

# Supplementary Material I:

## Proofs of results in “Parametric estimation of Ordinary Differential Equations with Orthogonality Conditions”

### Contents

<b>1</b>	<b>Ordinary Differential Equations</b>	<b>1</b>
<b>2</b>	<b>Assumptions</b>	<b>1</b>
<b>3</b>	<b>Consistency</b>	<b>2</b>
<b>4</b>	<b>Asymptotics</b>	<b>4</b>
<b>5</b>	<b>From optimal Weighting to a practical algorithm</b>	<b>7</b>

## 1 Ordinary Differential Equations

**Cauchy-Lipschitz theory for non-autonomous ODE.** *If the vector  $(t, x) \mapsto f(t, x, \theta)$  satisfies the following conditions on  $[0, 1] \times \mathcal{X}$ :*

- (a)  *$f$  is  $L^2$ -Caratheodory, i.e.  $x \mapsto f(t, x, \theta)$  is continuous for  $t$  a.e in  $[0, 1]$ ,  $t \mapsto f(t, x, \theta)$  is measurable for all  $x \in \mathcal{X}$  and  $\forall c > 0$ ,  $\exists h_c(\cdot, \theta) \in L^2 \setminus |x| \leq c \implies |f(t, x, \theta)| \leq h_c(t, \theta)$ ,*
- (b)  *$f$  is  $L^2$ -Lipschitz, i.e.  $\exists a(\cdot, \theta) \in L^2 / \forall x, x' \in \mathcal{X}$ ,  $|f(t, x, \theta) - f(t, x', \theta)| \leq a(t, \theta) |x - x'|$ .*

*Then, there exists a unique solution to the Initial Value Problem (theorem 3.4 in [4])*

$$\begin{cases} \dot{\phi}(t) &= f(t, \phi(t), \theta) \\ \phi(0) &= \phi_0 \end{cases} \quad (1)$$

## 2 Assumptions

**Condition C1** (a)  $\Theta$  is a compact set of  $\mathbb{R}^p$  and  $\theta^*$  is an interior point of  $\Theta$ ,  $\mathcal{X}$  is an open subset of  $\mathbb{R}^2$ ; (b)  $(t, x) \mapsto f(t, x, \theta^*)$  is  $L^2$ -Lipschitz and  $L^2$ -Caratheodory.

**Condition C2** (a)  $(Y_i, t_i)$  are i.i.d. with variance  $V(Y|T=t) = \Sigma_\epsilon = \sigma^2 I_2$ ; (b) For every  $K$ , there is a non-singular constant matrix  $B$  such that for  $P^K = B_p^K(t)$ ; (i) the smallest eigenvalue of  $E[P^K(T)P^K(T)^\top]$  is bounded away from zero uniformly in  $K$  and (ii) there is a sequence of constants  $\zeta_0(K)$  satisfying  $\sup_t |P^K(t)| \leq \zeta_0(K)$  and  $K = K(n)$  such that  $\zeta_0(K)^2 K/n \rightarrow 0$  as  $n \rightarrow \infty$ ; (c) There are  $\alpha, \mathbf{c}_{1,K}, \mathbf{c}_{2,K}$  such that  $\|\phi_j^* - p^K \mathbf{c}_{j,K}\|_\infty = \sup_{[0,1]} |\phi_j^*(t) - p^K(t)^\top \mathbf{c}_{j,K}| = O(K^{-\alpha})$ .

**Condition C3** There exists  $D > 0$ , such that the  $D$ -neighborhood of the solution range  $\mathcal{D} = \{x \in \mathbb{R}^2 | \exists t \in [0, 1], |x - \phi^*(t)| < D\}$  is included in  $\mathcal{X}$  and  $f$  is  $C^2$  in  $(x, \theta)$  on  $\mathcal{D} \times \Theta$  for  $t$  in  $[0, 1]$  a.e. Moreover, the derivatives of  $f$  w.r.t  $x$  and  $\theta$  (with obvious notations)  $f_x, f_\theta, f_{xx}, f_{x\theta}$  and  $f_{\theta\theta}$  are  $L^2$  uniformly bounded on  $\mathcal{D} \times \Theta$  by  $L^2$  functions  $\bar{h}_x, \bar{h}_\theta, \bar{h}_{x\theta}, \bar{h}_{xx}$  and  $\bar{h}_{\theta\theta}$  (respectively).

**Condition C4** Let  $(\varphi_\ell)_{\ell \geq 1}$  be an orthonormal sequence of  $C^1$  functions in  $H_0^1$ .

**Condition C5**  $\theta^*$  is the unique global minimizer of  $Q_{\mathcal{F}}^*$  and  $\inf_{|\theta - \theta^*| > \epsilon} Q_{\mathcal{F}}^*(\theta) > 0$ .

**Condition C6** There exists  $L_0$  such that for  $L \geq L_0$ ,  $\mathbf{J}_{\theta,L}(g, \theta)$  is full rank in a neighborhood of  $(\phi^*, \theta^*)$ .

**Condition C7** (a) The times  $T_1, \dots, T_n$  have a density  $\pi$  w.r.t. Lebesgue measure such  $0 < c < \pi < C < \infty$ ;  
(b)  $E[\epsilon^4] < \infty$ .

**Condition C8** For  $\ell = 1, \dots, L$ ,  $\theta \in \Theta$ , there exists  $\tilde{\beta}_{K_\ell}$  in  $\mathbb{R}^{K_\ell}$  with  $\|\frac{f_x(\cdot, \phi^*, \theta)\varphi_\ell + \dot{\varphi}_\ell}{\pi} - \tilde{\beta}_{K_\ell}^\top p^{K_\ell}\|_{L^2} \rightarrow 0$ .

**Condition C9** (a) The series estimator is a regression spline with a uniform knot sequence  $(\tau_{1,K}, \dots, \tau_{N_K,K})$  defining the spline basis  $\mathbf{p}^K$  satisfies  $\max_i |\tau_{i+1,K} - \tau_{i,K}| \rightarrow 0$  as  $K \rightarrow 0$ ; (b) For all  $\theta \in \Theta$ , for  $\ell = 1 \dots L$ ,  $v_\ell : t \mapsto \frac{f_x(t, \phi^*(t), \theta)\varphi_\ell(t) + \dot{\varphi}_\ell(t)}{\pi(t)}$  is  $C^1$ .

