Supplementary material
Optimal control and additive perturbations help in estimating ill-posed and uncertain dynamical systems

1 Notations and Norm inequalities

The definition of the Tracking estimator is based on the following optimization problem:

$$\begin{array}{ll}\text{Minimize: } \mathcal{C}(Y;x,u,x_0,\theta,U) = \int_0^T \|Cx(t)-Y(t)\|_2^2\,dt + \int_0^T \|u(t)\|_U^2\,dt \\ \text{Subject to: } \begin{cases} \dot{x}(t) = f(t,x(t),\theta) + Bu(t) \\ x(0) = x_0 \\ u \in L^2\left([0,\,T]\,,\mathbb{R}^{d_u}\right) \end{cases}\end{array} \tag{1.1}$$

The cost $\mathcal{S}$ is profiled on the admissible "perturbation" u:

$$\mathcal{S}(\hat{Y};\theta,U) = \min_{u\in L_a} \mathcal{C}(\hat{Y};u,\theta,U) \tag{1.2}$$

and the Tracking estimator is the global minimum of $\mathcal{S}$

$$\widehat{\theta}^T = \arg\min_{\theta\in\Theta} \mathcal{S}(\hat{Y};\theta,U)$$

For this reason, it is useful to introduce $X_{\theta,x_0,u}$ the solution of :

$$\begin{cases} \dot{x}(t) = f(t,x(t),\theta) + Bu(t) \\ x(0) = x_0 \end{cases} \tag{1.3}$$

We recall $\|.\|_U^2$ and $\|.\|_{L^2,U}$ are respectively the weighted version of the squared euclidean norm and L^2 norm (e.g $\|f(t)\|_U^2 = f(t)^T U f(t)$ and $\|f\|_{L^2,U}^2 = \int_0^T \|f(t)\|_U^2\,dt$), for $U = Id$ we simply use the classic notation $\|.\|^2$ and $\|.\|_{L^2}$. For matrices, we use the Frobenius norm $\|A\|_2 = \sqrt{\sum_{i,j}^d a_{i,j}^2}$. Continuity and differentiability have to be understood according to these norms.

We recall conditions C1-C11 introduced in the paper:

C1 The vector field f has a compact support Q w.r.t x, that is Q is compact and $f(t,x,\theta) = 0$ if $x \notin Q$. $\forall\theta\in\Theta$, there is a unique solution X_θ of the original ODE defined on $[0,\,T]$.

C2 $\forall\theta\in\Theta$, $(t,x)\longmapsto f(t,x,\theta)$ is continuous on $[0,\,T]\times Q$ and $\forall t\in[0,\,T]\,,\, x\longmapsto \frac{\partial f}{\partial x}(t,x,\theta)$ exists and $(t,x)\longmapsto \frac{\partial f}{\partial x}(t,x,\theta)$ is continuous on $[0,\,T]\times Q$.

C3 The signal $t\longmapsto Y(t)$ is continuous on $[0,\,T]$ (at least has a continuous representative).

C2 bis $\forall\theta\in\Theta$, $(t,x)\longmapsto f(t,x,\theta)$ is C^2 on $[0,\,T]\times Q$ and bounded.

C3 bis The signal $t\longmapsto Y(t)$ is C^2 on $[0,\,T]$.

C4 The model is structurally identifiable at (θ^*,x_0^*) i.e

$$\forall\,(\theta,x_0)\in\Theta\times Q\,;\,CX_{\theta,x_0}=CX_{\theta^*,x_0^*}\Longrightarrow(\theta,\,x_0)=(\theta^*,\,x_0^*)\,.$$

C5 The functions $(t,x,\theta)\longrightarrow\frac{\partial f}{\partial\theta}(t,x,\theta)$, $(t,x,\theta)\longrightarrow\frac{\partial^2 f}{\partial x\partial\theta}(t,x,\theta)$ are continuous on $[0,\,T]\times Q\times\Theta$ and bounded.

C6 The functions $(t,x,\theta)\longrightarrow\frac{\partial^2 f}{\partial^2\theta}(t,x,\theta)$, $(t,x,\theta)\longrightarrow\frac{\partial^3 f}{\partial x\partial^2\theta}(t,x,\theta)$,$(t,x,\theta)\longrightarrow\frac{\partial^3 f}{\partial^2 x\partial\theta}(t,x,\theta)$ and $(t,x,\theta)\longrightarrow\frac{\partial^3 f}{\partial^3 x}(t,x,\theta)$ are continuous on $[0,\,T]\times Q\times\Theta$ and bounded.

C7 $\frac{\partial^2\mathcal{S}(Y^*;\theta^*,U)}{\partial\theta^T\partial\theta}$ is non singular.

C8 Observations (t_i,Y_i) are i.i.d with $Var(Y_i\mid t_i)=\sigma I_{d'}$ with $\sigma<+\infty$ and the t_i are uniformely distributed on $[0\,,\,T]$.

C9 It exists $s\geq 1$ such $t\longmapsto f(t,X^*(t),\theta^*)$ is $C^{s-1}\left([0\,,\,T]\,,\mathbb{R}^d\right)$ and $\sqrt{n}K^{-s}\longrightarrow 0$ and $\frac{K^s}{n}\longrightarrow 0$.

C10 The meshize $\max_i|\tau_{i+1,K}-\tau_{i,K}|\longrightarrow 0$ when $K\longrightarrow+\infty$.

We recall also the following notations:

X_{θ,x_0}: solution of $\dot{x}(t)=f(t,x,\theta)$ with initial condition $X_{\theta,x_0}(0)=x_0$.

$X_{\theta,x_0,u}$: solution of $\dot{x}(t)=f(t,x,\theta)+Bu$ with initial condition $X_{\theta,x_0,u}(0)=x_0$.

$\Lambda_U(Y,t,x,u):=\|Cx-Y(t)\|_2^2+\|u\|_U^2$, running cost for the Optimal Control Problem (1.1).

$\overline{\phi}:=\max_{\Theta\times Q\times[0,\,T]}\|X_{\theta,x_0}(t)\|_2$

$\overline{f}_x \;:=\; \max_{[0,\,T]\times Q\times\Theta}\left\|\frac{\partial f}{\partial x}(t,x,\theta)\right\|_2$

$\overline{f}_{xx} \;:=\; \max_{[0,\,T]\times Q\times\Theta}\left\|\frac{\partial f^2}{\partial^2 x}(t,x,\theta)\right\|_2$

$D(Y):=\sup_{\theta\in\Theta}\|Y-CX_{\theta,u}(.Y)\|_{L^2}<\infty$

$E(Y):=2\sqrt{d}\,\|C\|_2\sqrt{\frac{e^{2\sqrt{d}\overline{f_x}T}-1}{2\sqrt{d}\overline{f_x}}}D(Y)$

$F(Y):=\frac{1}{2}\|B\|_2^2\,d^2T^2e^{2\sqrt{d}\overline{f_x}T}\left(\overline{f_{xx}}E(Y)+2\|C\|_2^2\right)$

$\lambda_1(Y):=\frac{\sqrt{d}T\|B\|_2^2E_Q(Y)}{4}\frac{e^{2\sqrt{d}\overline{f_x}T}}{\overline{f_x}}$

$\lambda_2(Y):=d\,\|B\|_2^2\left(\frac{\overline{f_{xx}}E(Y)+2\left\|C^TC\right\|_2}{8\overline{f_x}^2}\right)\sqrt{\left(e^{2\sqrt{d}\overline{f_x}T}-1-2\sqrt{d}\overline{f_x}T\right)\left(e^{2\overline{f_x}T}-1-2\overline{f_x}T\right)}$

$\lambda_3(\zeta)=F(Y^*)+\frac{1}{2}\left(L_1\zeta+\sqrt{\left(L_1\zeta+4\sqrt{d}\,\|C\|_2\,L_2\right)L_1\zeta}\right)$

$\lambda_4(\zeta)=F(Y^*)+L_3\zeta$

$L_1=\|B\|_2^2\,d^2\sqrt{d}T^2e^{2\sqrt{d}\overline{f_x}T}\overline{f_{xx}}\,\|C\|_2\sqrt{\frac{e^{2\sqrt{d}\overline{f_x}t}-1}{2\sqrt{d}\overline{f_x}}}$

$L_2=d\,\|B\|_2^2\,T^2e^{2\sqrt{d}\overline{f_x}T}\,\|C\|_2$

$L_3 = d^{\frac{3}{2}} \|B\|_2^2 T^2 e^{2\sqrt{d}T\overline{f_x}}\overline{f_{xx}} \|C\|_2 \sqrt{\frac{e^{2\sqrt{d}\overline{f_x}t}-1}{2\sqrt{d}\overline{f_x}}} \left(\frac{\sqrt{d}\|C\|_2 L_2}{\lambda - H(Y^*)} + 1\right)$

$B(Y^*, \zeta) = \left\{Y \in L^2\left([0,T], \mathbb{R}^{d'}\right) \; s.t \; \|Y - Y^*\|_{L^2} < \zeta\right\}$

In the following proofs, we use repeatedly the following norm inequalities: $\|AX\|_2 \leq \sqrt{d}\,\|A\|_2\,\|X\|_2$ and $\frac{d}{dt}\|\varphi(t)\|_2 \leq \left\|\frac{d\varphi}{dt}(t)\right\|_2$ where $t \mapsto \varphi(t)$ is a matrix valued function. Indeed, we have:

$$\begin{array}{rcl}
\frac{d}{dt}\|\varphi(t)\|_2 & = & \frac{d}{dt}\sqrt{\sum_{i=1}^{I}\sum_{j=1}^{J}\varphi_{i,j}(t)^2} \\
& = & \frac{2\sum_{i=1}^{I}\sum_{j=1}^{J}\varphi_{i,j}(t)\dot{\varphi}_{i,j}(t)}{2\sqrt{\sum_{i=1}^{I}\sum_{j=1}^{J}\varphi_{i,j}(t)^2}} \\
& = & \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\varphi_{i,j}(t)\dot{\varphi}_{i,j}(t)}{\|\varphi(t)\|_2} \\
& \leq & \frac{\|\varphi(t)\|_2\|\dot{\varphi}(t)\|_2}{\|\varphi(t)\|_2} = \|\dot{\varphi}(t)\|_2
\end{array}$$

The last inequality is obtained by Cauchy-Schwarz.

2 Extended simulation: Misspecified FitzHugh-Nagumo model

The FitzHugh-Nagumo is a nonlinear two-dimensional ODE introduced for modeling neurons. For well-chosen parameters and initial conditions, it exhibits a periodic behavior, with typical oscillations corresponding to a limit cycle.

$$\left\{\begin{array}{rcl}
\dot{V} & = & c\left(V - \frac{V^3}{3} + R\right) \\
\dot{R} & = & -\frac{1}{c}(V - a + bR)
\end{array}\right. . \tag{2.1}$$

The true parameters are $a^* = b^* = 0.2$ and $c^* = 3$ and $x_0^* = (V_0^*, R_0^*) = (-1, 1)$, and are taken from [7] where it was introduced as a benchmark for parameter estimation in ODE. In our case, the original model is altered by a step function Z defined by: $Z(t) = 0.3\mathbb{I}_{[5,\,10]}(t) + 0.3\mathbb{I}_{[15,\,20]}(t)$. This function is originally present in the model proposed by [4] to picture an exogenous stimuli. Hence, the true model is in fact

$$\left\{\begin{array}{rcl}
\dot{V} & = & c\left(V - \frac{V^3}{3} + R + Z\right) \\
\dot{R} & = & -\frac{1}{c}(V - a + bR)
\end{array}\right. \tag{2.2}$$

but we still use (2.1) as the true model during the estimation procedure. To give a clearer idea of the influence of Z in the resulting dynamics, we plot in figure 2.1 the solution of (2.1) and (2.2) for the same parameter value (a^*, b^*, c^*) and initial conditions x_0^*. We also plot an example of generated data with the parameter θ^*: the data are generated by adding a Gaussian noise to the trajectories of (V, R), for various sample sizes, see figures . This experiment gives an idea of the robustness of estimation with respect to model misspecification (the case of the estimation of a well-specified model is not discussed here as NLS, GP and Tracking behave similarly). The results are presented in table 2.2. In that case, Tracking or GP give notably superior estimates thanks to the use of approximate models: the estimates are obtained by profiling the possible perturbations of the model, and both the bias and variance of the estimators are reduced with respect to Nonlinear Least Squares. For $n = 100$, Tracking and Generalized Profiling are equivalent, but for a smaller size, Tracking gives a smaller bias and variance.

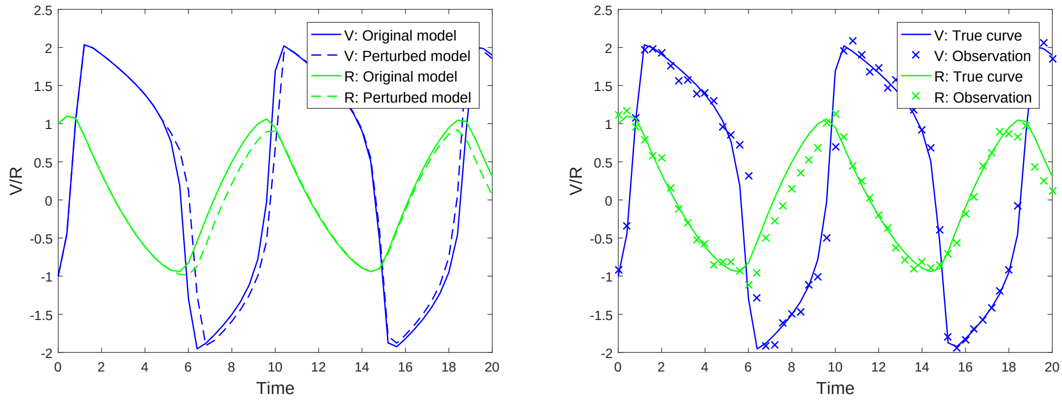

Fig. 2.1: Left: Solution of (2.1) and (2.2). Right:Example of simulated data and corresponding curves.

(n, σ)		$Bias(\widehat{\theta})$	$Tr\left(V(\widehat{\theta})\right)\times 10^{-4}$	MSE $(\times 10^{-2})$
$(100, 0.1)$	$\widehat{\theta}^{T}$	0.30	6.36	3.41
	$\widehat{\theta}^{NLS}$	0.43	13	7.13
	$\widehat{\theta}^{GP}$	0.30	7.26	3.48
$(50, 0.1)$	$\widehat{\theta}^{T}$	0.28	9.31	3.45
	$\widehat{\theta}^{NLS}$	0.42	17	6.52
	$\widehat{\theta}^{GP}$	0.36	11	5.47

Tab. 2.1: Bias, Variance, MSE, for parameter estimation for ill specified FitzHugh-Nagumo model

3 Pontryagin Maximum Principle and perturbed ODE

We want to find the optimal controls, i.e the solutions of the problem (1.1) in order to compute $\mathcal{S}$. By applying THEOREM 3.1 to the problem (1.1), we derive that the optimal processes $(X_{\theta,\overline{u}}, \overline{u})$ are obtained by solving a Boundary Value Problem.

Theorem 3.1. If conditions C1, C2 and C3 are satisfied, *then the optimal processes* $(X_{\theta,\overline{u}}, \overline{u})$ *for the problem (1.1) satisfy the Pontryagin Maximum Principle. That is, the optimal control and* $\mathcal{S}$ *are respectively equal to:*

$$\overline{u}(t) = \frac{1}{2}U^{-1}B^{T}p_{\theta}(t) \tag{3.1}$$

and

$$\mathcal{S}(Y;\theta,U) = \int_0^T \left\|CX_{\theta,\overline{u}}(t) - Y(t)\right\|_2^2 dt + \frac{1}{4}\int_0^T p_{\theta}(t)^T BU^{-1}B^T p_{\theta}(t)dt$$

where p_θ *is called the adjoint vector, an absolutely continuous vector valued function, defined such that* $(X_{\theta,\overline{u}}, p_\theta)$ *is the solution of the extended ODE with boundary constraint:*

$$\begin{cases} \dot{X}_{\theta,\overline{u}}(t) = f(t, X_{\theta,\overline{u}}(t), \theta) + \frac{1}{2}BU^{-1}B^{T}p_{\theta}(t) \\ \dot{p}_{\theta}(t) = -\frac{\partial f}{\partial x}(t, X_{\theta,\overline{u}}(t), \theta)^{T}p_{\theta}(t) + 2C^{T}\left(CX_{\theta,\overline{u}}(t) - Y(t)\right) \\ (X_{\theta,\overline{u}}(0), p_{\theta}(T)) = (x_0, 0) \end{cases} . \tag{3.2}$$

Proof. For applying the Pontryagin Maximum Principle, we need to check the required assumptions, given in [2].

The regularity assumptions C1 and C2 are enough in order to satisfy the “Classical regularity hypotheses" (22.1, page 427 in [2]): $F(t,x,u) := f(t,x) + Bu$ and $(t,x,u) \longmapsto \Lambda_U(Y,t,x,u) := \|Cx - Y(t)\|_2^2 + \|u\|_U^2$ are continuous, have derivatives w.r.t x continuous on $[0,\,T] \times Q$.

Since we do not know a priori if the optimal control is bounded, we need also to check hypothesis 22.16 p. 454 in [2]: there exists $\varepsilon > 0$, a constant c and a summable function d such that, for almost every $t \in [0,\,T]$, we have:

$$\|x - X_{\overline{u}}(t)\|_2 \leq \epsilon \Longrightarrow \left\| \frac{\partial\,(F, \Lambda_U)}{\partial x}\,(t, x, \overline{u}(t)) \right\|_2 \leq c\,\|(F, \Lambda_U)\,(t, x, \overline{u}(t))\|_2 + d(t).$$

The hypothesis 22.16 is also satisfied because

$$\begin{array}{rcl} \left\|\frac{\partial f}{\partial x}\,(t,x,u)\right\|_2 + \left\|\frac{\partial \Lambda_U}{\partial x}\,(t,x,u)\right\|_2 & \leq & \left\|\frac{\partial f}{\partial x}\,(t,x)\right\|_2 + 2\left\|C^T\,(Cx - Y(t))\right\|_2 \\ & \leq & \overline{f_x} + 2\left\|C^T\,(Cx - Y(t))\right\|_2 \end{array}$$

on $[0,\,T] \times Q$. We can use Pontryagin Maximum Principle to derive the existence of an arc $p\,:\,[0,\,T] \longrightarrow \mathbb{R}^d$ and a scalar η satisfying $\forall t \in [0,\,T]$, $(\eta, p(t)) \neq 0$ (non-triviality condition), $p(T) = 0$ (transversality condition),

$$-\dot{p}(t) = \frac{\partial}{\partial x} H^\eta(t, X_{\overline{u}}(t), p(t), \overline{u}(t))\ \text{(adjoint equation)}$$

and

$$H^\eta(t, X_{\overline{u}}(t), p(t), \overline{u}(t)) = \sup_u H^\eta(t, X_{\overline{u}}(t), p(t), u(t))\ \text{(maximum condition)}$$

where H^η is the Hamiltonian given by $H^\eta(t,x,p,u) = p^T f(t,x,u) - \eta\Lambda_U(Y,t,x,u)$.

The nontrivially condition imposes $\eta = 1$. Indeed, if $\eta = 0$, the adjoint vector p is the solution of the linear ODE

$$\begin{cases} \dot{p}(t) = -p(t)^T \frac{\partial f}{\partial x}(t, X_{\overline{u}}(t), \overline{u}(t)) \\ p(T) = 0 \end{cases}.$$

By uniqueness, this implies that it should satisfies $p(t) = 0$ for all t in $[0,T]$, which violates the nontrivially condition.

