MIT Open Access Articles

Simplified Least Squares Shadowing sensitivity analysis for chaotic ODEs and PDEs

Simplified Least Squares Shadowing sensitivity analysis for chaotic ODEs and PDEs

Mario Chater[a,*], Angxiu Ni[a], Qiqi Wang[a]

[a]*Aeronautics and Astronautics, MIT, 77 Massachusetts Ave, Cambridge, MA 02139, USA*

Abstract

This paper develops a variant of the Least Squares Shadowing (LSS) method, which has successfully computed the derivative for several chaotic ODEs and PDEs. The development in this paper aims to simplify Least Squares Shadowing method by improving how time dilation is treated. Instead of adding an explicit time dilation term as in the original method, the new variant uses windowing, which can be more efficient and simpler to implement, especially for PDEs.

Keywords: sensitivity analysis, linear response, adjoint equation, unsteady adjoint, chaos, least squares shadowing

1. Introduction

In many engineering and non-engineering applications, quantities of interest are time averages of some specific instantaneous output and this output may be affected by parameters of the physical system/model. For example, when evaluating the performance of a wing, the time averaged drag it produces is of great importance for aerodynamicists. Any shape parameter of this wing (width, length, curvature...) clearly has an impact on the produced drag and could be consequently considered a parameter of the system. Computing the derivative of the quantity of interest (time-averaged output) with respect to the system's parameters is crucial in:

- **Numerical optimization**: the derivative feeds a gradient-based algorithm which is then used to optimize the systems parameters [1],[2],[3]

- **Uncertainty quantification**: the derivative itself gives a useful assessment of the sensitivity and/or uncertainty of the system with respect to its parameters [4],[5]

For chaotic systems such as those encountered in turbulent fluid flows, the so-called *butterfly effect* makes conventional tangent/adjoint algorithms fail when

---

*Corresponding author.

Email addresses: `chaterm@mit.edu` (Mario Chater), `niangxiu@mit.edu` (Angxiu Ni), `qiqi@mit.edu` (Qiqi Wang)

computing the derivative of the quantity of interest. Indeed, the computed derivatives of infinite-time averaged quantities are, in general, orders of magnitude bigger than the real ones [6]. Some methods have been proposed to overcome this problem [7],[8],[9]. *Least Squares Shadowing* (LSS) is a relatively simple and efficient algorithm that evaluates the desired derivative by solving a constrained least squares problem [10]. In the original version, corrective time dilation factors have to be introduced to ensure its convergence but these same terms lead to a heavier and more expensive system of equations to solve. Indeed, the presence of time dilation terms prohibits the use of sparse solvers in a crucial step of the LSS algorithm which consists of inverting a matrix. Since this matrix scales up with the dimension of the problem, it is necessary to address this issue if we want to improve our algorithm for very large, industrial-scale problems [11]. In this paper, we introduce a modified version of LSS in which the time dilation factors are replaced by a simple *windowing* procedure: the result of the non corrected system of equations is multiplied by a specific windowing function. This procedure has been used for systems that have a periodic behavior [12], but we will show that the fundamental idea is still valid in our framework and can be transposed to chaotic dynamical systems. Since this *"windowing"* operation is almost costless, the new algorithm is much more efficient than standard LSS, specially for high-dimensional systems.

The paper is organized as follows: in the next section, we introduce the new algorithm in both its tangent and adjoint versions. Then, in the third section, the underlying theory of the algorithm is presented and we prove that the result given by simplified LSS converges to the one given by the original algorithm when integration time goes to infinity. Finally, the last section is dedicated to a numerical application of the new algorithm using the Lorenz 63 test case dynamical system.

2. A summary of the tangent and adjoint versions of simplified Least Squares Shadowing

Consider a family of ergodic dynamical systems satisfying a differential equation parameterized by s :

$$\frac{du}{dt} = f(u, s) \tag{1}$$

where f is smooth with respect to both u and s. Ergodicity means that an infinite-time averaged quantity

$$\langle J \rangle(s) := \lim_{T \to \infty} \frac{1}{T} \int_0^T J(u, s) dt$$

where J is a measurable function and $u = u(t; s)$ is a trajectory for parameter s, does not depend on the initial condition of the trajectory $u(0; s)$. The derivative of such infinite-time averaged quantity can be approximated by the following three steps:

2

1. Integrate equation (1) until the initial transient behavior has passed. Then further integrate the equation over a period of T, which should be a multiple of the longest time scales in its solution. Store the trajectory of the second time integration as $u(t), 0 \le t \le T$.

2. Find the function $\check{v}(t), 0 \le t \le T$ which minimizes $\int_0^T |\check{v}(t)|^2 dt$ and satisfies the linearized equation

$$\frac{d\check{v}}{dt} = f_u \check{v} + f_s, \tag{2}$$

where the time-dependent Jacobians $f_u := \frac{\partial f}{\partial u}$ and $f_s := \frac{\partial f}{\partial s}$ are based on the trajectory obtained in the previous step. This $\check{v}(t)$ can be found by solving a system of linear equations derived from the KKT conditions of the constrained least squares problem :

$$\begin{cases} \frac{d\check{v}}{dt} = f_u \check{v} + f_s \\ \frac{d\check{w}}{dt} = -f_u^T \check{w} + \check{v} \\ \check{w}(0) = \check{w}(T) = 0 \end{cases} \tag{3}$$

Note that this is the Least Squares Shadowing problem in previous literature [10], but without a time-dilation term.

3. Approximate the derivative of $\langle J \rangle(s)$ by computing a windowed time-average :

$$\frac{d\langle J \rangle}{ds} \approx \frac{1}{T} \int_0^T \mathbf{w}(\tfrac{t}{T}) \left(J_u \check{v} + J_s \right) dt \tag{4}$$

where $\mathbf{w}$, the window function, is a scalar function in $[0, 1]$ satisfying
   (a) $\mathbf{w}$ continuously differentiable,
   (b) $\mathbf{w}(0) = \mathbf{w}(1) = 0$,
   (c) $\int_0^1 \mathbf{w}(r) dr = 1$.
An example satisfying all three criteria is $\mathbf{w}(r) = 1 - \cos 2\pi r$.

This resulting derivative approximation converges to the true derivative as $T \to \infty$, as mathematically derived in Section 3 and under the same assumptions[1] as the original Least Squares Shadowing method [13].

   In the algorithm above, the cost of solving the constrained least squares problem (3) scales with the dimension of s, and is independent from the dimension of J. Such an algorithm, which favors a low-dimensional parameter s and a high-dimensional quantity of interest J, is called the tangent (or forward) version of simplified LSS.

A corresponding adjoint (or backward) version can be derived, whose computation cost favors a high-dimensional s and a low-dimensional J. This adjoint algorithm, also consists of three steps :

---

[1]The essential assumption is for the dynamical system to be uniformly hyperbolic.

1. Obtain a trajectory $u(t)$, $0 \leq t \leq T$ in the same way as Step 1 of the previous algorithm.

2. Solve the system of linear equations

$$\begin{cases} \frac{d\hat{w}}{dt} = f_u \hat{w} \\ \frac{d\hat{v}}{dt} = -f_u^T \hat{v} - \hat{w} - \mathbf{w}(\frac{t}{T}) J_u^T \\ \hat{v}(0) = \hat{v}(T) = 0 \end{cases} \tag{5}$$

where $\mathbf{w}(r)$, $0 \leq r \leq 1$ is a scalar windowing function satisfying the criteria described in Step 3 of the previous algorithm. This system of linear differential equations is the dual of the system in Step 2 of the previous algorithm, derived by combining it with Equation (4) and integrating by parts. A complete derivation of the adjoint version can be found in Appendix A.

