Technical communique

On the choice of dither in extremum seeking systems: A case study

Ying Tan*, Dragan Nešić, Iven Mareels

The Department of Electrical & Electronics Engineering, The University of Melbourne, Melbourne, VIC 3010, Australia

Received 28 March 2006; received in revised form 26 April 2007; accepted 11 October 2007

Available online 4 March 2008

Abstract

We discuss how the choice of dither (excitation signal) affects the performance of extremum seeking using a benchmark situation: a static scalar map; and a simple scalar extremum seeking scheme. Our comparisons are based on the performance of the system with different dithers in terms of three performance indicators: the speed of convergence, domain of attraction and accuracy (i.e. the ultimate bound on trajectories). Our analysis explicitly shows how the dither shape affects each of these performance indicators. Our study suggests that the practitioners using extremum seeking control should consider the dither shape as an important design parameter. Computer simulations support our theoretical findings.

© 2008 Elsevier Ltd. All rights reserved.

Keywords: Extremum seeking control; Dither choice; Performance indicators

1. Introduction

Extremum seeking (ES) control is a paradigm whose goal is to find an extremum value of an unknown nonlinear mapping. Although this method dates back to the early 1950’s and 1960’s, the first rigorous local stability analysis for an ES scheme was recently proved in Krstić and Wang (2000) and later extended to semi-global stability analysis in Tan, Nešić, and Mareels (2005, 2006). This has spurred a renewed interest in this research area (Ariyur & Krstić, 2003; Guay & Zhang, 2003; Guay, Dochain, & Perrier, 2004; Peterson & Stefanopoulou, 2004; Popović, Janković, Manger, & Teel, 2003) that has lead to numerous practical implementations of the scheme.

In our case study we focus on the following unknown nonlinear static map

\[ y = h(x). \] (1)

It is assumed that \( h(\cdot) \) has, for simplicity, a maximum \( h(x^*) \). An ES mechanism will drive the output \( y(t) \) to a small neighborhood of \( h(x^*) \). In this paper, for simplicity, we focus on the scalar case, i.e., both \( y \) and \( x \) are scalar, that is the simplest version of the extremum seeking scheme.

In this paper, we show how the choice of excitation signal affects the performance of extremum seeking. We emphasize that ES schemes in the literature employ only sinusoid waves as the excitation signals or “dithers” although it has been known that ES may work well with other types of dithers (Meerkov, 1967) for many years. Dither signals are widely used in the stabilization of nonlinear control systems (Gelig & Churilov, 1998; Iannelli, Johansson, Jonsson, & Vasca, 2003; Teel, Moreau, & Nešić, 2004; Zames & Shneydor, 1977). In Iannelli et al. (2003), the relations between dither shape and closed-loop performance were analyzed, and, in particular, it was shown that the amplitude distribution function of dither plays an important role in the closed-loop performance. In this paper, our attention is centered on the relationship between three performance indicators of ES mechanism and the choice of dither. In comparing the performance of the system with various dithers, we concentrate on three performance indicators: speed of convergence, domain of convergence and accuracy (i.e. the ultimate bound on trajectories). Some of our analysis is done using an appropriate average system that approximates well
the real closed-loop behavior if certain parameters are chosen sufficiently small.

First, it is shown that an auxiliary gradient system plays a crucial role in quantifying the performance of the extremum seeker with different dithers. Indeed, it is shown that with arbitrary dither shape it is possible to tune the parameters in the controller and the dither amplitude to recover (with arbitrarily small error) the domain of attraction and the accuracy of the auxiliary gradient system. On the other hand, the closed-loop system convergence speed depends on the convergence speed of the gradient system, as well as a scaling factor that is a product of four numbers: a controller parameter, the amplitude and frequency of the dither and the power of normalized dither (dither of the same shape with unit amplitude and period $2\pi$). This bound is tight and it explicitly states how the dither shape affects the convergence speed. Hence, our results demonstrate that the dither shape is an important degree of freedom in the design and tuning of extremum seeking controllers and its choice should be given careful consideration.

Our results benefit from the scalar extremum seeking scheme that was introduced for the first time in Tan et al. (2006). However, our main result improves upon results in Tan et al. (2006) that were presented only for sinusoidal dither and that presented more conservative estimates than the ones derived in this paper. Indeed, we use a different proof technique from the one used in Tan et al. (2006) in order to provide tighter stability estimates and to explicitly show how performance of the extremum seeking controller is affected by the choice of dither.