For all (t,x,p), $H(t,x,p,u)$ is strictly concave w.r.t u , hence it has an unique maximum given by the first order condition:

$$\frac{\partial H}{\partial u}(t,x,p,u) = 0 \Leftrightarrow u = \frac{1}{2}U^{-1}B^T p$$

For every t, we can compute the optimal control $\overline{u}$ with the maximum condition, which gives $\overline{u}(t) = \frac{1}{2}U^{-1}B^T p(t)$. Since we have:

$$\frac{\partial H}{\partial x}(t,x,p,u) = \frac{\partial f}{\partial x}(t,x)^T p - 2C^T\,(Cx - Y(t))$$

we know the adjoint vector is driven by the ODE

$$\dot{p}(t) = -\frac{\partial f}{\partial x}(t, X_{\overline{u}}(t))^T p(t) + 2C^T\,(CX_{\overline{u}}(t) - Y(t))$$

by merging this equation with the original one ruling $X_{\bar{u}}$ and with the optimal control expression, we obtain that $(X_{\overline{u}}(t), p(t))$ is solution of the extended ODE with boundary constraint (3.2). $\square$

4 Existence theorem

Theorem 4.1. *If conditions C1 and C2 are satisfied, then for all signals $Y \in L^2([0,\, T]\,, \mathbb{R}^{d'})$ and for all $\theta \in \Theta$, the problem (1.1) admits at least one solution. It exists a process $(X_{\theta,\overline{u}}, \overline{u})$ that minimizes the cost i.e $\mathcal{C}(Y; X_{\theta,\overline{u}}, \overline{u}, \theta, U) = \min_{u \in L^2_a} \mathcal{C}(Y; X_{\theta,u}, u, \theta, U)$.*

Proof. For the sake of notation clarity, we drop the dependence in θ for the vector field and the solution. Lemma (4.2) ensures the existence of admissible processes for problem (1.1), thus, we can consider an admissible minimizing sequence (X_{u_i}, u_i). Since we have:

$$\lambda \left\| u_i \right\|_{L^2}^2 \leq \mathcal{C}(Y; X_{u_i}, u_i, U)$$

with λ maximum eigenvalue of U, the sequence $\{u_i\}$ is uniformly bounded in $L^2([0,\, T]\,, \mathbb{R}^d)$, a reflexive Banach space, according to THEOREM III.27 in [1], its exists a subsequence converging weakly to a limit $\overline{u}$. Using Hölder inequality:

$$\left\| fg \right\|_{L^1} \leq \left\| f \right\|_{L^p} \left\| g \right\|_{L^q} \text{ with } \frac{1}{p} + \frac{1}{q} = 1$$

, we also know the subsequence is bounded in $L^1([0,\, T]\,, \mathbb{R}^d)$. (For the sake of notation here and in the following we still denote the subsequence by $\{u_i\}$). For the following, let us denote $\widetilde{u}$ the upper bound of the sequence $\{u_i\}$.

Knowing that:

$$\begin{aligned} \left\| \dot{X}_{u_i}(t) - \dot{X}(t) \right\|_2 &\leq \left\| f(t, X_{u_i}(t)) - f(t, X(t)) \right\|_2 + \left\| B u_i(t) \right\|_2 \\ &\leq \overline{f_x} \left\| X_{u_i}(t) - X(t) \right\|_2 + \sqrt{d_u} \left\| B \right\|_2 \left\| u_i(t) \right\|_2 \end{aligned}$$

(here $\overline{f_x} < \infty$ thanks to C2 as a continuous function on a compact subset). Gronwall's lemma gives us:

$$\left\| X_{u_i}(t) - X(t) \right\|_2 \leq \sqrt{d_u} \left\| B \right\|_2 \int_0^t e^{\overline{f_x}(t-s)} \left\| u_i(s) \right\|_2 ds \leq \sqrt{d_u} \left\| B \right\|_2 e^{\overline{f_x} t} \int_0^t \left\| u_i(s) \right\|_2 ds$$

and so:

$$\left\| X_{u_i}(t) \right\|_2 \leq \left\| X_{u_i}(t) - X(t) \right\|_2 + \left\| X(t) \right\|_2 \leq \sqrt{d_u} \left\| B \right\|_2 e^{\overline{f_x} t} \int_0^t \left\| u_i(s) \right\|_2 ds + \overline{\phi}$$

u_i, being bounded in $L^1([0,\, T]\,, \mathbb{R}^d)$, we deduce X_{u_i} (modulo a subsequence) is uniformly bounded on $[0,\, T]$ and since:

$$\begin{aligned} \left\| \dot{X}_{u_i}(t) \right\|_2 &\leq \left\| \dot{X}_{u_i}(t) - \dot{X}(t) \right\|_2 + \left\| \dot{X}(t) \right\|_2 \\ &\leq \left\| f(t, X_{u_i}(t)) - f(t, X(t)) \right\|_2 + \left\| B u_i(t) \right\|_2 + \left\| f(t, X(t)) \right\|_2 \\ &\leq \overline{f_x} \left\| X_{u_i}(t) - X(t) \right\|_2 + \left\| B u_i(t) \right\|_2 + \overline{f} \end{aligned}$$

we conclude from that $\dot{X}_{u_i}$ is bounded in $L^2([0,\, T]\,, \mathbb{R}^d)$, hence (again modulo a subsequence) $\dot{X}_{u_i}$ converges weakly to a limit $\overline{\dot{X}}$.

Since the sequence X_{u_i} is equicontinuous because

$$\left\| X_{u_i}(t) - X_{u_i}(t') \right\|_2 \leq \overline{f} \left| t - t' \right| + \sqrt{d_u} \left\| B \right\|_2 \widetilde{u} \sqrt{|t - t'|}$$

we can invoke Arzela-Ascoli theorem to obtain the uniform convergence (modulo a subsequence) of X_{u_i} toward a continuous function $\overline{X}$ on $[0,\, T]$. Using the identity, $X_{u_i}(t) = x_0 + \int_0^t \dot{X}_{u_i}(s)ds$ and by taking the limit we know $\overline{X}$ is an absolutely continuous function with $\dot{\overline{X}}(t) = \overline{X}(t)$ a.e.

We respect the hypothesis of THEOREM 6.38 in [2] we deduce from that:

$$\mathcal{C}(Y;\overline{X},\overline{u},U) \leq \liminf_{i\longrightarrow\infty} \mathcal{C}(Y;X_{u_i},u_i,U) = \inf \mathcal{C}(Y;x,u,U).$$

We now demonstrate $(\overline{X},\overline{u})$ is an admissible process (thus the infimum is reached). Using uniform convergence we have $\overline{X}(0) = x_0$. The last thing left to show is that $\overline{X}$ is a trajectory corresponding to $\overline{u}$, thus $\overline{X} = X_{\overline{u}}$. For any measurable subset A of $[0,\, T]$ we have:

$$\int_A \left(\dot{X}_{u_i}(t) - f(t,X_{u_i}(t)) - Bu_i(t)\right) dt = 0$$

by weak convergence we directly obtain $\int_A \dot{X}_{u_i}(t)dt \longrightarrow \int_A \overline{X}(t)dt$ and $\int_A Bu_i(t)dt \longrightarrow \int_A B\overline{u}(t)dt$. Using continuity of the vector field on the compact $[0,\, T]\times Q$ and invoking dominated convergence theorem: $\int_A f(t,X_{u_i}(t))dt \longrightarrow \int_A f(t,\overline{X}(t))dt$. By taking the limit we obtain:

$$\int_A \left(\dot{\overline{X}}(t) - f(t,\overline{X}(t)) - B\overline{u}(t)\right) dt = 0.$$

Hence, we have indeed demonstrate $\overline{u} \in L^2([0,\, T]\,,\mathbb{R}^d)$ and

$$\begin{cases} \dot{\overline{X}}(t) = f(t,\overline{X}(t)) + B\overline{u}(t) \,\text{a.e on}\, [0,\, T] \\ \overline{X}(0) = x_0 \end{cases}$$

which finishes the proof. □

To prove the existence of solutions for the optimal control problem (1.1) defining our estimator, we have considered in the proof a minimizing sequence (X_{u_i},u_i). But for doing so, we need to ensure the controlled ODE (1.3) has indeed solutions, it is the point of the following lemma.

Lemma 4.2. *Let us suppose conditions C1, C2 and, in the presence of functional parameters, that $z_{1,0} \in \Theta_f$. Then there exist admissible processes for the perturbed ODE (1.3) i.e $\exists u \neq 0$ s.t $u \in L^2_a$.*

Proof. In the first part of the lemma we will assume there is no functional parameter and $B = I_d$ i.e the perturbated ODE is simply $X = f(t,X) + u$.

We assume that no admissible process u exists, i.e we can not find a control u defined on $[0,\, T]$, with $\|u\|_{L^1} \neq 0$ s.t the corresponding solution X_u exists on $[0,T]$. Then each solution X_u value must leave every compact in finite time (Lemma 2.9 in [9]). Let us consider a compact C' which strictly contains X value. We define $\delta' > 0$ such that for a given x if $\exists t \in [0,\, T]$ $\|x - X(t)\|_2 < \delta' \Longrightarrow x \in C'$. Let us consider X_u the solution corresponding to a control u such that $0 < \|u\|_{L^2} < \frac{\delta'}{2e^{T\overline{f_x}}}$. Let us define $t' \in [0,\, T]$ the time the solution X_u leaves C' we have:

$$\begin{aligned} \left\|\dot{X}_u(t) - \dot{X}(t)\right\|_2 &\leq \|f(t,X_u(t)) - f(t,X(t))\|_2 + \|u(t)\|_2 \\ &\leq \overline{f_x}\,\|X_u(t) - X(t)\|_2 + \|u(t)\|_2 \end{aligned}$$

Gronwall's Lemma implies that

$$\begin{array}{rcl} \|X_u(t') - X(t')\|_2 & \leq & \int_0^{t'} e^{\overline{f_x}(t'-s)} \|u(s)\|_2 \, ds \\ & \leq & e^{T\overline{f_x}} \int_0^{t'} \|u(s)\|_2 \, ds \\ & \leq & \frac{\delta'}{2} \end{array}$$

So $X_u(t')$ is strictly contained in C', hence the contradiction.

Now let us deal with presence of functional parameters, we have to consider the extended ODE:

$$\left\{ \begin{array}{rcl} \dot{x} & = & f(t, x, z_1) \\ \dot{z}_1 & = & z_2 \\ \dot{z}_2 & = & v_2 \end{array} \right.$$

Here the autonomous linear subsystem:

$$\left\{ \begin{array}{rcl} \dot{z}_1 & = & z_2 \\ \dot{z}_2 & = & v_2 \end{array} \right.$$

fulfills the Kalman condition: according to [8] (Chapter 3, Theorem 3 p 89), this system is controllable at any time t, from any initial conditions. Hence, starting from an initial condition $z_{1,0} \in \Theta_f$ we can find controls (v_1, v_2) such that the resulting z_1 gives a solution of $\dot{x} = f(t, x, z_1)$ defined on $[0,\, T]$. Then, we can apply the first part of the lemma to conclude. ☐

5 Sufficiency of the Pontryagin maximum principle

The PMP only gives necessary conditions satisfied by optimal processes $(X_{\theta,\overline{u}},\, \overline{u}_\theta)$. In order to turn these conditions into sufficient ones, we need to impose a lower bound condition on λ.

1. A first lower bound for local optimality is derived by using the quadratic conditions devel-opped by Milyutin and Osmolovskii [5].
2. A second lower bound for global optimality is derived by finding a condition on λ ensuring uniqueness of the solution of BVP.

5.1 Local optimality of controls respecting the PMP

Theorem 5.1. *Let a process $(X_{\theta,\overline{u}}, \overline{u})$ be an admissible process respecting the PMP presented in theorem 3.2. If C1, C2bis and C3bis are satisfied, then for λ such that:*

$$\lambda > \frac{d^{\frac{3}{2}}}{2} \left\| t \longmapsto \frac{\partial^2}{\partial^2 x} \left(p_\theta^T(t) f(t, X_{\theta,\overline{u}}(t)) \right) \right\|_{L^2} \int_0^T \left(\int_0^t \|R_\theta(s) B\|_2^2 \, ds \right) dt$$

or:

$$\lambda > \overline{\lambda_1}(Y) \overline{f_{xx}}, \tag{5.1}$$

then $(X_{\theta,\overline{u}}, \overline{u})$ is a strong local minimum for the problem (1.1).

Proof. We use theorem 11.1 in [5] proving, by the means of the so-called quadratic conditions, sufficiency of the Pontryagin maximum principle for obtaining a bounded strong minimum. We use the same formalism and we consider the alternative cost, $J(x_0, y_0, x_f, y_f) = k_1 y_f$, the extended

ODE:

$$\begin{aligned}&\dot{x} = f(t,x) + Bu\\&\dot{y} = \Lambda_\lambda(Y,t,x,u)\\&(x(0),y(0)) = (x_0,y_0)\end{aligned}$$

and the constraint $K(x_0,y_0,x_f,y_f) = \begin{pmatrix} k_2\,(x_0 - x_0^*) \\ k_3 y_0 \end{pmatrix} = 0$. Here $k_1 > 0$ and $k_3 > 0$, k_2 depending on k_1 and k_3 (its expression is given by 5.2). Obviously any minimizer $(X_{\overline{u}}, \overline{y}, \overline{u})$ of the alternative cost is a minimizer of our original cost and reciprocally. We will show for λ large enough, the Pontryagin maximum principle is a sufficient condition for optimality for the alternative cost and hence for our original cost.

We define $\overline{l}(\alpha_0,\beta,\nu) = \alpha_0 J(\nu) + \beta^T K(\nu) = \alpha_0 k_1 y_f + \beta_1^T k_2\,(x_0 - x_0^*) + \beta_2 k_3 y_0$ with $\nu = (x_0,y_0,x_f,y_f)$, hence $\frac{\partial \overline{l}}{\partial v}(\alpha_0,\beta,\nu) = \begin{pmatrix} k_2\beta_1 & k_3\beta_2 & 0 & k_1\alpha_0 \end{pmatrix}$, we also introduce the extended Pontryagin function $\overline{H_\lambda}(p,p_1,w,t) = p^T(f(t,x)+Bu) + p_1\Lambda_\lambda(Y,t,x,u)$, with here $w = (x,y,u)$. With our regularity hypothesis, our cost and extended ODE vector field are twice differentiable as required by [5]. As in Milyutin et al., for a given process $(X_{\overline{u}}, \overline{y}, \overline{u})$, we denote M_0 the set of t-uples $\eta = (\alpha_0,\beta_1,\beta_2,p,p_1)$ verifying

$$\begin{cases} \alpha_0 \geq 0,\ \alpha_0 + |\beta_1| + |\beta_2| = 1 \\ \dot{p} = -\frac{\partial f}{\partial x}(t, X_{\overline{u}}(t))^T p - p_1 2C^T\left(CX_{\overline{u}}(t) - Y(t)\right) \\ \dot{p_1} = 0 \\ (p(0),\ p_1(0)) = (k_2\beta_1,\ k_3\beta_2) \\ (p(T),\ p_1(T)) = (0,\ -k_1\alpha_0) \\ \frac{\partial \overline{H}_\lambda}{\partial u}(p(t),p_1,X_{\overline{u}}(t),\overline{y}(t),\overline{u}(t),t) = 0 \\ \forall t \in [0,\,T]\,,\ \max_u \overline{H}_\lambda(p(t),p_1(t),X_{\overline{u}}(t),\overline{y}(t),u,t) = \overline{H}_\lambda(p(t),p_1(t),X_{\overline{u}}(t),\overline{y}(t),\overline{u}(t),t) \end{cases} . \tag{5.2}$$

Thanks to lemma 5.2, we know the set M_0 is non empty if and only if $(X_{\overline{u}}, \overline{u})$ respects the Pontryagin Maximum Principle version presented in theorem 3.2. Moreover, we have an expression for a tuple $\overline{\eta} = \left(\overline{\alpha_0}, \overline{\beta_1}, \overline{\beta_2}, \overline{p}, \overline{p_1}\right)$ respecting (5.2), it is under the form $\overline{\eta} = \left(\frac{1}{k_1}, \overline{\beta_1}, -\frac{1}{k_3}, \overline{p}, -1\right)$ where $\overline{\beta_1}$ the vector with each component equal to $\frac{1}{d}\left(1 - \frac{1}{k_1} - \frac{1}{k_3}\right)$ and $\overline{p}$ such that $(X_{\overline{u}},\ \overline{p})$ solution of the BVP presented in theorem 3.2

We now introduce the Quadratic Form:

$$\Omega_\lambda(\eta,\widetilde{w}) = -\int_0^T \widetilde{w}(t)^T \frac{\partial^2 \overline{H_\lambda}(p(t),p_1,X_{\overline{u}}(t),\overline{y}(t),\overline{u}(t),t)}{\partial^2 w}\widetilde{w}(t)dt$$

with $\widetilde{w} = (\widetilde{x},\widetilde{y},\widetilde{u})$ belonging to the critical cone $\mathcal{K}$ i.e, the points verifying:

$$\begin{matrix} \frac{\partial J}{\partial \nu}(\overline{\nu})\widetilde{\nu} \leq 0 \\ \frac{\partial K}{\partial \nu}(\overline{\nu})\widetilde{\nu} = 0 \end{matrix} \iff \begin{cases} \widetilde{y_f} \leq 0 \\ \widetilde{x_0} = 0 \\ \widetilde{y_0} = 0 \end{cases}$$

and the linear ODE:

$$\begin{aligned} \dot{\widetilde{x}} &= \frac{\partial f}{\partial x}(t, X_{\overline{u}}(t))\widetilde{x} + B\widetilde{u} \\ \dot{\widetilde{y}} &= \frac{\partial \Lambda_\lambda}{\partial x}(t, X_{\overline{u}}(t), \overline{u}(t))\widetilde{x} + \frac{\partial \Lambda_\lambda}{\partial u}(t, X_{\overline{u}}(t), \overline{u}(t))\widetilde{u} \end{aligned}$$

In our case, $\frac{\partial^2 \overline{H_\lambda}(p,p_1,w,t)}{\partial^2 w}$ is a sparse matrix:

$$\frac{\partial^2 \overline{H_\lambda}(p,p_1,w,t)}{\partial^2 w} = \begin{pmatrix} \frac{\partial^2}{\partial^2 x}\left(p^T f(t,x)\right) + 2p_1 C^T C & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2p_1 \lambda I_d \end{pmatrix}$$

so the Quadratic Form has a simpler expression:

$$\Omega_\lambda(\eta,\widetilde{w}) = -\int_0^T \widetilde{x}(t)^T \left(\frac{\partial^2}{\partial^2 x}\left(p^T f(t,X_{\overline{u}}(t))\right) + 2p_1 C^T C\right)\widetilde{x}(t)dt - 2\lambda p_1 \|\widetilde{u}\|_{L^2}^2$$

From this Quadratic Form, Milyutin et al. have expressed necessary and sufficient conditions for the given process $(X_{\overline{u}},\overline{y},\overline{u})$ to be a Pontryagin minimum.

Condition A: The set M_0 is nonempty and $\forall \widetilde{w} \in \mathcal{K}$, $\max_{\eta\in M_0}\Omega_\lambda(\eta,\widetilde{w}) \geq 0$.

According to theorem 10.1 in [5], condition A is a necessary condition for a Pontryagin minimum. But if we strengthened condition A, we can turn it into a sufficient condition for $(X_{\overline{u}},\overline{y},\overline{u})$ to be a bounded strong minimum.

We denote $Leg_+(M_0^+)$ the subset of M_0 respecting the additional conditions:

1) Strict maximum:

$$\forall t \in [0,\,T]\,, \forall u \neq \overline{u}(t),\ \overline{H_\lambda}(p(t),p_1,X_{\overline{u}}(t),\overline{y}(t),u,t) < \overline{H}_\lambda(p(t),p_1,X_{\overline{u}}(t),\overline{y}(t),\overline{u}(t),t).$$

2) Strengthened Legendre-Klebch condition: $\forall t \in [0,\,T]\,,\ \frac{\partial^2 \overline{H_\lambda}(p,p_1,X_{\overline{u}}(t),\overline{y}(t),\overline{u}(t),t)}{\partial^2 u}$ is negative definite.

Condition B: The set $Leg_+(M_0^+)$ is nonempty and it exists a nonempty compact set $M \subset Leg_+(M_0^+)$ and a constant $\epsilon > 0$ such that $\forall \widetilde{w} \in \mathcal{K}$, $\max_{\eta\in M}\Omega_\lambda(\eta,\widetilde{w}) \geq \epsilon \|\widetilde{u}\|_{L^2}^2$.

According to theorem 10.2 in [5], if condition B is fullfilled for a trajectory $X_{\overline{u}}$ then it is a bounded strong minimum. Since $\frac{\partial^2 \overline{H_\lambda}(p,p_1,X_{\overline{u}}(t),\overline{y}(t),\overline{u}(t),t)}{\partial^2 u} = 2p_1\lambda I_d$, choosing in M_0 the tuple $\overline{\eta}$ gives us $p_1 = -1$ and $\frac{\partial^2 \overline{H_\lambda}(p,p_1,X_{\overline{u}}(t),\overline{y}(t),\overline{u}(t),t)}{\partial^2 u}$ is negative definite in that case. Hence $\overline{\eta} \in Leg_+(M_0^+)$ and we choose $M = \{\overline{\eta}\}$.