3. Approximate the derivative of $\langle J \rangle(s)$ by the following equation, derived together with Equations (5) in Step 2 :

$$\frac{d\langle J \rangle}{ds} \approx \frac{1}{T} \int_0^T \left(f_s^T \hat{v} + J_s \right) dt \tag{6}$$

We can show that the adjoint version of LSS (equations 5,6) produces the same estimation of $\frac{d\langle J \rangle}{ds}$ than the one given by the tangent version (equations 3, 4) up to round-off errors [10]. The approximated $\frac{d\langle J \rangle}{ds}$ should be the same if the differential equations are solved exactly in both algorithms, and the integrals are evaluated exactly. With such exact numerics, the error in the approximation is solely due to the infeasibility of using an infinite T, and should diminish as $T \to \infty$. This error depends on the trajectory $u(t)$, but does not depend on whether the tangent or adjoint algorithm is used. Throughout this paper, we analyze the tangent version of LSS, but the conclusions are also valid for the adjoint version since both algorithms give the same result up to round-off errors.

3. How windowing mitigates the effect of time dilation

The method introduced in this paper is similar to the original Least Squares Shadowing method, except for two major differences. The first difference is the use of a smooth windowing function $\mathbf{w}$ satisfying

$$\mathbf{w}(0) = \mathbf{w}(1) = 0 \quad \text{and} \quad \int_0^1 \mathbf{w}(r) dr = 1;$$

which averages to 1 and tapers off to 0 at both ends of the interval $[0, 1]$. The second difference is the lack of a time dilation term in Equation (3) (Equation (5) for the adjoint version). Removal of the time dilation term can simplify the implementation of a chaotic sensitivity analysis capability for many solvers.

To understand the removal of the time dilation term, we should first understand the original Least Squares Shadowing formulation [10], and why it has a time dilation term. The Least Squares Shadowing method approximates the derivative of a long time average via an exchange of limit and derivative :

$$\frac{d\langle J\rangle}{ds} := \frac{d}{ds}\lim_{T\to\infty}\frac{1}{T}\int_0^T J(u(t;s),s)dt = \lim_{T\to\infty}\frac{d}{ds}\frac{1}{T}\int_0^T J(u(t;s),s)dt \quad (7)$$

As a sufficient condition for this exchange, we require the total derivative of J with respect to s,

$$\frac{d}{ds}J(u(t;s),s) = J_u\frac{\partial u}{\partial s} + J_s$$

to be uniformly continuous. For this to hold, we ask for J to have uniformly continuous partial derivatives, and $\frac{\partial u}{\partial s}$ to be uniformly bounded and continuous. Among these conditions, it is most difficult to make $\frac{\partial u}{\partial s}$ uniformly bounded, when u is governed by a family of chaotic dynamical systems. $\frac{\partial u}{\partial s}$ describes the difference between two solutions that satisfy the governing equations with infinitesimally different parameters s and $s + \delta s$. It must satisfy the linearized governing equation

$$\frac{\partial \check{v}}{\partial t} = f_u\check{v} + f_s. \quad (8)$$

It is well known, as the "butterfly effect", that solutions to this linearized equation can grow exponentially as t increases. Under some assumptions on the dynamical system, the shadowing lemma ([14], [15]) ensures the existence of a non-exponentially growing solution. Nevertheless, this solution is potentially linearly growing and thus not uniformly bounded. For a uniformly bounded solution to exist, Equation (8) has to be slightly modified to :

$$\frac{\partial v}{\partial t} = f_uv + f_s + \eta f. \quad (9)$$

for some uniformly bounded scalar function of time η, [16]. The solution to this equation, v, describes the difference between two solutions that satisfy the governing equation with both infinitesimally different parameter s and infinitesimally different rate of time marching. Specifically, if $u(t)$ satisfies the equation

$$\frac{du}{dt} = f(u,s),$$

then $u(t) + \epsilon v(t)$ satisfies the equation

$$\frac{d(u+\epsilon v)}{(1+\epsilon\eta)dt} = f(u+\epsilon v, s+\epsilon)$$

The relative difference between the rates of time marching is quantified by η. The shadowing lemma ensures that there exists a pair of v and η that are both uniformly bounded in time and satisfy Equation (9). The original Least Squares Shadowing method approximates this uniformly bounded pair by minimizing the

5

norm of v and η under the constraint of Equation (9). The desired derivative is then computed by modifying Equation (7) into :

$$\frac{d\langle J\rangle}{ds} = \langle J_u v\rangle + \langle J_s\rangle + \langle \eta J\rangle - \langle \eta\rangle\langle J\rangle, \tag{10}$$

to account for the time dilation effect of η.

The new method in this paper avoids time dilation by only computing an approximation of v instead of the uniformly bounded pair (v, η). This new quantity denoted $\check{v}_{\check{\tau}}$ is equal to :

$$\check{v}_{\check{\tau}}(t) = v(t) - f(t)\int_{\check{\tau}}^{t} \eta(r)dr \tag{11}$$

for some $\check{\tau}$, where $f(t)$ denotes $f(u(t; s), s) = \frac{du}{dt}$. The subscript $\check{\tau}$ indicates that $\check{v}_{\check{\tau}}$ is "anchored" to v at time $\check{\tau}$ since $\check{v}_{\check{\tau}}(\check{\tau}) = v(\check{\tau})$. As we will see later on in section 4 (figure 2), the norm of $\check{v}_{\check{\tau}}$ has a V-shape and the "bottom of the V" is reached around $t = \check{\tau}$ which explains the choice of the notation $\check{\tau}$. In general, $\check{v}_{\check{\tau}}$ has linear growth with respect to time due to the second term in equation (11). Time differentiating Equation (11), using (9) and the fact that $\frac{df}{dt} = f_u \frac{du}{dt} = f_u f$, we recover :

$$\frac{d\check{v}_{\check{\tau}}}{dt} = \frac{dv}{dt} - \frac{df}{dt}\int_{\check{\tau}}^{t} \eta dr - \eta f = f_u \check{v}_{\check{\tau}} + f_s,$$

meaning that $\check{v}_{\check{\tau}}$ satisfies the simple no time-dilated Equation (8).

Since this $\check{v}_{\check{\tau}}(t)$ defined by Equation (11) is not uniformly bounded it cannot be directly used to compute the desired derivative by commuting the limit and derivative as in Equation (7). Instead, we introduce an approximation that involves a window function $\mathbf{w} : [0, 1] \to \mathbf{R}$ to mitigate the linear growth of $\check{v}_{\check{\tau}}$. This approximation is

$$\frac{d\langle J\rangle}{ds} \approx \frac{1}{T}\int_{0}^{T} \mathbf{w}(\tfrac{t}{T})(J_u \check{v}_{\check{\tau}} + J_s)dt \tag{12}$$

for any $\check{\tau} \in [0, T]$. The validity of this approximation is established through the following theorem.

