition 1. This implies that the joint process (X_i, y_i) is strictly stationary. Let $(y_i)_{i\in\mathbb{Z}}$ be its extension to a process indexed by $\mathbb{Z}$.

5.1 Link with an ARMA model

We generalize the result of Genon-Catalot *et al.* (2003) to the bidimensional case and also to the multidimensional case (see Appendix E).

***Proposition* 2** *The process $(y_i)_{i\in\mathbb{Z}}$ is centered Gaussian and ARMA(2,2).*

Proof. Evidently (y_i) is centered Gaussian. We easily check that

$$y_i - (\theta_1 + \theta_2)y_{i-1} + \theta_1\theta_2 y_{i-2} = \xi_i \tag{15}$$

where ξ_i is defined by

$$\xi_i = HA(\theta)\eta_{i-1} + H\eta_i + \sigma\varepsilon_i - (\theta_1 + \theta_2)H\eta_{i-1} - (\theta_1 + \theta_2)\sigma\varepsilon_{i-1} + \theta_1\theta_2\sigma\varepsilon_{i-2}$$

As the $(\eta_i)_i$ and $(\varepsilon_i)_i$ are mutually independent, we get that:

$$Cov(\xi_i, \xi_{i+k}) = 0, \ \forall k \geq 3.$$

This implies that (ξ_i) is MA(2). Hence the result. □

***Proposition* 3** *The spectral density $f(u, \vartheta)$ of (y_i) has the explicit form:*

$$f(u,\vartheta) = \sigma^2 + \frac{H(ARA' + (1 + (\theta_1 + \theta_2)^2)R)H' + 2\cos(u)(HA - (\theta_1 + \theta_2)H)RH'}{1 + (\theta_1 + \theta_2)^2 + \theta_1^2\theta_2^2 - 2(\theta_1 + \theta_2)(1 + \theta_1\theta_2)\cos(u) + 2\cos(2u)\theta_1\theta_2}$$

with $A = A(\theta), R = R(\theta)$.

Proof. Let $\gamma(k) = Cov(\xi_i, \xi_{i+k})$. Elementary computations show that

$$\begin{aligned}
\gamma(0) &= HARA'H' + HRH'(1 + (\theta_1 + \theta_2)^2) + \sigma^2(1 + (\theta_1 + \theta_2)^2 + \theta_1^2\theta_2^2) \\
\gamma(1) &= (HA - (\theta_1 + \theta_2)H)RH' - \sigma^2(\theta_1 + \theta_2)(1 + \theta_1\theta_2) \\
\gamma(2) &= \sigma^2\theta_1\theta_2 \\
\gamma(k) &= 0 \ \forall \ k \geq 3
\end{aligned}$$

The spectral density (with respect to $\frac{du}{2\pi}$) $h(u, \vartheta)$ of (ξ_i) is

$$h(u,\vartheta) = \sum_{n\in\mathbb{Z}} \gamma(n)\exp(-inu) = \gamma(0) + \gamma(1)2\cos(u) + \gamma(2)2\cos(2u).$$

For the AR(2) part, we set: $p(x) = 1 - (\theta_1 + \theta_2)x + \theta_1\theta_2 x^2$ (recall that $\theta_1, \theta_2 < 1$). Then

$$\begin{aligned}
f(u,\vartheta) &= \frac{h(u,\vartheta)}{|p(\exp(-iu))|^2} \\
&= \frac{\gamma(0) + \gamma(1)2\cos(u) + \gamma(2)2\cos(2u)}{1 + (\theta_1 + \theta_2)^2 + \theta_1^2\theta_2^2 - 2(\theta_1 + \theta_2)(1 + \theta_1\theta_2)\cos(u) + 2\cos(2u)\theta_1\theta_2}
\end{aligned}$$

See Brockwell and Davis (1991) for technical details. □

5.2 Asymptotic properties of maximum likelihood estimator

We now deduce the asymptotic properties of the maximum likelihood estimators when (y_i) is stationary. We can only estimate consistently the parameters which are identifiable on the spectral density. The identifiable parameters are $(\theta_1 + \theta_2, \theta_1\theta_2, \sigma^2, H(ARA'+(1+(\theta_1+\theta_2)^2)R)H', (HA-(\theta_1+\theta_2)H)RH')$. This means that we can only identify five parameters out of the six. Therefore, we can choose to estimate either $(\theta_1, \ldots, \theta_5)$, or σ^2 and four among $\theta_1, \ldots, \theta_5$. Moreover, since we identify only the sum $\theta_1 + \theta_2$ and the product $\theta_1\theta_2$, by convention in our algorithms, we choose $\theta_1 < \theta_2$.
We denote by θ^0 the true value of parameter. We assume that the parameter set Θ is an open subset of $\mathbb{R}^5$. The asymptotic information matrix $I(\theta)$ is

$$\text{for } i,j \in \{1,\ldots,5\} \quad I(\theta)_{i,j} = \int_{-\pi}^{\pi} \frac{\partial}{\partial\theta_i} \log f(u,\theta) \frac{\partial}{\partial\theta_j} \log f(u,\theta) \frac{du}{2\pi}$$