Thus we have to find a lower bound under the form $\epsilon \|\widetilde{u}\|_{L^2}^2$for $\Omega_U(\overline{\eta},\widetilde{w})$. For $p_1 = -1$ we have:

$$\Omega_\lambda(\eta,\widetilde{w}) > 2\lambda \|\widetilde{u}\|_{L^2}^2 - \int_0^T \widetilde{x}(t)^T \frac{\partial^2}{\partial^2 x}\left(p_\theta^T(t) f(t,X_{\theta,\overline{u}}(t))\right)\widetilde{x}(t)dt \tag{5.3}$$

By using Cauchy-Schwarz and norm inequality, we have:

$$\begin{aligned} \int_0^T \widetilde{x}(t)^T \tfrac{\partial^2}{\partial^2 x}\left(p_\theta^T(t) f(t,X_{\theta,\overline{u}}(t))\right)\widetilde{x}(t)dt &\leq \|\widetilde{x}\|_{L^2}\left\|t \longmapsto \tfrac{\partial^2}{\partial^2 x}\left(p_\theta^T(t) f(t,X_{\theta,\overline{u}}(t))\right)\widetilde{x}(t)\right\|_{L^2} \\ &\leq \sqrt{d}\left\|t \longmapsto \tfrac{\partial^2}{\partial^2 x}\left(p_\theta^T(t) f(t,X_{\theta,\overline{u}}(t))\right)\right\|_{L^2}\|\widetilde{x}\|_{L^2}^2 \end{aligned}$$

All we have left to do is to control $\|\widetilde{x}\|_{L^2}^2$ w.r.t $\|\widetilde{u}\|_{L^2}$. For this, we recall that $\widetilde{x}$ follows a linear ODE, so we can use Duhamel's formula to obtain:

$$\begin{aligned} \widetilde{x}(t) &= \int_0^t R_\theta(s) B\widetilde{u}(s)ds \\ \dot{R}_\theta(t) &= \tfrac{\partial f}{\partial x}(t,X_{\overline{u}}(t))R_\theta(t) \\ R_\theta(0) &= I_d \end{aligned}$$

so:

$$\begin{array}{rcl}\|\widetilde{x}(t)\|_2 & \leq & \sqrt{d}\int_0^t \|R_\theta(s)B\|_2 \|\widetilde{u}(s)\|_2\, ds \\ & \leq & \sqrt{d}\sqrt{\int_0^t \|R_\theta(s)B\|_2^2\, ds}\, \|\widetilde{u}\|_{L^2}\end{array}$$

Taking the L^2-norm gives us: $\|\widetilde{x}\|_{L^2}^2 \leq d\int_0^T \left(\int_0^t \|R_\theta(s)B\|_2^2\, ds\right) dt\, \|\widetilde{u}\|_{L^2}^2$ and finally we obtain the bound:

$$\Omega_\lambda(\eta,\widetilde{w}) > \left(2\lambda - d^{\frac{3}{2}}\left\|t \longmapsto \frac{\partial^2}{\partial^2 x}\left(p_\theta^T(t)f(t,X_{\theta,\overline{u}}(t))\right)\right\|_{L^2}\int_0^T\left(\int_0^t \|R_\theta(s)B\|_2^2\, ds\right)dt\right)\|\widetilde{u}\|_{L^2}^2$$

which gives a tractable and computable in practice a posteriori criteria.

However, if we want a theoretical a priori upper bound for λ (i.e which do not depends on the found control $\overline{u}$), we need lemma lemma 7.4 which gives the uniform upper bound $\|p_\theta(t,Y)\|_2 \leq E(Y)$, to obtain:

$$\int_0^T \widetilde{x}(t)^T p(t)^T \frac{\partial^2 f}{\partial^2 x}(t,X_{\overline{u}}(t))\widetilde{x}(t)dt \leq \sqrt{d}E(Y)\overline{f_{xx}}\,\|\widetilde{x}\|_{L^2}^2 \tag{5.4}$$

and the Gronwall's lemma to have the bound: $\|\widetilde{x}\|_{L^2}^2 \leq T\|B\|_2^2 \frac{e^{2\sqrt{d}\overline{f_x}T}}{2\overline{f_x}}\|\widetilde{u}\|_{L^2}^2$. By reinjecting these inequalities in (5.3), we can find a lower bound for $\Omega_U(\overline{\eta},\widetilde{w})$ only expressed in terms of the controls:

$$\Omega_U(\overline{\eta},\widetilde{w}) \quad \geq \quad \left(2\lambda - E(Y)\sqrt{d}T\|B\|_2^2 \frac{e^{2\sqrt{d}\overline{f_x}T}}{2\overline{f_x}}\overline{f_{xx}}\right)\|\widetilde{u}\|_{L^2}^2$$

Hence taking λ larger than the values given by (5.1) ensures the existence of a constant $\epsilon > 0$ such that $\forall \widetilde{w} \in \mathcal{K}$, $\Omega_\lambda(\overline{\eta},\widetilde{w}) \geq \epsilon\|\widetilde{u}\|_{L^2}^2$, and so $(X_{\overline{u}},\overline{y},\overline{u})$ is a bounded strong minimum.

We now use theorem 9.4 in [5] to show that bounded strong minimum is a strong one for our cost. For this, we need to fullfill condition $\mathfrak{B}_H^\infty$ presented in page 275 in [5]: it exists $\epsilon > 0$, $\rho > 0$ such that for all $t \in [0, T]$ and $x \in \mathbb{R}^d$ such that $\|x - X_{\overline{u}}(t)\|_2 < \epsilon$ imply $-\overline{H_\lambda}(p,-1,x,y,u,t) \geq -\overline{H_\lambda}(p,-1,x,y,\overline{u}(t),t) + \rho\|u - \overline{u}(t)\|_2$.

Here $\overline{H_\lambda}(p,-1,x,y,\overline{u}(t),t) = p^T(f(t,x) + B\overline{u}(t)) - \Lambda_\lambda(Y,t,x,\overline{u}(t))$ and we have already shown that $\forall t \in [0,T]$, $x \in \mathbb{R}^d$, we have $\frac{\partial \overline{H_\lambda}}{\partial u}(p,-1,x,y,\overline{u}(t),t) = 0$ and $\frac{\partial^2 \overline{H_\lambda}}{\partial^2 u}(p,-1,x,y,\overline{u}(t),t) = -2\lambda I_d$. We can conclude using theorem 9.4 in [5]. □

Lemma 5.2. *$(X_{\overline{u}},\overline{y},\overline{u})$ respects the simplified Pontryagin maximum principle presented in theorem 3.2 if and only if it respects Pontryagin Maximum Principle presented in Milyutin et al. for the alternative cost with constraint:*

$$\begin{array}{l}J(x_0,y_0,x_f,y_f) = k_1 y_f \\ K(x_0,y_0,x_f,y_f) = \begin{pmatrix} k_2\,(x_0 - x_0^*) \\ k_3 y_0\end{pmatrix} = 0 \\ \dot{x} = f(t,x) + Bu \\ \dot{y} = \Lambda_\lambda(t,x,u) \\ (x(0),y(0)) = (x_0,y_0)\end{array}$$

where $k_2 = \frac{d\overline{p}(0)}{1-\frac{1}{k_1}-\frac{1}{k_3}}$ with $\overline{p}$ such that $(X_{\overline{u}},\overline{p})$ is the solution of the extended ODE with boundary constraint:

$$\begin{array}{l}\dot{X}_{\overline{u}}(t) = f(t,X_{\overline{u}}(t)) + \frac{1}{2\lambda}BB^T\overline{p}(t) \\ \dot{\overline{p}}(t) = -\frac{\partial f}{\partial x}(t,X_{\overline{u}}(t))^T\overline{p}(t) + 2C^T\left(CX_{\overline{u}}(t) - Y(t)\right) \\ (X_{\overline{u}}(0),\overline{p}(T)) = (x_0^*,0)\end{array} \tag{5.5}$$

Moreover we have an expression for a tuple $\overline{\eta} \in M_0$ under the form: $\overline{\eta} = \left(\frac{1}{k_1}, \overline{\beta_1}, -\frac{1}{k_3}, \overline{p}, -1\right)$ with $\overline{\beta_1}$ the vector with each component equal to $\frac{1}{d}\left(1 - \frac{1}{k_1} - \frac{1}{k_3}\right)$.

Proof. We consider $\overline{w} = (X_{\overline{u}}, \overline{y}, \overline{u})$ respecting our simplified Pontryagin maximum principle, i.e the optimal control is equal to $\overline{u}(t) = \frac{1}{2\lambda} B^T \overline{p}(t)$, such $(X_{\overline{u}}, \overline{p})$ is the solution of the extended ODE with boundary constraint 5.5. Again we denote M_0 the set of tuples $\eta = (\alpha_0, \beta_1, \beta_2, p_e, p_1)$ verifying (5.2). The Pontryagin maximum principle presented in Milyutin for the alternative cost is equivalent to have M_0 nonempty. Starting from the simplified maximum principle let us contruct a tuple η respecting (5.2).

Firstly let us choose $\alpha_0 = \frac{1}{k_1}$ obviously $\alpha_0 > 0$ and p_1 beeing constant it imposes $p_1(t) = -1 = k_3\beta_2$ so we take $\beta_2 = -\frac{1}{k_3}$. The adjoint equation become:

$$\begin{aligned}
&\dot{p}(t) = -\tfrac{\partial f}{\partial x}(t, X_{\overline{u}}(t))^T p(t) + 2C^T \left(CX_{\overline{u}}(t) - Y(t)\right) \\
&p(T) = 0
\end{aligned}$$

which is indeed respected by $\overline{p}(t)$. The non triviality condition $\alpha_0 + |\beta_1| + |\beta_2| = 1$ imposes $|\beta_1| = 1 - \frac{1}{k_1} - \frac{1}{k_3}$. We can choose for β_1 the vector with each component equal to $\frac{1}{d}\left(1 - \frac{1}{k_1} - \frac{1}{k_3}\right)$. In order to respect the constraint $\overline{p}(0) = k_2\beta_1$ we have to set $k_2 = \frac{d\overline{p}(0)}{1 - \frac{1}{k_1} - \frac{1}{k_3}}$. Since $\overline{H_\lambda}(p, p_1, w, t) = p^T(f(t,x) + Bu) + p_1\Lambda_\lambda(t, x, u)$ is the same for the two problems, the maximality constraints:

$$\begin{aligned}
&\tfrac{\partial \overline{H_\lambda}}{\partial u}(\overline{p}, -1, \overline{w}, t) = 0 \\
&\max_u \overline{H_\lambda}(\overline{p}(t), -1, X_{\overline{u}}(t), \overline{y}(t), u, t) = \overline{H_\lambda}(\overline{p}(t), -1, X_{\overline{u}}(t), \overline{y}(t), \overline{u}(t), t)
\end{aligned}$$

are already fullfillled. Hence the tuple: $\overline{\eta} = \left(\frac{1}{k_1}, \overline{\beta_1}, -\frac{1}{k_3}, \overline{p}, -1\right)$ with $\overline{\beta_1}$ the vector with each component equal to $\frac{1}{d}\left(1 - \frac{1}{k_1} - \frac{1}{k_3}\right)$ belongs to M_0.

Reciprocal is obtained by substituting $\overline{\eta}$ in (5.2). □

5.2 Uniqueness of BVP solutions

We need to introduce the reversed time solutions $p^i_\theta(t) = p_\theta(T - t)$ and $X^i_{\theta,\overline{u}}(t) = X_{\theta,\overline{u}}(T - t)$.

Theorem 5.3. *If C1, C2bis are satisfied, then for $\lambda > \lambda_2(Y)$, we have uniqueness of the solution of*

$$\begin{aligned}
&\dot{X}_{\theta,\overline{u}}(t) = f(t, X_{\theta,\overline{u}}(t), \theta) + \tfrac{1}{2\lambda} BB^T p_\theta(t) \\
&\dot{p}_\theta(t) = -\tfrac{\partial f}{\partial x}(t, X_{\theta,\overline{u}}(t), \theta)^T p_\theta(t) + 2C^T \left(CX_{\theta,\overline{u}}(t) - Y(t)\right) \\
&(X_{\theta,\overline{u}}(0), p_\theta(T)) = (x_0, 0)
\end{aligned} \tag{5.6}$$

Proof. Let us define $\left(X^1, p^1\right)$ and $\left(X^2, p^2\right)$ two solutions of:

$$\begin{aligned}
&\dot{X}(t) = f(t, X(t)) + \tfrac{1}{2\lambda} BB^T p(t) \\
&\dot{p}(t) = -\tfrac{\partial f}{\partial x}(t, X(t))^T p(t) + 2C^T \left(CX(t) - Y(t)\right) \\
&(X(0), p(T)) = (x_0, 0)
\end{aligned}$$

First let us control $\left\|X^1 - X^2\right\|_{L^2}$ w.r.t $\left\|p^1 - p^2\right\|_{L^2}$, we have:

$$\dot{X^1}(t) - \dot{X^2}(t) = f(t, X^1(t)) - f(t, X^2(t)) + \frac{1}{2\lambda} BB^T \left(p^1(t) - p^2(t)\right)$$

Taking the Frobenius norm gives us:

$$\begin{array}{rcl} \frac{d}{dt}\left\|X^1(t)-X^2(t)\right\|_2 & \leq & \left\|f(t,X^1(t))-f(t,X^2(t))\right\|_2 \\ & + & \frac{d\|B\|_2^2}{2\lambda}\left\|p^1(t)-p^2(t)\right\|_2 \\ & \leq & \overline{f_x}\left\|X^1(t)-X^2(t)\right\|_2+\frac{d\|B\|_2^2}{2\lambda}\left\|p^1(t)-p^2(t)\right\|_2 \end{array}$$

Using Gronwall's lemma we obtain:

$$\left\|X^1(t)-X^2(t)\right\|_2 \leq \frac{d\left\|B\right\|_2^2}{2\lambda}\int_0^t e^{\overline{f_x}(t-s)}\left\|p^1(s)-p^2(s)\right\|_2 ds$$

and Cauchy-Schwarz inequality gives us:

$$\begin{array}{rcl} \left\|X^1(t)-X^2(t)\right\|_2 & \leq & \frac{d\|B\|_2^2}{2\lambda}\sqrt{\int_0^t e^{2\overline{f_x}(t-s)}ds}\sqrt{\int_0^t \|p^1(s)-p^2(s)\|_2^2\, ds} \\ & \leq & \frac{d\|B\|_2^2}{2\lambda}\sqrt{\frac{e^{2\overline{f_x}t}-1}{2\overline{f_x}}}\left\|p^1-p^2\right\|_{L^2} \end{array}$$

We finally obtain the following upper bound for $\left\|X^1-X^2\right\|_{L^2}$:

$$\left\|X^1-X^2\right\|_{L^2} \leq \frac{d\left\|B\right\|_2^2}{2\lambda}\frac{\sqrt{e^{2\overline{f_x}T}-2\overline{f_x}T-1}}{2\overline{f_x}}\left\|p^1-p^2\right\|_{L^2} \tag{5.7}$$

Now we have to bound $\left\|p^1-p^2\right\|_{L^2}$ w.r.t $\left\|X^1-X^2\right\|_{L^2}$ for the reversed time solutions we have the differential relation:

$$\begin{array}{rcl} p^{\dot{1},i}(t)-p^{\dot{2},i}(t) & = & -\left(\frac{\partial f}{\partial x}(T-t,X^{1,i}(t))-\frac{\partial f}{\partial x}(T-t,X^{2,i}(t))\right)^T p^{1,i}(t) \\ & & -\frac{\partial f}{\partial x}(T-t,X^{2,i}(t))^T\left(p^{1,i}(t)-p^{2,i}(t)\right)+2C^TC\left(X^{2,i}(t)-X^{1,i}(t)\right) \end{array} \tag{5.8}$$

Taking the Frobenius norm we have the following inequality:

$$\begin{array}{rcl} \frac{d}{dt}\left\|p^{2,i}(t)-p^{1,i}(t)\right\|_2 & \leq & \sqrt{d}\left\|\frac{\partial f}{\partial x}(T-t,X^{2,i}(t))\right\|_2\left\|p^{2,i}(t)-p^{1,i}(t)\right\|_2 \\ & + & 2\sqrt{d}\left\|C^TC\right\|_2\left\|X^{2,i}(t)-X^{1,i}(t)\right\|_2 \\ & + & \sqrt{d}\left\|p^{1,i}(t)\right\|_2\left\|\frac{\partial f}{\partial x}(T-t,X^{1,i}(t))-\frac{\partial f}{\partial x}(T-t,X^{2,i}(t))\right\|_2 \\ & \leq & \sqrt{d\overline{f_x}}\left\|p^{2,i}(t)-p^{1,i}(t)\right\|_2 \\ & + & \sqrt{d}\left(\overline{f_{xx}}\left\|p^{1,i}(t)\right\|_2+2\left\|C^TC\right\|_2\right)\left\|X^{2,i}(t)-X^{1,i}(t)\right\|_2 \end{array}$$

Using Gronwall's lemma we obtain:

$$\begin{array}{rcl} \left\|p^{2,i}(t)-p^{1,i}(t)\right\|_2 & \leq & \sqrt{d}\int_0^t e^{\sqrt{d\overline{f_x}}(t-s)}\left(\overline{f_{xx}}\left\|p^{1,i}(t)\right\|_2+2\left\|C^TC\right\|_2\right)\left\|X^{2,i}(s)-X^{1,i}(s)\right\|_2 ds \\ & \leq & \sqrt{d}\left(\overline{f_{xx}}E(Y)+2\left\|C^TC\right\|_2\right)\int_0^t e^{\sqrt{d\overline{f_x}}(t-s)}\left\|X^{2,i}(s)-X^{1,i}(s)\right\|_2 ds \end{array}$$

Cauchy-Schwarz inequality gives us:

$$\begin{array}{rcl} \left\|p^{2,i}(t)-p^{1,i}(t)\right\|_2 & \leq & \sqrt{d}\left(\overline{f_{xx}}E(Y)+2\left\|C^TC\right\|_2\right)\sqrt{\int_0^t e^{2\sqrt{d\overline{f_x}}(t-s)}ds}\sqrt{\int_0^t\|X^{2,i}(s)-X^{1,i}(s)\|_2^2\, ds} \\ & \leq & \sqrt{d}\sqrt{\frac{e^{2\sqrt{d\overline{f_x}}t}-1}{2\sqrt{d\overline{f_x}}}}\left(\overline{f_{xx}}E(Y)+2\left\|C^TC\right\|_2\right)\left\|X^1-X^2\right\|_{L^2} \end{array}$$

Taking the L^2 norm finally gives us the desired upper bound:

$$\left\|p^1-p^2\right\|_{L^2} \quad \leq \quad \sqrt{e^{2\sqrt{d\overline{f_x}}T}-2\sqrt{d\overline{f_x}}T-1}\,\frac{\left(\overline{f_{xx}}E(Y)+2\left\|C^TC\right\|_2\right)}{2\overline{f_x}}\left\|X^1-X^2\right\|_{L^2} \tag{5.9}$$

By plugging (5.7) into this inequality, we get

$$\left\|p^1-p^2\right\|_{L^2} \leq \frac{d\|B\|_2^2}{\lambda}\left(\frac{\overline{f_{xx}}E(Y)+2\left\|C^TC\right\|_2}{8\overline{f_x}^2}\right)\sqrt{\left(e^{2\sqrt{d\overline{f_x}}T}-1-2\sqrt{d\overline{f_x}}T\right)\left(e^{2\overline{f_x}T}-1-2\overline{f_x}T\right)}\left\|p^1-p^2\right\|_{L^2}$$

Hence the solution uniqueness follows if $\lambda > \lambda_2(Y)$. □

6 Consistency

$X_{\theta,u}$ and p_θ depend on the data through Y, to emphasize they are functions of Y, we sometimes denote $X_{\theta,u}(t,Y)$ and $p_\theta(t,Y)$ for the controlled trajectory and adjoint variable for a given θ and Y. For the sake of simplicity, we denote $\hat{X}_{\theta,u}(t) := X_{\theta,u}(t,\widehat{Y})$, $\hat{p}_\theta(t) := p_\theta(t,\widehat{Y})$ and $X^*_{\theta,u}(t) := X_{\theta,u}(t,Y^*)$, $p^*_\theta(t) := p_\theta(t,Y^*)$. Since we have defined an M-estimator, we need to prove:

1. $\mathcal{S}(Y^*;\theta,\lambda)$ has a global well-separated minimum at $\theta=\theta^*$ (proposition 6.1),
2. uniform convergence of $\mathcal{S}(\widehat{Y};\theta,\lambda)$ toward $\mathcal{S}(Y^*;\theta,\lambda)$ (proposition 6.2),

to ensure consistency of $\widehat{\theta}^T_\lambda$ by using the theorem 5.7 in [10].