2. A benchmark example

We denote the set of real numbers as $\mathbb{R}$. The continuous function $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is of class $\mathcal{KL}$ if for each fixed $t \geq 0$ the function $\beta(\cdot, t)$ is zero at zero, and strictly increasing and for each fixed $s > 0$ it is decreasing to zero as $t \rightarrow \infty$. Given a sufficiently smooth function $h : \mathbb{R} \rightarrow \mathbb{R}$, we denote its $i$th derivative as $D^i h(x)$. When $i = 1$ we write simply $Dh(x) := D^1 h(x)$.

We consider a static mapping $h(\cdot)$ with the first-order extremum seeking controller (see Fig. 1). This control law is a simplified version of the schemes considered in Krstić and Wang (2000) that was introduced in Tan et al. (2006).

In particular, this is the simplest possible variant of schemes considered in Krstić and Wang (2000) that can still illustrate that the choice of dither affects the extremum seeking in several different aspects. The model of the closed-loop system in Fig. 1 is given by:

$$\dot{x} = \delta \cdot \omega \cdot h(x + d(t)) \cdot d(t),$$

(2)

where $h : \mathbb{R} \rightarrow \mathbb{R}$ is sufficiently smooth. The signal $d(\cdot)$ is referred to as “dither” and $\delta > 0$ and $\omega > 0$ are parameters that the designer can choose. We use the following assumptions:

**Assumption 1.** There exists a maximum $x^*$ of $h(\cdot)$ such that $Dh(x^*) = 0$; $D^2 h(x^*) < 0.$

**Assumption 2.** Dither signals $d(\cdot)$ are periodic functions of period $T > 0$ (and frequency $\omega = \frac{2\pi}{T}$) that satisfy:

$$\int_0^T d(s)ds = 0; \quad \frac{1}{T} \int_0^T d^2(s)ds > 0; \quad \max_{s \in [0, T]} |d(s)| = a;$$

where $a > 0$ is the amplitude of the dither.

The parameter $\omega$ in (2) is chosen to be the same as the frequency of the dither signal.

For comparison purposes, three special kinds of dither are used repeatedly in our examples: sine wave, square wave and triangle wave. The sine wave is defined in the usual manner. The square wave and triangle wave of unit amplitude and period $2\pi$ are defined as follows:

$$sq(t) := \begin{cases} 1, & t \in [2\pi k, \pi (2k + 1)] \\ -1, & t \in [\pi (2k + 1), 2\pi (k + 1)] \end{cases}$$

$$tri(t) := \begin{cases} \frac{2}{\pi} (t - 2\pi k), & t \in \left[2\pi k, \pi \left(2k + \frac{1}{2}\right)\right] \\ \frac{2}{\pi} (-t - \pi (2k - 1)), & t \in \left[\pi (2k + \frac{1}{2}), \pi \left(2k + \frac{3}{2}\right)\right] \\ \frac{2}{\pi} (t - \pi (2k + 2)), & t \in \left[\pi (2k + \frac{3}{2}), 2\pi (k + 1)\right] \end{cases}$$

Note that by definition, the signals $sin(t)$, $sq(t)$ and $tri(t)$ are of unit amplitude and period $2\pi$. We can generate similar signals of arbitrary amplitude $a$ and frequency $\omega$, e.g., $a \cdot sq(\omega t)$. We will often use the “power” (average of the square of the signal) of the normalized dithers (unit amplitude and period $2\pi$). In general, we use $P_d$ to denote the power of normalized dithers $d(\cdot)$ and $P_d := \frac{1}{T} \int_0^T d^2(s)ds$. By computation, we have

$$P_{sq} = \frac{1}{2}; \quad P_{sin} = \frac{1}{2}; \quad P_{tri} = \frac{1}{3}.$$ (4)

The power for dither $d(\cdot)$ with amplitude $a \neq 1$ is equal to $a^2 P_d$. We emphasize that our results apply to arbitrary dithers satisfying Assumption 2.