### 3 Consistency

**Theorem 3.1.** *If conditions C1 to C6 are satisfied, then*

$$\hat{\theta}_{n,L} - \theta_L^* = O_P(1)$$

and the bias  $\mathbf{B}_L = \theta_L^* - \theta^*$  tends to zero as  $L \rightarrow \infty$ .

In particular, if we use the sine basis and if  $\mathcal{E}(\phi^*, \theta)$  is in  $H^1$  for all  $\theta$ , then  $\mathbf{B}_L = o(\frac{1}{L})$ .

*Proof.* The classical proof for the consistency of an  $M$ -estimator  $\hat{\theta}_{n,L} = \arg \min_{\theta \in \Theta} Q_{n,L}(\theta)$  such as the orthogonal conditions estimator relies basically on two stages: the uniform convergence of  $Q_{n,L}(\theta)$  towards  $Q_L^*(\theta) = \sum_{\ell=1}^L |e_\ell(\phi^*, \theta)|^2$  for  $\theta$  in  $\Theta$ , and the fact that the true parameter  $\hat{\theta}_{n,L}$  is a unique isolated global maximum (by theorem 5.7 in [6]). In a first step, we assume that  $Q_L^*(\theta)$  has a unique global minimum denoted  $\theta_L^*$ , and we will show that this is indeed the case, and that  $\theta_L^*$  is not very far from the true parameter  $\theta^*$ . We have to show that  $\sup_\theta |Q_{n,L}(\theta) - Q_L^*(\theta)| \rightarrow 0$  as  $n \rightarrow \infty$ . From the simple additive expression of  $Q_{n,L}(\theta)$ , we see that it suffices to show the uniform convergence of  $|e_\ell(\hat{\phi}, \theta)|^2$  to  $|e_\ell(\phi^*, \theta)|^2$ . From the inequality  $|a^2 - b^2| \leq |a - b|(|a| + |b|)$ , we have

$$\left| |e_\ell(\hat{\phi}, \theta)|^2 - |e_\ell(\phi^*, \theta)|^2 \right| \leq |e_\ell(\hat{\phi}, \theta) - e_\ell(\phi^*, \theta)| (|e_\ell(\hat{\phi}, \theta)| + |e_\ell(\phi^*, \theta)|).$$

As  $\hat{\phi}$  is such that for  $j = 1, 2$  we have  $\|\hat{\phi}_j - \phi_j^*\|_\infty = O_P(\zeta_0(K)(\sqrt{K/n} + K^{-\alpha}))$  this means that  $\hat{\phi}(t) \in \mathcal{D} = \{x \in \mathbb{R}^2 \mid \exists t \in [0, 1], |x - \phi^*(t)| \leq D\}$  with a probability tending to 1. This shows that with a probability tending to 1, we have  $|f(t, \hat{\phi}(t), \theta)| \leq h(t, \theta) \leq H(t)$  (because  $f$  is uniformly  $L^2$ -Caratheodory), moreover  $\hat{\phi}$  is also bounded by  $M + D > 0$  (because  $\phi^*$  is bounded as a continuous function on  $[0, 1]$ ) and

$$\begin{aligned} |e_\ell(\hat{\phi}, \theta)| + |e_\ell(\phi^*, \theta)| &\leq |\langle \mathcal{E}(\hat{\phi}, \theta), \varphi_\ell \rangle| + |\langle \hat{\phi} - \phi^*, \dot{\varphi}_\ell \rangle| + |\langle \mathcal{E}(\phi^*, \theta), \varphi_\ell \rangle| + |\langle \phi^*, \dot{\varphi}_\ell \rangle| \\ &\leq 2\|H\|_{L^2} + \|\dot{\varphi}_\ell\|_{L^2}(2M + D) \end{aligned}$$

Hence  $|e_\ell(\hat{\phi}, \theta)| + |e_\ell(\phi^*, \theta)|$  is uniformly bounded in probability, so that the uniform convergence of the criterion boils down to the uniform convergence of  $|e_\ell(\hat{\phi}, \theta) - e_\ell(\phi^*, \theta)|$ . We can re-write  $e_\ell(\hat{\phi}, \theta) - e_\ell(\phi^*, \theta) = \langle \mathcal{E}(\hat{\phi}, \theta) - \mathcal{E}(\phi^*, \theta), \varphi_\ell \rangle + \langle \hat{\phi} - \phi^*, \dot{\varphi}_\ell \rangle$  and it is clear that the second right-hand side term converges uniformly to zero in probability for all  $\ell \geq 1$ . Consequently, we just have to check that  $\mathcal{E}(\hat{\phi}, \theta) - \mathcal{E}(\phi^*, \theta)$  converges uniformly in  $\theta$  to 0 in probability. First of all, we remark that  $g \mapsto \langle \mathcal{E}(g, \theta), \varphi_\ell \rangle$  is a continuous function from  $B_\infty(\phi^*, D) = \{g \in C([0, 1]) \mid \|g - \phi^*\|_\infty \leq D\}$  to  $\mathbb{R}$  (w.r.t the sup-norm), because

$$\begin{aligned} |\langle \mathcal{E}(g, \theta), \varphi_\ell \rangle - \langle \mathcal{E}(g', \theta), \varphi_\ell \rangle| &\leq \langle a(\cdot, \theta) | g - g' |, \varphi_\ell \rangle \\ &\leq \langle a(\cdot, \theta), \varphi_\ell \rangle \|g - g'\| \end{aligned}$$

with  $a(\cdot, \theta) \in L^2$  for all  $\theta$ , as the vector field is  $L^2$ -Lipschitz. By the continuous mapping theorem, we get the point-wise convergence of  $\langle \mathcal{E}(\hat{\phi}, \theta) - \mathcal{E}(\phi^*, \theta), \varphi_\ell \rangle$ , and the hard part consist in the uniform convergence. This