Consider the two assumptions (which can be checked up to some technicities)

A1 $\theta \longmapsto f(.,\theta)$ is one to one

A2 $(u,\theta) \longmapsto f(u,\theta)$ is a $\mathcal{C}^3$-function on a neighborhood of $[-\pi,\pi] \times \Theta$

As (y_i) is a ARMA(2,2) process, its spectral density is positive for every $(u,\theta) \in [-\pi,\pi] \times \Theta$.

***Proposition* 4** (Information matrix) *We have $\mathbb{P}_{\theta^0}$-a.s.*

$$\lim_{n\to\infty} \left(-\frac{1}{n} \frac{\partial^2}{\partial\theta_i \partial\theta_j} l_{0:n}\left(\theta^0\right) \right)_{1\le i,j\le 5} = I(\theta^0) \tag{16}$$

***Proposition* 5** (Consistency and asymptotic normality of the MLE) *Let $\hat{\theta}_n$ be a maximum likelihood estimator of θ^0 based on $y_{0:n}$. Then, $\hat{\theta}_n \to \theta^0$ a.s. as $n \to \infty$. Moreover, if $I(\theta^0)$ is invertible, $\sqrt{n}(\hat{\theta}_n - \theta^0)$ converges in distribution:*

$$\sqrt{n}(\hat{\theta}_n - \theta^0) \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0, I^{-1}(\theta^0))$$

Proof. This result may be found *e.g.* in Brockwell and Davis (1991). □

6 Simulation study

We compare the performances of the two methods of estimation on simulated data. The exact maximum likelihood estimators are computed as described in Section 3, using Kalman-based recursions. The EM estimators are computed as described in Section 4. Data are simulated using equally spaced observation times ($\Delta = 0.2$), $n = 5000$ observations and the following parameter values

$$\theta_1 = 0.3,\ \theta_2 = 0.8,\ \theta_3 = 0.5,\ \theta_4 = 1,\ \theta_5 = 0.1,$$

which corresponds to initial parameters equal to $\beta = 4.86,\ \lambda = 0.88,\ k = 2.27,\ \sigma_1 = 3.31,\ \sigma_2 = 1.92$.

To evaluate the impact of the noise level σ^2 on the estimation, we simulate 20 data sets with $\sigma^2 = 0.2$ and 20 data sets with $\sigma^2 = 1$. The parameter σ^2 is fixed to its true value, and θ is estimated by the exact MLE and EM algorithm. Results are reported in Table 1. Both methods estimate accurately the parameter θ_2 and its corresponding variance parameter θ_4. The noise level σ^2 does not affect the exact MLE while the EM algorithm is more accurate when the observation noise is large.

[Table 1 about here.]

Then we estimate σ^2 and fix θ_5 to its true value (recall that only 5 parameters are identifiable). Results are reported in Table 2. The estimation of θ_2 is satisfactory. However, the estimation of the other four parameters is deteriorated for both methods, especially θ_1.

[Table 2 about here.]

To evaluate the impact of the sampling interval, we simulate 20 data sets with successively $\Delta = 0.1, n = 10000$, $\Delta = 0.2, n = 5000$ and $\Delta = 1, n = 1000$, each design corresponding to $n \times \Delta = 1000$. Result are reported in Table 3. Note that the true values of the parameters are changed according to the value of Δ. The estimation is more precise for smaller Δ.

[Table 3 about here.]

The Matlab code are given at http://www.mi.parisdescartes.fr/~favetto.

7 Conclusion

The Kalman filter is classical in the field of noisy, discretely and partially observed stochastic differential equations. In this paper, we have shown that it can be used for the estimation of the parameters by maximum likelihood. In the particular case of an Ornstein-Uhlenbeck process, this method computes the exact likelihood, its gradient and hessian. We have also shown that the EM algorithm combined with a smoother algorithm can be used for the parameter estimation.