Theorem 1. *If the following are true:*

1. $\mathbf{w}$ *is continuously differentiable,*
2. $\mathbf{w}(0) = \mathbf{w}(1) = 0$, *and*
3. $\int_{0}^{1} \mathbf{w}(r)dr = 1$,
4. $\check{\tau}$ *is a function of T satisfying $0 \leq \check{\tau} \leq T$ and $\lim\limits_{T\to\infty} \dfrac{\check{\tau}}{T}$ exists.*

6

then,

$$\frac{d}{ds} \lim_{T\to\infty} \frac{1}{T} \int_0^T J(u(t;s),s)dt = \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{w}(\tfrac{t}{T})(J_u \check{v}_{\tilde{\tau}} + J_s)dt . \qquad (13)$$

This equality is nontrivial. To prove that it is true, we first define the following.

Definition 1. For any continuous function $\mathbf{w} : [0,1] \to \mathbf{R}$, the mean of the window is

$$\overline{\mathbf{w}} := \int_0^1 \mathbf{w}(r)dr;$$

the infinitely-long windowed time average of a signal $x(t)$ is

$$\langle x \rangle_{\mathbf{w}} := \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{w}(\tfrac{t}{T}) x(t) dt,$$

A special case of the window function is $\mathbf{w} \equiv 1$ called *square* window. The mean of this window is 1; the infinitely-long windowed time average of a signal $x(t)$ is simply its ergodic average, which we already denoted as :

$$\langle x \rangle := \lim_{T\to\infty} \frac{1}{T} \int_0^T x(t)dt.$$

Lemma 2. *If $x(t)$ is bounded and $\langle x \rangle$ exists, then $\langle x \rangle_{\mathbf{w}}$ also exists for any continuous $\mathbf{w}$, and the following equality is true:*

$$\langle x \rangle_{\mathbf{w}} = \overline{\mathbf{w}} \langle x \rangle.$$

The proof is given in Appendix B.

Note that Lemma 2 does not apply to the windowed average on the right hand side of Equation (13) in Theorem 1. This is because the lemma requires $x(t)$ to be independent of the averaging length T. But in Theorem 1, $\check{\tau}$ and thus $J_u \check{v}_{\tilde{\tau}} + J_s$ depends on T. To apply Lemma 2, we must first decompose $\check{v}_{\tilde{\tau}}$ according to Equation (11). Lemma 2 can then be applied to all but one of the components:

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{w}(\tfrac{t}{T})(J_u \check{v}_{\tilde{\tau}} + J_s)dt$$

$$= \langle J_u v \rangle_{\mathbf{w}} + \langle J_s \rangle_{\mathbf{w}} - \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{w}(\tfrac{t}{T}) J_u f(t) (\int_{\tilde{\tau}}^t \eta(r)dr)dt \qquad (14)$$

$$= \langle J_u v \rangle + \langle J_s \rangle - \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{w}(\tfrac{t}{T}) \left(\int_{\tilde{\tau}}^t \eta(r)dr \right) \frac{dJ}{dt} dt$$

if $\overline{\mathbf{w}} = 1$. Here we used the fact that :

$$\frac{dJ}{dt} = J_u \frac{du}{dt} + J_s \frac{d\cancel{s}}{\cancel{dt}}^{\,0} = J_u f$$

7

We then apply integration by parts to the remaining windowed average, and use the assumption that $\mathbf{w}(0) = \mathbf{w}(1) = 0$ in Theorem 1, to obtain :

$$
-\lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{w}(\tfrac{t}{T})\Big(\int_{\check{\tau}}^t \eta(r)dr\Big)\frac{dJ}{dt}\,dt
$$

$$
= \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{w}(\tfrac{t}{T})\eta(t)J(t)\,dt + \frac{1}{T}\int_0^T \frac{d}{dt}\mathbf{w}(\tfrac{t}{T})\Big(\int_{\check{\tau}}^t \eta(r)dr\Big)J(t)\,dt \qquad (15)
$$

$$
= \langle \eta J \rangle + \lim_{T\to\infty} \frac{1}{T}\int_0^T \mathbf{w}'(\tfrac{t}{T})\tfrac{1}{T}\Big(\int_{\check{\tau}}^t \eta(r)dr\Big)J(t)\,dt
$$

Here $\mathbf{w}'$ is the derivative of the window function $\mathbf{w}$. By substituting Equation (15) into Equation (14), then comparing with Equation (7), we see that we can prove Theorem 1 by proving the equality

$$
\lim_{T\to\infty} \frac{1}{T}\int_0^T \mathbf{w}'(\tfrac{t}{T})\tfrac{1}{T}\Big(\int_{\check{\tau}}^t \eta(r)dr\Big)J(t)\,dt = -\langle\eta\rangle\langle J\rangle \qquad (16)
$$

We now establish this equality, thereby proving Theorem 1, through two lemmas:

Lemma 3. *If η is bounded and $\langle\eta\rangle$ exists, then*

$$
\lim_{T\to\infty}\left(\sup_{\check{\tau},t\in[0,T]}\left(\tfrac{1}{T}\Big(\int_{\check{\tau}}^t \eta(r)dr\Big) - \langle\eta\rangle\frac{t-\check{\tau}}{T}\right)\right) = 0.
$$

The proof of this lemma is given in Appendix C. This lemma establishes the equality that

$$
\lim_{T\to\infty} \frac{1}{T}\int_0^T \mathbf{w}'(\tfrac{t}{T})\tfrac{1}{T}\Big(\int_{\check{\tau}}^t \eta(r)dr\Big)J(t)\,dt = \langle\eta\rangle \lim_{T\to\infty} \frac{1}{T}\int_0^T \mathbf{w}'(\tfrac{t}{T})\frac{t-\check{\tau}}{T}J(t)\,dt
$$

The remaining task in proving Equation (16) is achieved by the following lemma:

Lemma 4. *If J is bounded, $\mathbf{w} \in C^1[0,1]$, $\mathbf{w}(0) = \mathbf{w}(1) = 0$, and $\lim\limits_{T\to\infty}\dfrac{\check{\tau}}{T}$ exists, then*

$$
\lim_{T\to\infty} \frac{1}{T}\int_0^T \mathbf{w}'(\tfrac{t}{T})\frac{t-\check{\tau}}{T}J(t)\,dt = -\overline{\mathbf{w}}\langle J\rangle
$$

PROOF. Let $\check{\tau} = \lim\limits_{T\to\infty}\dfrac{\check{\tau}}{T}$. Because both J and $\mathbf{w}'$ are bounded ($\mathbf{w}$ is continuously differentiable in a closed interval),

$$
\lim_{T\to\infty} \frac{1}{T}\int_0^T \mathbf{w}'(\tfrac{t}{T})\frac{t-\check{\tau}}{T}J(t)\,dt = \lim_{T\to\infty} \frac{1}{T}\int_0^T \mathbf{w}'(\tfrac{t}{T})\left(\tfrac{t}{T}-\check{\tau}\right)J(t)\,dt
$$

Define $\mathbf{w}_{\check{\tau}}(r) = \mathbf{w}'(r)(r-\check{\tau})$, then Lemma 2 can turn the equality above into,

$$
\lim_{T\to\infty} \frac{1}{T}\int_0^T \mathbf{w}'(\tfrac{t}{T})\frac{t-\check{\tau}}{T}J(t)\,dt = \langle J\rangle_{\mathbf{w}_{\check{\tau}}} = \overline{\mathbf{w}_{\check{\tau}}}\langle J\rangle,
$$

8

in which

$$\overline{\mathbf{w}_{\check{\tau}}} := \int_0^1 \mathbf{w}'(r)(r-\check{\tau})dr = \int_0^1 \frac{d\mathbf{w}}{dr}r\, dr - \int_0^1 \frac{d\mathbf{w}}{dr}\check{\tau}\, dr = -\int_0^1 w(r)dr = -\overline{\mathbf{w}}$$

$\square$

In our case, not only $\lim\limits_{T\to\infty}\dfrac{\check{\tau}}{T}$ exist but we even have $\lim\limits_{T\to\infty}\dfrac{\check{\tau}}{T} \to \dfrac{1}{2}$ as shown in Appendix D. This result comes from the fact that the computed $\check{v}_{\check{\tau}}$ minimizes $\int_0^T |\check{v}_{\check{\tau}}(t)|^2 dt$.