can be proven by controlling the oscillations in  $\theta$  of the process  $\mathcal{E} = (\mathcal{E}(\theta))_{\theta \in \Theta} = \left( \langle \mathcal{E}(\hat{\phi}, \theta) - \mathcal{E}(\phi^*, \theta), \varphi_\ell \rangle \right)_{\theta \in \Theta}$  and by using theorem 18.11 in [6], as convergence in probability towards a constant is equivalent to weak convergence. In order to show that the process  $\mathcal{E}$  converges weakly to 0 in the space  $C(\Theta)$  of continuous functions on  $\Theta$  equipped with the supremum norm, we have to check that for all  $k$  and all  $\theta_1, \dots, \theta_k$  in  $\Theta$

$$(\mathcal{E}(\theta_1), \dots, \mathcal{E}(\theta_k)) \rightsquigarrow (0 \dots 0) \quad (2)$$

and that for all  $\epsilon, \alpha > 0$  there exists a partition of  $\Theta_1, \dots, \Theta_K$  of  $\Theta$  such that

$$\limsup_{n \rightarrow \infty} P \left( \sup_k \sup_{\theta, \theta' \in \Theta_k} |\mathcal{E}(\theta) - \mathcal{E}(\theta')| \geq \alpha \right) \leq \epsilon. \quad (3)$$

The first condition (2) is a direct consequence of the point-wise convergence in probability of  $\mathcal{E}(\theta)$  in  $\Theta$ . Concerning the second condition, from we have for  $t$  in  $[0, 1]$  a.e.

$$\begin{aligned} \left| f(t, \hat{\phi}(t), \theta) - f(t, \hat{\phi}(t), \theta') \right| &\leq \sum_{i=1}^p \left| f_{\theta_i}(t, \hat{\phi}(t), \tilde{\theta}_n(t)) \right| \left| \theta_i - \theta'_i \right| \\ &\leq \bar{\mathbf{a}}'(t) \left| \theta - \theta' \right| \end{aligned}$$

with a probability tending to 1 ( $\tilde{\theta}_n(t)$  being a parameter between  $\theta$  and  $\theta'$ ). As a consequence the following inequality

$$|\mathcal{E}(\theta) - \mathcal{E}(\theta')| \leq \left\langle \bar{\mathbf{a}}' \left| \theta - \theta' \right|, |\varphi_\ell| \right\rangle \leq \left| \theta - \theta' \right|$$

is true with a probability tending to 1. Since  $\Theta$  is a compact set, it is possible to find a finite partition  $\Theta_1, \dots, \Theta_K$  of  $\Theta$  such that the diameter of  $\Theta_i$  is smaller than an arbitrary  $\alpha$  independently of  $n$ . This ensures that condition (3) is also satisfied and the uniform convergence of  $Q_{n,L}(\theta)$  to  $Q_L^*(\theta)$  can be claimed. Now, we relate  $\theta_L^* - \theta^*$  to the approximation quality of  $\mathbf{E}_{\mathcal{F}}(\phi^*, \theta)$  by the basis  $(\phi_\ell)_{\ell \geq 1}$ . We remark first that we have  $|\mathbf{E}_{\mathcal{F}}(\phi^*, \theta)| \leq 2\bar{\mathbf{h}}(t)$ : this implies that for all  $\ell$ , and all  $\theta$  we have  $|e_\ell^*(\theta)| \leq 2\bar{\mathbf{h}}_\ell$  with  $\bar{\mathbf{h}}_\ell = \langle \bar{\mathbf{h}}, \varphi_\ell \rangle$ . The global rate of convergence of the series  $\sum_\ell e_\ell^2(\theta)$  is controlled uniformly by the rate of  $\sum_{\ell \geq 1} H_\ell^2$ , denoted  $r_L^2 = \sum_{\ell > L} H_\ell^2$ . By orthogonality, we can write

$$\|\mathbf{E}_{\mathcal{F}}(\phi^*, \theta)\|_{L^2}^2 = \|\mathbf{E}_L(\phi^*, \theta)\|_{L^2}^2 + \|\mathbf{R}_L(\phi^*, \theta)\|_{L^2}^2$$

which means that is  $\|\mathbf{E}_L(\phi^*, \theta)\|_{L^2}^2$  is a perturbation of the function  $\theta \mapsto \|\mathbf{E}_{\mathcal{F}}(\phi^*, \theta)\|_{L^2}^2$  by the function  $\theta \mapsto -\|\mathbf{R}_L(\phi^*, \theta)\|_{L^2}^2$ . This perturbation  $\|\mathbf{R}_L(\phi^*, \theta)\|_{L^2}^2$  is uniformly dominated by  $r_L$ , hence it becomes possible to relate the two minima. From assumption **C5**, we know  $\|\mathbf{R}_L(\phi^*, \theta)\|_{L^2}^2$  is differentiable and we compute a series decomposition of its gradient thanks to  $\partial_\theta \mathbf{R}_L(\phi^*, \theta) = \sum_{\ell > L} \partial_\theta e_\ell(\phi^*, \theta) \varphi_\ell$ :

$$\partial_\theta \|\mathbf{R}_L(\phi^*, \theta)\|_{L^2}^2 = 2\mathbf{R}_L(\phi^*, \theta) \partial_\theta \mathbf{R}_L(\phi^*, \theta).$$

and we recall that  $\mathbf{R}_L(\phi^*, \theta)$  converges uniformly to 0 and  $\partial_\theta \mathbf{R}_L(\phi^*, \theta)$  is uniformly bounded in  $\theta$  on  $\Theta$  (as a continuous function on the compact set  $\Theta$ ). Starting from this last remark, we use the Implicit Function Theorem to the continuously differentiable function  $G(\epsilon, \theta) = \partial_\theta \|\mathbf{E}_{\mathcal{F}}(\phi^*, \theta)\|_{L^2}^2 - 2\epsilon \partial_\theta \mathbf{R}_L(\phi^*, \theta)$ , in order to get a Taylor expansion. We denote  $\theta_\epsilon$  the solution to  $G(\epsilon, \theta_\epsilon) = 0$ , and we remark that in particular  $G(0, \theta^*) = 0$ . Thus, there exists  $\epsilon_0, \delta_0 > 0$  and a function  $\psi : ]-\epsilon_0, \epsilon_0[ \rightarrow B(\theta^*, \delta_0)$  such that  $\psi(\epsilon) = \theta_\epsilon$  (i.e.  $G(\epsilon, \psi(\epsilon)) = 0$ ). We can also compute the first order variation of  $\psi$ :  $\psi(\epsilon) = \psi(0) + \epsilon \psi'(0) + o(\epsilon)$  where

