Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

A More Optimal Stochastic Extremum Seeking Control Using Fractional Dithering For A Class of Smooth Convex Functions ⋆

Derek Hollenbeck * **YangQuan Chen** *

* *University of California, Merced, Merced, CA 95343 USA (e-mail: ychen53@ucmerced.edu, dhollenbeck@ucmerced.edu).*

Abstract: Adaptive control such as extremum seeking control (ESC) can be a very useful tool to optimize problems with smooth convex functions. However, some systems can be noisy or contain non-convex regions where the solution may result in a local minimum or maximum. Using perturbation based ESC with appropriate amplitude, one can overcome local regions to find the global extremum. This work introduces fractional dithering noise into stochastic ESC to improve performance for a class of smooth convex functions. Noise with long range dependent behavior can yield a more optimal solution.

Keywords: Stochastic adaptive control,Extremum seeking and model free adaptive control,Fractional systems, fractional-order systems, Dithering

1. INTRODUCTION

Source seeking is an important topic when it comes to gas and oil industries. Where even small leaks can cause big (and sometimes fatal) problems, such as the 2010 San Bruno incident in southern California. These leaks are sometimes referred to as fugitive leaks and are difficult to resolve. There has been work done to localize leaks over the years. Matthes et al. (2005) looked at spatially distributed systems and inverted the steady state solution. Mesquita et al. (2008) looked at Optimotaxis using a stochastic switching method based on *Escherichia Coli* which mimics a chemotatic approach using a run and tumble controller. Other researchers like Nehorai et al. (1995) used an information based approached which uses the a combination of the maximum likelihood principle and the Fischer information matrix to calculate the gradient of the Cramer Rao bound. This points the searching in the direction of the source in a similar way to the flux measurements shown in Zarzhitsky et al. (2004). Most of the work done using source seeking is approximate or requires heavy computational resources to calculate the analytical solution or numerically simulate. Thus trying source seeking from a control perspective with model free approach seems desirable.

However, in source seeking the plant information is generally unknown and the measured output of the plant is noisy and sparse. Therefore, these systems are not explicitly convex. Considering that the average system is convex, such as the Gaussian plume model, but contains local extrema due to turbulence or some other perturbation. One can possibly leverage convex optimizers such as the extremum seeking control (ESC) to minimize some cost function J for the system.

In this work we look at fractional dithering into the ESC framework by introducing fractional Gaussian noise (fGn) and Symmetric alpha stable (SαS) as a stochastic perturbation. The paper is organized as follows: section 2, overviews of ESC framework; section 3, covers the definition of fractional Gaussian noise; section 4, class of convex functions; section 5, goes over averaging principle and the non-robust stability analysis; section 6, simulation; section 7, results and performance; section 8 discusses conclusions and future work.

⋆ National Science Foundation Research Traineeship Grant DGE - 1633722.

2. ESC OVERVIEW

The idea of ESC has been shown as early as Leblanc (1922) on electric railways. There are many different types of ESC including; Analog optimizers, Numerical optimization, Sliding mode control, and adaptive control (Zhang and Ordóñez (2011)). These variations can even include state machines. In this paper we are focusing on a specific type of perturbation based ESC called stochastic ESC (SESC). The ideas of SESC can be found in the works of (Liu and Krstic (2012); Manzie and Krstic (2009)) and under constraints (Coito et al. (2005)). Even under a discrete setting (Liu and Krstić (2014)

The principle of ESC relies on the minimization of a cost function given some dynamical system. The cost function takes the output of the plant as an input and its output exploited using a perturbation strategy. This involves taking the fluctuations of the cost function by way of high pass filter and multiplying the perturbed control signal to it. This high pass filter can be thought of as a taking the derivative of the signal as to explore the slope of the cost function. The multiplication then will give rise to a mainly negative or mainly positive signal. Typically, a deterministic sinusoidal signal is used as the perturbation (such as $\nu = A\sin\omega t$ in Fig. 1). Here we do not explicitly use a cost function.

Coito et al. (2005); Liu and Krstic (2012) introduced stochastic probing signals into ESC ($\nu = \eta$, in Fig. 1) due to the bounded nature of the sine function and the possibility to restrict the algorithms region of attraction. The nature of the sinusoidal perturbations require the mean to be zero and variance to be positive. Using Gaussian white noise this is preserved. Additionally the stochastic ESC can be bounded as $\nu = A\sin\eta$ in Fig. 1.