Lemma 3-4 combines to prove Equation (16), which combines with (14) and (15) to :

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T \mathbf{w}(\tfrac{t}{T})(J_u\check{v}_{\check{\tau}} + J_s)dt = \langle J_u v\rangle + \langle J_s\rangle + \langle \eta J\rangle - \langle \eta\rangle\langle J\rangle \ . \qquad (17)$$

This, together with Equation (10), derived in previous literature ([10], [16]), proves that the desired derivative $\frac{d\langle J\rangle}{ds}$ can be computed via a windowed average of $J_u\check{v}_{\check{\tau}} + J_s$, where $\check{v}$ is a solution to the tangent equation without time dilation.

4. Numerical results

Now, we are going to apply simplified LSS to a test case dynamical system known as *Lorenz 63*. It is a 3-dimensional autonomous differential equation parameterized by σ, β, ρ and the dynamics happen to be chaotic when the parameters belong to a certain range of values. The governing equations are the following :

$$\begin{cases} \frac{dx}{dt} = \sigma(y-x) \\ \frac{dy}{dt} = x(\rho - z) - y \\ \frac{dz}{dt} = xy - \beta z \end{cases}$$

Edward Lorenz introduced them in 1963 to model the atmospheric convection. The quantity of interest in this example is $\langle z\rangle = \lim_{T\to\infty}\frac{1}{T}\int_{t=0}^T z(t)dt$, the time average of the component z and the algorithm is used to compute $\frac{d\langle z\rangle}{d\rho}$, its derivative with respect to ρ. In this case, the design paramater s is equal to ρ and $J(u,\rho) = z$ where $u = (x,y,z)$. While a variety of values for ρ will be tested, the other two parameters are set to $\sigma = 10$ and $\beta = \frac{8}{3}$. It has been shown that $\frac{d\langle z\rangle}{d\rho}$ is approximately equal to 1 for a wide range of values of ρ (our numerical applications will stay in this range)[6].

The chosen windowing functions for this test case are (see figure 1) :

- Square window :

$$\mathbf{w}(t) = 1, \quad t \in [0,1]$$

9

- Sine window:
$$\mathbf{w}(t) = \sin(\pi t), \quad t \in [0, 1]$$

- Sine squared window :
$$\mathbf{w}(t) = \sin^2(\pi t), \quad t \in [0, 1]$$

- Sine to the power four also known as Hann-squared window :
$$\mathbf{w}(t) = \sin^4(\pi t), \quad t \in [0, 1]$$

- Bump window :
$$\mathbf{w}(t) = \exp(\frac{-1}{t - t^2}), \quad t \in [0, 1]$$

defined up to normalization constants.

All windows respect the criteria of theorem 1 with the exception of the square window which is equivalent to not using any window. The major difference between the last four functions is their Taylor expansion in the neighborhood of $t = 0$ and $t = 1$. Actually, we can expect the areas around the extremities to be the most "sensible" ones, leading the the biggest error terms in the approximation of $\frac{d\langle z \rangle}{d\rho}$ since $\check{v}_{\tilde{\tau}}$ is maximal in these areas. Thus, choosing a windowing function that collapses quite fast on the extremities is expected to be the more accurate. The simple sine has non zero derivatives on the extremities, the squared one has a zero derivative on both sides, the Hann-squared function has zero coefficients until the 3rd order derivative while the bump function has zero coefficients for all derivative orders. A deep and rigorous analysis of this intuitive idea has been carried out by Krakos et al. for the periodic case (periodic dynamical system)[12].

For all simulations, the equations were discretized using an implicit second-order Crank-Nicolson scheme[2] solved up to machine precision and a uniform timestep of $\Delta t = 0.02$. The burn-in period for computing $u(t)$ was set to 10 time units.

First, figure 2 shows the norm of the computed $\check{v}_{\tilde{\tau}}$ with respect to t where $\rho = 28$ and the time integration length is $T = 100$. We can clearly notice the linear growth of the "envelope" of $\check{v}_{\tilde{\tau}}$ as t increases or decreases as predicted in equation (11) :
$$\check{v}_{\tilde{\tau}}(t) = v(t) - f(t) \int_{\tilde{\tau}}^{t} \eta(r) dr$$

---

[2]It is worth noting that both integration scheme for $u(t)$ and discretization of equations (3) should have the same accuracy order for stability reasons.
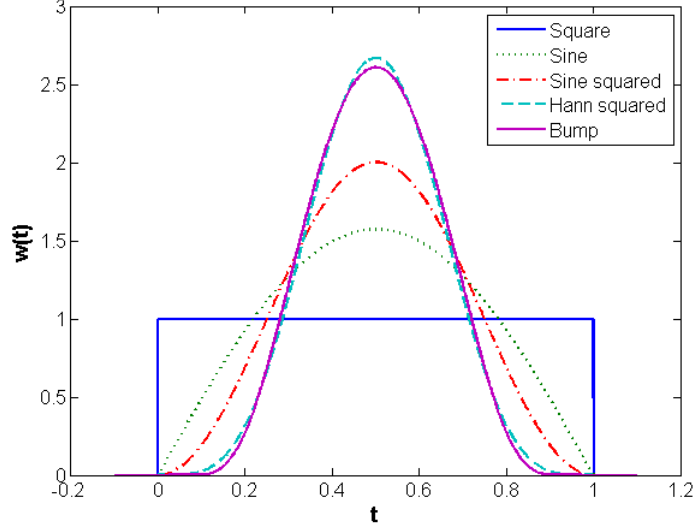
Figure 1: Shape of the different windowing functions

Furthermore, $\check{v}_{\check{\tau}}$ is minimal around $t = \frac{T}{2} = 50$ which is in agreement with the following result :

$$\lim_{T \to \infty} \frac{\check{\tau}}{T} \to \frac{1}{2}$$

Then, we have computed $\frac{d\langle z \rangle}{d\rho}$ for $\rho \in [25, 35]$ and $T = 50$ using different windowing functions as well as the original Least Squared Shadowing algorithm (figure 3). In this range of the parameter ρ, the dynamical system is known to be quasi-hyperbolic and the analysis we carried out in the previous section remains valid.