$$\psi'(0) = -2 \left( \partial_{\theta\theta} \|\mathbf{E}_{\mathcal{F}}(\phi^*, \theta)\|_{L^2}^2 \right)^{-1} \partial_\theta \mathbf{R}_L(\phi^*, \theta^*).$$

Since  $\mathbf{R}_L(\phi^*, \theta)$  converges uniformly to 0, there exists  $L_0 > 0$ , such that for  $L > L_0$ ,  $|\mathbf{R}_L(\phi^*, \theta)| \leq \epsilon_0$  so that we can apply the linearization above to  $G(\mathbf{R}_L(\phi^*, \theta), \theta) = \partial_\theta \|\mathbf{E}_{\mathcal{F}}(\phi^*, \theta)\|_{L^2}^2 - \partial_\theta \|\mathbf{R}_L(\phi^*, \theta)\|_{L^2}^2$ . We obtain that the minima  $\theta_L^*$  and  $\theta^*$  are such that

$$\theta_L^* = \theta^* + \mathbf{R}_L(\phi^*, \theta) 2 \left( \partial_{\theta\theta} \|\mathbf{E}_{\mathcal{F}}(\phi^*, \theta)\|_{L^2}^2 \right)^{-1} \partial_\theta \mathbf{R}_L(\phi^*, \theta^*) + o(r_L).$$

This implies that  $|\theta_L^* - \theta^*| = O(r_L)$ . If we use the sine basis, this means that  $\mathbf{P}_{\mathcal{F}} H = \sum_\ell \bar{\mathbf{h}}_\ell \sqrt{2} \sin(\pi \ell t)$  but it is also in  $H_0^1$  (as it is in  $H^1$ ), then we have  $\sum_{\ell \geq 1} \ell^2 \bar{\mathbf{h}}_\ell^2$  and  $r_L^2 = o\left(\frac{1}{L^2}\right)$ .  $\square$

## 4 Asymptotics

The asymptotics of the estimators  $\hat{\theta}_{n,L}$  are obtained by two (successive) linearizations of the orthogonal conditions  $\mathbf{e}_L(\hat{\phi}, \hat{\theta}_{n,L})$  around  $(\phi^*, \theta^*)$ . We detail the proof of proposition 4.1 and theorem 4.1.

### Section 4.1: Proofs

**Proposition 4.1.** *If conditions C1-C6 are satisfied, then*

$$\left[ \mathbf{J}_{\theta,L}(\hat{\phi}, \hat{\theta}_{n,L})^\top \tilde{\mathbf{J}}_{\theta,L} \right]^{-1} \mathbf{J}_{\theta,L}(\hat{\phi}, \hat{\theta}_{n,L})^\top \xrightarrow{P} \mathbf{M}_L^* = [\mathbf{J}_{\theta,L}^* \mathbf{J}_{\theta,L}^*]^{-1} \mathbf{J}_{\theta,L}^{*\top} \quad (4)$$

where the matrix  $\tilde{\mathbf{J}}_{\theta,L}$  is the Jacobian  $\mathbf{J}_{\theta,L}$  evaluated at a point  $\tilde{\theta}$  between  $\theta^*$  and  $\hat{\theta}_{n,L}$ . Moreover, we have

$$\hat{\theta}_{n,L} - \theta_L^* = -\mathbf{M}_L^* \mathbf{e}_L(\hat{\phi}, \theta_L^*) + o_P(1). \quad (5)$$

*Proof.* If conditions C1-C6 are satisfied, then

$$\left[ \mathbf{J}_{\theta,L}(\hat{\phi}, \hat{\theta}_{n,L})^\top \tilde{\mathbf{J}}_{\theta,L} \right]^{-1} \mathbf{J}_{\theta,L}(\hat{\phi}, \hat{\theta}_{n,L})^\top \xrightarrow{P} \mathbf{M}_L^* = [\mathbf{J}_{\theta,L}^{*\top} \mathbf{J}_{\theta,L}^*]^{-1} \mathbf{J}_{\theta,L}^{*\top} \quad (6)$$

and we have

$$\hat{\theta}_{n,L} - \theta_L^* = -\mathbf{M}_L^* \mathbf{e}_L(\hat{\phi}, \theta^*) + o_{\mathbb{P}}(1). \quad (7)$$

The first order condition implies that  $\hat{\theta}_{n,L}$  satisfies

$$\mathbf{J}_{\theta,L}(\hat{\phi}, \hat{\theta}_{n,L})^\top \mathbf{e}_L(\hat{\phi}, \hat{\theta}_{n,L}) = 0. \quad (8)$$

We develop a Taylor expansion of  $f(t, \hat{\phi}(t), \theta)$  around  $\theta^*$  of order 1 for  $t$  a.e. in  $[0, 1]$ : the orthogonal condition  $e_\ell(\hat{\phi}, \theta)$  can be decomposed as  $e_\ell(\hat{\phi}, \theta^*) + \langle f_{\theta}(\cdot, \hat{\phi}, \tilde{\theta}), \varphi_\ell \rangle$  where  $\tilde{\theta}$  is on straight line between  $\theta^*$  and  $\hat{\theta}_{n,L}$ . We can write it in vector form

$$\mathbf{e}_L(\hat{\phi}, \hat{\theta}_{n,L}) = \mathbf{e}_L(\hat{\phi}, \theta^*) + \tilde{\mathbf{J}}_{\theta,L}(\hat{\theta}_{n,L} - \theta_L^*). \quad (9)$$