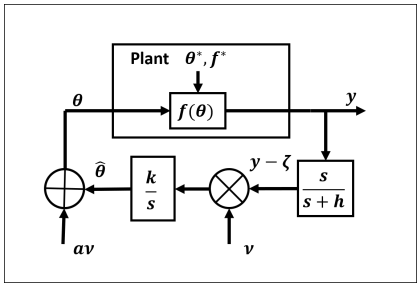

Fig. 1. Control diagram of ESC.

3. DITHERING NOISES

In digital audio, dithering refers to the low level noise added to reduce errors during changing of bit depth. The word dither means "nervous vibration" and can be used to improve reduced images as well. In the context of ESC the stochastic perturbations can be seen as a type of dithering. However, not all noise is the same and thus we introduce fractional dithering to improve performance.

3.1 Fractional Gaussian Noise

Fractional Gaussian Noise (fGn) can be derived in a similar manner as Gaussian Noise (Gn). By taking the difference between successive steps in Brownian motion (Bm) one can arrive at Gn. The Riemann-Liouville fractional integral can be used to define fractional Brownian motion (fBm) and is given below,

$$B_H(t) = \frac{1}{\Gamma(H+1/2)} \int_0^t (t-s)^{H-1/2} dB(s).$$

Here the term $dB(s)$ is the general definition of white noise, the term H is the Hurst parameter and $\Gamma(\cdot)$ is the usual gamma function. We can see that depending on Hurst parameter the motion can be,

- Brownian motion with $H = 1/2$
- Positively correlated $H > 1/2$
- Negatively correlated $H < 1/2$.

For $H > 0.5$ the process exhibits long-range dependence such that,

$$\sum_{n=1}^{\infty} E[B_H(1)(B_H(n+1) - B_H(n))] = \infty.$$

We can denote the fGn based on Hurst parameter as, $fGn_H(k) = B_H(k+1) - B_H(k)$. Where k here is used as a discrete time step and not the gain used in the control diagram. In our case we use low pass filtered white noise with high cutoff ratio initially. This process is sometimes referred to as an Ornstein-Uhlenbeck (OU) process. Replacing white noise with fGn creates a fractional OU process with new tuning parameter H. To calculate or compute the fractional integral one can use the Gaussian quadrature, Cholesky decomposition method, or circulant embedding by Dietrich and Newsam (1997). Observation of fGn by itself may be difficult to abstract the correlations in-between time steps. These characteristics can be quite easily seen by looking at fBm (see Fig. 2). The probability density function (PDF) in Fig. 3 shows the that the step lengths still obey a Gaussian distribution. However, by examining the power spectrum density (PSD) we can observe the persistent behavior of fGn at values of $H >= 0.5$. If $H < 0.5$ the slope of the PSD

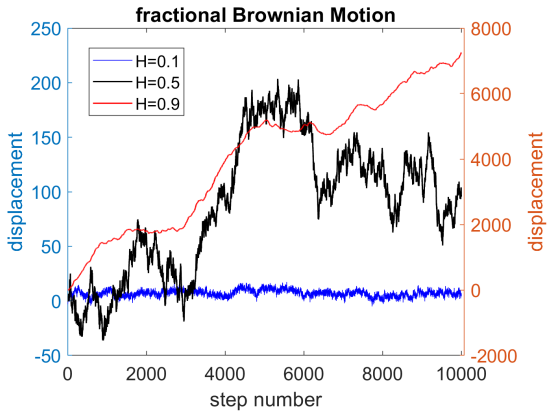

Fig. 2. Plot of fBm showing (left axis) neutral $H = 0.5$, and anti-persistent behavior $H = 0.1$ and (right axis) persistent behavior $H = 0.9$.

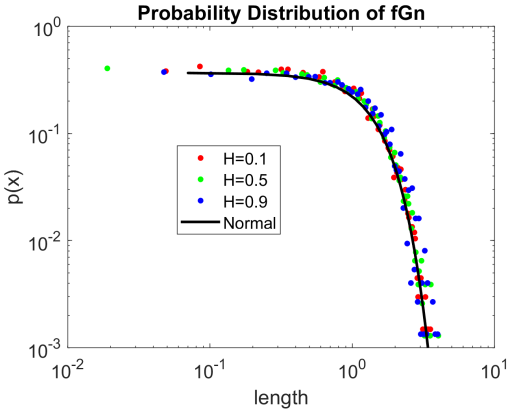

Fig. 3. Step length distribution for different values of H.