Proposition 6.1. *If C1, C2bis,C3 and C4 are satisfied, then for all $\lambda > F(Y^*)$ we have:*

$$\mathcal{S}(Y^*;\theta,\lambda)=0 \Longleftrightarrow \theta=\theta^*$$

Proof. In that case our asymptotic criteria become: $\mathcal{C}(Y^*;u,\theta,\lambda)=\int_0^T \|CX_{\theta,u}(t)-CX^*(t)\|_2^2\,dt+\lambda\|u\|_{L^2}^2$. If $\theta=\theta^*$ then the process $(X^*_{\theta^*,u}, u\equiv 0)$ ensures $\mathcal{C}(Y^*;u,\theta,\lambda)=0$, is solution of the asymptotic ODE with boundary value and is unique thanks to lemma 7.5, since $\lambda > F(Y^*)$, hence $\mathcal{S}(Y^*;\theta^*,\lambda)=0$.

We denote θ^0 s.t $\mathcal{S}(Y^*;\theta^0,\lambda)=0$ since we have reached the minimum, we have by definition of $\mathcal{S}(Y^*;\theta,\lambda)$ that the related optimal control $\overline{u}$ is equal to the null function. So $X^*_{\theta^0,\overline{u}}$ is solution of the unpertubated ODE, thus we have$\int_0^T\left\|CX^*_{\theta^0,\overline{u}}(t)-CX^*(t)\right\|_2^2 dt=0$, identifiability condition imposes $\theta^0=\theta^*$. □

Proposition 6.2. *If C1 and C2bis are satisfied, then for all $\lambda > F(Y^*)$ we have:*

$$\begin{aligned}\left|\mathcal{S}(Y^*;\theta,\lambda)-\mathcal{S}(\widehat{Y};\theta,\lambda)\right| &\leq \left(D(Y^*)+D(\widehat{Y})\right)\left(\frac{M_1\|C\|_2}{\lambda-F(Y^*)}+1\right)\left\|Y^*-\widehat{Y}\right\|_{L^2}\\ &+ \frac{1}{2\lambda}\left(E(Y^*)+E(\widehat{Y})\right)\left(\frac{M_2}{\lambda-F(Y^*)}+M_3\right)\left\|Y^*-\widehat{Y}\right\|_{L^2}\end{aligned}$$

$$\begin{aligned} & M_1 = d\|B\|_2^2 T^2 e^{2\sqrt{d\overline{f_x}}T}\|C\|_2\\ \textit{with}\quad & M_2 = T^2\|B\|_2^2 d^{\frac{3}{2}} e^{\sqrt{d\overline{f_x}}T} M_1\left(\overline{f_{xx}}E(Y^*)+2\|C\|_2^2\right)\\ & M_3 = 2T^2\|B\|_2^2 d^{\frac{3}{2}} e^{\sqrt{d\overline{f_x}}T}\|C\|_2\end{aligned}$$

Proof. We have using Cauchy-Schwarz inequality:

$$\begin{aligned}\left|\mathcal{S}(Y^*;\theta,\lambda)-\mathcal{S}(\widehat{Y};\theta,\lambda)\right| &\leq \int_0^T\left|\left(CX_{\theta,\overline{u}}(t)-CX^*_{\theta,\overline{u}}(t)+Y^*(t)-\widehat{Y}(t)\right)^T\left(CX_{\theta,\overline{u}}(t)-\widehat{Y}(t)\right)\right|dt\\ &+ \int_0^T\left|\left(CX^*_{\theta,\overline{u}}(t)-Y^*(t)\right)^T\left(CX_{\theta,\overline{u}}(t)-CX^*_{\theta,\overline{u}}(t)+Y^*(t)-\widehat{Y}(t)\right)\right|dt\\ &+ \frac{d\|B\|_2^2}{2\lambda}\int_0^T\|p_\theta(t)+p^*_\theta(t)\|_2\|p_\theta(t)-p^*_\theta(t)\|_2\,dt\\ &\leq \|C\|_2\left(\left\|CX_{\theta,\overline{u}}-\widehat{Y}\right\|_{L^2}+\left\|CX^*_{\theta,\overline{u}}-Y^*\right\|_{L^2}\right)\left\|X_{\theta,\overline{u}}-X^*_{\theta,\overline{u}}\right\|_{L^2}\\ &+ \left(\left\|CX_{\theta,\overline{u}}-\widehat{Y}\right\|_{L^2}+\left\|CX^*_{\theta,\overline{u}}-Y^*\right\|_{L^2}\right)\left\|Y^*-\widehat{Y}\right\|_{L^2}\\ &+ \frac{d\|B\|_2^2}{2\lambda}\|p_\theta+p^*_\theta\|_{L^2}\|p_\theta-p^*_\theta\|_{L^2}\end{aligned}$$

By using lemma 7.4, which gives us $\|p_\theta(t,Y)\|_2 \le E(Y)$, we obtain the following upper bound:

$$\begin{array}{rcl} \left|\mathcal{S}(Y^*;\theta,\lambda)-\mathcal{S}(\widehat{Y};\theta,\lambda)\right| & \le & \left(D(Y^*)+D(\widehat{Y})\right)\left(\|C\|_2\left\|X_{\theta,\overline{u}}-X^*_{\theta,\overline{u}}\right\|_{L^2}+\left\|Y^*-\widehat{Y}\right\|_{L^2}\right) \\ & + & \frac{dT\|B\|_2^2}{2\lambda}\left(E(Y^*)+E(\widehat{Y})\right)\|p_\theta-p^*_\theta\|_{L^2} \end{array} \tag{6.1}$$

Using lemma 7.5, we know we have for $\lambda > F(Y^*)$,

$$\|X_{\theta,\overline{u}}-X_{\theta,\overline{u}}(.,Y^*)\|_{L^2} \le \frac{M_1}{\lambda-F(Y^*)}\left\|Y^*-\widehat{Y}\right\|_{L^2}$$

and

$$\|p_\theta-p_\theta(.,Y^*)\|_{L^2} \le \sqrt{d}Te^{\sqrt{df_x}T}\left(\frac{(\overline{f_{xx}}E(Y^*)+2\|C\|_2^2)K_1}{\lambda-F(Y^*)}+2\|C\|_2\right)\left\|Y^*-\widehat{Y}\right\|_{L^2}.$$

By reinjecting theses inequalities, we obtain the desired result. □

These two results have allowed to conclude about the consistency:

Theorem 6.3. *If conditions C1, C2bis, C3, C4 are satisfied and* $\widehat{Y}\xrightarrow[n\longrightarrow\infty]{L^2}Y^*$ *in probability, then for any* $\lambda > F(Y^*)$*, we have:*

$$\widehat{\theta}^T_\lambda \xrightarrow{P} \theta^*$$

Proof. Using proposition 6.2 we have the following uniform upper bound on Θ:

$$\begin{array}{rcl} \left|\mathcal{S}(Y^*;\theta,\lambda)-\mathcal{S}(\widehat{Y};\theta,\lambda)\right| & \le & \left(D(Y^*)+D(\widehat{Y})\right)\left(\frac{M_1\|C\|_2}{\lambda-F(Y^*)}+1\right)\left\|Y^*-\widehat{Y}\right\|_{L^2} \\ & + & \frac{1}{2\lambda}\left(E(Y^*)+E(\widehat{Y})\right)\left(\frac{M_2}{\lambda-F(Y^*)}+M_3\right)\left\|Y^*-\widehat{Y}\right\|_{L^2} \end{array}$$

We can conclude that if is $\widehat{Y}$ is consistent we have $sup_{\theta\in\Theta}\left|\mathcal{S}(Y^*;\theta,\lambda)-\mathcal{S}(\widehat{Y};\theta,\lambda)\right| = o_P(1)$.

Application of proposition 6.1 gives us the identifiability criteria. Hence we conclude by using the theorem 5.7 in [10]. □

7 Asymptotics proof

We obtain the asymptotic normality with $\sqrt{n}$-rate of $\widehat{\theta}^T_\lambda$, by proving that, as long as $\widehat{Y}\in B(Y^*,\zeta)$ and $\lambda > \max(\overline{\lambda_3}(\zeta), \overline{\lambda_4}(\zeta))$,

1. $\widehat{\theta}^T_\lambda - \theta^*$ behaves like the difference $\Gamma(\widehat{Y})-\Gamma(Y^*)$, where Γ is a continuous function,
2. $\Gamma(\widehat{Y})-\Gamma(Y^*)$ is asymptotically normal by using the plug-in properties of regression splines.

7.1 Asymptotic representation

Proposition 7.1. *Let* ζ *such that* $Y\in B(Y^*,\zeta)$*. Under conditions 1-6 and* $\lambda > \max(\overline{\lambda_3}(\zeta), \overline{\lambda_4}(\zeta))$*, we have:*

$$\widehat{\theta}^T-\theta^* = -2\frac{\partial^2\mathcal{S}(Y^*;\theta^*,\lambda)}{\partial\theta^T\partial\theta}^{-1}\nabla_\theta\mathcal{S}(\widehat{Y};\theta^*,\lambda)+o_P(1)$$

Proof. For the sake of notation we will simply denote the estimator $\widehat{\theta}$. In the multidimensional case the Hessian expression is cumbersome in matrician form so for the demonstration we will consider $d=1$ but the demonstration stay the same in the multidimensional case. First of all we need to show $\theta\longmapsto\mathcal{S}(Y;\theta,\lambda)$ is C^2 for all $Y\in B(Y^*,\zeta)$ and $(\theta,Y)\longmapsto\frac{\partial^2\mathcal{S}(Y;\theta,\lambda)}{\partial^2\theta}$ is continuous on $\Theta\times B(Y^*,\zeta)$.

According to lemma 7.5 if λ is greater than $\overline{\lambda_3}(\zeta)$ we have:

$$\begin{array}{lcl}
\left\|X_{\theta,\overline{u}}(.,Y)-X_{\theta',\overline{u}}(.,Y')\right\|_{L^2} & \leq & \frac{K_2+\left\|Y'-Y^*\right\|_{L^2}K_6+\lambda K_3}{\lambda-F(Y^*)-K_5\left\|Y'-Y^*\right\|_{L^2}}\left\|\theta-\theta'\right\|_2\\
& + & \frac{\sqrt{T}\left(\sqrt{dT\overline{f_x}}e^{\sqrt{d\overline{f_x}}T}+1\right)}{\lambda-F(Y^*)-\left\|Y'-Y^*\right\|_{L^2}K_5}\left\|x_0-x_0'\right\|_2\\
& + & \frac{K_1}{\lambda-F(Y^*)-\left\|Y'-Y^*\right\|_{L^2}K_5}\left\|Y-Y'\right\|_{L^2}\\
\left\|p_\theta(.,Y)-p_{\theta'}(.,Y')\right\|_{L^2} & \leq & K_7\frac{\left(\overline{f_{xx}}\left(E(Y^*)+K_4\left\|Y'-Y^*\right\|_{L^2}\right)+2\|C\|_2^2\right)\left(K_2+\left\|Y'-Y^*\right\|_{L^2}K_6+\lambda K_3\right)}{\left(\lambda-F(Y^*)-K_5\left\|Y'-Y^*\right\|_{L^2}\right)}\left\|\theta-\theta'\right\|_2\\
& + & K_8\left(E(Y^*)+K_4\left\|Y'-Y^*\right\|_{L^2}\right)\left\|\theta-\theta'\right\|_2\\
& + & \frac{K_9\left(\overline{f_{xx}}\left(E(Y^*)+K_4\left\|Y'-Y^*\right\|_{L^2}\right)+2\|C\|_2^2\right)}{\lambda-F(Y^*)-\left\|Y'-Y^*\right\|_{L^2}K_5}\left\|x_0-x_0'\right\|_2\\
& + & \left(K_{10}+\frac{K_{11}\left(\overline{f_{xx}}\left(E(Y^*)+K_4\left\|Y'-Y^*\right\|_{L^2}\right)+2\|C\|_2^2\right)}{\lambda-F(Y^*)-\left\|Y'-Y^*\right\|_{L^2}K_5}\right)\left\|Y-Y'\right\|_{L^2}
\end{array}$$

Hence $(\theta,Y)\longmapsto X_{\theta,\overline{u}}(.,Y)$ and $(\theta,Y)\longmapsto p_\theta(.,Y)$ are continuous on $\Theta\times B(Y^*,\zeta)$ and so $\theta\longmapsto\mathcal{S}(Y;\theta,\lambda)$ is continuous for all $Y\in B(Y^*,\zeta)$.

We now show $\theta\longmapsto\mathcal{S}(Y;\theta,\lambda)$ is C^1 for all $Y\in B(Y^*,\zeta)$, the main point is to show $(\theta,\,Y)\longmapsto\left(\frac{\partial X_{\theta,\overline{u}}(.,Y)}{\partial\theta},\frac{\partial p_\theta(.,Y)}{\partial\theta}\right)$ is continuous. We have $\left(\frac{\partial X_{\theta,\overline{u}}(t,Y)}{\partial\theta},\frac{\partial p_\theta(t,Y)}{\partial\theta}\right)$ solution of the ODE:

$$\frac{d}{dt}\begin{pmatrix}\frac{\partial X_{\theta,\overline{u}}(t,Y)}{\partial\theta}\\ \frac{\partial p_\theta(t,Y)}{\partial\theta}\end{pmatrix} = H(t,\theta,Y)\begin{pmatrix}\frac{\partial X_{\theta,\overline{u}}(t,Y)}{\partial\theta}\\ \frac{\partial p_\theta(t,Y)}{\partial\theta}\end{pmatrix} + G(t,\theta,Y) \tag{7.1}$$

with:

$$\begin{array}{l}
H(t,\theta,Y)=\begin{pmatrix}\frac{\partial f}{\partial x}(t,X_{\theta,\overline{u}}(t,Y),\theta) & \frac{1}{2\lambda}BB^T\\ -\frac{\partial^2 f}{\partial^2 x}(t,X_{\theta,\overline{u}}(t,Y),\theta)p_\theta(t,Y)+2C^TC & -\frac{\partial f}{\partial x}(t,X_{\theta,\overline{u}}(t,Y),\theta)\end{pmatrix}\\
G(t,\theta,Y)=\begin{pmatrix}\frac{\partial f}{\partial\theta}(t,X_{\theta,\overline{u}}(t,Y),\theta)\\ -\frac{\partial^2 f}{\partial x\partial\theta}(t,X_{\theta,\overline{u}}(t,Y),\theta)p_\theta(t,Y)\end{pmatrix}
\end{array}$$

Condition C1, C2bis, C5, C6 gives us the continuity of $(\theta,\,Y)\longmapsto H(.,\theta,Y)$, $(\theta,\,Y)\longmapsto G(.,\theta,Y)$ and $\|G(.,\theta,Y)\|_{L^2}$ uniform boundedness on $\Theta\times B(Y^*,\zeta)$. Using lemma 7.6 if

$$\frac{\|B\|_2^2\,d^2T^2}{2\lambda}\left(\overline{f_{xx}}\left(E(Y^*)+K_4\zeta\right)+2\|C\|_2^2\right)e^{2\sqrt{dT\overline{f_x}}}<1$$

and so if $\lambda>\overline{\lambda_4}(\zeta)$, we have continuity of $(\theta,\,Y)\longmapsto\left(\frac{\partial X_{\theta,\overline{u}}(.,Y)}{\partial\theta},\frac{\partial p_\theta(.,Y)}{\partial\theta}\right)$. In particular we derive from that $\theta\longmapsto\mathcal{S}(Y;\theta,\lambda)$ is C^1 for all $Y\in B(Y^*,\zeta)$ and:

$$\begin{array}{lcl}
\nabla_\theta\mathcal{S}(Y;\theta,\lambda) & = & 2\int_0^T\frac{\partial X_{\theta,\overline{u}}(t,Y)}{\partial\theta}^T C^T\left(CX_{\theta,\overline{u}}(t,Y)-Y(t)\right)dt\\
& + & \frac{1}{2\lambda}\int_0^T\frac{\partial p_\theta(t,Y)}{\partial\theta}^T BB^Tp_\theta(t,Y)dt
\end{array}$$

We now demonstrate $\theta\longmapsto\mathcal{S}(Y;\theta,\lambda)$ is C^2 for all $Y\in B(Y^*,\zeta)$, again we need to prove $(\theta,\,Y)\longmapsto\left(\frac{\partial^2 X_{\theta,\overline{u}}(.,Y)}{\partial^2\theta},\frac{\partial^2 p_\theta(.,Y)}{\partial^2\theta}\right)$ shares the same degree of regularity w.r.t θ. Here $\left(\frac{\partial^2 X_{\theta,\overline{u}}(t,Y)}{\partial^2\theta},\frac{\partial^2 p_\theta(t,Y)}{\partial^2\theta}\right)$ is solution of the ODE:

$$\frac{d}{dt}\begin{pmatrix}\frac{\partial^2 X_{\theta,\overline{u}}(t,Y)}{\partial^2\theta}\\ \frac{\partial^2 p_\theta(t,Y)}{\partial^2\theta}\end{pmatrix} = H(t,\theta,Y)\begin{pmatrix}\frac{\partial^2 X_{\theta,\overline{u}}(t,Y)}{\partial^2\theta}\\ \frac{\partial^2 p_\theta(t,Y)}{\partial^2\theta}\end{pmatrix} +I(t,\theta,Y) \tag{7.2}$$

with:

$$I(t,\theta,Y) = \begin{pmatrix} I_1(t,\theta,Y) \\ I_2(t,\theta,Y) \end{pmatrix} = \frac{\partial H}{\partial \theta}(t,\theta,Y) \begin{pmatrix} \frac{\partial X_{\theta,\overline{u}}(t,Y)}{\partial \theta} \\ \frac{\partial p_{\theta}(t,Y)}{\partial \theta} \end{pmatrix} + \frac{\partial G}{\partial \theta}(t,\theta,Y)$$

$$\frac{\partial H}{\partial \theta}(t,\theta,Y) = \begin{pmatrix} \frac{\partial H_1}{\partial \theta}(t,\theta,Y) & 0 \\ \frac{\partial H_3}{\partial \theta}(t,\theta,Y) & \frac{\partial H_4}{\partial \theta}(t,\theta,Y) \end{pmatrix}$$

$$\frac{\partial G}{\partial \theta}(t,\theta,Y) = \begin{pmatrix} \frac{\partial G_1}{\partial \theta}(t,\theta,Y) \\ \frac{\partial G_2}{\partial \theta}(t,\theta,Y) \end{pmatrix}$$

where the components of $\frac{\partial H}{\partial \theta}(t,\theta,Y)$ and $\frac{\partial G}{\partial \theta}(t,\theta,Y)$ are respectively equal to:

$$\begin{array}{rcl} \frac{\partial H_1}{\partial \theta}(t,\theta,Y) & = & \frac{\partial^2 f}{\partial^2 x}(t,X_{\theta,\overline{u}}(t,Y),\theta)\frac{\partial X_{\theta,\overline{u}}(t,Y)}{\partial \theta} + \frac{\partial^2 f}{\partial x \partial \theta}(t,X_{\theta,\overline{u}}(t,Y),\theta) \\ \frac{\partial H_3}{\partial \theta}(t,\theta,Y) & = & -\frac{\partial^3 f}{\partial^3 x}(t,X_{\theta,\overline{u}}(t,Y),\theta)\frac{\partial X_{\theta,\overline{u}}(t,Y)}{\partial \theta}p_{\theta}(t,Y) \\ & - & \frac{\partial^2 f}{\partial^2 x}(t,X_{\theta,\overline{u}}(t,Y),\theta)\frac{\partial p_{\theta}(t,Y)}{\partial \theta} - \frac{\partial^3 f}{\partial^2 x \partial \theta}(t,X_{\theta,\overline{u}}(t,Y),\theta)p_{\theta}(t,Y) \\ \frac{\partial H_4}{\partial \theta}(t,\theta,Y) & = & -\frac{\partial^2 f}{\partial^2 x}(t,X_{\theta,\overline{u}}(t,Y),\theta)\frac{\partial X_{\theta,\overline{u}}(t,Y)}{\partial \theta} - \frac{\partial^2 f}{\partial x \partial \theta}(t,X_{\theta,\overline{u}}(t,Y),\theta) \end{array}$$

and:

$$\begin{array}{rcl} \frac{\partial G_1}{\partial \theta}(t,\theta,Y) & = & \frac{\partial^2 f}{\partial x \partial \theta}(t,X_{\theta,\overline{u}}(t,Y),\theta)\frac{\partial X_{\theta,\overline{u}}(t,Y)}{\partial \theta} + \frac{\partial^2 f}{\partial^2 \theta}(t,X_{\theta,\overline{u}}(t,Y),\theta) \\ \frac{\partial G_2}{\partial \theta}(t,\theta,Y) & = & -\frac{\partial^3 f}{\partial^2 x \partial \theta}(t,X_{\theta,\overline{u}}(t,Y),\theta)\frac{\partial X_{\theta,\overline{u}}(t,Y)}{\partial \theta}p_{\theta}(t,Y) \\ & - & \frac{\partial^2 f}{\partial x \partial \theta}(t,X_{\theta,\overline{u}}(t,Y),\theta)\frac{\partial p_{\theta}(t,Y)}{\partial \theta} - \frac{\partial^3 f}{\partial x \partial^2 \theta}(t,X_{\theta,\overline{u}}(t,Y),\theta) \end{array}$$

Condition C1, C2bis, C5, C6 gives us the continuity of $(\theta,\, Y) \longmapsto I\,(.,\theta,Y)$ and $\|I\,(.,\theta,Y)\|_{L^2}$ uniform boundedness on $\Theta \times B\,(Y^*,\zeta)$. So, by using again lemma 7.6, if $\lambda > \overline{\lambda_4}(\zeta)$ we have continuity of $(\theta,\, Y) \longmapsto \left(\frac{\partial^2 X_{\theta,\overline{u}}(.,Y)}{\partial \theta^T \partial \theta}, \frac{\partial^2 p_{\theta}(.,Y)}{\partial \theta^T \partial \theta}\right)$. In particular we derive from that $\theta \longmapsto \mathcal{S}(Y;\theta,\lambda)$ is C^2 for all $Y \in B\,(Y^*,\zeta)$ and:

$$\begin{array}{rcl} \frac{\partial^2 \mathcal{S}(Y;\theta,\lambda)}{\partial^2 \theta} & = & 2\sum_{i=1}^{d}\int_0^T \frac{\partial^2 X^i_{\theta,\overline{u}}(t,Y)}{\partial^2 \theta}^T \left(C^T\left(CX_{\theta,\overline{u}}(t,Y) - Y(t)\right)\right)_i dt \\ & + & 2\int_0^T \frac{\partial X_{\theta,\overline{u}}(t,Y)}{\partial \theta}^T C^T C \frac{\partial X_{\theta,\overline{u}}(t,Y)}{\partial \theta} dt \\ & + & \frac{1}{2\lambda}\sum_{i=1}^{d}\int_0^T \frac{\partial^2 p^i_{\theta}(t,Y)}{\partial^2 \theta} BB^T p^i_{\theta}(t,Y) dt \\ & + & \frac{1}{2\lambda}\int_0^T \frac{\partial p_{\theta}(t,Y)}{\partial \theta}^T BB^T \frac{\partial p_{\theta}(t,Y)}{\partial \theta} dt \end{array}$$

Moreover continuity on $\varTheta \times B\,(Y^*,\zeta)$ of the derivative form of $(\theta,Y) \longmapsto X_{\theta,\overline{u}}(.,Y)$ and $(\theta,Y) \longmapsto p_{\theta}(.,Y)$ ensures continuity of $(\theta,\, Y) \longmapsto \frac{\partial^2 \mathcal{S}(Y;\theta,\lambda)}{\partial^2 \theta}$ as well.