We study some theoretical properties of the model. We show that only five out of the six parameters are identifiable and we deduce the asymptotic properties of the maximum likelihood estimate. We illustrate the two methods on simulated data. The identifiability problem is confirmed on the simulation study: the observed Fisher information matrix computed by the Kalman method is not invertible when we estimate the six parameters. Furthermore, we recommend to estimate σ^2 in a preliminary phase and then to estimate the five other parameters.

The next step of this work is its application to real data in anti-cancer therapy. However, as real data are obtained from medical images such as MRI or CT-scan, the number of data is limited in order not to ray the patient. Furthermore, these images are measured at non-equidistant times. Therefore, the two methods have to be adapted to the case of non-equidistant observation times.

Acknowledgments

The authors thank Valentine Genon-Catalot for her constructive advices and help. The authors thank Professor Charles-Andre Cuenod for her helpful discussions on the biological problem. This project was supported by a grant BQR from University Paris Descartes.

References

Bibby, B. and Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli* **1**, 17–39.

Brochot, C., Bessoud, B., Balvay, D., Cuenod, C., Siauve, N. and Bois, F. (2006). Evaluation of antiangiogenic treatment effects on tumors' microcir-

culation by bayesian physiological pharmacokinetic modeling and magnetic resonance imaging. *Magn Reson Imaging.* **24**, 1059–1067.

Brockwell, P. and Davis, R. (1991). *Time Series: Theory and Methods.* Springer.

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

Dempster, A., Laird, N. and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Jr. R. Stat. Soc. B* **39**, 1–38.

Ditlevsen, S. and De Gaetano, A. (2005). Stochastic vs. deterministic uptake of dodecanedioic acid by isolated rat livers. *Bull. Math. Biol.* **67**, 547–561.

Ditlevsen, S., Yip, K. and Holstein-Rathlou, N. (2005). Parameter estimation in a stochastic model of the tubuloglomerular feedback mechanism in a rat nephron. *Math Biosc* **194**, 49–69.

Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.* **29**, 119–151.

Genon-Catalot, V., Jeantheau, T. and Laredo, C. (2003). Conditional likelihood estimators for hidden Markov models and stochastic volatility models. *Scand. J. Statist.* **30**, 297–316.

Genon-Catalot, V. and Laredo, C. (2006). Leroux's method for general hidden Markov models. *Stochastic Process. Appl.* **116**, 222–243.

Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**, 211–229.

Picchini, U., Ditlevsen, S. and De Gaetano, A. (2006). Modeling the euglycemic hyperinsulinemic clamp by stochastic differential equations. *J. Math. Biol.* **53**, 771–796.

Picchini, U., Ditlevsen, S. and De Gaetano, A. (2008). Maximum likelihood estimation of a time-inhomogeneous stochastic differential model of glucose dynamics. *Mathematical Medicine and Biology* **25**, 141–155.

Segal, M. and Weinstein, E. (1989). A new method for evaluating the log-likelihood gradient, the hessian, and the fisher information matrix for linear dynamic systems. *IEEE Trans. Inf. Theory* **35**, 682–687.

Shumway, R. and Stoffer, D. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Series Anal* **3**, 253–264.

Stoer, J. and Bulirsch, R. (1993). *Introduction to numerical analysis*, vol. 12 of *Texts in Applied Mathematics*. Springer-Verlag, New York, second edn.

Wu, C. (1983). On the convergence property of the EM algorithm. *Ann. Statist.* **11**, 95–103.

The corresponding author for negotiations concerning the manuscript is
Adeline Samson (adeline.samson@parisdescartes.fr)
Laboratoire MAP5
Universite Paris Descartes
45 rue des St Peres
75006 Paris
France

A Physiological model

We focus on the evaluation of anti-angiogenesis treatments in anti-cancer therapy. These treatments take effect on the vascularization of tissue. The *in vivo* evaluation of their efficacy is based on the estimation of the tissue microvascularization parameters. The experiment consists in the injection of a contrast agent to the patient, followed by the recording of a medical images sequence which measures the evolution of the concentration of contrast agent along time. The contrast agent pharmacokinetic is modeled by a bidimensional differential system. The contrast agent pulsates in the plasma and interstitium cells. Let $a(t)$, $P(t)$ and $I(t)$ denote respectively the quantity of contrast agent at time t in the artery, the plasma and the interstitium and $1-h$, V_P and V_I the volume of artery, plasma and interstitium (h is the hematocrit rate). The initial condition at time $t_0 = 0$ is $P(0) = 0, I(0) = 0$. The contrast agent is injected in vein at time t_0, transits in the artery and arrives in plasma, with a tissue perfusion flow equal to F_{tp}. The contrast agent is eliminated from plasma with the perfusion flow F_{tp}, proportionally to the concentration of contrast agent in plasma. The quantity of contrast agent exchanging from plasma through interstitium is equal to K_{trans} times the concentration of contrast agent in plasma, where K_{trans} is the volume transfer constant. Inversely, the quantity of contrast agent exchanging from interstitium through plasma is equal to K_{trans} times the concentration