**Remark 1.** Assumption 2 is needed in averaging based analysis of (2). We note that most extremum seeking literature (see Ariyur and Krstić (2003)) uses dither signals of the form $d(t) = a \cdot sin(\omega t)$ which obviously satisfy our Assumption 2.
Introducing the coordinate change \( \tilde{x} = x - x^* \), we can rewrite (2) as follows:

\[
\dot{\tilde{x}} = \delta \cdot \omega \cdot h(\tilde{x} + x^* + d(t)) \cdot d(t) =: \delta \cdot \omega \cdot f(t, \tilde{x}, d).
\] (5)

It was shown in Tan et al. (2006, Theorem 1) that under a stronger version of Assumption 1 (uniqueness of the maximum) and with the sinusoidal dither \( d(t) = a \cdot \sin(\omega t) \), where \( \omega = 1 \) (that satisfies Assumption 2) and \( \tilde{x}_0 = \tilde{x}(t_0) \), we have that for any compact set \( D \) and any \( \nu > 0 \) we can choose parameters \( a, \delta > 0 \) and find a class \( K_c \) function \( \beta \), which depends on \( \delta \) and \( d(\cdot) \), such that the solutions of the closed-loop system (5) satisfy:

\[
|\tilde{x}_0| \in D \Rightarrow |\tilde{x}(t)| \leq \beta(|\tilde{x}_0|, t - t_0) + \nu, \quad \forall t \geq t_0 \geq 0.
\] (6)

Note that \( D, \nu \) and \( \beta \) are performance indicators since they quantify different aspects of the performance of the ES algorithm. We will show later that each of these indicators is affected by our choice of dither \( d(\cdot) \) and the parameter \( \delta > 0 \). In particular, we have that:

- **Speed of convergence** of the algorithm is captured by the function \( \beta \).
- **Domain of convergence** is quantified by the set \( D \).
- **Accuracy** of the algorithm is quantified by the number \( \nu > 0 \) since all trajectories starting in the set \( D \) eventually end up in the ball \( B_\nu \), where we have that \( |x(t) - x^*| \leq \nu \).

Obviously, faster convergence speed, larger domain of convergence and smaller \( \nu \) are always preferred in applications.

It turns out that a direct analysis of the system (5) to estimate \( D, \nu, \beta \) is hard but the system can be analyzed via an appropriate auxiliary averaged system. We will carry out such an analysis in the next section.

### 3. Main result

In this section, we present the main result (Theorem 1) that describes how different dithers affect the domain of attraction and speed of convergence, as well as the accuracy of extremum seeking control. It is shown that the square wave produces the fastest convergence among all signals with the same amplitude and frequency, if the amplitude \( a \) and the parameter \( \delta \) in the controller are sufficiently small. Moreover, it is shown that in the limit as the amplitude is reduced to zero, all dithers yield almost the same domain of attraction and accuracy.

Consider the following auxiliary gradient system:

\[
\dot{\xi} = Dh(\xi + x^*).
\] (7)

Because of Assumption 1, the system (7) has the property that \( x^* \) is an asymptotically stable equilibrium.\(^1\) Let \( D \) denote the domain of attraction of \( x^* \) for the system (7) and note that since \( h(\cdot) \) is assumed smooth, the set \( D \) is a neighborhood of \( x^* \). In other words, a consequence of Assumption 1 is that there exist \( \beta \in K_c \) and a set \( D \) such that for all \( t \geq 0 \) the solutions of (7) satisfy:

\[
\xi_0 \in D \Rightarrow |\xi(t)| \leq \beta(|\xi_0|, t).
\] (8)

Using this auxiliary system, we can state our main result:

**Theorem 1.** Suppose that Assumption 1 holds and consider the closed-loop system (2) with an arbitrary dither \( d(\cdot) \) for which Assumption 2 holds, where \( a > 0 \) is the dither amplitude. Let \( D \) and \( \beta \) come from (8). Then, for any strict compact subset \( \hat{D} \) of \( D \) and any \( \nu > 0 \), there exist \( a^* > 0 \) and \( \delta^* > 0 \) such that for any \( a \in (0, a^*], \delta \in (0, \delta^*] \) and any \( \omega > 0 \) we have that solutions of (2) satisfying:

\[
\tilde{x}_0 \in \hat{D} \Rightarrow |\tilde{x}(t)| \leq \beta(|\tilde{x}_0|, t - t_0) + \nu.
\] (9)

**Sketch of proof of Theorem 1.** Consider the system (2). First, let \( \omega > 0 \) be arbitrary frequency of the dither and introduce the change of time scale \( \tau := \omega t \). Then, we can rewrite (2) as follows:

\[
\frac{d\tilde{x}}{d\tau} = \delta \cdot \omega \cdot h(\tilde{x} + x^* + d\left(\frac{\tau}{\omega}\right)) \cdot d\left(\frac{\tau}{\omega}\right),
\]

where \( d\left(\frac{\tau}{\omega}\right) \) has period \( 2\pi \) in time scale \( \tau \). We apply the Taylor series expansion for \( h\left(\tilde{x} + x^* + d\left(\frac{\tau}{\omega}\right)\right) \) around \( \tilde{x} + x^* \) and using averaging method in Tan et al. (2006) and Sanders and Verhulst (1985), we obtain the average system in the form:

\[
\frac{d\tilde{x}}{d\tau} = \delta \cdot \omega^2 \cdot P_d \cdot Dh(\tilde{x} + x^*).
\] (10)