Now we can obtain the desired asymptotic representation. According to first order optimality condition:

$$\nabla \mathcal{S}(\widehat{Y};\widehat{\theta},\lambda) = 0$$

We have shown $\theta \longmapsto \mathcal{S}(\widehat{Y};\theta,\lambda)$ is C^2 on Θ and hence it exists a point $\widetilde{\theta}$ between $\widehat{\theta}$ and θ^* s.t $\widetilde{\theta} \longrightarrow \theta^*$ when $n \longrightarrow +\infty$ and:

$$-\nabla_{\theta}\mathcal{S}(\widehat{Y};\theta^*,\lambda) = \frac{1}{2}\frac{\partial^2 \mathcal{S}(\widehat{Y};\widetilde{\theta},\lambda)}{\partial \theta^T \partial \theta}\left(\widehat{\theta} - \theta^*\right)$$

and thanks to the continuous mapping theorem we have:

$$\frac{\partial^2 \mathcal{S}(\widehat{Y};\widetilde{\theta},\lambda)}{\partial \theta^T \partial \theta} \rightarrow \frac{\partial^2 \mathcal{S}(Y^*;\theta^*,\lambda)}{\partial \theta^T \partial \theta}$$

in probability because of $(\theta,\, Y) \longmapsto \frac{\partial^2 \mathcal{S}(Y;\theta,\lambda)}{\partial^2 \theta}$ continuity. Since condition C6 imposes $\frac{\partial^2 \mathcal{S}}{\partial \theta^T \partial \theta}(\theta^*, Y^*, \lambda)$

non singularity, we obtain the desired asymptotic representation:

$$\widehat{\theta}-\theta^{*}=-2\frac{\partial^{2}\mathcal{S}(Y^{*};\theta^{*},\lambda)}{\partial\theta^{T}\partial\theta}^{-1}\nabla_{\theta}\mathcal{S}(\widehat{Y};\theta^{*},\lambda)+o_{P}(1)$$

□

7.2 Linear representation of the differential of $Y\longmapsto\nabla_{\theta}\mathcal{S}\left(Y\,;\theta,\lambda\right)$

Lemma 7.2. *Let us suppose* C1, C2bis, C5, C6 and $\lambda>\max(\overline{\lambda_{3}}(\zeta),\,\overline{\lambda_{4}}(\zeta))$*. Then* $\forall\theta\in\Theta$*,* $Y\longmapsto\nabla_{\theta}\mathcal{S}(Y;\theta,\lambda)$ *is differentiable on* $B\left(Y^{*},\zeta\right)$*,* $(Y_{1},Y_{2})\longmapsto D\left(\nabla_{\theta}S\left(\,.\,;\theta,\lambda\right)\right)(Y_{1}).Y_{2}$ *is continuous on* $B\left(Y^{*},\zeta\right)\times B\left(Y^{*},\zeta\right)$ and can be represented as a scalar product in L^{2} i.e:

$$D\left(\nabla_{\theta}\mathcal{S}\left(\,.\,;\theta,\lambda\right)\right)(Y_{1}).Y_{2}\quad=\quad<V(Y_{1},\theta),Y_{2}>$$

with $V(Y_{1},\theta)\in C^{1}\left(\left[0,\,T\right],\mathbb{R}^{p\times d^{'}}\right)$.

Proof. Taking λ greater than $\overline{\lambda_{3}}(\zeta)$ gives us continuity of $Y\longmapsto\begin{pmatrix}X_{\theta,\overline{u}}(.,Y)\\ p_{\theta}(.,Y)\end{pmatrix}$ on $B\left(Y^{*},\zeta\right)$. Assuming conditions C1, C2bis , $Y\longmapsto\begin{pmatrix}X_{\theta,\overline{u}}(.,Y)\\ p_{\theta}(.,Y)\end{pmatrix}$ is differentiable on $B\left(Y^{*},\zeta\right)$ with $\begin{pmatrix}dX_{\theta}(Y_{1},Y_{2})\\ dP_{\theta}(Y_{1},Y_{2})\end{pmatrix}=\begin{pmatrix}D\left(X_{\theta,\overline{u}}\right)(Y_{1}).Y_{2}\\ D\left(p_{\theta}\right)(Y_{1}).Y_{2}\end{pmatrix}$ solution of:

$$\begin{cases}\frac{d}{dt}\begin{pmatrix}dX_{\theta}(Y_{1},Y_{2})(t)\\ dP_{\theta}(Y_{1},Y_{2})(t)\end{pmatrix}&=\;H\left(t,\theta,Y_{1}\right)\begin{pmatrix}dX_{\theta}(Y_{1},Y_{2})(t)\\ dP_{\theta}(Y_{1},Y_{2})(t)\end{pmatrix}-\begin{pmatrix}0\\ 2C^{T}Y_{2}(t)\end{pmatrix}\\ dX_{\theta}(Y_{1},Y_{2})(0)=0\\ dP_{\theta}(Y_{1},Y_{2})(T)=0\end{cases}\tag{7.3}$$

with:

$$H\left(t,\theta,Y\right)=\begin{pmatrix}\frac{\partial f}{\partial x}(t,X_{\theta,\overline{u}}(t,Y),\theta)&\frac{1}{2\lambda}BB^{T}\\ -\frac{\partial^{2}f}{\partial^{2}x}(t,X_{\theta,\overline{u}}(t,Y),\theta)p_{\theta}(t,Y)+2C^{T}C&-\frac{\partial f}{\partial x}(t,X_{\theta,\overline{u}}(t,Y),\theta)\end{pmatrix}$$

Using lemma 7.6 we have continuity of $(Y_{1},Y_{2})\longmapsto\begin{pmatrix}dX_{\theta}(Y_{1},Y_{2})\\ dP_{\theta}(Y_{1},Y_{2})\end{pmatrix}$ on $B\left(Y^{*},\zeta\right)\times B\left(Y^{*},\zeta\right)$ as soon as $\lambda>\overline{\lambda_{4}}(\zeta)$, in that case we have also continuity of $Y\longmapsto\begin{pmatrix}\frac{\partial X_{\theta,\overline{u}}(.,Y)}{\partial\theta}\\ \frac{\partial p_{\theta}(.,Y)}{\partial\theta}\end{pmatrix}$ on $B\left(Y^{*},\zeta\right)$. Moreover by assuming condition C5 and C6 $Y\longmapsto\begin{pmatrix}\frac{\partial X_{\theta,\overline{u}}(.,Y)}{\partial\theta}\\ \frac{\partial p_{\theta}(.,Y)}{\partial\theta}\end{pmatrix}$ is differentiable on $B\left(Y^{*},\zeta\right)$ with $\begin{pmatrix}d\partial X_{\theta}(Y_{1},Y_{2})\\ d\partial P_{\theta}(Y_{1},Y_{2})\end{pmatrix}=\begin{pmatrix}D\left(\frac{\partial X_{\theta,\overline{u}}}{\partial\theta}\right)(Y_{1}).Y_{2}\\ D\left(\frac{\partial p_{\theta}}{\partial\theta}\right)(Y_{1}).Y_{2}\end{pmatrix}$ solution of:

$$\begin{cases}\frac{d}{dt}\begin{pmatrix}d\partial X_{\theta}(Y_{1},Y_{2})(t)\\ d\partial P_{\theta}(Y_{1},Y_{2})(t)\end{pmatrix}&=\;H\left(t,\theta,Y_{1}\right)\begin{pmatrix}d\partial X_{\theta}(Y_{1},Y_{2})(t)\\ d\partial P_{\theta}(Y_{1},Y_{2})(t)\end{pmatrix}\;+\;I(\theta,Y_{1},Y_{2})(t)\\ d\partial X_{\theta}(Y_{1},Y_{2})(0)=0\\ d\partial P_{\theta}(Y_{1},Y_{2})(T)=0\end{cases}\tag{7.4}$$

with:

$$I(\theta,Y_1,Y_2)(t)=H_Y\left(t,\theta,Y_1,Y_2\right)\begin{pmatrix}\frac{\partial X_{\theta,\overline{u}}(.,Y_1)}{\partial\theta}\\ \frac{\partial p_\theta(.,Y_1)}{\partial\theta}\end{pmatrix}+G_Y(t,\theta,Y_1,Y_2)$$

and:

$$\begin{array}{l}H_Y\left(t,\theta,Y_1,Y_2\right)=\begin{pmatrix}H_Y^1\left(t,\theta,Y_1,Y_2\right) & 0\\ H_Y^3\left(t,\theta,Y_1,Y_2\right) & H_Y^4\left(t,\theta,Y_1,Y_2\right)\end{pmatrix}\\ G_Y(t,\theta,Y_1,Y_2)=\begin{pmatrix}G_Y^1(t,\theta,Y_1,Y_2)\\ G_Y^2(t,\theta,Y_1,Y_2)\end{pmatrix}\end{array}$$

where the components of H_Y and G_Y are respectively equal to:

$$\begin{array}{lcl}H_Y^1\left(t,\theta,Y_1,Y_2\right) & = & \frac{\partial^2 f}{\partial^2 x}(t,X_{\theta,\overline{u}}(t,Y_1),\theta)dX_\theta(Y_1,Y_2)(t)\\ H_Y^3\left(t,\theta,Y_1,Y_2\right) & = & -\frac{\partial^3 f}{\partial^3 x}(t,X_{\theta,\overline{u}}(t,Y_1),\theta)dX_\theta(Y_1,Y_2)(t)p_\theta(t,Y_1)-\frac{\partial^2 f}{\partial^2 x}(t,X_{\theta,\overline{u}}(t,Y_1),\theta)dP_\theta(Y_1,Y_2)(t)\\ H_Y^4\left(t,\theta,Y_1,Y_2\right) & = & \frac{\partial^2 f}{\partial^2 x}(t,X_{\theta,\overline{u}}(t,Y_1),\theta)dX_\theta(Y_1,Y_2)(t)\end{array}$$

and:

$$\begin{array}{lcl}G_Y^1(t,\theta,Y_1,Y_2) & = & \frac{\partial^2 f}{\partial x\partial\theta}(t,X_{\theta,\overline{u}}(t,Y_1),\theta)dX_\theta(Y_1,Y_2)(t)\\ G_Y^2(t,\theta,Y_1,Y_2) & = & -\frac{\partial^3 f}{\partial^3 x}(t,X_{\theta,\overline{u}}(t,Y_1),\theta)dX_\theta(Y_1,Y_2)(t)p_\theta(t,Y)-\frac{\partial^2 f}{\partial^2 x}(t,X_{\theta,\overline{u}}(t,Y_1),\theta)dP_\theta(Y_1,Y_2)(t)\end{array}$$

Again using lemma 7.6 we have continuity of $(Y_1,Y_2)\longmapsto\begin{pmatrix}d\partial X_\theta(Y_1,Y_2)\\ d\partial P_\theta(Y_1,Y_2)\end{pmatrix}$ on $B\left(Y^*,\zeta\right)\times B\left(Y^*,\zeta\right)$ as soon as $\lambda>\overline{\lambda_4}(\zeta)$.

Hence $Y\longmapsto\nabla_\theta\mathcal{S}(Y;\theta,\lambda)$ is differentiable on $B\left(Y^*,\zeta\right)$ and equal to:

$$\begin{array}{lcl}D\left(\nabla_\theta\mathcal{S}\left(\,.\,;\theta,\lambda\right)\right)(Y_1).Y_2 & = & 2\int_0^T d\partial X_\theta(Y_1,Y_2)(t)^T C^T\left(CX_{\theta,\overline{u}}(t,Y_1)-Y_1(t)\right)dt\\ & + & 2\int_0^T\frac{\partial X_{\theta,\overline{u}}(t,Y_1)}{\partial\theta}^T C^T\left(dX_\theta(Y_1,Y_2)(t)-Y_2(t)\right)dt\\ & + & \frac{1}{2\lambda}\int_0^T d\partial P_\theta(Y_1,Y_2)(t)^T BB^T p_\theta(t,Y_1)dt\\ & + & \frac{1}{2\lambda}\int_0^T\frac{\partial p_\theta(t,Y_1)}{\partial\theta}^T BB^T dP_\theta(Y_1,Y_2)(t)dt\end{array}\tag{7.5}$$

and thanks to previous regularity results we derive $(Y_1,Y_2)\longmapsto D\left(\nabla_\theta\mathcal{S}\left(\,.\,;\theta,\lambda\right)\right)(Y_1).Y_2$ continuity on $B\left(Y^*,\zeta\right)\times B\left(Y^*,\zeta\right)$.

Using Duhamel formula we have the explicit expression respectively for $\begin{pmatrix}dX_\theta(Y_1,Y_2)\\ dP_\theta(Y_1,Y_2)\end{pmatrix}$ and $\begin{pmatrix}d\partial X_\theta(Y_1,Y_2)\\ d\partial P_\theta(Y_1,Y_2)\end{pmatrix}$:

$$\begin{array}{lcl}\begin{pmatrix}dX_\theta(Y_1,Y_2)(t)\\ dP_\theta(Y_1,Y_2)(t)\end{pmatrix} & = & R_Y(t,\theta,Y_1)\begin{pmatrix}0\\ dP_\theta(Y_1,Y_2)(0)\end{pmatrix}-R_Y(t,\theta,Y_1)\int_0^t R_Y(s,\theta,Y_1)^{-1}\begin{pmatrix}0\\ 2C^TY_2(s)\end{pmatrix}ds\\ \begin{pmatrix}d\partial X_\theta(Y_1,Y_2)(t)\\ d\partial P_\theta(Y_1,Y_2)(t)\end{pmatrix} & = & R_Y(t,\theta,Y_1)\begin{pmatrix}0\\ d\partial P_\theta(Y_1,Y_2)(0)\end{pmatrix}+R_Y(t,\theta,Y_1)\int_0^t R_Y(s,\theta,Y_1)^{-1}I(\theta,Y_1,Y_2)(s)ds\end{array}$$

with $R_Y(t,\theta,Y_1)$ solution of:

$$\begin{cases}\frac{d}{dt}R_Y(t,\theta,Y_1)=H\left(t,\theta,Y_1\right)R_Y(t,\theta,Y_1)\\ R_Y(0,\theta,Y_1)=I_{2d}\end{cases}\tag{7.6}$$

In particular

$$\left(\begin{array}{c} dX_\theta(Y_1,Y_2)(T) \\ 0 \end{array}\right) = \left(\begin{array}{c} R_Y^2(T,\theta,Y_1)dP_\theta(Y_1,Y_2)(0) \\ R_Y^4(T,\theta,Y_1)dP_\theta(Y_1,Y_2)(0) \end{array}\right) - R_Y(T,\theta,Y_1)\int_0^T R_Y(s,\theta,Y_1)^{-1}\left(\begin{array}{c} 0 \\ 2C^TY_2(s) \end{array}\right)ds \tag{7.7}$$

and

$$\left(\begin{array}{c} d\partial X_\theta(Y_1,Y_2)(T) \\ 0 \end{array}\right) = \left(\begin{array}{c} R_Y^2(T,\theta,Y_1)d\partial P_\theta(Y_1,Y_2)(0) \\ R_Y^4(T,\theta,Y_1)d\partial P_\theta(Y_1,Y_2)(0) \end{array}\right) + R_Y(T,\theta,Y_1)\int_0^T R_Y(s,\theta,Y_1)^{-1}I(\theta,Y_1,Y_2)(s)ds \tag{7.8}$$

In order to represent the gradient differential as a scalar product:

$$D\left(\nabla_\theta \mathcal{S}\left(\,.\,;\theta,\lambda\right)\right)(Y_1).Y_2 \;=\; < V(Y_1,\theta), Y_2 >$$

with $V(Y_1,\theta)$ a C^1 function for each couple (Y_1,θ), we need to demonstrate $R_Y^4(T,\theta,Y_1)$ is invertible.

Let us choose x such that $R_Y^4(T,\theta,Y_1)x=0$. Hence $\left(\begin{array}{c} R_Y^2(t,\theta,Y_1)x \\ R_Y^4(t,\theta,Y_1)x \end{array}\right)$ respects the ODE with boundary condition:

$$\begin{cases} \frac{d}{dt}\left(\begin{array}{c} R_Y^2(t,\theta,Y_1)x \\ R_Y^4(t,\theta,Y_1)x \end{array}\right) = H\left(t,\theta,Y_1\right)\left(\begin{array}{c} R_Y^2(t,\theta,Y_1)x \\ R_Y^4(t,\theta,Y_1)x \end{array}\right) \\ R_Y^2(0,\theta,Y_1)x = 0 \\ R_Y^4(T,\theta,Y_1)x = 0 \end{cases}$$

here $\left(\begin{array}{c} R_Y^2(t,\theta,Y_1)x \\ R_Y^4(t,\theta,Y_1)x \end{array}\right) = \left(\begin{array}{c} 0 \\ 0 \end{array}\right)$ is an obvious solution. Using lemma 7.6, we know for $\lambda > \overline{\lambda_4}(\zeta)$ it is the only one which implies $R_Y^4(0,\theta,Y_1)x=0$ by resolvant definition we know $R_Y^4(0,\theta,Y_1)=I_d$ which necessarily implies $x=0$, hence the invertibility of $R_Y^4(T,\theta,Y_1)$.