of contrast agent in interstitium. Lastly, the two-compartment model is:

$$\left\{\begin{array}{lcl} \frac{dP(t)}{dt} & = & \frac{F_{tp}}{1-h}a(t) - \frac{K_{trans}}{V_P}P(t) + \frac{K_{trans}}{V_I}I(t) - \frac{F_{tp}}{V_P}P(t) \\ \frac{dI(t)}{dt} & = & \frac{K_{trans}}{V_P}P(t) - \frac{K_{trans}}{V_I}I(t) \end{array}\right. \tag{17}$$

For statistical accommodations, we use a new parameterization and set:

$$\alpha = \frac{F_{tp}}{1-h}, \beta = \frac{F_{tp}}{V_P}, \lambda = \frac{K_{trans}}{V_P}, k = \frac{K_{trans}}{V_P} + \frac{K_{trans}}{V_I}$$

Model (17) can thus be transformed as follows:

$$\left\{\begin{array}{lcl} \frac{dP(t)}{dt} & = & \alpha a(t) - \lambda P(t) + kI(t) - \beta P(t) \\ \frac{dI(t)}{dt} & = & \lambda P(t) - kI(t) \end{array}\right. \tag{18}$$

B Proof of Proposition 1

The process $X(t) = P^{-1}U(t)$ is solution of:

$$dX(t) = (DX(t) + P^{-1}F(t))dt + P^{-1}\Sigma dW_t,\ X(0) = X_0 = P^{-1}U_0.$$

Applying Ito's formula, we obtain

$$X(t) = e^{Dt}X_0 + e^{Dt}\int_0^t e^{-Ds}P^{-1}F(s)ds + e^{Dt}\int_0^t e^{-Ds}\Gamma dW_s.$$

From this equation, we deduce:

$$X(t+h) = e^{Dh}X(t) + B(t, t+h) + Z(t, t+h)$$

where $B(t, t+h)$ and $Z(t, t+h)$ are given in Proposition 1. Using that W_1, W_2 are independent and that X_0 is independent of (W_1, W_2), we obtain the conditional law of $X(t+h)|(X(s), s \leq t)$.
The stationary distribution can be deduced from equation (3) with $F(t) = 0$. As the two elements of D are negative, we have

$$\begin{array}{rcl} \lim\limits_{t\to+\infty} \mathbb{E}(X(t)) & = & \lim\limits_{t\to+\infty} e^{Dt}\mathbb{E}(X_0) = 0 \\ \lim\limits_{t\to+\infty} Var(X(t)) & = & \lim\limits_{t\to+\infty} R(0,t) = \left(\frac{1}{-(\mu_k + \mu_{k'})}\left(\Gamma\Gamma'\right)^{kk'}\right)_{1\leq k,k'\leq 2} = V. \end{array}$$

If $X_0 \sim \mathcal{N}_2(0, V)$, an elementary computation shows that $(X(t))$ is strictly

stationary. The case $\lim_{t\to\infty} a(t) = c$ is straightforward. □

C Gradient and hessian of the log-likelihood

The gradient and the hessian of the loglikelihood are computed with explicit recursions. We denote $\vartheta_6 = \sigma^2$. The first order derivatives of $A(\vartheta)$ are equal to:

$$\frac{\partial A(\vartheta)}{\partial \vartheta_1} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \frac{\partial A(\vartheta)}{\partial \vartheta_2} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } \frac{\partial A(\vartheta)}{\partial \vartheta_q} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, q = 3, 4, 5, 6.$$

and for $R(\vartheta)$ we get:

$$\frac{\partial R(\vartheta)}{\partial \vartheta_1} = \frac{\partial R(\vartheta)}{\partial \vartheta_2} = \frac{\partial R(\vartheta)}{\partial \vartheta_6} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