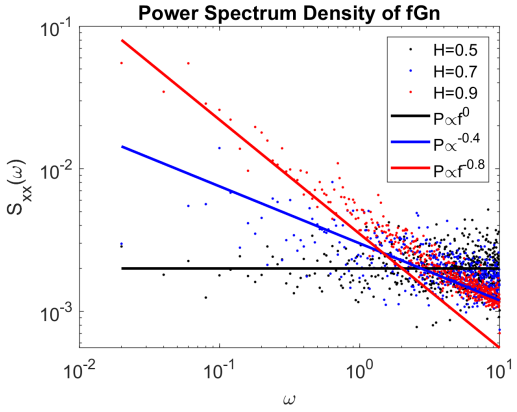

Fig. 4. The PSD shows the behavior in fGn for $H > 0.5$ which $S_{xx}(\omega) \propto \omega^{-(2H-1)}$.

line will become positive, indicating that there is much more high frequency behavior (anti-persistent).

3.2 Symmetric α-Stable Noise

Alpha stable noise can be written in terms of four key parameters α, β, γ, and b which can change the distribution uniquely. α is called the characteristic exponent and measures the "thickness" of the tails in the distribution. β, is a symmetry parameter that can skew the distribution to the left or right, here

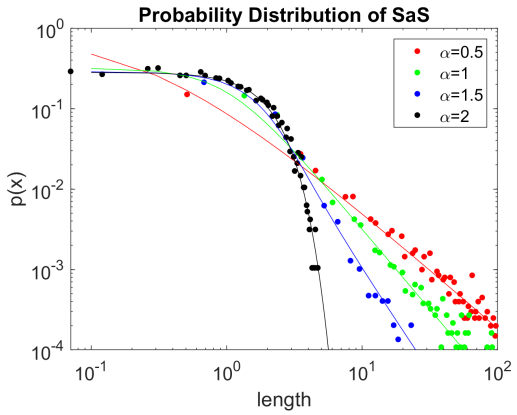

Fig. 5. The long range dependence can be seen in the heavy tails of the probability distribution function of SαS for various α.

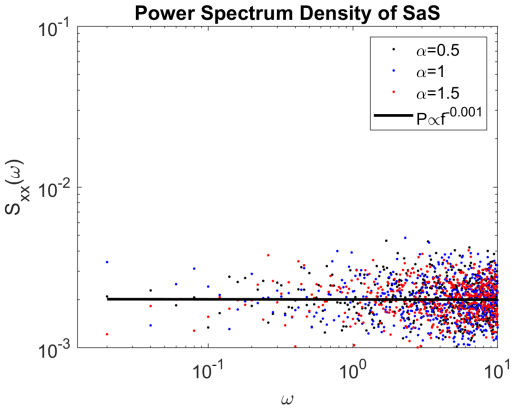

Fig. 6. The PSD of SαS can be seen as a constant for various α indicating no kind persistence or memory in the signal.

we let $\beta = 0$ which results in a symmetric α-stable or SαS. γ is the scale parameter, also called dispersion, and is similar to the variance in the normal Gaussian case when $\alpha = 2$. The last parameter is b and is referred to as the location parameter. It is the mean when $1 < \alpha \leq 2$ and the median when $0 < \alpha < 1$.

The standard SαS density function is given below where $\alpha > 0$.

$$f_\alpha(x) \begin{cases} \frac{1}{\pi x}\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k!}\Gamma(\alpha k+1)|x|^{-\alpha k}\sin\frac{k\alpha\pi}{2}, & \alpha < 1 \\ \frac{1}{\pi\alpha}\sum_{k=1}^{\infty}\frac{(-1)^{k}}{2k!}\Gamma(\frac{2k+1}{\alpha})x^{2k}, & 1 \leq \alpha \leq 2. \end{cases}$$

The mean is defined and zero for $1 < \alpha \leq 2$ and variance is ∞ for $\alpha < 2$. Fractional lower order moments (FLOM) can be used to further analyze SαS for $\alpha > 1$. The p-th order moment of SαS must be less than α otherwise the moment is ∞. See Shao and Nikias (1993); Nikias and Shao (1995), for more details.