Since $R_Y^4(T,\theta,Y_1)$ is invertible and $t \longrightarrow R_Y(t,\theta,Y_1)$ is C^1 (by using classic regularity results about ODE solutions), using equation (7.7) we have access to a function $V_P(\theta,Y_1) \in C^1\left([0,\,T]\,,\,\mathbb{R}^{d\times d'}\right)$ such that:

$$dP_\theta(Y_1,Y_2)(0) = \int_0^T V_P(\theta,Y_1)(t)Y_2(t)dt$$

and so:

$$\begin{aligned} \left(\begin{array}{c} dX_\theta(Y_1,Y_2)(t) \\ dP_\theta(Y_1,Y_2)(t) \end{array}\right) &= R_Y(t,\theta,Y_1)\left(\begin{array}{c} -2\int_0^t R_Y^2(s,\theta,Y_1)^{-1}C^TY_2(s)ds \\ \int_0^T V_P(\theta,Y_1)(s)Y_2(s)ds - 2\int_0^t R_Y^2(s,\theta,Y_1)^{-1}C^TY_2(s)ds \end{array}\right) \\ &= \left(\begin{array}{c} J^1(t,\theta,Y_1)\int_0^t R_Y^2(s,\theta,Y_1)^{-1}C^TY_2(s)ds + R_Y^2(t,\theta,Y_1)\int_0^T V_P(\theta,Y_1)(s)Y_2(s)ds \\ J^2(t,\theta,Y_1)\int_0^t R_Y^2(s,\theta,Y_1)^{-1}C^TY_2(s)ds + R_Y^4(t,\theta,Y_1)\int_0^T V_P(\theta,Y_1)(s)Y_2(s)ds \end{array}\right) \end{aligned}$$

with:

$$\begin{aligned} J^1(t,\theta,Y_1) &= -2\left(R_Y^1(t,\theta,Y_1) + R_Y^2(t,\theta,Y_1)\right) \\ J^2(t,\theta,Y_1) &= -2\left(R_Y^3(t,\theta,Y_1) + R_Y^4(t,\theta,Y_1)\right) \end{aligned}$$

The same holds for $Y_2 \longmapsto d\partial P_\theta(Y_1,Y_2)(0)$, hence for each $(\theta,\,Y_1)$ we know it exists an unique C^1 function $V_{\partial P}(\theta,Y_1)$ such that:

$$d\partial P_\theta(Y_1,Y_2)(0) = \int_0^T V_{\partial P}(\theta,Y_1)(t)Y_2(t)dt$$

Similar computation lead us to

$$\begin{pmatrix} d\partial X_\theta(Y_1,Y_2)(t) \\ d\partial P_\theta(Y_1,Y_2)(t) \end{pmatrix} = R_Y(t,\theta,Y_1)\begin{pmatrix} J^3(t,\theta,Y_1,Y_2) \\ J^4(t,\theta,Y_1,Y_2) \end{pmatrix}$$

where $J^3(t,\theta,Y_1,Y_2)$ and $J^3(t,\theta,Y_1,Y_2)$ are continuous linear functions in Y_2 and differentiable in t given by:

$$\begin{array}{rcl} J^3(t,\theta,Y_1,Y_2) &=& J^5(t,\theta,Y_1)\int_0^t R_Y^2(s,\theta,Y_1)^{-1}C^TY_2(s)ds - \int_0^t J^6(s,\theta,Y_1)Y_2(s)ds \\ &+& \int_0^t J^7(s,\theta,Y_1)ds.\int_0^T V_P(\theta,Y_1)(u)Y_2(u)du \\ J^4(t,\theta,Y_1,Y_2) &=& J^8(t,\theta,Y_1)\int_0^t R_Y^2(s,\theta,Y_1)^{-1}C^TY_2(s)ds - \int_0^t J^9(s,\theta,Y_1)Y_2(s)ds \\ &+& \int_0^T V_{\partial P}(\theta,Y_1)(s)Y_2(s)dt + J^{10}(t,\theta,Y_1)\int_0^T V_P(\theta,Y_1)(u)Y_2(u)du \end{array}$$

where the J^i, $i \in [\![5,\,10]\!]$ are functions differentiable in t obtained from the expression (7.8). Then using the following formula obtained by integration by part:

$$\begin{array}{rcl} \int_0^T u(t)\left(\int_0^t v(s)ds\right)dt &=& \int_0^T u(t)dt\int_0^T v(t)dt - \int_0^T\left(\int_0^t u(s)ds\right).v(t)dt \\ &=& \int_0^T\left(\int_0^T u(s)ds\right)v(t)dt - \int_0^T\left(\int_0^t u(s)ds\right).v(t)dt \\ &=& \int_0^T\left(\int_0^T u(s)ds - \int_0^t u(s)ds\right)v(t)dt \\ &=& \int_0^T\left(\int_t^T u(s)ds\right)v(t)dt \end{array}$$

and the gradient expression given by (7.5), we know for each couple (Y_1,θ) it exists a C^1 function $V(Y_1,\theta)$ such that:

$$D\left(\nabla_\theta \mathcal{S}\left(\,.\,;\theta,\lambda\right)\right)(Y_1).Y_2 \;=\; < V(Y_1,\theta), Y_2 >$$

□

7.3 Theorem

Theorem 7.3. *Assuming that $\widehat{Y}$ is a regression splines estimator and assuming conditions 1, 2bis and 3-10 we have for $\lambda > F(Y^*)$ that $\widehat{\theta^T} - \theta^*$ is asymptotically normal and*

$$\widehat{\theta}^T - \theta^* = O_P(n^{-1/2})$$

Proof. The use of proposition 7.1 and lemma 7.2 thanks to conditions 1-6 gives us for $\lambda > \max(\overline{\lambda_3}(\zeta),\,\overline{\lambda_4}(\zeta))$, the following asymptotic representation:

$$\widehat{\theta} - \theta^* = -2\frac{\partial^2\mathcal{S}(Y^*;\theta^*,\lambda)}{\partial\theta^T\partial\theta}^{-1}\left(\nabla_\theta\mathcal{S}(\widehat{Y};\theta^*,\lambda) - \nabla_\theta\mathcal{S}(Y^*;\theta^*,\lambda)\right) + o_P(1)$$

(since first order optimality condition imposes: $\nabla_\theta\mathcal{S}(Y^*;\theta^*,\lambda) = 0$) and continuous differentiability of $Y \longmapsto \nabla_\theta\mathcal{S}(Y;\theta,\lambda)$ on $B\left(Y^*,\zeta\right)$ with the inner product representation in L^2 :

$$D\left(\nabla_\theta \mathcal{S}\left(\,.\,;\theta,\lambda\right)\right)(Y_1).Y_2 \;=\; < V(Y_1,\theta), Y_2 >$$

if ζ is such that $\widehat{Y} \in B\left(Y^*,\zeta\right)$. According to Theorem 7 in [6] $\widehat{Y}$ is a consistent estimator of Y^*. Hence, asymptotically we can take ζ as small as we want and the previous results holds for $\lambda > F(Y^*)$.

We will now use theorem 9 in [6] in order to obtain the asymptotic normality with $\sqrt{n}$ rate of

$\nabla_\theta \mathcal{S}(\widehat{Y};\theta^*,\lambda) - \nabla_\theta \mathcal{S}(Y^*;\theta^*,\lambda)$. We have to prove that:

1) (t_i, Y_i) are i.i.d with $Var(Y \mid t)$ bounded.

2) $\mathbb{E}((Y - Y^*(t))^4 \mid t)$ is bounded, and $Var(Y \mid t)$ is bounded away from 0.

3) The support of t is a compact interval on which t has a probability density function bounded away from 0.

4) There is $v(t)$ such that $\mathbb{E}(v(t)v(t)^T)$ is finite and non-singular such that: $D\left(\nabla_\theta \mathcal{S}\left(\,.\,;\theta,\lambda\right)\right)(Y^*).Y^* = \mathbb{E}(v(t)Y^*(t))$ and $D\left(\nabla_\theta \mathcal{S}\left(\,.\,;\theta^*,\lambda\right)\right)(Y^*).p_{kK} = \mathbb{E}(v(t)p_{kK}(t))$ for all k and K and there is c_K with $\mathbb{E}(\|v(t) - c_K p_K(t)\|_2^2) \to 0$.

5) $CX^*(t) = \mathbb{E}(Y \mid t)$ is derivable of order s on the support of t.

Requirement 1, 2, 3 are simple consequence of condition 8 and the fact the solution is defined on the closed interval $[0\,,\,T]$ and requirement 5 is a simple consequence of the condition 9. Using proposition 7.2 for requirement 4 we can take

$$v(t) = V(Y^*,\theta^*)(t)$$

to obtain the scalar product representation:

$$D\left(\nabla_\theta \mathcal{S}\left(\,.\,;\theta^*,\lambda\right)\right)(Y^*).Y_2 = < V(Y^*,\theta^*), Y_2 >$$

In order to prove the existence of c_K such that $\mathbb{E}(\|v(t) - c_K p_K(t)\|_2^2) \to 0$ we need to remark that v is C^1 on $[0,\,T]$ thanks to proposition 7.2, then we can use condition 9-10 and theorem 7.3 in [3] to conclude. □

7.4 Auxiliary lemma

For the preceeding proofs, we need to ensure the continuity of $(Y,\theta) \longmapsto (X_{\theta,\overline{u}}(.,Y),\, p_\theta(.,Y))$ and of its derivatives w.r.t Y and/or θ. For this, we have

1. to uniformly control the discrepancy between two solutions of the BVP presented in theorem 3.2 (lemma 7.5),
2. to derive conditions ensuring continuity of $\theta \longrightarrow P_\theta$ where P_θ is the solution of a linear BVP depending of a parameter θ (lemma 7.6).

Lemma 7.4. *Let us suppose conditions C1, C2.* $\forall \theta \in \Theta$ let us consider $(X_{\theta,\overline{u}}, p_\theta)$ an admissible solution of:

$$\begin{cases} \dot{X}_{\theta,\overline{u}}(t) = f(t, X_{\theta,\overline{u}}(t),\theta) + \frac{1}{2\lambda}BB^T p_\theta(t) \\ \dot{p}_\theta(t) = -\frac{\partial f}{\partial x}(t, X_{\theta,\overline{u}}(t),\theta)^T p_\theta(t) + 2C^T\left(CX_{\theta,\overline{u}}(t) - Y(t)\right) \\ \left(X_{\theta,\overline{u}}\left(0\right), p_\theta(T)\right) = (x_0, 0) \end{cases}$$

then $\|p_\theta(t,Y)\|_2 \leq E(Y)$.

Proof. Using Gronwall's lemma we have:

$$\begin{aligned} \left|p^i(t)\right| &\leq 2\sqrt{d}\,\|C\|_2 \int_0^t e^{\sqrt{df_x}(t-s)} \left\|Y(T-s) - CX_{\overline{u}}^i(s)\right\|_2 ds \\ &\leq 2\sqrt{d}\,\|C\|_2 \sqrt{\tfrac{e^{2\sqrt{df_x}t}-1}{2\sqrt{df_x}}} \sqrt{\int_0^t \left\|Y(T-s) - CX_{\overline{u}}^i(s)\right\|_2^2 ds} \\ &\leq 2\sqrt{d}\,\|C\|_2 \sqrt{\tfrac{e^{2\sqrt{df_x}t}-1}{2\sqrt{df_x}}}\, \|Y - CX_{\overline{u}}\|_{L^2} \,\leq\, E(Y) \end{aligned}$$

□

Lemma 7.5. *For λ such that $\lambda > F(Y^*)$, we have for all $Y \in L^2\left([0, T], \mathbb{R}^{d'}\right)$ the bound*

$$\|p_\theta(t, Y)\|_2 \leq E(Y^*) + K_4 \|Y - Y^*\|_{L^2}$$

and for all $\left(Y, Y^{'}\right) \in L^2\left([0, T], \mathbb{R}^{d'}\right)$ and λ such that:

$$\lambda > F(Y^*) + \frac{1}{2}\left(K_0 \left\|Y^{'} - Y^*\right\|_{L^2} + \sqrt{\left(K_0 \|Y' - Y^*\|_{L^2} + 4\sqrt{d}\,\|C\|_2\, K_1\right) K_0 \|Y' - Y^*\|_{L^2}}\right)$$

we have the discrepancy:

$$\begin{array}{rcl}
\left\|X_{\theta,\overline{u}}(.,Y) - X_{\theta',\overline{u}}(.,Y^{'})\right\|_{L^2} & \leq & \frac{K_2 + \left\|Y^{'} - Y^*\right\|_{L^2} K_6 + \lambda K_3}{\lambda - F(Y^*) - K_5 \left\|Y^{'} - Y^*\right\|_{L^2}} \left\|\theta - \theta^{'}\right\|_2 \\
& + & \frac{\sqrt{T}\left(\sqrt{dT}\overline{f_x} e^{\sqrt{d f_x} T} + 1\right)}{\lambda - F(Y^*) - \left\|Y^{'} - Y^*\right\|_{L^2} K_5} \left\|x_0 - x_0^{'}\right\|_2 \\
& + & \frac{K_1}{\lambda - F(Y^*) - \left\|Y^{'} - Y^*\right\|_{L^2} K_5} \left\|Y - Y^{'}\right\|_{L^2} \\
\left\|p_\theta(.,Y) - p_{\theta'}(.,Y^{'})\right\|_{L^2} & \leq & K_7 \frac{\left(\overline{f_{xx}}\left(E(Y^*) + K_4 \left\|Y^{'} - Y^*\right\|_{L^2}\right) + 2\|C\|_2^2\right)\left(K_2 + \left\|Y^{'} - Y^*\right\|_{L^2} K_6 + \lambda K_3\right)}{\left(\lambda - F(Y^*) - K_5 \left\|Y^{'} - Y^*\right\|_{L^2}\right)} \left\|\theta - \theta^{'}\right\|_2 \\
& + & K_8 \left(E(Y^*) + K_4 \left\|Y^{'} - Y^*\right\|_{L^2}\right) \left\|\theta - \theta^{'}\right\|_2 \\
& + & \frac{K_9\left(\overline{f_{xx}}\left(E(Y^*) + K_4 \left\|Y^{'} - Y^*\right\|_{L^2}\right) + 2\|C\|_2^2\right)}{\lambda - F(Y^*) - \left\|Y^{'} - Y^*\right\|_{L^2} K_5} \left\|x_0 - x_0^{'}\right\|_2 \\
& + & \left(K_{10} + \frac{K_{11}\left(\overline{f_{xx}}\left(E(Y^*) + K_4 \left\|Y^{'} - Y^*\right\|_{L^2}\right) + 2\|C\|_2^2\right)}{\lambda - F(Y^*) - \left\|Y^{'} - Y^*\right\|_{L^2} K_5}\right) \left\|Y - Y^{'}\right\|_{L^2}
\end{array}$$

with:

$$\begin{array}{rcl}
K_0 & = & \|B\|_2^2 d^2 \sqrt{d} T^2 e^{2\sqrt{d f_x} T} \overline{f_{xx}} \|C\|_2 \sqrt{\frac{e^{2\sqrt{d f_x} t} - 1}{2\sqrt{d f_x}}} \\
K_1 & = & d \|B\|_2^2 T^2 e^{2\sqrt{d f_x} T} \|C\|_2 \\
K_2 & = & \frac{1}{2} \|B\|_2^2 d^2 T^2 e^{2\sqrt{d f_x} T} \overline{f_{x\theta}} E(Y^*) \\
K_3 & = & \overline{f}_\theta T e^{\overline{f_x} T} \\
K_4 & = & 2\sqrt{d}\,\|C\|_2 \sqrt{\frac{e^{2\sqrt{d f_x} t} - 1}{2\sqrt{d f_x}}} \left(\frac{\sqrt{d}\|C\|_2 K_1}{\lambda - F(Y^*)} + 1\right) \\
K_5 & = & \frac{1}{2} \|B\|_2^2 d^2 T^2 e^{2\sqrt{d f_x} T} \overline{f_{xx}} K_4
\end{array}
\qquad
\begin{array}{rl}
K_6 & = \frac{1}{2} \|B\|_2^2 d^2 T^2 e^{2\sqrt{d f_x} T} \overline{f_{x\theta}} K_4 \\
K_7 & = \sqrt{d} T e^{\sqrt{d f_x} T} \\
K_8 & = \sqrt{d} T \overline{f_{x\theta}} e^{\sqrt{d f_x} T} \\
K_9 & = \sqrt{d} T^{\frac{3}{2}} \overline{f_x} e^{\sqrt{d f_x} T} \left(\sqrt{dT}\overline{f_x} e^{\sqrt{d f_x} T} + 1\right) \\
K_{10} & = 2\sqrt{d} T \|C\|_2 e^{\sqrt{d f_x} T} \\
K_{11} & = \sqrt{d} T e^{\sqrt{d f_x} T} K_1
\end{array}$$

Proof. Firstly we will bound $\left\|X_{\theta,\overline{u}}(.,Y) - X_{\theta',\overline{u}}(.,Y^{'})\right\|_{L^2}$ w.r.t $\|p_\theta(.,Y) - p_\theta(.,Y')\|_{L^2}$ and $\left\|\theta - \theta^{'}\right\|_2$.

$$\begin{array}{rcl}
\left\|\dot{X}_{\theta,\overline{u}}(t,Y) - \dot{X}_{\theta',\overline{u}}(t,Y^{'})\right\|_2 & \leq & \left\|f(t, X_{\theta,\overline{u}}(t,Y), \theta) - f(t, X_{\theta',\overline{u}}(t,Y^{'}), \theta^{'})\right\|_2 \\
& + & \frac{d\|B\|_2^2}{2\lambda} \left\|p_\theta(t,Y) - p_{\theta'}(t,Y^{'})\right\|_2 \\
& \leq & \sqrt{d f_x} \left\|X_{\theta,\overline{u}}(t,Y) - X_{\theta',\overline{u}}(t,Y^{'})\right\|_2 \\
& + & \overline{f}_\theta \left\|\theta - \theta^{'}\right\|_2 + \frac{\sqrt{d} d\|B\|_2^2}{2\lambda} \left\|p_\theta(t,Y) - p_{\theta'}(t,Y^{'})\right\|_2
\end{array}$$

Gronwall lemma and Cauchy-Schwarz inequality gives us:

$$\begin{array}{rcl}
\left\|X_{\theta,\overline{u}}(t,Y) - X_{\theta',\overline{u}}(t,Y^{'})\right\|_2 & \leq & \frac{\sqrt{d} d\|B\|_2^2}{2\lambda} \int_0^t e^{\sqrt{d f_x}(t-s)} \left\|p_\theta(s,Y) - p_{\theta'}(s,Y^{'})\right\|_2 ds \\
& + & \left\|x_0 - x_0^{'}\right\|_2 + \int_0^t e^{\sqrt{d f_x}(t-s)} \left(\sqrt{d f_x} s \left\|x_0 - x_0^{'}\right\|_2 + \overline{f}_\theta \left\|\theta - \theta^{'}\right\|_2\right) ds \\
& \leq & \frac{\sqrt{d} d\|B\|_2^2}{2\lambda} \sqrt{T} e^{\sqrt{d f_x} T} \left\|p_\theta(.,Y) - p_{\theta'}(.,Y^{'})\right\|_{L^2} \\
& + & \overline{f}_\theta \sqrt{T} e^{\sqrt{d f_x} T} \left\|\theta - \theta^{'}\right\|_2 + \left(\sqrt{d f_x} e^{\sqrt{d f_x} T} t + 1\right) \left\|x_0 - x_0^{'}\right\|_2
\end{array}$$

We finally obtain:

$$\begin{aligned}\left\|X_{\theta,\overline{u}}(.,Y)-X_{\theta^{'},\overline{u}}(.,Y^{'})\right\|_{L^2} &\leq \frac{\sqrt{d}d\|B\|_2^2}{2\lambda}Te^{\sqrt{d\overline{f}_x}T}\left\|p_{\theta}(.,Y)-p_{\theta^{'}}(.,Y^{'})\right\|_{L^2}\\ &+ \overline{f}_{\theta}Te^{\sqrt{d\overline{f}_x}T}\left\|\theta-\theta^{'}\right\|_2+\sqrt{T}\left(\sqrt{dT}\overline{f}_x e^{\sqrt{d\overline{f}_x}T}+1\right)\left\|x_0-x_0^{'}\right\|_2\end{aligned} \tag{7.9}$$

Secondly we will bound $\|p_{\theta}(.,Y)-p_{\theta}(.,Y')\|_{L^2}$ w.r.t$\left\|\theta-\theta^{'}\right\|_2$ and $\left\|Y-Y^{'}\right\|_{L^2}$. For this we have for the reversed time adjoint equations:

$$\begin{aligned}\dot{p_{\theta}^i}(t,Y)-\dot{p^i}_{\theta^{'}}(t,Y^{'}) &= \frac{\partial f}{\partial x}(T-t,X^i_{\theta,\overline{u}}(t,Y),\theta)^T\left(p_{\theta}^i(t,Y)-p^i_{\theta^{'}}(t,Y^{'})\right)\\ &+ \left(\frac{\partial f}{\partial x}(T-t,X^i_{\theta,\overline{u}}(t,Y),\theta)-\frac{\partial f}{\partial x}(T-t,X^i_{\theta^{'},\overline{u}}(t,Y^{'}),\theta^{'})\right)^T p^i_{\theta^{'}}(t,Y^{'})\\ &+ 2C^TC\left(X^i_{\theta^{'},\overline{u}}(t,Y^{'})-X^i_{\theta,\overline{u}}(t,Y)\right)\\ &+ 2C^T\left(Y^i(t)-Y^{i'}(t)\right)\end{aligned} \tag{7.10}$$