In order to have a rough estimation of $\frac{d\langle z \rangle}{d\rho}$ for this range of ρ, we computed $\langle z \rangle(\rho = 25)$, $\langle z \rangle(\rho = 35)$ for 50 different initial conditions and constituted the finite difference approximation :

$$\frac{d\langle z \rangle}{d\rho} \approx \frac{\langle z \rangle(\rho = 35) - \langle z \rangle(\rho = 25)}{35 - 25} \tag{18}$$

Then, we represented in figure 3 the mean of these estimations as well as their 3-sigma confidence interval:

$$\text{CI} = \left[\mu_{50}^{\text{fd}} \pm 3 \frac{\sigma_{50}^{\text{fd}}}{\sqrt{50}} \right] \tag{19}$$

where μ_{50}^{fd} is the average of the 50 samples of $\frac{d\langle z \rangle}{d\rho}$ computed using finite difference and σ_{50}^{fd} is their estimated standard deviation. After comparing these bounds
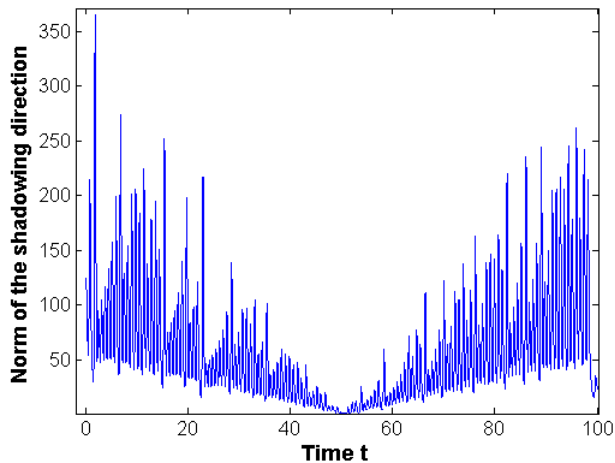
11

Figure 2: Norm of $\check{v}_{\check{\tau}}$ with respect to t for $\rho = 28$

with the results obtained with our algorithms, we notice that original LSS as well as the windowed algorithms give correct and very similar estimations of $\frac{d\langle z \rangle}{d\rho}$. Since the different versions of LSS have been tried on a single trajectory and for a finite integration time $T = 50$, there is no reason for the computed derivatives to be unbiased. Furthermore, all windows do not seem to have the same performance: the results obtained with the Hann-squared and the bump window are smoother and more self-consistent than the ones coming from the sine squared and much more smoother than the ones given by the simple sine window. In order to point out this phenomenon, a complementary analysis has been done: for $\rho = 28$, $\rho = 50$ and integration time lengths of $T = 25$, $T = 50$ and $T = 100$, 2000 simulations (each one with a different initial condition) were run to compute a 95% confidence interval of the standard deviation σ of $\frac{d\langle z \rangle}{d\rho}$ for each one of the windows (see figure 4). The 95% confidence intervals were approximated as follows :

$$\text{CI} = \left[\left(\sigma_{2000}^2 \pm 1.96 \sqrt{\frac{\sigma_{2000}^4}{2000} \left((\kappa_{2000} - 1) + \frac{2}{2000 - 1} \right)} \right)^{\frac{1}{2}} \right] \qquad (20)$$

where σ_{2000} and κ_{2000} are, respectively, the estimates of the 2000 samples standard deviation and kurtosis.

The results confirm our previous remark: for a fixed T and $\rho = 28$, the Hann-squared and bump windows have the lowest standard deviations meaning they are more robust than the other windows. Among the valid windowing functions, the simple sine gives the worst results. Then, for a fixed window, we notice that increasing the time integration length decreases the standard deviation which is an intuitive result since when T gets bigger the influence of the initial condition
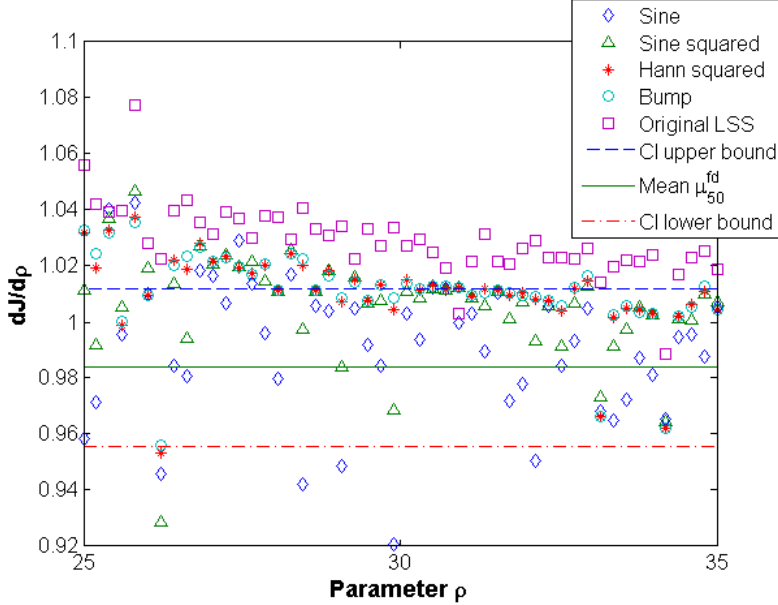
12

Figure 3: $\frac{d\langle z\rangle}{d\rho}$ using different windowing functions and original LSS for integration time $T = 50$

on the dynamical system fades out (due to ergodicity). Finally, for $\rho = 50$, all windows give bad results (standard deviation of order 1) which can be explained by the fact that the dynamical system is no longer quasi hyperbolic for this value of ρ. The theory we have developed doesn't hold anymore and there is no reason for simplified LSS to converge to the true value of $\frac{d\langle z\rangle}{d\rho}$ (for $T = 100$, the confidence intervals were so big and uninformative that we only represented the standard deviation estimator).

5. Conclusion

Simplified LSS is an improvement of original LSS in the sense that the time dilation factors have been removed from the formulation of the problem leading to a much simpler system of equations to solve. When running the original version of LSS, we are lead to solve a slightly more complex set of KKT equations than (3) which can be found in [17]. This set of equations is solved by inverting the Schur complement of the KKT matrix which contains the concatenation of smaller matrices of the form $f_{u_i} \times f_{u_i}^T + f(u_i) \times f(u_i)^T$ where $\{u_i\}$ is a finite discretization of a trajectory. The first term is a multiplication of two $n \times n$ matrices where n is the phase space dimension. For PDEs, n can be extremely large. However, since the finite stencil used to discretize the PDE is local, f_{u_i} is very sparse, so is $f_{u_i} \times f_{u_i}^T$. The second term, which is totally due to the time
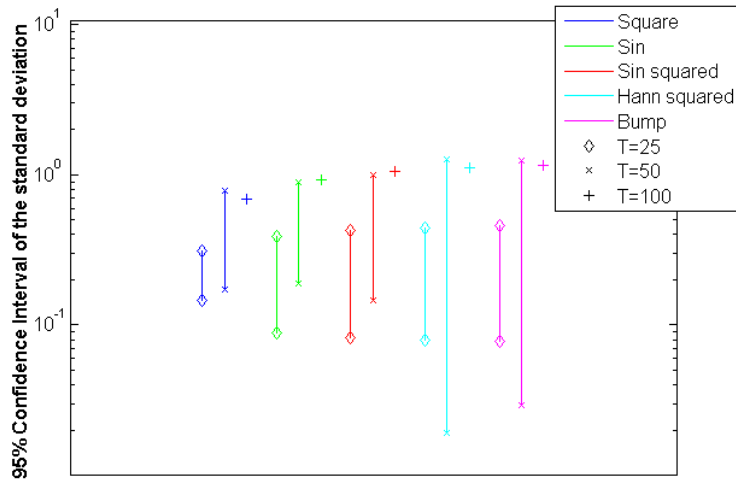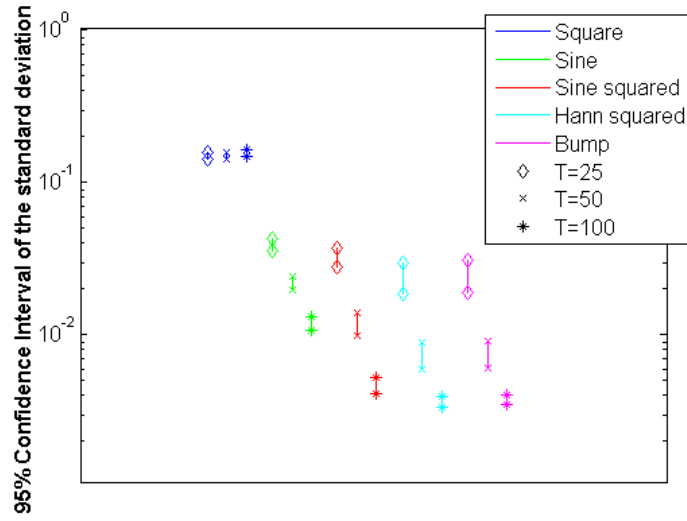
Figure 4: Standard deviation confidence intervals for $\rho = 28$ (upper graph) and $\rho = 50$ (lower graph) using different windowing functions