The PDF of SαS is given in Fig. 5. The heavy tail behavior can be observed as they decay at a much slower rate than the traditional normal distribution as α decreases. The PSD of SαS is shown in Fig. 6 to be flat. Thus the SαS noise behaves like white noise with a heavy tail.

4. NON-SMOOTH CONVEX FUNCTIONS

Consider a class of convex functions that can be noisy, contain local minimums and maximums, and is separable into an average function $\bar{f}$ and deviation function $\tilde{f}$.

$$f(\boldsymbol{x},u) = \bar{f}(\boldsymbol{x},u) + \tilde{f}(\boldsymbol{x},u).$$

The function $\bar{f}(\boldsymbol{x},u))$ can be given in the time interval T such that it is convex $\forall \boldsymbol{x} \in \Omega$. The expectation of $\tilde{f}(\boldsymbol{x},u)$ is zero and the standard deviation is given as $\sigma_{\boldsymbol{x}}$ for some time interval T. The standard deviation σ in this class of functions represents the amplitude of noise or turbulence in the function.

5. STABILITY ANALYSIS

Using the averaging principle, as illustrated nicely in Liu and Krstic (2012), the error dynamics of the averaged system can be shown to be exponential stable in the Lyapunov sense.

5.1 Averaging principle

Given a system with a small positive parameter ϵ and a function $f(x,u)$ that is T-periodic in t.

$$\frac{dZ_t^\epsilon}{dt} = \epsilon f(Z^\epsilon(t+T),\xi_t) = \epsilon f(Z(t),\xi_t), \quad Z_0^\epsilon = x.$$

We make the change of variables $X_t^\epsilon = Z_{t/\epsilon}^\epsilon$ such that,

$$\frac{dX_t^\epsilon}{dt} = f(X_t^\epsilon,\xi_{t/\epsilon}), \quad X_0^\epsilon = x.$$

Considering that the system is bounded, continuous and satisfies global Lipschitz condition the average dynamics can be written as

$$\frac{d\bar{X}_t^\epsilon}{dt} = \bar{f}(\bar{X}_t), \quad \bar{X}_0 = x,$$

given, ξ_t is periodic or the sum of periodic functions. Then by using the Gronwall-Bellman Lemma (Khalil (2015) and in Appendix. A) the error dynamics can be written as,

$$|X_t^\epsilon - \bar{X}_t| \leq K\int_0^t |X_s^\epsilon - \bar{X}_s|ds + C(\epsilon)$$

$$\sup_{0\leq t\leq T}|X_t^\epsilon - \bar{X}_t| \leq C(\epsilon)e^{KT}$$

$$C(\epsilon) \triangleq \sup_{0\leq t\leq T}\left|\int_0^t [f(\bar{X}_s^\epsilon,\xi_{s/\epsilon}) - \bar{f}(\bar{X}_s)]ds\right|.$$

It can be shown that $C(\epsilon) \to 0$ as $\epsilon \to 0$. Thus the averaged system will converge to the original system.

5.2 Extension from Gaussian Stability

To show the stability of the stochastic ESC with Gaussian noise, we consider a control law $u = g(x,\theta)$ and smooth function $x = l(\theta)$ s.t. $f(x,u) = f(\theta)$. Defining parameter θ as,

$$\theta(t) = a\eta(t) + \hat{\theta}(t)$$

Here $\eta(t)$ is the stochastic unbounded noise that is added according to the transfer function where h is the high frequency cutoff.

$$\eta(t) = \frac{\sqrt{\epsilon}q}{\epsilon s + h}[Gn].$$

We can define the error of parameter $\tilde{\theta}$ as,

$$\tilde{\theta}(t) = \theta^* - \hat{\theta}(t),$$

where the $\hat{\theta}(t)$ is the estimate of the parameter θ. Thus, the error dynamics then become,

$$\dot{\tilde{\theta}}(t) = -\dot{\hat{\theta}}(t) = k\eta(t)f(\theta(t)).$$

We now need to find the function $f(\theta)$ which can be approximated by using the Taylor series expansion around θ^*.

$$f(\theta) = f(a\eta+\hat{\theta})$$
$$\approx f(\theta^*)+f'(\theta^*)(a\eta-\tilde{\theta})+\frac{1}{2}f''(\theta^*)(a\eta-\tilde{\theta})^2.$$