$$\frac{\partial R(\vartheta)}{\partial \vartheta_3} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \frac{\partial R(\vartheta)}{\partial \vartheta_4} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } \frac{\partial R(\vartheta)}{\partial \vartheta_5} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The second order derivatives of $A(\vartheta)$ and $R(\vartheta)$ are null. The first order derivatives of $\hat{X}_i^-(\vartheta)$ and $P_i^-(\vartheta)$ with respect to ϑ_q, $q = 1, \ldots, 6$ can be deduced:
$\frac{\partial \hat{X}_i^-(\vartheta)}{\partial \vartheta_q} = \frac{\partial A(\vartheta)}{\partial \vartheta_q}\hat{X}_{i-1}(\vartheta) + A(\vartheta)\frac{\partial \hat{X}_{i-1}(\vartheta)}{\partial \vartheta_q}$
$\frac{\partial P_i^-(\vartheta)}{\partial \vartheta_q} = \frac{\partial A(\vartheta)}{\partial \vartheta_q}P_{i-1}(\vartheta)A(\vartheta)' + A(\vartheta)\frac{\partial P_{i-1}(\vartheta)}{\partial \vartheta_q}A(\vartheta)' + A(\vartheta)P_{i-1}(\vartheta)\frac{\partial A(\vartheta)'}{\partial \vartheta_q} + \frac{\partial R(\vartheta)}{\partial \vartheta_q}$.
Then we get the derivatives of the mean and the covariance of the filter: $\frac{\partial \hat{X}_{i-1}(\vartheta)}{\partial \vartheta_q} =$
$\frac{\partial \hat{X}_{i-1}^-(\vartheta)}{\partial \vartheta_q} + \frac{\partial P_{i-1}^-(\vartheta)}{\partial \vartheta_q}\frac{H'W_{i-1}(\vartheta)}{V_{i-1}(\vartheta)} + \frac{P_{i-1}^-(\vartheta)H'}{V_{i-1}(\vartheta)}\left(\frac{\partial W_{i-1}(\vartheta)}{\partial \vartheta_q} - \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_q}\frac{W_{i-1}(\vartheta)}{V_{i-1}(\vartheta)}\right)\frac{\partial P_{i-1}(\vartheta)}{\partial \vartheta_q} =$
$\left(I - \frac{P_{i-1}^-H'H}{V_{i-1}(\vartheta)}\right)\frac{\partial P_{i-1}^-(\vartheta)}{\partial \vartheta_q} - \left(\frac{\partial P_{i-1}^-(\vartheta)}{\partial \vartheta_q}\frac{H'}{V_{i-1}(\vartheta)} - \frac{P_{i-1}^-(\vartheta)H'}{V_{i-1}(\vartheta)^2}\frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_q}\right)HP_{i-1}^-(\vartheta)$
The computation of second order derivatives can be deduced. Because second order derivatives of A and R are null, we have for $q, r = 1, \ldots, 6$, :

$$\frac{\partial^2 \hat{X}_i^-(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} = A(\vartheta)\frac{\partial^2 \hat{X}_{i-1}(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} + \frac{\partial A(\vartheta)}{\partial \vartheta_r}\frac{\partial \hat{X}_{i-1}(\vartheta)}{\partial \vartheta_q} + \frac{\partial A(\vartheta)}{\partial \vartheta_q}\frac{\partial \hat{X}_{i-1}(\vartheta)}{\partial \vartheta_r}$$
$$\frac{\partial^2 P_i^-(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} = \left(\frac{\partial A(\vartheta)}{\partial \vartheta_r}\frac{\partial P_{i-1}(\vartheta)}{\partial \vartheta_q} + \frac{\partial A(\vartheta)}{\partial \vartheta_q}\frac{\partial P_{i-1}(\vartheta)}{\partial \vartheta_r}\right)A(\vartheta)' + \frac{\partial A(\vartheta)}{\partial \vartheta_r}P_{i-1}(\vartheta)\frac{\partial A(\vartheta)'}{\partial \vartheta_q}$$
$$+A(\vartheta)\left(\frac{\partial^2 P_{i-1}(\vartheta)}{\partial \vartheta_q \partial \vartheta_r}A(\vartheta)' + \frac{\partial P_{i-1}(\vartheta)}{\partial \vartheta_r}\frac{\partial A(\vartheta)'}{\partial \vartheta_q} + \frac{\partial P_{i-1}(\vartheta)}{\partial \vartheta_q}\frac{\partial A(\vartheta)'}{\partial \vartheta_r}\right) + \frac{\partial A(\vartheta)}{\partial \vartheta_q}P_{i-1}(\vartheta)\frac{\partial A(\vartheta)'}{\partial \vartheta_r}$$