dilation terms, is a $n \times 1$ by $1 \times n$ multiplication which leads to a rank one but fully dense $n \times n$ matrix. Consequently, $f_{u_i} \times f_{u_i}^T + f(u_i) \times f(u_i)^T$ is dense as well and the Schur complement becomes hard, if not impossible to inverse for high-dimensional systems.

With windowing, the time dilation factors are no longer needed and the Schur complement is now only formed of $f_{u_i} \times f_{u_i}^T$ submatrices. We can consequently use sparse solvers and preconditioners such as incomplete LU factorization to solve the KKT system. Simplified LSS can be efficiently applied to very large, industrial-scale problems which were very hard if not impossible to tackle with original LSS [11]. Additionally, simplified LSS is easier to implement.

The simplification has been made possible thanks to the introduction of a windowing function that mitigates the linear growth of the shadowing direction $\check{v}_{\tilde{\tau}}$. This function is required to be continuously differentiable, null at its extremities and with integral equal to 1. Once picked in this class of functions, the major characteristic of the window is the way it decays at its extremities: a rapidly decaying function such as the bump window or the Hann-squared one will be less sensitive to transient behavior and give more robust results. Since all windowing functions lead to the same computational cost, our study strongly recommends the choice of a fastly decaying window. In the same way, increasing the time integration length decreases the influence of the initial condition making the algorithm more robust and accurate.

Acknowledgements

References

References

[1] J. NOCEDAL, S. WRIGHT, *Numerical optimization*, Springer Science and Business Media, 2006.

[2] J. REUTHER, A. JAMESON, J. ALONSO, M. RIMLINGER, D. SAUNDERS, *Constrained multipoint aerodynamic shape optimization using an adjoint formulation and parallel computers*, Journal of aircraft(1), 51-60, 1999.

[3] R.M. HICKS, P.A. HENNE, *Wing design by numerical optimization*, Journal of Aircraft, 15.7: 407-412, 1978.

[4] R.C. SMITH, *Uncertainty quantification: theory, implementation, and applications.*, Vol. 12. SIAM, 2013.

[5] Q. Wang, *Uncertainty quantification for unsteady fluid flow using adjoint-based approaches*, Ph.D. thesis, Stanford University, Stanford, CA, 2009.

[6] D. Lea, M. Allen, T. Haine, *Sensitivity analysis of the climate of a chaotic system*, Tellus, 52A:523-532, 2000.

[7] G. Eyink, T. Haine, D. Lea, *Ruelle's linear response formula, ensemble adjoint schemes and Lvy flights*, Nonlinearity 17, 1867-1889, 2004.

[8] R.V. Abramov, A.J. Majda, *Blended response algorithms for linear fluctuation-dissipation for complex nonlinear dynamical systems*, Nonlinearity, 20(12):2793, 2007.

[9] F. Cooper, P. Haynes, *Climate sensitivity via a nonparametric fluctuation-dissipation theorem*, Journal of the Atmospheric Sciences, 68(5):937-953, 2011.

[10] Q. Wang, R. Hu, P. Blonigan, *Least Squares Shadowing Sensitivity analysis of chaotic limit cycle oscillations*, Journal of Computational Physics, 2014.

[11] P. J. Blonigan, Q. Wang, E. J. Nielsen, B. Diskin, *Least Squares Shadowing Sensitivity Analysis of Chaotic Flow around a Two-Dimensional Airfoil.*, 54th AIAA Aerospace Sciences Meeting, 2016.

[12] J. A. Krakos, Q. Wang, S. R. Hall, D. L. Darmofal, *Sensitivity analysis of limit cycle oscillations*, Journal of Computational Physics, doi:10.1016/ j.jcp.2012.01.001, 2012.

[13] Q. Wang, *Convergence of the Least Squares Shadowing method for computing derivative of ergodic averages*, SIAM, J. Numer. Anal. 52 (1) 156-170, 2014.

[14] S.P. Kuznetsov, *Hyperbolic Chaos: A Physicist's View*, Springer Berlin Heidelberg, 2012.

[15] C.Bonatti, L. Diaz, M. Viana, *Dynamics Beyond Uniform Hyperbolicity : A Global Geometric and Probabilistic Perspective*, Encyclopaedia of Mathematical Sciences, Springer, 2010.

[16] M. Chater, A. Ni, P. J. Blonigan, Q. Wang, *Least squares shadowing method for sensitivity analysis of differential equations*, submitted to SIAM, in revision, arXiv:1509.02882v1 [math.DS] 9 Sep 2015.

[17] P. Blonigan, S. Gomez, Q. Wang, *Least Squares Shadowing Sensitivity Analysis of Chaotic and Turbulent Fluid Flows.*, APS Meeting Abstracts, Vol. 1, 2013.

Appendix A. Derivation of the adjoint version of *simplified LSS*

The adjoint version is directly obtained from the equations of the tangent version. We should have in mind that the objective function we want to compute is :

$$\frac{d\langle J\rangle}{ds} = \frac{1}{T}\int_0^T \mathbf{w}(\tfrac{t}{T})\left(J_u\check{v} + J_s\right)dt \qquad (A.1)$$

In order to obtain the adjoint formulation, we need to replace the first term $\mathbf{w}(\tfrac{t}{T})J_u\check{v}$ which is based on the solution $\check{v}$ to the forward problem by a new term based on a backward solution $\hat{v}$. First, we reformulate the *forward* set of equations (3):

$$\begin{cases} \frac{d\check{v}}{dt} - f_u\check{v} - f_s = 0 \\ \frac{d\check{w}}{dt} + f_u^T\check{w} - \check{v} = 0 \\ \check{w}(0) = \check{w}(T) = 0 \end{cases} \qquad (A.2)$$

Then, we introduce the dual variables $\hat{v}$ and $\hat{w}$ for the first two equations and obtain the scalar equation :

$$\int_0^T \hat{v}^T(\frac{d\check{v}}{dt} - f_u\check{v} - f_s) + \hat{w}^T(\frac{d\check{w}}{dt} + f_u^T\check{w} - \check{v})dt = 0 \qquad (A.3)$$

where the superscript T stands for the transpose operation. After integrating by parts and thanks to the third equation of (A.2) :

$$0 = \hat{v}^T\check{v}(T) - \hat{v}^T\check{v}(0) + \int_0^T \hat{v}^T(-f_u\check{v} - f_s) - \frac{d\hat{v}}{dt}^T\check{v} + \hat{w}^T(f_u^T\check{w} - \check{v}) - \frac{d\hat{w}}{dt}^T\check{w}dt \qquad (A.4)$$

$$= \check{v}^T\hat{v}(T) - \check{v}^T\hat{v}(0) + \int_0^T (-f_u^T\hat{v} - \frac{d\hat{v}}{dt} - \hat{w})^T\check{v} + (f_u\hat{w} - \frac{d\hat{w}}{dt})^T\check{w} - f_s^T\hat{v}dt \qquad (A.5)$$