Given that θ^* is a minimum to $f(\theta)$ then $f'(\theta^*)=0$. Using this finding and expanding the terms in the previous equation we can substitute back into the equation for the error dynamics.

$$\dot{\tilde{\theta}}(t) \approx k\eta(t)\Big[f(\theta^*)+\frac{1}{2}f''(\theta^*)(a^2\eta^2-2a\eta\tilde{\theta}+\tilde{\theta}^2)\Big].$$

By regrouping the terms of $\eta(t)$ we can then take the expectation and average the system. The error dynamics can then be broken into three terms: first moment, second moment, and third moment.

$$\dot{\tilde{\theta}}(t) \approx k\Big[\eta(t)[f(\theta^*)+\frac{1}{2}f''(\theta^*)]-\eta^2(t)af''(\theta^*)\tilde{\theta}^2(t)$$
$$+\eta^3(t)\frac{a^2}{2}f''(\theta^*)\Big]$$

$$\lim_{t\to\infty}E[\eta(t)]=0,\ \lim_{t\to\infty}E[\eta^2(t)]=\frac{q^2}{2},\ \lim_{t\to\infty}E[\eta^3(t)]=0.$$

We can then define the Lyapunov function $V=\frac{1}{2}\tilde{\theta}^2$ where,

$$\dot{V}=\tilde{\theta}\dot{\tilde{\theta}}\approx-\frac{kaq^2}{2}f''(\theta^*)\tilde{\theta}^2. \tag{1}$$

Thus in order for the error dynamics to be exponentially stable in the averaged system we need $f(x,u)$ to be convex around the equilibrium point, $f''(\theta^*)>0$. We also need the quantity, $ka>0$.

Since the fGn has zero mean and variance tends to infinity as $H\in(0.5,1)$, we can substitute it into η such that $\epsilon d\eta_H = -\eta_H dt+\sqrt{\epsilon}qdB_H$. Consider that for values of $H>0.5$ the result in (1) still holds.

For SαS, the first moment is zero for $1<\alpha<2$ and undefined for $0<\alpha<1$. Here we assume that it is zero. The third moment is undefined due to the second moment being ∞. FLOM can be used to characterize SαS but would not be directly applicable to the above analysis. Consider that by truncating the heavy tail of a SαS and normalizing the sample variance we can achieve a computational variance $<\infty$.

6. SIMULATION

When given a function map to optimize a minimum or maximum we want the SESC algorithm to be robust to the event where we may get stuck in a local minimum or the map is not sufficiently smooth (i.e. in source seeking problems). Here we explore the use of a static map. However, a dynamic map can be used given the dynamics change very slowly, otherwise, it will invoke internal model principle. Using the Rastrigin function as the function map,

$$f(\boldsymbol{x})=An+\sum_{i+1}^{n}[x_i^2-A\cos(2\pi x_i)]$$

gives direct knowledge of the well width, ε, and amplitude, A. In the source seeking framework this can be thought of as the spatial distribution snapshot of a fluid subject to turbulence. We then initialize $\theta(t=0)=\theta_0=-3$ such that we have multiple extremums before the algorithm can reach the global extremum at $\theta=0$ (see Fig. 7). To understand the characteristics for each type of noise under these conditions we ran 1000 trials per configuration choice of H and α. The values of H and α range from $0.1<H<0.9$ and $0.5<\alpha<2$.

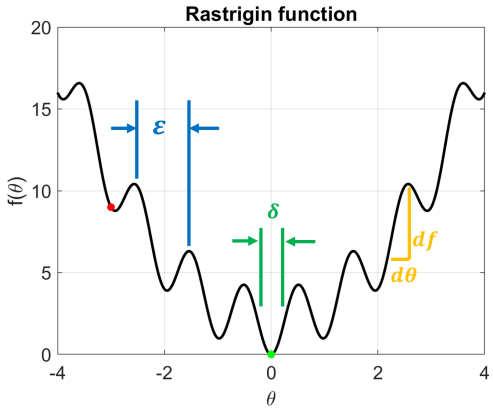

Fig. 7. Rastrigin function.