$\tilde{\mathbf{J}}_{\theta,L}$  is a matrix  $\mathbb{R}^{L \times p}$  with entries  $\langle f_{\theta_i}(\cdot, \hat{\phi}, \tilde{\theta}), \varphi_\ell \rangle$ . Thus, if we premultiply (9) by  $\mathbf{J}_{\theta,L}(\hat{\phi}, \hat{\theta}_{n,L})^\top$ , we get the asymptotic expansion

$$0 = \mathbf{J}_{\theta,L}(\hat{\phi}, \hat{\theta}_{n,L})^\top \mathbf{e}_L(\hat{\phi}, \theta_L^*) + \mathbf{J}_{\theta,L}(\hat{\phi}, \hat{\theta}_{n,L})^\top \tilde{\mathbf{J}}_{\theta,L}(\hat{\theta}_{n,L} - \theta_L^*). \quad (10)$$

This shows that the key result for relating the behavior of  $(\hat{\theta}_{n,L} - \theta_L^*)$  to the behavior of  $\mathbf{e}_L(\hat{\phi}, \theta^*)$  is the convergence in probability of  $\mathbf{J}_{\theta,L}(\hat{\phi}, \hat{\theta}_{n,L})^\top$ . This is indeed the case because the matrix value function  $(g, \theta) \mapsto \mathbf{J}_{\theta,L}(g, \theta)$  is continuous w.r.t the uniform norm on  $\mathcal{D} \times \Theta$  (because functions  $\frac{\partial}{\partial \theta_i} f$  are uniformly Lipschitz) and  $(\hat{\phi}, \hat{\theta}_{n,L})$  converges in probability to  $(\phi^*, \theta^*)$ . Finally, since the matrix  $\mathbf{J}_{\theta,L}^{*\top} \mathbf{J}_{\theta,L}^*$  is nonsingular, by the continuous mapping theorem, we see that

$$\left[ \mathbf{J}_{\theta,L}(\hat{\phi}, \hat{\theta}_{n,L})^\top \tilde{\mathbf{J}}_{\theta,L} \right]^{-1} \mathbf{J}_{\theta,L}(\hat{\phi}, \hat{\theta}_{n,L})^\top$$

converges in probability to  $[\mathbf{J}_{\theta,L}^* \mathbf{J}_{\theta,L}^*]^{-1} \mathbf{J}_{\theta,L}^{*\top} = \mathbf{M}_L^*$ . Thus, this implies that  $\hat{\theta}_{n,L} - \theta_L^* = -\mathbf{M}_L^* \mathbf{e}_L(\theta^*) + o_P(1)$ .  $\square$

### section 4.3: Proofs

For ease of reading, we restate theorem 4.1, and we give a sketch of the proof that relies essentially on the properties of plug-in estimators for series estimators obtained in [3]. First, we give a preliminary result about the asymptotics of the Orthogonal Conditions  $\mathbf{e}_L(\hat{\phi}, \theta) - \mathbf{e}_L(\phi^*, \theta)$  for all  $\theta \in \Theta$ .

**Proposition 4.2 - Asymptotic Normality of Orthogonal Conditions for general series estimators.** Under conditions **C1-C3**, **C7** and **C8** and if  $K$  is taken such that  $\sqrt{n}K^{-\alpha} \rightarrow 0$  and

1.  $f$  is linear (in state  $x$ ),
2. **OR**,  $f$  is nonlinear (in state  $x$ ) and  $\frac{\zeta(K)^4 K^2}{n} \rightarrow 0$ ,

then for all  $\theta \in \Theta$ ,

$$\sqrt{n} \left( \mathbf{e}_L(\hat{\phi}, \theta) - \mathbf{e}_L(\phi^*, \theta) \right) \rightsquigarrow N(0, V_{e,L}(\theta))$$

Moreover, the asymptotic variance can be estimated by  $\widehat{V_{e,L}(\theta)} = \mathbf{A}(\phi^*, \theta) \widehat{V(\hat{\mathbf{c}}_1)} \mathbf{A}(\phi^*, \theta)^\top + \mathbf{B}(\phi^*, \theta) \widehat{V(\hat{\mathbf{c}}_2)} \mathbf{B}(\phi^*, \theta)^\top$ .

*Proof.* We have to show a kind of “delta method” result. Hence we need to compute the Fréchet derivative of the functional  $g \mapsto \mathbf{e}_L(g, \theta)$  around  $\phi^*$  for each  $\theta$ . For simplicity, we consider the scalar functional  $e_\ell(g, \theta)$ . From section 4.3, we know that the Fréchet derivative for the sup norm in  $\mathcal{D} = \{x \in \mathbb{R}^2 | \exists t \in [0, 1], |x - \phi^*(t)| < D\}$  is given, for all  $h \in L^2$

$$De_\ell(g, \theta).h = \int_0^1 \{f_x(t, g(t), \theta) \varphi(t) + \dot{\varphi}(t)\} h(t) dt.$$

Hence, the Fréchet derivative can be represented as a scalar product with a gradient vector  $v(g)$ , w.r.t. the  $L^2$  inner product with respect to the distribution  $\pi$  of times  $T$  (bounded away from 0, see condition **C7**), i.e

$$\begin{aligned} De_\ell(g, \theta).h &= \int_0^1 \frac{\{f_x(t, g(t), \theta) \varphi(t) + \dot{\varphi}(t)\}}{\pi(t)} h(t) \pi(t) dt \\ &= \langle v(g), h \rangle_{L^2(\pi)} \end{aligned}$$

with  $v(g)(t) = \frac{\{f_x(t, g(t), \theta) \varphi(t) + \dot{\varphi}(t)\}}{\pi(t)}$ . Moreover, this gradient is in  $L^2(\pi)$ , because  $f_x$  is dominated,  $\varphi, \dot{\varphi}$  and  $\pi^{-1}$  are bounded above.