Taking the Frobenius norm we have the following inequality:

$$\begin{aligned}\left\|\dot{p_{\theta}^i}(t,Y)-\dot{p^i_{\theta^{'}}}(t,Y^{'})\right\|_2 &\leq \sqrt{d}\left\|\frac{\partial f}{\partial x}(T-t,X^i_{\theta,\overline{u}}(t,Y),\theta)\right\|_2\left\|p_{\theta}^i(t,Y)-p^i_{\theta^{'}}(t,Y^{'})\right\|_2\\ &+ \sqrt{d}\left\|p^i_{\theta^{'}}(t,Y^{'})\right\|_2\left\|\frac{\partial f}{\partial x}(T-t,X^i_{\theta,\overline{u}}(t,Y),\theta)-\frac{\partial f}{\partial x}(T-t,X^i_{\theta^{'},\overline{u}}(t,Y^{'}),\theta^{'})\right\|_2\\ &+ 2\sqrt{d}\left\|C\right\|_2^2\left\|X^i_{\theta^{'},\overline{u}}(t,Y^{'})-X^i_{\theta,\overline{u}}(t,Y)\right\|_2\\ &+ 2\sqrt{d}\left\|C\right\|_2\left\|Y^i(t)-Y^{i'}(t)\right\|_2\\ &\leq \sqrt{d}\left(\left\|p_{\theta}^i(t,Y^{'})\right\|_2\overline{f_{xx}}+2\left\|C\right\|_2^2\right)\left\|X^i_{\theta,\overline{u}}(t,Y)-X^i_{\theta^{'},\overline{u}}(t,Y^{'})\right\|_2\\ &+ \sqrt{d}\left(\overline{f_x}\left\|p_{\theta}^i(t,Y)-p^i_{\theta^{'}}(t,Y^{'})\right\|_2+\left\|p^i_{\theta^{'}}(t,Y^{'})\right\|_2\overline{f_{x\theta}}\left\|\theta-\theta^{'}\right\|_2\right)\\ &+ 2\sqrt{d}\left\|C\right\|_2\left\|Y^i(t)-Y^{i'}(t)\right\|_2\end{aligned}$$

By using Gronwall's lemma we obtain:

$$\begin{aligned}\left\|p_{\theta}^i(t,Y)-p^i_{\theta^{'}}(t,Y^{'})\right\|_2 &\leq \sqrt{d}e^{\sqrt{d\overline{f}_x}T}\int_0^t\left(\overline{f_{xx}}\left\|p_{\theta}^i(s,Y^{'})\right\|_2+2\left\|C\right\|_2^2\right)\left\|X^i_{\theta,\overline{u}}(s,Y)-X^i_{\theta^{'},\overline{u}}(s,Y^{'})\right\|_2 ds\\ &+ \sqrt{d\overline{f}_{x\theta}}e^{\sqrt{d\overline{f}_x}T}\left\|\theta-\theta^{'}\right\|_2\int_0^t\left\|p_{\theta}^i(s,Y^{'})\right\|_2 ds\\ &+ 2\sqrt{d}\left\|C\right\|_2 e^{\sqrt{d\overline{f}_x}T}\int_0^t\left\|Y^i(s)-Y^{i'}(s)\right\|_2 ds\end{aligned}$$

We need a bound for $\|p_{\theta}(s,Y)\|_2$ with uniform control w.r.t by $E(Y^*)$ and the discrepancy $\|Y-Y^*\|_{L^2}$ but so far all we have is $\|p_{\theta}(s,Y)\|_2\leq E(Y)$. For finding such a bound we need to consider the special case where $Y^{'}=Y^*$.

Using the previous inequality we have:

$$\begin{aligned}\left\|p_{\theta}^i(t,Y)-p^i_{\theta^{'}}(t,Y^*)\right\|_2 &\leq \sqrt{d}e^{\sqrt{d\overline{f}_x}T}\int_0^t\left(\overline{f_{xx}}\left\|p^i_{\theta^{'}}(t,Y^*)\right\|_2+2\left\|C\right\|_2^2\right)\left\|X^i_{\theta,\overline{u}}(s,Y)-X^i_{\theta^{'},\overline{u}}(s,Y^*)\right\|_2 ds\\ &+ \sqrt{d\overline{f}_{x\theta}}e^{\sqrt{d\overline{f}_x}T}\left\|\theta-\theta^{'}\right\|_2\int_0^t\left\|p^i_{\theta^{'}}(s,Y^*)\right\|_2 ds\\ &+ 2\sqrt{d}\left\|C\right\|_2 e^{\sqrt{d\overline{f}_x}T}\int_0^t\left\|Y^i(s)-Y^{i*}(s)\right\|_2 ds\\ &\leq \sqrt{dT}e^{\sqrt{d\overline{f}_x}T}\left(\overline{f_{xx}}E(Y^*)+2\left\|C\right\|_2^2\right)\left\|X^i_{\theta,\overline{u}}(.,Y)-X^i_{\theta^{'},\overline{u}}(.,Y^*)\right\|_{L^2}\\ &+ \sqrt{dT\overline{f}_{x\theta}}e^{\sqrt{d\overline{f}_x}T}E(Y^*)\left\|\theta-\theta^{'}\right\|_2\\ &+ 2\sqrt{dT}\left\|C\right\|_2 e^{\sqrt{d\overline{f}_x}T}\left\|Y-Y^*\right\|_{L^2}\end{aligned}$$

And so we obtain the L^2-discrepancy:

$$\begin{aligned}\|p_\theta(.,Y) - p_{\theta'}(.,Y^*)\|_{L^2} &\leq \sqrt{d}Te^{\sqrt{df_x}T}\left(\overline{f_{xx}}E(Y^*) + 2\|C\|_2^2\right)\left\|X^i_{\theta,\overline{u}}(.,Y) - X^i_{\theta',\overline{u}}(.,Y^*)\right\|_{L^2} \\ &+ \sqrt{d}T\overline{f_{x\theta}}e^{\sqrt{df_x}T}E(Y^*)\left\|\theta - \theta'\right\|_2 \\ &+ 2\sqrt{d}T\|C\|_2 e^{\sqrt{df_x}T}\|Y - Y^*\|_{L^2}\end{aligned} \tag{7.11}$$

By reinjecting in (7.9) we have:

$$\begin{aligned}\left\|X_{\theta,\overline{u}}(.,Y) - X_{\theta',\overline{u}}(.,Y^*)\right\|_{L^2} &\leq \tfrac{F(Y^*)}{\lambda}\left\|X_{\theta,\overline{u}}(.,Y) - X_{\theta',\overline{u}}(.,Y^*)\right\|_{L^2} \\ &+ \left(\tfrac{K_2}{\lambda} + K_3\right)\left\|\theta - \theta'\right\|_2 + \tfrac{K_1}{\lambda}\|Y - Y^*\|_{L^2} \\ &+ \sqrt{T}\left(\sqrt{dT}\overline{f_x}e^{\sqrt{df_x}T} + 1\right)\left\|x_0 - x_0'\right\|_2\end{aligned}$$

and thus for $\lambda > F(Y^*)$ we obtain:

$$\left\|X_{\theta,\overline{u}} - X^*_{\theta',\overline{u}}\right\|_{L^2} \leq \frac{K_2 + \lambda K_3}{\lambda - F(Y^*)}\left\|\theta - \theta'\right\|_2 + \frac{\sqrt{T}\left(\sqrt{dT}\overline{f_x}e^{\sqrt{df_x}T} + 1\right)}{\lambda - F(Y^*)}\left\|x_0 - x_0'\right\|_2 + \frac{K_1}{\lambda - F(Y^*)}\|Y - Y^*\|_{L^2} \tag{7.12}$$

From this previous inequality and Gronwall's lemma we can now obtain a bound for $\left\|p^i_\theta(t,Y)\right\|_2$:

$$\begin{aligned}\left\|p^i_\theta(t,Y)\right\|_2 &\leq 2\sqrt{d}\|C\|_2\int_0^t e^{\sqrt{df_x}(t-s)}\left\|Y(T-s) - CX^i_{\theta,\overline{u}}(s,Y)\right\|_2 ds \\ &\leq 2\sqrt{d}\|C\|_2\sqrt{\tfrac{e^{2\sqrt{df_x}t}-1}{2\sqrt{df_x}}}\sqrt{\int_0^t\left\|Y(T-s) - CX^i_{\theta,\overline{u}}(s,Y)\right\|_2^2 ds} \\ &\leq 2\sqrt{d}\|C\|_2\sqrt{\tfrac{e^{2\sqrt{df_x}t}-1}{2\sqrt{df_x}}}\|Y - CX_{\theta,\overline{u}}(.,Y)\|_{L^2} \\ &\leq 2\sqrt{d}\|C\|_2\sqrt{\tfrac{e^{2\sqrt{df_x}t}-1}{2\sqrt{df_x}}}\left(\|Y - Y^*\|_{L^2} + \left\|Y^* - CX^*_{\theta,\overline{u}}\right\|_{L^2} + \left\|CX^*_{\theta,\overline{u}} - CX_{\theta,\overline{u}}(.,Y)\right\|_{L^2}\right) \\ &\leq 2\sqrt{d}\|C\|_2\sqrt{\tfrac{e^{2\sqrt{df_x}t}-1}{2\sqrt{df_x}}}\left(D(Y^*) + \left(\tfrac{\sqrt{d}\|C\|_2K_1}{\lambda - F(Y^*)} + 1\right)\|Y - Y^*\|_{L^2}\right) \\ &\leq E(Y^*) + K_4\|Y - Y^*\|_{L^2}\end{aligned}$$

In the last inequality, the discrepancy $\left\|CX^*_{\theta,\overline{u}} - CX_{\theta,\overline{u}}(.,Y)\right\|$ has been bound by using the inequality (7.12).

We can now control $\left\|p^i_\theta(t,Y) - p^i_{\theta'}(t,Y')\right\|_2$ in the general case:

$$\begin{aligned}\left\|p^i_\theta(t,Y) - p^i_{\theta'}(t,Y')\right\|_2 &\leq \sqrt{d}e^{\sqrt{df_x}T}\int_0^t\left(\overline{f_{xx}}\left\|p^i_\theta(s,Y')\right\|_2 + 2\left\|C^TC\right\|_2\right)\left\|X^i_{\theta,\overline{u}}(s,Y) - X^i_{\theta',\overline{u}}(s,Y')\right\|_2 ds \\ &+ \sqrt{d}\overline{f_{x\theta}}\left\|\theta - \theta'\right\|_2 e^{\sqrt{df_x}T}\int_0^t\left\|p^i_{\theta'}(t,Y')\right\|_2 ds \\ &+ 2\sqrt{d}\|C\|_2 e^{\sqrt{df_x}T}\int_0^t\left\|Y^i(s) - Y^{i'}(s)\right\|_2 ds \\ &\leq \sqrt{dT}\overline{f_{xx}}\left(E(Y^*) + K_4\left\|Y' - Y^*\right\|_{L^2}\right)e^{\sqrt{df_x}T}\left\|X^i_{\theta,\overline{u}}(.,Y) - X^i_{\theta',\overline{u}}(.,Y')\right\|_{L^2} \\ &+ 2\sqrt{dT}\|C\|_2^2 e^{\sqrt{df_x}T}\left\|X^i_{\theta,\overline{u}}(.,Y) - X^i_{\theta',\overline{u}}(.,Y')\right\|_{L^2} \\ &+ \sqrt{dT}\overline{f_{x\theta}}\left\|\theta - \theta'\right\|_2 e^{\sqrt{df_x}T}\left(E(Y^*) + K_4\left\|Y' - Y^*\right\|_{L^2}\right) \\ &+ 2\sqrt{dT}\|C\|_2 e^{\sqrt{df_x}T}\left\|Y - Y'\right\|_{L^2}\end{aligned}$$

Taking the L^2 norm finally gives us:

$$\begin{aligned}
\left\|p_\theta(.,Y) - p_{\theta'}(.,Y^{'})\right\|_{L^2} &\leq \sqrt{d}T\overline{f_{xx}}\left(E(Y^*) + K_4\left\|Y^{'} - Y^*\right\|_{L^2}\right)e^{\sqrt{d f_x}T}\left\|X^i_{\theta,\overline{u}}(.,Y) - X^i_{\theta',\overline{u}}(.,Y^{'})\right\|_{L^2} \\
&+ 2\sqrt{d}T\left\|C\right\|_2^2 e^{\sqrt{d f_x}T}\left\|X^i_{\theta,\overline{u}}(.,Y) - X^i_{\theta',\overline{u}}(.,Y^{'})\right\|_{L^2} \\
&+ \sqrt{d}T\overline{f_{x\theta}}\left\|\theta - \theta^{'}\right\|_2 e^{\sqrt{d f_x}T}\left(E(Y^*) + K_4\left\|Y^{'} - Y^*\right\|_{L^2}\right) \\
&+ 2\sqrt{d}T\left\|C\right\|_2 e^{\sqrt{d f_x}T}\left\|Y - Y^{'}\right\|_{L^2}
\end{aligned} \tag{7.13}$$

By reinjecting in (7.9) we obtain:

$$\begin{aligned}
\left\|X_{\theta,\overline{u}}(.,Y) - X_{\theta',\overline{u}}(.,Y^{'})\right\|_{L^2} &\leq \frac{F(Y^*)+K_5\left\|Y^{'}-Y^*\right\|_{L^2}}{\lambda}\left\|X_{\theta,\overline{u}}(.,Y) - X_{\theta',\overline{u}}(.,Y^{'})\right\|_{L^2} \\
&+ \left(\frac{K_2+\left\|Y^{'}-Y^*\right\|_{L^2}K_6}{\lambda} + K_3\right)\left\|\theta - \theta^{'}\right\|_2 + \frac{K_1}{\lambda}\left\|Y - Y^{'}\right\|_{L^2} \\
&+ \sqrt{T}\left(\sqrt{d}T\overline{f_x}e^{\sqrt{d f_x}T} + 1\right)\left\|x_0 - x_0^{'}\right\|_2
\end{aligned}$$

By taking $\lambda > F(Y^*) + \left\|Y^{'} - Y^*\right\|_{L^2} K_5$ we have:

$$\begin{aligned}
\left\|X_{\theta,\overline{u}}(.,Y) - X_{\theta',\overline{u}}(.,Y^{'})\right\|_{L^2} &\leq \frac{K_2+\left\|Y^{'}-Y^*\right\|_{L^2}K_6+\lambda K_3}{\lambda - F(Y^*) - K_5\left\|Y^{'}-Y^*\right\|_{L^2}}\left\|\theta - \theta^{'}\right\|_2 \\
&+ \frac{\sqrt{T}\left(\sqrt{d}T\overline{f_x}e^{\sqrt{d f_x}T}+1\right)}{\lambda - F(Y^*) - \left\|Y^{'}-Y^*\right\|_{L^2}K_5}\left\|x_0 - x_0^{'}\right\|_2 \\
&+ \frac{K_1}{\lambda - F(Y^*) - \left\|Y^{'}-Y^*\right\|_{L^2}K_5}\left\|Y - Y^{'}\right\|_{L^2}
\end{aligned}$$

and by reinjecting in (7.13) we obtain:

$$\begin{aligned}
\left\|p_\theta(.,Y) - p_{\theta'}(.,Y^{'})\right\|_{L^2} &\leq K_7\frac{\left(\overline{f_{xx}}\left(E(Y^*)+K_4\left\|Y^{'}-Y^*\right\|_{L^2}\right)+2\left\|C\right\|_2^2\right)\left(K_2+\left\|Y^{'}-Y^*\right\|_{L^2}K_6+\lambda K_3\right)}{\lambda - F(Y^*) - K_5\left\|Y^{'}-Y^*\right\|_{L^2}}\left\|\theta - \theta^{'}\right\|_2 \\
&+ K_8\left(E(Y^*) + K_4\left\|Y^{'} - Y^*\right\|_{L^2}\right)\left\|\theta - \theta^{'}\right\|_2 \\
&+ \frac{K_9\left(\overline{f_{xx}}\left(E(Y^*)+K_4\left\|Y^{'}-Y^*\right\|_{L^2}\right)+2\left\|C\right\|_2^2\right)}{\lambda - F(Y^*) - K_5\left\|Y^{'}-Y^*\right\|_{L^2}}\left\|x_0 - x_0^{'}\right\|_2 \\
&+ \left(K_{10} + \frac{K_{11}\left(\overline{f_{xx}}\left(E(Y^*)+K_4\left\|Y^{'}-Y^*\right\|_{L^2}\right)+2\left\|C\right\|_2^2\right)}{\lambda - F(Y^*) - \left\|Y^{'}-Y^*\right\|_{L^2}K_5}\right)\left\|Y - Y^{'}\right\|_{L^2}
\end{aligned}$$

We know K_5 is a decreasing function $w.r.t$ λ and the condition $\lambda > F(Y^*) + \left\|Y^{'} - Y^*\right\|_{L^2} K_5$ becomes:

$$\lambda > F(Y^*) + \frac{1}{2}\left(K_0\left\|Y^{'} - Y^*\right\|_{L^2} + \sqrt{\left(K_0\left\|Y^{'} - Y^*\right\|_{L^2} + 4\sqrt{d}\left\|C\right\|_2 K_1\right)K_0\left\|Y^{'} - Y^*\right\|_{L^2}}\right)$$

$\square$

Lemma 7.6. *Let us consider* $\begin{pmatrix} \partial X(t,\theta) \\ \partial P(t,\theta) \end{pmatrix}$ *solution of the linear ODE with boundary condition:*

$$\begin{cases} \frac{d}{dt}\begin{pmatrix} \partial X(t,\theta) \\ \partial P(t,\theta) \end{pmatrix} = \begin{pmatrix} A_1(t,\theta) & A_2(t,\theta) \\ A_3(t,\theta) & A_4(t,\theta) \end{pmatrix}\begin{pmatrix} \partial X(t,\theta) \\ \partial P(t,\theta) \end{pmatrix} + B(t,\theta) \\ \begin{pmatrix} \partial X(0,\theta) \\ \partial P(T,\theta) \end{pmatrix} = \begin{pmatrix} \partial X_0 \\ \partial P_T \end{pmatrix} \end{cases}$$

with θ a (possibly infinite dimensional) parameter belonging to a normed space $(S_\theta,\ \|.\|_\theta)$.

If $\theta \longmapsto \begin{pmatrix} A_1(.,\theta) & A_2(.,\theta) \\ A_3(.,\theta) & A_4(.,\theta) \end{pmatrix}$ is continuous with $\overline{A_i} = \sup_{[0,\,T]\times S_\theta} |A_i(t,\theta)| < +\infty$, $\theta \longmapsto B(.,\theta)$ is continuous with $\|B(.,\theta)\|_{L^2}$ uniformly bounded on S_θ and $dT^2\overline{A_2}\,\overline{A_3}e^{\sqrt{d}T\left(\overline{A_1}+\overline{A_4}\right)} < 1$ then the solution $\begin{pmatrix} \partial X(.,\theta) \\ \partial P(.,\theta) \end{pmatrix}$ is unique for each θ, is uniformly bounded on $[0,\,T]\times S_\theta$ and $\theta \longmapsto \begin{pmatrix} \partial X(.,\theta) \\ \partial P(.,\theta) \end{pmatrix}$ is continuous on S_θ.

Proof. We have:

$$\begin{aligned} \tfrac{d}{dt}\|\partial X(t,\theta)\|_2 &\leq \sqrt{d}\overline{A_1}\,\|\partial X(t,\theta)\|_2 + \sqrt{d}\overline{A_2}\,\|\partial P(t,\theta)\|_2 + \|B_1(t,\theta)\|_2 \\ \tfrac{d}{dt}\|\partial P(t,\theta)\|_2 &\leq \sqrt{d}\overline{A_3}\,\|\partial X(t,\theta)\|_2 + \sqrt{d}\overline{A_4}\,\|\partial P(t,\theta)\|_2 + \|B_2(t,\theta)\|_2 \end{aligned}$$

Gronwall's lemma gives us:

$$\begin{aligned} \|\partial X(t,\theta)\|_2 &\leq e^{\sqrt{d}\overline{A_1}T}\int_0^t \left(\sqrt{d}\overline{A_2}\,\|\partial P(t,\theta)\|_2 + \|B_1(t,\theta)\|_2\right) ds \\ \|\partial P(T-t,\theta)\|_2 &\leq e^{\sqrt{d}\overline{A_4}T}\int_0^t \left(\sqrt{d}\overline{A_3}\,\|\partial X(T-t,\theta)\|_2 + \|B_2(T-t,\theta)\|_2\right) ds \end{aligned}$$

By taking the L^2 norm we obtain:

$$\begin{aligned} \|\partial X(.,\theta)\|_{L^2} &\leq \sqrt{d}T\overline{A_2}e^{\sqrt{d}\overline{A_1}T}\,\|\partial P(.,\theta)\|_{L^2} + Te^{\sqrt{d}\overline{A_1}T}\,\|B_1(.,\theta)\|_{L^2} \\ \|\partial P(.,\theta)\|_{L^2} &\leq \sqrt{d}T\overline{A_3}e^{\sqrt{d}\overline{A_4}T}\,\|\partial X(.,\theta)\|_{L^2} + Te^{\sqrt{d}\overline{A_4}T}\,\|B_2(.,\theta)\|_{L^2} \end{aligned}$$

The condition $dT^2\overline{A_2}\,\overline{A_3}e^{\sqrt{d}T\left(\overline{A_1}+\overline{A_4}\right)} < 1$ implies the uniform boundedness of $\|\partial X(.,\theta)\|_{L^2}$ on S_θ and so of $\|\partial P(.,\theta)\|_{L^2}$.