Based on (A.1), we impose the *adjoint* (or *backward*) set of equations :

$$\begin{cases} -f_u^T\hat{v} - \frac{d\hat{v}}{dt} - \hat{w} = \mathbf{w}(\tfrac{t}{T})J_u^T \\ f_u\hat{w} - \frac{d\hat{w}}{dt} = 0 \\ \hat{v}(0) = \hat{v}(T) = 0 \end{cases} \qquad (A.6)$$

After replacing the equalities of (A.6) into (A.5), we get:

$$\frac{1}{T}\int_0^T (-f_u^T\hat{v} - \frac{d\hat{v}}{dt} - \hat{w})^T\check{v}dt = \frac{1}{T}\int_0^T \mathbf{w}(\tfrac{t}{T})J_u\check{v}dt = \frac{1}{T}\int_0^T f_s^T\hat{v}dt \qquad (A.7)$$

Combining (A.1) and (A.7), the new objetive function becomes:

$$\frac{d\langle J\rangle}{ds} = \frac{1}{T}\int_0^T \left(f_s^T\hat{v} + J_s\right)dt \qquad (A.8)$$

17

Appendix B. Lemma 2

Given any $\mathbf{w} \in C[0,1], \mathbf{w} \geq 0$, and bounded function $x(t)$ such that $\langle x \rangle = \lim_{T \to \infty} \frac{1}{T} \int_0^T x(t)dt$ exists, there is the relation:

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T \mathbf{w}(\tfrac{t}{T})x(t)dt = \overline{\mathbf{w}} \cdot \langle x \rangle \tag{B.1}$$

PROOF. Given any $\epsilon > 0$, let $\epsilon_1 = \epsilon/2B_x$, here B_x is the bound for $x(t)$. $\mathbf{w}$ is continuous on a compact set, hence its range is also bounded, say, by B_w. Moreover, $\mathbf{w}$ is uniformly continuous, hence $\exists K$, s.t. $\forall \tau_1, \tau_2 \in [0,1], |\tau_1 - \tau_2| \leq 1/K$, we have $|\mathbf{w}(\tau_1) - \mathbf{w}(\tau_2)| \leq \epsilon_1$. Construct the partition $P = \{0, \frac{1}{K}, \frac{2}{K}, ..., 1\}$. Now $\forall i \in \{1, 2, ..., K\}$, let $\mathbf{w}_i = K \cdot \int_{(i-1)/K}^{i/K} w(\tau)d\tau$, then $\overline{\mathbf{w}} = \frac{1}{K} \sum_{i=1}^K \mathbf{w}_i$. Let $\xi_i(\tau) = \mathbf{w}(\tau) - \mathbf{w}_i$, $\tau \in I_i = [(i-1)/K, i/K]$.

$\forall i, \mathbf{w}$ is continuous on the segment I_i, which is a compact set, hence $\mathbf{w}$ achieves its maximum and minimum value on this segment. Denote its maximum and minimum values on this segment by $\mathbf{w}_{i,max}$ and $\mathbf{w}_{i,min}$ respectively. By selection of K, $\mathbf{w}_{i,max} - \mathbf{w}_{i,min} \leq \epsilon_1$. By definition of $\mathbf{w}_i$, $\mathbf{w}_{i,min} \leq \mathbf{w}_i \leq \mathbf{w}_{i,max}$. Hence we have $\xi_i(\tau) \leq \epsilon_1, \tau \in I_i$.

With above K, let $\epsilon_2 = \epsilon/2KB_w$. Since $\lim_{T \to \infty} \frac{1}{T} \int_0^T x(t)dt$ exists, $\exists T_0$, such that $\forall M \geq T_0$, $\langle x \rangle - \frac{1}{M} \int_0^M x(t)dt \leq \epsilon_2$. Hence we have:

$$\left| \langle x \rangle - \frac{1}{iM} \int_0^{iM} x(t)dt \right| \leq \epsilon_2, \quad i = 1, 2, 3, ...$$

the i-th inequality is equivalent to:

$$-i\epsilon_2 \leq i\langle x \rangle - \frac{1}{M} \int_0^M x(t)dt... - \frac{1}{M} \int_{(i-1)M}^{iM} x(t)dt \leq i\epsilon_2, \quad i = 1, 2, 3, ...$$

the (i+1)-th inequality is equivalent to:

$$-(i+1)\epsilon_2 \leq -(i+1)\langle x \rangle + \frac{1}{M} \int_0^M x(t)dt... + \frac{1}{M} \int_{iM}^{(i+1)M} x(t)dt \leq (i+1)\epsilon_2, \quad i = 0, 1, 2, ...$$

add above two inequalities together,

$$-(2i+1)\epsilon_2 \leq \langle x \rangle - \frac{1}{M} \int_{iM}^{(i+1)M} x(t)dt \leq (2i+1)\epsilon_2, \quad i = 0, 1, 2, ...$$

Now $\forall T \geq KT_0$, let $M = T/K$, then $M \geq T_0$, and the difference between

the two sides in B.1, when T is finite, could be represented by:

$$\overline{\mathbf{w}} \cdot \langle x \rangle - \frac{1}{T} \int_0^T \mathbf{w}(\tfrac{t}{T}) x(t) dt$$

$$= \frac{1}{K} \sum_{i=1}^K \left[\mathbf{w}_i \cdot \langle x \rangle - \frac{1}{M} \int_{(i-1)M}^{iM} (\mathbf{w}_i + \xi_i(\tfrac{t}{T})) x(t) dt \right]$$

$$= \frac{1}{K} \sum_{i=1}^K \left[-\frac{1}{M} \int_{(i-1)M}^{iM} \xi_i\left(\tfrac{t}{T}\right) x(t) dt + \mathbf{w}_i \cdot \left(\langle x \rangle - \frac{1}{M} \int_{(i-1)M}^{iM} x(t) dt \right) \right]$$

the first term is confined by:

$$\left| \frac{1}{M} \int_{(i-1)M}^{iM} \xi_i(\tfrac{t}{T}) x(t) dt \right| \leq \epsilon_1 \cdot \frac{1}{M} \int_{(i-1)M}^{iM} |x(t)| \, dt$$

the second term is confined by:

$$\left| \mathbf{w}_i \cdot \left(\langle x \rangle - \frac{1}{M} \int_{(i-1)M}^{iM} x(t) dt \right) \right| \leq \mathbf{w}_i \cdot (2i-1)\epsilon_2$$

As a result,

$$\left| \overline{\mathbf{w}} \cdot \langle x \rangle - \frac{1}{T} \int_0^T \mathbf{w}(\tfrac{t}{T}) x(t) dt \right|$$

$$\leq \frac{1}{K} \sum_{i=1}^K \left[\epsilon_1 \cdot \frac{1}{M} \int_{(i-1)M}^{iM} |x(t)| dt + \mathbf{w}_i \cdot (2i-1)\epsilon_2 \right]$$

$$\leq \epsilon_1 \cdot B_x + \epsilon_2 \cdot B_w K \leq \epsilon$$

Now we find $T^* = KT_0$, s.t. $\forall T \geq T^*$, $\left| \frac{1}{T} \int_0^T \mathbf{w}(\tfrac{t}{T}) x(t) dt - \overline{\mathbf{w}} \cdot \langle x \rangle \right| \leq \epsilon$. $\quad \square$