It is possible to derive a central limit theorem if the gradient function  $v(\phi^*)$  can be itself approximated by a series expansion in the same basis  $p_{kK}$ ,  $k = 1, \dots, K$  as the function  $\phi^*$ . We need to be sure that the approximation is still valid under a random design, which is immediate, as  $\pi$  is equivalent to Lebesgue measure  $c \leq \pi \leq C$ . Indeed, let  $\tilde{v}_K = \beta_K^\top p^K$  being an approximation of  $v$  (according to condition **C8**), then we have

$$\int_0^T (v(t) - \tilde{v}_K(t))^2 \pi(t) dt \leq C \int_0^T (v(t) - \tilde{v}_K(t))^2 dt \rightarrow 0$$

In section 4.3, we have checked that the Fréchet derivative satisfies the following properties (thanks to the regularity of the vector field, condition **C3**):

1.  $\theta \in \Theta$ ,  $|e_\ell(g + h, \theta) - e_\ell(g, \theta) - De_\ell(g, \theta).h| \leq C \|h\|_\infty^2$ , and
2.  $|De_\ell(g, \theta).h - De_\ell(g', \theta).h| \leq C \|h\|_\infty \|g - g'\|_\infty$  with  $C$ , a constant independent of  $\theta, \epsilon$  and  $g, g'$ .

The same conditions are satisfied for the vector of conditions  $\mathbf{e}_L(\hat{\phi}, \theta)$ , so conditions of theorem 3 in [3] are satisfied, and we can claim that even for a nonlinear functional, we have a root- $n$  rate. Let  $\mathbf{v}_L(\theta)$  be the Jacobian of the vector of orthogonal conditions  $\mathbf{e}_L(g, \theta)$  evaluated at  $\phi^*$ , with entries  $v_\ell(\theta) = \frac{f_x(\cdot, \phi^*, \theta) \varphi_\ell + \dot{\varphi}_\ell}{\pi}$ . Then, we can claim that

$$\sqrt{n} \left( \mathbf{e}_L(\hat{\phi}, \theta) - \mathbf{e}_L(\phi^*, \theta) \right) \rightsquigarrow N(0, V_{e,L}(\theta))$$

and the asymptotic variance is given by  $V_{e,L}(\theta) = E[\mathbf{v}_L(\theta) \mathbf{v}_L(\theta)^\top \sigma^2]$ . Moreover, the asymptotic variance can be estimated by the plugin estimate as described in section 4.3 i.e  $\widehat{V_{e,L}(\theta)} = \mathbf{A}(\phi^*, \theta) \widehat{V(\hat{\mathbf{c}}_1)} \mathbf{A}(\phi^*, \theta)^\top + \mathbf{B}(\phi^*, \theta) \widehat{V(\hat{\mathbf{c}}_2)} \mathbf{B}(\phi^*, \theta)^\top$ .  $\square$

When splines with a uniform sequence of knots on  $[0, 1]$  are used for estimating  $\phi^*$ , it is possible to replace condition **C8** by an easier condition to check based on classical regularity assumptions. Indeed, splines are good universal approximating space for smooth functions: typically, if the meshsize  $\delta_K$  of the knots of the splines basis  $\mathbf{p}^K$  is such that  $\delta_K \rightarrow 0$  and the function  $v_\ell : t \mapsto \frac{f_x(t, \phi^*(t), \theta) \varphi_\ell(t) + \dot{\varphi}_\ell(t)}{\pi(t)}$  is  $C^1$ , then one can find a sequence of splines  $s_\ell^K$  such that  $\|v_\ell - s_\ell^K\|_{L^2} \rightarrow 0$  as  $K \rightarrow \infty$ , see theorem 7.3 in [1]. In that case, we can replace condition **C8** by condition **C9**.

**Proposition 4.3 - Asymptotic Normality of Orthogonal Conditions for splines.** *We suppose that the function  $\phi^*$  is  $C^s$  with  $s \geq 3$  and that conditions **C1**, **C2(a)**, **C3**, **C7** and **C9** are satisfied, then if  $\sqrt{n}K^{-s} \rightarrow 0$  and*

1.  *$f$  is linear (in state  $x$ ) and  $\frac{K^2}{n} \rightarrow 0$ ,*
2. ***OR**,  $f$  is nonlinear (in state  $x$ ) and  $\frac{K^4}{n} \rightarrow 0$*

then for all  $\theta \in \Theta$ ,

$$\sqrt{n} \left( \mathbf{e}_L(\hat{\phi}, \theta) - \mathbf{e}_L(\phi^*, \theta) \right) \rightsquigarrow N(0, V_{e,L}(\theta))$$

and the asymptotic variance can be estimated by  $\widehat{V_{e,L}(\theta)} = \mathbf{A}(\phi^*, \theta) \widehat{V(\hat{c}_1)} \mathbf{A}(\phi^*, \theta)^\top + \mathbf{B}(\phi^*, \theta) \widehat{V(\hat{c}_2)} \mathbf{B}(\phi^*, \theta)^\top$ .

*Proof.* This proposition is an application of proposition 4.2 to splines with uniform knots. It corresponds also to theorem 9 in [3], which is a general theorem for root-n consistency estimation of functional estimated with (uniform) splines. With this series estimators, conditions **C2** is satisfied, as the condition **C8** (with condition **C9**).  $\square$

With propositions 4.1, 4.2 and 4.3, the proof of the theorem 4.1 is rather straightforward as it follows the lines of classical proof for Generalized Moments Estimators.