The inequalities:

$$\begin{aligned} \|\partial X(t,\theta)\|_2 &\leq \sqrt{dT}\overline{A_2}e^{\sqrt{d}\overline{A_1}T}\,\|\partial P(.,\theta)\|_{L^2} + \sqrt{T}e^{\sqrt{d}\overline{A_1}T}\,\|B_1(.,\theta)\|_{L^2} \\ \|\partial P(T-t,\theta)\|_2 &\leq \sqrt{dT}\overline{A_3}e^{\sqrt{d}\overline{A_4}T}\,\|\partial X(.,\theta)\|_{L^2} + \sqrt{T}e^{\sqrt{d}\overline{A_4}T}\,\|B_2(.,\theta)\|_{L^2} \end{aligned}$$

hence implies the existence of uniform bounds $\overline{\partial X}$ and $\overline{\partial P}$.

We have:

$$\begin{aligned} \dot{\partial X}(t,\theta) - \dot{\partial X}(t,\theta') &= A_1(t,\theta)\partial X(t,\theta) - A_1(t,\theta')\partial X(t,\theta') \\ &+ A_2(t,\theta)\partial P(t,\theta) - A_2(t,\theta')\partial P(t,\theta') \\ &+ B_1(t,\theta) - B_1(t,\theta') \end{aligned}$$

and so:

$$\begin{array}{rcl}\frac{d}{dt}\left\|\partial X(t,\theta)-\partial X(t,\theta^{'})\right\|_2 & \leq & \sqrt{dA_1}\left\|\partial X(t,\theta)-\partial X(t,\theta^{'})\right\|_2+\sqrt{d\partial X}\left\|A_1(t,\theta)-A_1(t,\theta^{'})\right\|_2 \\ & + & \sqrt{dA_2}\left\|\partial P(t,\theta)-\partial P(t,\theta^{'})\right\|_2+\sqrt{d\partial P}\left\|A_2(t,\theta)-A_2(t,\theta^{'})\right\|_2 \\ & + & \left\|B_1(t,\theta)-B_1(t,\theta^{'})\right\|_2\end{array}$$

Gronwall's lemma gives us:

$$\begin{array}{rcl}\frac{d}{dt}\left\|\partial X(t,\theta)-\partial X(t,\theta^{'})\right\|_2 & \leq & \sqrt{d\partial X}e^{\sqrt{dA_1}T}\int_0^t\left\|A_1(t,\theta)-A_1(t,\theta^{'})\right\|_2 ds \\ & + & \sqrt{dA_2}e^{\sqrt{dA_1}T}\int_0^t\left\|\partial P(t,\theta)-\partial P(t,\theta^{'})\right\|_2 ds \\ & + & \sqrt{d\partial P}e^{\sqrt{dA_1}T}\int_0^t\left\|A_2(t,\theta)-A_2(t,\theta^{'})\right\|_2 ds \\ & + & e^{\sqrt{dA_1}T}\int_0^t\left\|B_1(t,\theta)-B_1(t,\theta^{'})\right\|_2 ds\end{array}$$

By passing through similar computations we obtain the following bound for $\left\|\partial P(t,\theta)-\partial P(t,\theta^{'})\right\|_2$:

$$\begin{array}{rcl}\left\|\partial P(t,\theta)-\partial P(t,\theta^{'})\right\|_2 & \leq & \sqrt{d\partial X}e^{\sqrt{dA_4}T}\int_0^t\left\|A_3(s,\theta)-A_3(s,\theta^{'})\right\|_2 ds \\ & + & \sqrt{dA_3}e^{\sqrt{dA_4}T}\int_0^t\left\|\partial X(s,\theta)-\partial X(s,\theta^{'})\right\|_2 ds \\ & + & \sqrt{d\partial P}e^{\sqrt{dA_4}T}\int_0^t\left\|A_4(s,\theta)-A_4(s,\theta^{'})\right\|_2 ds \\ & + & e^{\sqrt{dA_4}T}\int_0^t\left\|B_2(s,\theta)-B_2(s,\theta^{'})\right\|_2 ds\end{array}$$

Using Cauchy-Schwarz inequality and taking the L^2 norm eventually lead us to:

$$\begin{array}{rcl}\left\|\partial X(.,\theta)-\partial X(.,\theta^{'})\right\|_{L^2} & \leq & \sqrt{dT\overline{\partial X}}e^{\sqrt{dA_1}T}\left\|A_1(.,\theta)-A_1(.,\theta^{'})\right\|_{L^2} \\ & + & \sqrt{dT\overline{A_2}}e^{\sqrt{dA_1}T}\left\|\partial P(.,\theta)-\partial P(.,\theta^{'})\right\|_{L^2} \\ & + & \sqrt{dT\overline{\partial P}}e^{\sqrt{dA_1}T}\left\|A_2(.,\theta)-A_2(.,\theta^{'})\right\|_{L^2} \\ & + & Te^{\sqrt{dA_1}T}\left\|B_1(.,\theta)-B_1(.,\theta^{'})\right\|_{L^2} \\ & \leq & \sqrt{dT\overline{A_2}}e^{\sqrt{dA_1}T}\left\|\partial P(.,\theta)-\partial P(.,\theta^{'})\right\|_{L^2}+K_X\left\|\theta-\theta'\right\|\end{array}$$

with K_X a constant due to continuity and uniform boundedness on S_θ of A and B. We can also find a bound for $\left\|\partial P(.,\theta)-\partial P(.,\theta^{'})\right\|_{L^2}$:

$$\begin{array}{rcl}\left\|\partial P(.,\theta)-\partial P(.,\theta^{'})\right\|_{L^2} & \leq & \sqrt{dT\overline{\partial X}}e^{\sqrt{dA_4}T}\left\|A_3(.,\theta)-A_3(.,\theta^{'})\right\|_{L^2} \\ & + & \sqrt{dT\overline{A_3}}e^{\sqrt{dA_4}T}\left\|\partial X(.,\theta)-\partial X(.,\theta^{'})\right\|_{L^2} \\ & + & \sqrt{d\overline{\partial P}}e^{\sqrt{dA_4}T}\left\|A_4(.,\theta)-A_4(.,\theta^{'})\right\|_{L^2} \\ & + & Te^{\sqrt{dA_4}T}\left\|B_2(.,\theta)-B_2(.,\theta^{'})\right\|_{L^2} \\ & \leq & \sqrt{dT\overline{A_3}}e^{\sqrt{dA_4}T}\left\|\partial X(.,\theta)-\partial X(.,\theta^{'})\right\|_{L^2}+K_P\left\|\theta-\theta'\right\|\end{array}$$

with K_P a constant, by reinjecting the bound found for $\left\|\partial P(.,\theta)-\partial P(.,\theta^{'})\right\|_{L^2}$ in the one for $\left\|\partial X(.,\theta)-\partial X(.,\theta^{'})\right\|_{L^2}$ we obtain:

$$\begin{array}{rcl}\left\|\partial X(.,\theta)-\partial X(.,\theta^{'})\right\|_{L^2} & \leq & dT^2\overline{A_2}\,\overline{A_3}e^{\sqrt{dT}\left(\overline{A_1}+\overline{A_4}\right)}\left\|\partial X(.,\theta)-\partial X(.,\theta^{'})\right\|_{L^2} \\ & + & \left(\sqrt{dT\overline{A_2}}e^{\sqrt{dA_1}T}K_P+K_X\right)\left\|\theta-\theta'\right\|\end{array}$$

which gives the uniqueness of the solution and the continuity of $\theta \longmapsto \partial X(.,\theta)$ and hence of $\theta \longmapsto \partial P(.,\theta)$ using the upper bound of $\left\|\partial P(.,\theta) - \partial P(.,\theta')\right\|_{L^2}$ for $dT^2\overline{A_2}\overline{A_3}e^{\sqrt{dT}\left(\overline{A_1}+\overline{A_4}\right)} < 1$. $\square$

8 Adjoint method for gradient Computation

8.1 Profiled cost and ODE (3.2) reformulation

In order to use adjoint method we reformulate the profiled cost under the integral form $\mathcal{S}(\hat{Y};\theta,\lambda) = \int_0^T l(R_\theta(s),s)ds$ as well as the ODE (3.2) under the form: $\dot{R}_\theta(t) = F(R_\theta(t),\theta,t)$.

By posing:

$$\begin{array}{lcl} R_\theta(t) & = & \left(X_{\theta,\overline{u}}(t)^T, p_\theta(t)^T\right)^T \\ l(R_\theta(t),t) & = & \left(CX_{\theta,\overline{u}}(t) - \widehat{Y}(t)\right)^T \left(CX_{\theta,\overline{u}}(t) - \widehat{Y}(t)\right) + \frac{1}{4\lambda}p_\theta(t)^T BB^T p_\theta(t) \\ F(R_\theta(t),\theta,t) & = & \left(\begin{array}{c} f(t,X_{\theta,\overline{u}}(t),\theta) + \frac{1}{2\lambda}BB^T p_\theta(t) \\ -\frac{\partial f}{\partial x}(t,X_{\theta,\overline{u}}(t),\theta)^T p_\theta(t) + 2C^T\left(CX_{\theta,\overline{u}}(t) - \widehat{Y}(t)\right) \end{array}\right) \end{array}.$$

We also need to compute their derivatives:

$$\frac{\partial F}{\partial \theta}(R_\theta(t),\theta,t) = \left(\begin{array}{c} \frac{\partial f}{\partial \theta}(t,X_{\theta,\overline{u}}(t),\theta) \\ -\sum_j \left(\frac{\partial f_j}{\partial \theta^T \partial X}(t,X_{\theta,\overline{u}}(t),\theta)p_{j,\theta}(t)\right) \end{array}\right).$$

$$\frac{\partial F}{\partial R}(R_\theta(t),\theta,t) = \left(\begin{array}{cc} \frac{\partial f}{\partial X}(t,X_{\theta,\overline{u}}(t),\theta) & \frac{1}{2\lambda}BB^T \\ -\sum_j \left(\frac{\partial f_j}{\partial X^T \partial X}(t,X_{\theta,\overline{u}}(t),\theta)p_{j,\theta}(t)\right) + 2C^TC & -\frac{\partial f}{\partial X}(t,X_{\theta,\overline{u}}(t),\theta)^T \end{array}\right).$$

$$\frac{\partial l(R_\theta(t),t)}{\partial R} = \left(\begin{array}{c} 2C^T\left(CX_{\theta,\overline{u}}(t) - \widehat{Y}(t)\right) \\ \frac{1}{2\lambda}BB^T p_\theta(t) \end{array}\right)^T.$$

8.2 Adjoint method

8.2.1 Known initial condition

The gradient of $\mathcal{S}(\hat{Y};\theta,\lambda)$ is expressed under the form:

$$\nabla_\theta \mathcal{S}(\hat{Y};\theta,\lambda) = \int_0^T \frac{\partial l}{\partial R}(R_\theta(s),s)\frac{\partial R_\theta(s)}{\partial \theta}ds$$

with $\frac{\partial R_\theta(t)}{\partial \theta}$ solution of the sensitivity equation:

$$\frac{d}{dt}\left(\frac{\partial R_\theta(t)}{\partial \theta}\right) = \frac{\partial F}{\partial R}(R_\theta(t),\theta,t)\frac{\partial R_\theta(t)}{\partial \theta} + \frac{\partial F}{\partial \theta}(R_\theta(t),\theta,t).$$

If we premultiply the right and left term of the previous ODE by the $2d-$sized adjoint vector $P(t) = (P_1(t),P_2(t))$ and then integrate we obtain

$$\int_0^T P(t).\frac{d}{dt}\left(\frac{\partial R_\theta(t)}{\partial \theta}\right)dt = \int_0^T P(t).\frac{\partial F}{\partial R}(R_\theta(t),\theta,t)\frac{\partial R_\theta(t)}{\partial \theta}dt + \int_0^T P(t).\frac{\partial F}{\partial \theta}(R_\theta(t),\theta,t)dt.$$

Integration by part gives us

$$\int_0^T P(t).\frac{d}{dt}(\frac{\partial R_\theta(t)}{\partial\theta})dt = P(T).\frac{\partial R_\theta(T)}{\partial\theta} - P(0).\frac{\partial R_\theta(0)}{\partial\theta} - \int_0^T \dot{P}(t).\frac{\partial R_\theta(t)}{\partial\theta}dt.$$

Since $\frac{\partial X_{\theta,\overline{u}}(0)}{\partial\theta} = 0$ and $\frac{\partial p_\theta(T)}{\partial\theta} = 0$ by developing by block we have:

$$\int_0^T P(t).\frac{d}{dt}(\frac{\partial R_\theta(t)}{\partial\theta})dt = P_1(T).\frac{\partial X_{\theta,\overline{u}}(T)}{\partial\theta} - P_2(0).\frac{\partial p_\theta(0)}{\partial\theta} - \int_0^T \dot{P}(t).\frac{\partial R_\theta(t)}{\partial\theta}dt.$$

So if we take $P_1(T) = 0$ and $P_2(0) = 0$ we obtain the variational relation:

$$\int_0^T \left(\dot{P}(t) + P(t).\frac{\partial F}{\partial R}(R_\theta(t),\theta,t)\right)\frac{\partial R_\theta(t)}{\partial\theta}dt + \int_0^T P(t).\frac{\partial F}{\partial\theta}(R_\theta(t),\theta,t)dt = 0$$

and by imposing:

$$\dot{P}(t) + P(t).\frac{\partial F}{\partial R}(R_\theta(t),\theta,t) = \frac{\partial l(R_\theta(t),t)}{\partial R}$$

we derive the expression:

$$\int_0^T \frac{\partial l(R_\theta(t),t)}{\partial R}\frac{\partial R_\theta(t)}{\partial\theta}dt = -\int_0^T P(t).\frac{\partial F}{\partial\theta}(R_\theta(t),\theta,t)dt$$

and so

$$\nabla_\theta \mathcal{S}(\hat{Y};\theta,\lambda) = -\int_0^T P(t).\frac{\partial F}{\partial\theta}(R_\theta(t),\theta,t)dt.$$

We can now compute $\nabla_\theta \mathcal{S}(\hat{Y};\theta,\lambda)$ by considering:

$$\begin{aligned}
&\nabla_\theta \mathcal{S}(\hat{Y};\theta,\lambda) = -\textstyle\int_0^T P(t).\frac{\partial F}{\partial\theta}(R_\theta(t),\theta,t)dt\\
&\begin{cases}
\dot{P}(t) = \frac{\partial l(R_\theta(t),t)}{\partial R} - P(t).\frac{\partial F}{\partial R}(R_\theta(t),\theta,t)\\
P_1(T) = 0\\
P_2(0) = 0
\end{cases}
\end{aligned}.$$

8.2.2 Unknown initial condition

In that case we have to consider the extended parameter set (θ, x_0) and the extended gradient:

$$\nabla_{(\theta,x_0)}\mathcal{S}(\hat{Y};\theta,x_0,\lambda) = \left(\nabla_\theta \mathcal{S}(\hat{Y};\theta,x_0,\lambda),\ \nabla_{x_0}\mathcal{S}(\hat{Y};\theta,x_0,\lambda)\right).$$

We have already obtained $\nabla_\theta \mathcal{S}(\hat{Y};\theta,x_0,\lambda)$ we now compute

$$\nabla_{x_0}\mathcal{S}(\hat{Y};\theta,x_0,\lambda) = \int_0^T \frac{\partial l}{\partial R}(R_\theta(s),s)\frac{\partial R_\theta(s)}{\partial x_0}ds$$

using adjoint method. If we premultiply the right and left term of the sensitivity ODE w.r.t x_0:

$$\frac{d}{dt}\left(\frac{\partial R_\theta(t)}{\partial x_0}\right) = \frac{\partial F}{\partial R}(R_\theta(t),\theta,t)\frac{\partial R_\theta(t)}{\partial x_0}$$

by the $2d-$sized adjoint vector $P(t) = (P_1(t), P_2(t))$ and then integrate we obtain:

$$\int_0^T P(t).\frac{d}{dt}(\frac{\partial R_\theta(t)}{\partial x_0})dt = \int_0^T P(t).\frac{\partial F}{\partial R}(R_\theta(t), \theta, t)\frac{\partial R_\theta(t)}{\partial x_0}dt.$$

Integration by part gives us

$$\int_0^T P(t).\frac{d}{dt}(\frac{\partial R_\theta(t)}{\partial x_0})dt = P(T).\frac{\partial R_\theta(T)}{\partial x_0} - P(0).\frac{\partial R_\theta(0)}{\partial x_0} - \int_0^T \dot{P}(t).\frac{\partial R_\theta(t)}{\partial x_0}dt.$$

Reminding that $\frac{\partial R_\theta(T)}{\partial x_0} = \left(\frac{\partial X_{\theta,\overline{u}}(0)}{\partial x_0}^T, 0\right)^T$ and $\frac{\partial R_\theta(0)}{\partial x_0} = \left(I_d, \frac{\partial p_\theta(0)}{\partial x_0}^T\right)^T$ we know if we take $P_1(T) = 0$ and $P_2(0) = 0$ we obtain the variational relation:

$$\int_0^T \left(\dot{P}(t) + P(t).\frac{\partial F}{\partial R}(R_\theta(t), \theta, t)\right)\frac{\partial R_\theta(t)}{\partial x_0}dt + P_1(0) = 0$$

and by imposing:

$$\dot{P}(t) + P(t).\frac{\partial F}{\partial R}(R_\theta(t), \theta, t) = \frac{\partial l(R_\theta(t), t)}{\partial R}$$

we derive from that

$$\int_0^T \frac{\partial l(R_\theta(t), t)}{\partial R}\frac{\partial R_\theta(t)}{\partial x_0}dt = -P_1(0)$$

and so

$$\nabla_{x_0}\mathcal{S}(\hat{Y}; \theta, x_0, \lambda) = -P_1(0).$$

We can now compute $\nabla_{x_0}\mathcal{S}(\hat{Y}; \theta, x_0, \lambda)$ by considering:

$$\begin{array}{l} \nabla_{x_0}\mathcal{S}(\hat{Y}; \theta, x_0, \lambda) = -P_1(0) \\ \begin{cases} \dot{P}(t) = \frac{\partial l(R_\theta(t), t)}{\partial R} - P(t).\frac{\partial F}{\partial R}(R_\theta(t), \theta, t) \\ P_1(T) = 0 \\ P_2(0) = 0 \end{cases} \end{array}.$$

Since the adjoint ODE has already been solved for parameter gradient computation, it does not require any extra cost to compute $\nabla_{x_0}\mathcal{S}(\hat{Y}; \theta, x_0, \lambda)$.

References

[1] H. Brezis. *Functional Analysis.* Dunod, 1983.

[2] Francis Clarke. *Functional Analysis, Calculus of Variations and Optimal Control.* Graduate Texts in Mathematics. Springer-Verlag London, 2013.

[3] R.A. DeVore and G.G. Lorentz. *Constructive Approximation (Grundlehren der mathematischen Wissenschaften)*, volume 303. Springer-Verlag, 1993.

[4] R. FitzHugh. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal*, 6:445–466, 1961.

[5] A.A. Milyutin and N.P. Osmolovskii. *Calculus of Variation and Optimal control.* Mathematical Monographs. American mathematical society, 1998.

[6] W. K. Newey. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79:147–168, 1997.

[7] J.O. Ramsay, G. Hooker, J. Cao, and D. Campbell. Parameter estimation for differential equations: A generalized smoothing approach. *Journal of the Royal Statistical Society (B)*, 69:741–796, 2007.

[8] E. Sontag. *Mathematical Control Theory: Deterministic finite-dimensional systems*. Springer-Verlag (New-York), 1998.

[9] G. Teschl. *Ordinary Differential Equations and Dynamical Systems*. Graduate Studies in Mathematics, Volume 140. American Mathematical Society, 2012.

[10] A.W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilities Mathematics. Cambridge University Press, 1998.