The second order derivatives of mean and covariance of Kalman filter are:

$$\frac{\partial^2 \hat{X}_{i-1}(\vartheta)}{\partial\vartheta_q\partial\vartheta_r} = \frac{\partial^2 \hat{X}^-_{i-1}(\vartheta)}{\partial\vartheta_q\partial\vartheta_r} + \frac{\partial^2 P^-_{i-1}(\vartheta)}{\partial\vartheta_q\partial\vartheta_r}\frac{H'W_{i-1}(\vartheta)}{V_{i-1}(\vartheta)} + P^-_{i-1}(\vartheta)\frac{H'}{V_{i-1}(\vartheta)}\left(\frac{\partial^2 W_{i-1}(\vartheta)}{\partial\vartheta_q\partial\vartheta_r}\right.$$
$$-\frac{\partial V_{i-1}(\vartheta)}{\partial\vartheta_q}\frac{1}{V_{i-1}(\vartheta)}\frac{\partial W_{i-1}(\vartheta)}{\partial\vartheta_r} - \frac{\partial^2 V_{i-1}(\vartheta)}{\partial\vartheta_q\partial\vartheta_r}\frac{W_{i-1}(\vartheta)}{V_{i-1}(\vartheta)} + 2\frac{\partial V_{i-1}(\vartheta)}{\partial\vartheta_r}\frac{\partial V_{i-1}(\vartheta)}{\partial\vartheta_q}\frac{W_{i-1}(\vartheta)}{V^2_{i-1}(\vartheta)}$$
$$\left.-\frac{\partial V_{i-1}(\vartheta)}{\partial\vartheta_r}\frac{1}{V_{i-1}(\vartheta)}\frac{\partial W_{i-1}(\vartheta)}{\partial\vartheta_q}\right) + \frac{\partial P^-_{i-1}(\vartheta)}{\partial\vartheta_r}\frac{H'}{V_{i-1}(\vartheta)}\left(\frac{\partial W_{i-1}(\vartheta)}{\partial\vartheta_q} - \frac{\partial V_{i-1}(\vartheta)}{\partial\vartheta_q}\frac{W_{i-1}(\vartheta)}{V_{i-1}(\vartheta)}\right)$$
$$+\frac{\partial P^-_{i-1}(\vartheta)}{\partial\vartheta_q}\frac{H'}{V_{i-1}(\vartheta)}\left(\frac{\partial W_{i-1}(\vartheta)}{\partial\vartheta_r} - \frac{\partial V_{i-1}(\vartheta)}{\partial\vartheta_r}\frac{W_{i-1}(\vartheta)}{V_{i-1}(\vartheta)}\right)$$

$$\frac{\partial^2 P_{i-1}(\vartheta)}{\partial\vartheta_q\partial\vartheta_r} = \left(I - \frac{P^-_{i-1}(\vartheta)H'H}{V_{i-1}(\vartheta)}\right)\frac{\partial^2 P^-_{i-1}(\vartheta)}{\partial\vartheta_q\partial\vartheta_r} - \left[\frac{\partial^2 P^-_{i-1}(\vartheta)}{\partial\vartheta_q\partial\vartheta_r} - \frac{P^-_{i-1}(\vartheta)}{V_{i-1}(\vartheta)}\frac{\partial^2 V_{i-1}(\vartheta)}{\partial\vartheta_q\partial\vartheta_r}\right.$$
$$\left.-\left(\frac{\partial P^-_{i-1}(\vartheta)}{\partial\vartheta_r}\frac{\partial V_{i-1}(\vartheta)}{\partial\vartheta_q} + \frac{\partial P^-_{i-1}(\vartheta)}{\partial\vartheta_q}\frac{\partial V_{i-1}(\vartheta)}{\partial\vartheta_r}\right)\frac{1}{V_{i-1}(\vartheta)} + \frac{2P^-_{i-1}(\vartheta)}{V^2_{i-1}(\vartheta)}\frac{\partial V_{i-1}(\vartheta)}{\partial\vartheta_q}\frac{\partial V_{i-1}(\vartheta)}{\partial\vartheta_r}\right]\frac{H'HP^-_{i-1}(\vartheta)}{V_{i-1}(\vartheta)}$$
$$-\left(\frac{\partial P^-_{i-1}(\vartheta)}{\partial\vartheta_r}H' - \frac{P^-_{i-1}(\vartheta)H'}{V_{i-1}(\vartheta)}\frac{\partial V_{i-1}(\vartheta)}{\partial\vartheta_r}\right)\frac{H}{V_{i-1}(\vartheta)}\frac{\partial P^-_{i-1}(\vartheta)}{\partial\vartheta_q}$$
$$-\left(\frac{\partial P^-_{i-1}(\vartheta)}{\partial\vartheta_q}H' - P^-_{i-1}(\vartheta)\frac{H'}{V_{i-1}(\vartheta)}\frac{\partial V_{i-1}(\vartheta)}{\partial\vartheta_q}\right)\frac{H}{V_{i-1}(\vartheta)}\frac{\partial P^-_{i-1}(\vartheta)}{\partial\vartheta_r}$$