Using a transformation of coordinates given in Tan et al. (2006), we can show that the actual system (2) can be regarded as the average system (10) that is additively perturbed with the regular perturbations that can be reduced arbitrarily by reducing \( \delta \) and \( a \) simultaneously. Hence, we can show appropriate closeness of solutions on compact time intervals between the actual and average systems. We can use the trajectory based proofs for stability via averaging (Teel, Peuteman, & Aeyels, 1999) to complete the proof. \( \diamond \)

**Remark 2.** We emphasize that the auxiliary gradient system (7) plays a crucial role in terms of achievable performance of the extremum seeking controller as \( D \) and \( \beta \) are independent of the choice of dither.

**Remark 3.** We now discuss Theorem 1 in more detail to explain how dither shape affects the domain of attraction, accuracy and convergence speed of the closed-loop system. We note that the controller parameter \((a, \delta)\) needs to be tuned appropriately in order for Theorem 1 to hold.

**Domain of attraction:** It is shown that any dither satisfying Assumption 2 can yield a domain of attraction \( \hat{D} \) that is an arbitrary strict subset of the domain of attraction of the gradient system (7) if \( a \) and \( \delta \) are sufficiently small. We emphasize that \( \omega > 0 \) can be arbitrary and \( a \) and \( \delta \) do not depend on it.

**Accuracy:** The ultimate bound \( \nu \) approaches to zero for any dither satisfying Assumption 2 if \( a \) and \( \delta \) tend to zero. Hence,
in the limit, all dithers perform equally well in terms of domain of attraction and accuracy.

**Convergence speed:** We emphasize that $\beta$ in (9) is the same as $\beta$ in (8) for any dither $d(\cdot)$. The main difference in speed of convergence comes from the scaling factor within the function $\beta$:

$$\delta \cdot \omega \cdot a^2 \cdot P_d, \quad \text{(11)}$$

where $\delta$ is a controller parameter, $a$ and $\omega$ are respectively the amplitude and frequency of dither and $P_d$ is the power of the normalized dither. Note also that $\omega \delta$ is the integrator constant in Fig. 1. Also, note that Theorem 1 holds for sufficiently small $a$ and $\delta$ that are independent of $\omega$ which is an arbitrary positive number. Obviously, if the product (11) is larger (smaller) than 1 then the closed-loop system (2) converges faster (slower) than the auxiliary gradient system (7).

According to Remark 3, for fixed $a$, $\delta$ and $P_d$, we have that the larger the $\omega$, the faster the convergence. Simulations in Example 1 verify our analysis.

Suppose now that $a$, $\delta$ and $\omega$ are fixed and we are only interested in how the shape of dither affects the convergence. As we change dither, its (normalized) power $P_d$ changes and as we can see from (4) that the square wave will yield twice larger normalized power than the sine wave and three times larger power than the triangle wave. Consequently, we can expect twice faster convergence with the square wave than with the sine wave and three times faster convergence than with the triangle wave. Simulation results in Example 2 that we present in the sequel are consistent with the above analysis.

**Remark 4.** A weaker version of Theorem 1 was proved in Tan et al. (2006) for the sine wave dither only. Indeed, the results in Tan et al. (2006) do not consider arbitrary dither and the domain of attraction and convergence estimates are not as sharp as in Theorem 1. For instance, the relationship of convergence rate and the domain of attraction to the auxiliary system (7) was not shown in Tan et al. (2006) as this was impossible to do using the Lyapunov based proofs used in this reference. On the other hand, using the trajectory based proofs adopted in this paper, we can prove tight estimates as outlined in Theorem 1. Moreover, in Tan et al. (2006) it was not clear how the dither power $P_d$ affects the convergence rate of the average system. Note that the function $\beta$ in (9) is the same for any dither satisfying Assumption 2 and the only difference comes from the parameters in (11). However, the values of $a^*$ and $\delta^*$ are typically different for different dithers.

**Remark 5.** We note that one can state and prove a more general version of Theorem 1 that applies to general dynamical plants and in this case $h(\cdot)$ is an appropriate reference-to-output map. With extra assumptions on the plant dynamics, one can use singular perturbation theory to prove this more general result (see for instance Tan et al. (2006) for a Lyapunov based proof in the case of sine wave dither). However, in this case we will need to require that $\omega$ is sufficiently small.