**Theorem 4.1 - Root-n consistency & asymptotic normality of OC Estimators.** *If either the following conditions are satisfied:*

**$\mathbf{p}^K$  is a general series estimators** *Under conditions **C1-C8** and if  $f$  is a linear vector field or,  $f$  is a nonlinear vector field and  $K$  is chosen such that  $\frac{\zeta_0(K)^4 K^2}{n} \rightarrow 0$*

**$\mathbf{p}^K$  is a uniform knot splines** *Under conditions **C1-C2(a)**, **C3-C7**, **C9** and if  $f$  is a linear vector field and  $\frac{K^2}{n} \rightarrow 0$ , or  $f$  is a nonlinear vector field and  $\frac{K^4}{n} \rightarrow 0$*

Then  $\hat{\theta}_{n,L}$  is such that

$$\sqrt{n} \left( \hat{\theta}_{n,L} - \theta_L^* \right) \rightsquigarrow N(0, \mathbf{V}_L^*) \quad (11)$$

with

$$\mathbf{V}_L^* = \mathbf{M}_L^* \mathbf{V}_{e,L}^* \mathbf{M}_L^{*\top}. \quad (12)$$

where  $\mathbf{V}_{e,L}^* = V_{e,L}(\theta_L^*)$ . The asymptotic variance can be estimated by  $\hat{\mathbf{M}}_L V_{e,L}(\hat{\theta}_{n,L}) \hat{\mathbf{M}}_L^\top \xrightarrow{P} \mathbf{V}_L^*$ . In particular, if we use regression splines and  $t \mapsto f(t, \phi^*(t), \theta)$  is  $C^s$  on  $[0, 1]$  with  $s \geq 3$ , then (11) holds with  $K$  such that  $\sqrt{n}K^{-s} \rightarrow 0$  and  $n^{-1}K^4 \rightarrow 0$ . Moreover, if  $L = L(n) \rightarrow \infty, n \rightarrow \infty$  is chosen such that the bias  $\mathbf{B}_{L(n)} = O(n^{-1/2})$ , then we have

$$\hat{\theta}_{n,L(n)} - \theta^* = O_P(n^{-1/2}) \quad (13)$$

In particular, this is the case when the test functions  $\varphi_\ell$  are the sine basis, and  $L(n) = O(n^\alpha)$  with  $\alpha > 1/2$ .

*Proof.* We start with the asymptotic representation given in proposition 4.1:

$$\hat{\theta}_{n,L} - \theta_L^* = -\mathbf{M}_L^* \mathbf{e}_L(\hat{\phi}, \theta_L^*) + o_P(1)$$

As the parameter  $\theta_L^*$  is the global minima of  $\theta \mapsto Q_L^*(\theta) = \|\mathbf{e}_L(\phi^*, \theta)\|_2^2$ , the first order condition implies that  $\mathbf{J}_{\theta,L}(\phi^*, \theta_L^*)^\top \mathbf{e}_L(\phi^*, \theta_L^*) = 0$ . By definition of  $\mathbf{M}_L^*$ , we obtain that  $\mathbf{M}_L^* \mathbf{e}_L(\phi^*, \theta_L^*) = 0$  and we can write

$$\left( \hat{\theta}_{n,L} - \theta_L^* \right) = -\mathbf{M}_L^* \left( \mathbf{e}_L(\hat{\phi}, \theta_L^*) - \mathbf{e}_L(\phi^*, \theta_L^*) \right) + o_P(1)$$

and the application of proposition 4.3 gives directly the root-n consistency and normality of  $\hat{\theta}_{n,L}$  and the asymptotic variance  $\lim_n \sqrt{n}V(\hat{\theta}_{n,L})$  is directly derived from the asymptotic variance of the orthogonal conditions i.e it equals  $\mathbf{M}_L^* V_{e,L}(\theta_L^*) \mathbf{M}_L^{*\top}$ . The latter can be estimated by the plugin estimates  $V_{e,L}(\hat{\theta}_{n,L})$  and  $\mathbf{M}_L^*$  is estimated by using the observed Jacobian matrix  $J_{\theta,L}(\hat{\phi}, \hat{\theta}_{n,L})$  (following the estimator of proposition 4.2).

Finally, we can include the bias  $r_L = \theta_L^* - \theta^*$ , given in theorem 1, in the asymptotics by assuming that we can chose  $L = L(n)$  such that  $r_n = o(n^{-1/2})$ . From lemma 5 (and theorem 1), we can have an estimate of the bias due to the truncation of the Fourier expansion and  $r_L = o(\frac{1}{L})$  i.e  $r_n = o(n^{-1/2})$ . Then, the influence of the bias is negligible as

$$\begin{aligned} \hat{\theta}_{n,L(n)} - \theta^* &= \hat{\theta}_{n,L(n)} - \theta_{L(n)}^* + \theta_{L(n)}^* - \theta^* \\ &= O_P(n^{-1/2}) + o(n^{-1/2}) \\ &= O_P(n^{-1/2}) \end{aligned}$$

and asymptotic behavior (13) can hold.  $\square$

## 5 From optimal Weighting to a practical algorithm

Theorem 4.1 is of practical interest as it provides a closed-form expression for the asymptotic variance of  $\hat{\theta}_{n,L}$  for general ODE. We can make a parallel between the orthogonal condition estimator and the weighted nonlinear least-squares:

$$\hat{\theta}_c^{WLS} = \arg \min \sum_{i=1}^n w(t_i) |y_i - \phi(t_i, \theta_c)|^2$$

where  $w(\cdot)$  is a (positive) weight function and  $\theta_c = (\phi_0, \theta)$ .  $\hat{\theta}_c^{WLS}$  is consistent and  $\sqrt{n}(\hat{\theta}_c^{WLS} - \theta_c^*) \rightsquigarrow N(0, \mathbf{V}^{AWLS})$  under classical regularity assumptions on  $\mathbf{f}$ , see [5]. The asymptotic variance  $\mathbf{V}^{AWLS}$  is directly computed from the sensitivity equations, and the optimal weight function  $w(\cdot)$  is proportional to the variance function  $\sigma^2(\cdot)$ , meaning that the unweighted least-squares estimator is optimal in the homoscedastic case. It is clear that  $\mathbf{V}_L^* \neq \mathbf{V}^{AWLS}$ , which means that we are sure that the orthogonal condition estimator is not an efficient estimator (at least in the Gaussian case, because NLS and MLE are the same in that case). A striking difference between  $\mathbf{V}_L^*$  and  $\mathbf{V}^{AWLS}$  is that the least squares involves the Jacobian of the solution w.r.t. the initial values and parameters, whereas the orthogonal conditions involve the Jacobian of the vector field. It is then hard to compare these two matrices in generality and to evaluate the loss of efficiency. Nevertheless, we can compare the influence of a weight matrix for the minimization of the orthogonal conditions. Indeed, if we introduce a positive definite matrix  $W$  in  $\mathbb{R}^{L \times L}$ , we can define the weighted criterion