D Smoother algorithm

The Kalman smoother is calculated recursively with a forward-backward algorithm (see *e.g.* Cappé *et al.* (2005)). The forward algorithm is the classical Kalman filter which computes $M_{i|0:i-1} = \hat{X}^-_i = \mathbb{E}(X_i|y_{0:i-1})$, $\Sigma_{i|0:i-1} = P^-_i = Var(X_i|y_{0:i-1})$, $M_{i|0:i} = \hat{X}_i = \mathbb{E}(X_i|y_{0:i})$ and $\Sigma_{i|0:i} = P_i = Var(X_i|y_{0:i})$. Then, in order to calculate $M_{i|0:n} = \mathbb{E}(X_i|y_{0:n})$, $\Sigma_{i|0:n} = Var(X_i|y_{0:n})$, $\Sigma_{i-1,i|0:n} = Cov(X_{i-1}, X_i|y_{0:n})$, one performs the set of backward recursions $i = n, n-1, \ldots, 1$:

$$\begin{aligned} J_{i-1} &= \Sigma_{i-1|0:i-1}A'(\Sigma_{i|0:i-1})^{-1} \\ M_{i-1|0:n} &= M_{i-1|0:i-1} + J_{i-1}(M_{i|0:n} - M_{i|0:i-1}) \\ \Sigma_{i-1|0:n} &= \Sigma_{i-1|0:i-1} + J_{i-1}(\Sigma_{i|0:n} - \Sigma_{i|0:i-1})J'_{i-1} \end{aligned}$$

To calculate $\Sigma_{i-1,i|0:n}$, we have

$$\Sigma_{n-1,n|0:n} = (I - K_n H)A\Sigma_{n-1|0:n-1}$$

and the following backward recursions, for $i = n-1, n-2, \ldots, 1$

$$\Sigma_{i-1,i|0:n} = \Sigma_{i|0:i}J'_{i-1} + J_i(\Sigma_{i,i+1|0:n} - A\Sigma_{i|0:i})J'_{i-1}$$

E ARMA property of multidimensional process

In our model, $(y_i)_{i\in\mathbb{Z}}$ is an ARMA(2,2) process and asymptotic properties of the maximum likelihood estimator are derived. This result can be generalised to the case where X_i is p-dimensional under weak assumptions. We consider the model:

$$y_i = HX_i + \sigma\varepsilon_i, X_i = AX_{i-1} + \eta_i, \;\; X_0 \sim \nu$$

where X_i is p-dimensional, A is a diagonal matrix with diagonal coefficients $(\theta_k, k = 1, \ldots, p)$ such that $\theta_k \neq \theta_l$ for $k \neq l$ and $\theta_i \in (0,1)$ for $i = 1 \ldots p$, $(\eta_i)_{i\geq 0}$ is a sequence of independent $\mathcal{N}_p(0, R)$ random variables, H is a $(1,p)$-matrix and (ε_i) is a sequence of i.i.d $\mathcal{N}(0,1)$ random variables. Up to a transformation of (X_i), H is equal to $H = (1 \ldots 1\ 0 \ldots 0)$ with its first d coordinates equal to 1 and its $p-d$ next coordinates equal to 0 $(1 \leq d \leq p)$. Consequently, we observe with additive noise the partial sum of the first d coordinates of X_i. Since A is diagonal and $\theta_i \in (0,1)$ for $i = 1 \ldots p$, the process $(X_i)_{i\geq 0}$ admits a stationary distribution ν. Then with $X_0 \sim \nu$, the process $(X_i)_{i\geq 0}$ is stationary. Denote $(y_i)_{i\in\mathbb{Z}}$ the extension of y to $\mathbb{Z}$ by stationarity.

***Proposition* 6** *The process $(y_i)_{i\in\mathbb{Z}}$ is ARMA(d,d).*

Proof. Denote S_j, $j = 1, \ldots, d$, the j-th symmetric function of $\theta_1, \ldots, \theta_d$:

$$S_j = \sum_{1\leq i_1 < \cdots < i_j \leq d} \theta_{i_1} \ldots \theta_{i_j}$$

and $S_0 = 1$. Define the polynomial P such that $P(\theta_1) = \ldots = P(\theta_d) = 0$:

$$P(x) = \sum_{k=0}^{d} (-1)^k S_k x^{d-k}$$

Set L the one-lag operator : $Ly_i = y_{i-1}$, $L\varepsilon_i = \varepsilon_{i-1}$. Set $\xi_i = P(L)(y_i) = \sum_{k=0}^{d}(-1)^k S_k y_{i-k}$. By recursive computation for $0 \leq k \leq d$, we have

$$y_{i-k} = HA^{d-k}X_{i-d} + \sum_{j=0}^{d-k-1} HA^j \eta_{i-k-j} + \sigma\varepsilon_{i-k}$$