**Remark 6.** It is obvious that the power of the normalized square wave is larger than or equal to the power of any other normalized dither satisfying Assumption 2. In other words, for fixed $\delta$ and $a$ for which (9) holds, the square wave is guaranteed to produce the fastest convergence over all dithers with the same amplitude and frequency.

### 4. Discussions and examples

It has been shown in Theorem 1 that the convergence speed of the ES systems depends on the choice the dither shape $P_d$, amplitude $a$ and frequency $\omega$ as well as $\delta$. It is also shown that the domain of the attraction and accuracy of all dithers are almost the same when both $a$ and $\delta$ approach to zero. In this part, we use two examples to illustrate various behaviors and simulations to confirm our theoretical findings. Our results should motivate the users of extremum seeking control to experiment with different dithers in order to achieve the desired trade-off between convergence, domain of attraction or accuracy. The first example illustrates that increasing the frequency of dither while keeping $a$, $\delta$ and $P_d$ the same yields faster convergence.

**Example 1.** Consider the quadratic mapping

$$h(x) = -(x + 4)^2, \quad \text{(12)}$$

where $Dh(\tilde{x} + x^*) = -2\tilde{x}$, where $\tilde{x} = x - x^*$ and $x^* = -4$. It is trivial to see that in this case $D = R$ and $\beta(s, t) = se^{-2t}$ (see Theorem 1). The dither is chosen to be $d(t) = asin(\omega t)$. Hence, from (4) we have $P_{sin} = 1/2$. The output response with different frequencies is shown in Fig. 2. It is clear that the larger the $\omega$ is, the faster the convergence.

Next, we will discuss the situation when $h(\cdot)$ is a quadratic mapping, where we assume $\omega = 1$. Consider the simplest possible case of quadratic maps: $h(x) = -x^2 + ax + a_0$. Consequently, we have that $Dh(x) = -2x + a_1$ and since $x^* = \frac{a_0}{2}$, in new coordinate $Dh(\tilde{x} + x^*) = -2\tilde{x}$, where $\tilde{x} := x - x^*$, we have

$$Dh(\tilde{x} + x^*) = -2\tilde{x}.$$ 

Hence, the auxiliary gradient system (7) is $\dot{\xi} = -2\xi$. It is not hard to show that in this case for arbitrary dither $d(\cdot)$ satisfying...
Assumption 2 we have that the average system is of the form:
\[ \dot{x} = -2a^2 P_0 \delta \tilde{x}, \]  
with the solution \[ \tilde{x}(t) = e^{-2a^2 P_0 \delta t} \tilde{x}_0. \] That is the average system for any dither satisfying Assumption 2 is globally exponentially stable. For square wave, sine wave and triangular wave, we have from (4) that the following holds for all \( x_0 \in \mathbb{R}, t \geq 0, \) respectively: \[ |\tilde{x}_{sq}(t)| = e^{-2a^2 P_0 \delta t} |\tilde{x}_0|; \] \[ |\tilde{x}_{sin}(t)| = e^{-a^2 P_0 \delta t} |\tilde{x}_0|; \text{ and} \] \[ |\tilde{x}_{tri}(t)| = e^{-\frac{2}{3}a^2 P_0 \delta t} |\tilde{x}_0|. \] The square wave produces the fastest speed of convergence for the average system among all dithers with the same amplitude. The same can be concluded for the actual system using the proof of Theorem 1. This suggests that the square-wave dither should be the prime candidate to use in the ES system for fast convergence speed, although this dither is rarely considered in the literature (Ariyur & Krstić, 2003). The simulation results shown in Example 2 illustrates that the convergence speed of ES with the square wave is fastest among all dithers with the same amplitude.

Example 2. The simulation is done in system (12) where \( \omega = 1. \) The averaged system is given in (13). The simulation result is shown in Fig. 3. Simulations show that the extremum seeking controller with the square-wave dither converges fastest.

In Example 2, it is also observed that the domain of the attraction and accuracy of three dithers are almost same when \( a \) and \( \delta \) are both small.

5. Conclusions

We have presented results that illustrate how the choice of dither affects the performance of an extremum seeking scheme. Our results demonstrate that the dither is an important design parameter to be considered when tuning the extremum seeking controller. For instance, it is shown that for small amplitudes the square wave provides the best convergence rate among all dithers of the same amplitude and frequency. Our examples illustrate further the flexibility that may be gained through dither design and simulations support our theoretical findings.

References