Appendix C. Lemma 3

If η is bounded and $\langle \eta \rangle$ exists, then

$$\lim_{T \to \infty} \left(\sup_{\check{\tau}, t \in [0,T]} \left(\tfrac{1}{T} \left(\int_{\check{\tau}}^t \eta(r) dr \right) - \langle \eta \rangle \frac{t - \check{\tau}}{T} \right) \right) = 0.$$

PROOF. First, we write:

$$\tfrac{1}{T} \left(\int_{\check{\tau}}^t \eta(r) dr \right) - \langle \eta \rangle \frac{t - \check{\tau}}{T} = \frac{t - \check{\tau}}{T} \left(\frac{1}{t - \check{\tau}} \left(\int_{\check{\tau}}^t \eta(r) dr \right) - \langle \eta \rangle \right) \qquad \text{(C.1)}$$

Based on the two assumptions $t, \check{\tau} \in [0, T]$ and $\eta \leq \|\eta\|^\infty < \infty$, we obtain the following bounds on the two terms of the product we have just derived:

19

$$\left|\frac{t-\check{\tau}}{T}\right| \leq 1 \qquad \text{as well as} \qquad \left|\frac{1}{t-\check{\tau}}\left(\int_{\check{\tau}}^{t}\eta(r)dr\right) - \langle\eta\rangle\right| \leq 2\|\eta\|^{\infty} \qquad (C.2)$$

Let us define $c = \max\left(1, 2\|\eta\|^{\infty}\right)$. Based on the ergodicity of η, for any $\epsilon > 0$ there is a constant $M > 0$ such that for all t_0 and $T \geq M$:

$$\left|\frac{1}{T}\int_{r=t_0}^{t_0+T}\eta(r)dr - \langle\eta\rangle\right| < \epsilon \qquad (C.3)$$

Consequently, for all $\epsilon > 0$ and for all $T > \frac{cM}{\epsilon}$, we have:

- if $|t - \check{\tau}| < M$, then $\left|\frac{t-\check{\tau}}{T}\right| \leq \frac{\epsilon}{c}$ which means that :

$$\left|\frac{t-\check{\tau}}{T}\left(\frac{1}{t-\check{\tau}}\left(\int_{\check{\tau}}^{t}\eta(r)dr\right) - \langle\eta\rangle\right)\right| \leq \frac{\epsilon}{c}\left|\frac{1}{t-\check{\tau}}\left(\int_{\check{\tau}}^{t}\eta(r)dr\right) - \langle\eta\rangle\right| \leq \epsilon \quad (C.4)$$

- if $|t - \check{\tau}| \geq M$, thanks to relation (C.3):

$$\left|\frac{1}{t-\check{\tau}}\left(\int_{\check{\tau}}^{t}\eta(r)dr\right) - \langle\eta\rangle\right| \leq \epsilon \qquad (C.5)$$

which again implies :

$$\left|\frac{t-\check{\tau}}{T}\left(\frac{1}{t-\check{\tau}}\left(\int_{\check{\tau}}^{t}\eta(r)dr\right) - \langle\eta\rangle\right)\right| \leq \epsilon \qquad (C.6)$$

that concludes the proof.

Appendix D. Proof of convergence of $\lim_{T\to\infty}\frac{\check{\tau}}{T}$

When minimizing the cost function $\int_0^T \|\check{v}_{\check{\tau}}(t)\|^2 dt$, the choice of $\check{\tau}$ is arbitrary, thus :

$$\frac{d}{d\check{\tau}}\left(\int_0^T \|\check{v}_{\check{\tau}}(t)\|^2 dt\right) = 0 \qquad (D.1)$$

at the global minimum. We have :

$$\int_0^T \|\check{v}_{\check{\tau}}(t)\|^2 dt = \int_0^T \left(\|v(t)\|^2 + \|f(t)\|^2\left(\int_{\check{\tau}}^{t}\eta(r)dr\right)^2 - 2v(t)^T f(t)\int_{\check{\tau}}^{t}\eta(r)dr\right)dt \qquad (D.2)$$

Thus :

$$\frac{d}{d\check{\tau}}\left(\int_0^T \|\check{v}_{\check{\tau}}(t)\|^2 dt\right) = -2\int_0^T \|f(t)\|^2\eta(\check{\tau})\left(\int_{\check{\tau}}^{t}\eta(r)dr\right)dt + 2\int_0^T v(t)^T f(t)\eta(t)dt \qquad (D.3)$$

$$= 2\eta(\check{\tau})\left(\int_0^T v(t)^T f(t)dt - \int_0^T \|f(t)\|^2\left(\int_{\check{\tau}}^{t}\eta(r)dr\right)dt\right) \qquad (D.4)$$

20

It can be shown that the $\check{\tau}$ such that $\eta(\check{\tau}) = 0$ if any, don't correspond to a global minimum. We consequently should have :

$$\int_0^T v(t)^T f(t)dt - \int_0^T \|f(t)\|^2 \Big(\int_{\check{\tau}}^t \eta(r)dr\Big)dt = 0 \tag{D.5}$$

On one side :

$$\Big|\int_0^T v(t)^T f(t)dt\Big| \leq \int_0^T \|v\|^\infty \|f\|^\infty dt \tag{D.6}$$

$$\leq T\|v\|^\infty \|f\|^\infty \tag{D.7}$$

On the other, $\int_{\check{\tau}}^t \eta(r)dr \sim (t - \check{\tau})\langle\eta\rangle$ which implies :

$$\Big|\int_0^T \|f(t)\|^2 \Big(\int_{\check{\tau}}^t \eta(r)dr\Big)dt\Big| \sim \Big|\int_0^T \|f(t)\|^2 (t - \check{\tau})\langle\eta\rangle dt\Big| \tag{D.8}$$

after integrating by parts :

$$\Big|\int_0^T \|f(t)\|^2 (t - \check{\tau})\langle\eta\rangle dt\Big| = |\langle\eta\rangle|\Big|(T - \check{\tau})\int_0^T \|f(s)\|^2 ds - \int_0^T \int_0^t \|f(s)\|^2 dsdt\Big| \tag{D.9}$$

$$\sim |\langle\eta\rangle|\Big|T(T - \check{\tau})\langle\|f\|^2\rangle - \frac{T^2}{2}\langle\|f\|^2\rangle\Big| \tag{D.10}$$

$$\sim \Big|\langle\eta\rangle\langle\|f\|^2\rangle T^2\Big(\frac{1}{2} - \frac{\check{\tau}}{T}\Big)\Big| \tag{D.11}$$

which means that :

$$\Big|\int_0^T \|f(t)\|^2 \Big(\int_{\check{\tau}}^t \eta(r)dr\Big)dt\Big| \sim |\langle\eta\rangle\langle\|f\|^2\rangle| \times T^2|\Big(\frac{1}{2} - \frac{\check{\tau}}{T}\Big)| \tag{D.12}$$

This relation shows that if $\frac{\check{\tau}}{T} \not\to \frac{1}{2}$ as $T \to \infty$, the second term grows as T^2 while the first term grows, at most, as T. This contradicts equation (D.5) and concludes the proof.