We deduce

$$\xi_i \quad = \quad HP(A)X_{i-d} + \sum_{k=0}^{d}(-1)^k S_k \left(\sum_{j=0}^{d-k-1} HA^j \eta_{i-k-j} \right) + \sigma P(L)(\varepsilon_i)$$

But $HP(A) = (P(\theta_1) \dots P(\theta_d)\ 0 \dots 0) = (0 \dots\dots 0)$. Thus ξ_i only depends on $(\eta_j, \varepsilon_j)_{j \leq i}$. Therefore (y_i) verifies an AR(d) equation. Moreover, as the $(\eta_i)_i$ and $(\varepsilon_i)_i$ are mutually independent, we get that:

$$Cov(\xi_i, \xi_{i+k}) = 0,\ \forall k \geq d.$$

This implies that (ξ_i) is MA(d). Hence the result.

	$\sigma^2 = 0.2$		$\sigma^2 = 1$	
Method	MLE	EM	MLE	EM
θ_1	0.28 (0.02)	0.21 (0.02)	0.25 (0.05)	0.29 (0.03)
θ_2	0.80 (0.00)	0.80 (0.00)	0.80 (0.00)	0.80 (0.00)
θ_3	0.52 (0.01)	0.61 (0.03)	0.53 (0.04)	0.55 (0.02)
θ_4	0.98 (0.04)	0.97 (0.04)	0.98 (0.10)	0.96 (0.06)
θ_5	0.09 (0.00)	0.15 (0.00)	0.08 (0.01)	0.08 (0.00)

Table 1: Mean estimated values (with estimated standard deviations in bracket) obtained with the exact MLE and the EM algorithms, evaluated on 20 simulated data with $n = 5000$ observations and $\sigma^2 = 0.2$ or $\sigma^2 = 1$ (σ^2 fixed to its true value).

	$\sigma^2 = 0.2$		$\sigma^2 = 1$	
Method	MLE	EM	MLE	EM
θ_1	0.47 (0.06)	0.26 (0.02)	0.48 (0.09)	0.49 (0.07)
θ_2	0.80 (0.00)	0.80 (0.00)	0.79 (0.00)	0.80 (0.00)
θ_3	0.46 (0.05)	0.62 (0.05)	0.53 (0.23)	0.64 (0.07)
θ_4	0.91 (0.02)	0.96 (0.04)	0.94 (0.01)	0.78 (0.07)
σ^2	0.29 (0.54)	0.34 (0.02)	1.00 (0.15)	1.02 (0.01)

Table 2: Mean estimated values (with estimated standard deviations in bracket) obtained with the exact MLE and the EM method, evaluated on simulated data with $n = 5000$ observations and $\sigma^2 = 0.2$ or $\sigma^2 = 1$ (θ_5 fixed to its true value).

	$\Delta = 0.1, n = 10000$		$\Delta = 0.2, n = 5000$		$\Delta = 1, n = 1000$	
value	true	estimated	true	estimated	true	estimated
MLE						
θ_1	0.30	0.31 (0.01)	0.09	0.26 (0.04)	0.00	0.00 (0.00)
θ_2	0.80	0.80 (0.00)	0.64	0.68 (0.00)	0.11	0.17 (0.02)
θ_3	0.50	0.53 (0.01)	0.54	0.72 (0.03)	0.55	1.04 (0.24)
θ_4	1.00	0.96 (0.04)	1.64	1.04 (0.35)	2.75	1.07 (2.79)
θ_5	0.10	0.09 (0.00)	0.12	0.36 (0.06)	0.13	0.74 (0.38)
EM						
θ_1	0.30	0.32 (0.01)	0.09	0.20 (0.20)	0.00	0.01 (0.01)
θ_2	0.80	0.80 (0.00)	0.64	0.66 (0.01)	0.11	0.16 (0.01)
θ_3	0.50	0.55 (0.01)	0.54	0.66 (0.03)	0.55	1.09 (0.32)
θ_4	1.00	0.94 (0.04)	1.64	1.25 (0.22)	2.75	1.25 (2.27)
θ_5	0.10	0.09 (0.01)	0.12	0.27 (0.03)	0.13	0.59 (0.22)

Table 3: Influence of Δ when $\sigma^2 = 1$: for each Δ, true values of θ, mean estimated values (with estimated standard deviations in bracket) are presented.