$$Q_{n,L}^W(\theta) = \mathbf{e}_L(\hat{\phi}, \theta)^\top W \mathbf{e}_L(\hat{\phi}, \theta)$$

and the corresponding estimator

$$\hat{\theta}_{n,L}^W = \arg \min_{\theta \in \Theta} Q_{n,L}^W(\theta).$$

The results of theorems 3 and 4 are then still true under straightforward adaptations. In particular,  $\hat{\theta}_{n,L}^W$  is consistent and asymptotically normal with asymptotic variance  $\mathbf{M}_{W,L}^* \mathbf{V}_{e,L}^* \mathbf{M}_{W,L}^{*\top}$  where  $\mathbf{M}_{W,L}^* = [\mathbf{J}_L^{*\top} W \mathbf{J}_L^*]^{-1} \mathbf{J}_L^{*\top} W$ . Eventually, we can then ask for the best weighting matrix  $W$  giving the smallest asymptotic variance. This is a classical result for Generalized Moment Estimators that we recall in the following proposition (see for instance section 3.6 in [2]):

**Proposition 5.1. Optimal weighting matrix**

*The minimal asymptotic variance for  $\hat{\theta}_{n,L}^W$  is obtained with  $W^{opt} = \mathbf{V}_{e,L}^{*-1}$  and*

$$\mathbf{V}_L^{opt}(\theta^*) = \left( \mathbf{J}_{\theta,L}^{*\top} \mathbf{V}_{e,L}^{*-1} \mathbf{J}_{\theta,L}^* \right)^{-1} \quad (14)$$

Even if we have an homoscedastic model, we have an interest in using a weighted estimator. Interestingly, the problem of choosing the best weighting matrix is directly related to the choice of the best set of test functions  $\varphi_1, \dots, \varphi_L$ . Indeed, the diagonalization of  $\mathbf{V}_{e,L}^* = \mathbf{U} \Lambda \mathbf{U}^\top$  permits the introduction of the eigenvalues

$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_L)$  (with  $\lambda_1 \geq \dots \geq \lambda_L \geq 0$ ) and  $U = (\mathbf{u}_1 | \dots | \mathbf{u}_L)$  is an orthogonal matrix. If  $\Lambda$  is nonsingular,  $Q_{n,L}^{opt}(\theta) \doteq Q_{n,L}^{W^{opt}}(\theta) = \widetilde{\mathbf{e}}_L(\hat{\phi}, \theta)^\top \Lambda^{-1} \widetilde{\mathbf{e}}_L(\hat{\phi}, \theta)$ , with  $\widetilde{\mathbf{e}}_L(\hat{\phi}, \theta) = U^\top \mathbf{e}_L(\hat{\phi}, \theta)$ . By linearity, the new orthogonal conditions  $\widetilde{\mathbf{e}}_L(\hat{\phi}, \theta)$  can be written as orthogonal conditions with the test functions  $\psi_1, \dots, \psi_L$  derived from the eigenvectors of  $\mathbf{V}_{e,L}^*$  as  $\psi_\ell = \sum_{k=1}^L u_{k\ell} \varphi_k$ . The use of the optimal weighting matrix is then equivalent to choose the best test functions in  $\mathcal{F}_L$  and the diagonalization show that some care must be taken (typically  $L$  have to be reasonably small to avoid  $\lambda_L \approx 0$ ).

The optimal weighting matrix  $W^{opt}$  given in proposition 4.2 is depends on the true unknown parameter: obviously, this makes this result is hard to use in practice. We suggest to approximate  $\mathbf{V}_{e,L}^{*-1}$  by computing a sequence of weighted estimates, as described in **Algorithm 1**.

---

**Algorithm 1** Iteratively Reweighted Orthogonal Conditions - IRWOC

---

**Require:**  $\hat{\phi}, \theta_{init} \in \Theta, \epsilon > 0$ ,

    Compute the unweighted estimator

$$\hat{\theta}_{n,L}^{(0)} = \arg \min_{\theta} Q_{n,L}(\theta).$$

    Compute the asymptotic covariance of  $\mathbf{e}_L(\hat{\phi}, \hat{\theta}_{n,L}^{(0)})$  with equation in Proposition 4.3.

**while**  $|\hat{\theta}_{n,L}^{(k)} - \hat{\theta}_{n,L}^{(k-1)}| > \epsilon$  **do**

$$\begin{cases} W^{(k)} &= \left( \widehat{V_{e,L}(\hat{\theta}_{n,L}^{(k)})} \right)^{-1} \\ \hat{\theta}_{n,L}^{(k+1)} &= \arg \min_{\theta} Q_{n,L}^{W^{(k)}}(\theta) \end{cases} . \quad (15)$$

**end while**

---

This is known as iterated Generalized Method of Moments (GMM), and consists in successive re-weighting, by using the consistent estimator of the variance  $\widehat{V_{e,L}(\hat{\theta}_{n,L}^{(k)})}$ , [2]. There is no theoretical guarantee for the (numerical) convergence of this algorithm but if the model is correct, this sequence of iterations tends to ameliorate the quality of the GMM estimator. Possibly, during this stage, it could be necessary to select appropriately  $L$ , as  $\widehat{V_{e,L}(\hat{\theta}_{n,L}^{(k)})}$  can be close to be nonsingular at an iteration  $k$ .

## References

- [1] R.A. DeVore and G.G. Lorentz. *Constructive Approximation*, volume 303 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, 1993.
- [2] A.R. Hall. *Generalized Method of Moments*. Oxford University Press, 2005.
- [3] W. K. Newey. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79:147–168, 1997.
- [4] D. O'Regan. *Existence theory for nonlinear ordinary differential equations*. Mathematics and its applications. Kluwer, 1997.
- [5] L. Pronzato. Optimal experimental design and some related control problems. *Automatica*, 44:303–325, 2008.
- [6] A.W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilities Mathematics. Cambridge University Press, 1998.