6.1 Parameter Selection

In order to compare the different noises in both the bounded and unbounded case we needed a metric. Using the root mean square (RMS) of a sinusoidal signal ($A_{RMS}=A/\sqrt{2}$) we can scale the amplitude to be proportional to well width, ε.

$$\begin{cases} aA\sin\sigma_\eta/\sqrt{2}\geq\sqrt{2}\varepsilon, & \text{Stochastic Bounded} \\ a\sigma_\eta/\sqrt{2}\geq 2\varepsilon, & \text{Stochastic Unbounded.}\end{cases}$$

Here the term σ_η represents the scaled standard deviation of the generated noise Y to satisfy the above inequality, such that $Y=X\sigma_\eta/\sigma_X$ where X is either fGn or SαS. Due to the periodic behavior of the sine function we chose to have the σ_η for the bounded case to be $\pi/2$ to maximize the output to be $A[-1,1]$. If the σ_η is chosen too big we see harmonics show up in the results as H or α is changed. Alternatively, if σ_η is chosen too small the output is too small to jump outside of the well width.

The value of k depends proportionally to the inverse of the slope of the function.

$$k\propto\frac{1}{\max_{x_i\in\Omega}|\frac{\partial f}{\partial\theta_i}|}=\frac{1}{m},\quad\text{for}\quad i=1,2,...n.$$

Here we chose $k=k_m/m$ with $k_m=20$. The value of a can be used to scale the perturbation on the filtered output and θ. Here we choose it to be 1 for simplicity. However, results can be further improved by tuning a.

The filter parameter ϵ is chosen to be 0.25, where for small ϵ the $\lim_{t\to\infty}E[\eta^2]=q^2/2$. The gain of the filter becomes $\sqrt{\epsilon}q/h$ and the cutoff frequency is $h/2\pi\epsilon$. For simplicity we choose $h=5$ and $q=1$ for all configurations leaving the noise parameters to be optimized.

6.2 Performance Metrics

The time convergence t^* is given as,

$$t^*=\min_{t\to T}t,\quad\text{s.t.}\quad|\theta^*-\hat{\theta}|<\text{tol}.$$

The stopping criteria tolerance is chosen to be 5% of ε. Selection of a small tolerance can mean no convergence. The rigorous proof of Krstic and Wang (1997) in Theorem 5.1 says that their exist a ball of initial conditions and under the assumption that $f(\theta)$ is smooth, $\dot{\theta}=f(\theta)$ is locally exponentially

stable for $\theta \in \Omega$ and $f'(\theta^*) = 0$ and $f''(\theta^*) < 0$ that solution exponentially converges to a neighborhood of order $O(A+\omega+$ small positive constant). Since our A is tuned based on well width ε it seems appropriate to set the tolerance relative to the system not the perturbation.

The δ-neighborhood is defined as the standard deviation of the estimate of θ,

$$\delta = \sqrt{E[(\hat{\theta} - \theta^*)^2]}_{t^*<t<T}.$$

The reported value of θ estimate is given as,

$$\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N} \hat{\theta}_i, \quad \text{for} \quad i = 1,2,3...N,$$

where $\hat{\theta}_i$ is the value of $\hat{\theta}_{t=T}$ and T is the length of the simulation. Here we choose $T = 100$ seconds with time step $\Delta t = 0.01$.

7. RESULTS

From the performance plots of the SESC with appropriate choices of A and k (see Figs. 8-13), suggests that the type of noise matters. We can observe that the performance of α decreases around $\alpha = 1.5$ before approaching a similar performance for the $\alpha = 2$ case. Intuitively it seems that this should not be the case. However, when scaling the SαS noise the signal is comprised mainly of low frequency parts with the occasional long jump. Due to the low pass filter the high frequency jumps get attenuated. The result is a more narrow frequency band of jump sizes and because the amplitude of the perturbations affect the SESC performance for jumping from one well to another it makes sense t^* is reduced. For fGn, the anti-persistent behavior results tends to prohibit the well jumps to happen. For $H < 0.4$ and $H > 0.8$ the global convergence is not met (Figs 10 and 13). For larger values of H (persistent behavior) it can be seen in Figs. 9 and 12 that the δ-neighborhood grows exponentially. A more optimal selection of noise can be found by minimizing a cost function based on the performance metrics mentioned earlier for $H \in (0.1, 0.9)$ and $\alpha \in (0.5, 2)$,

$$\min J \quad \text{s.t.} \quad J = t^* + \delta/A + \hat{\theta}$$

The optimal value for bounded/unbounded cases of fGn is found to be around $H = 0.63$ in this case. For SαS, the optimal values for bounded/unbounded cases is $\alpha = 2, 0.5$, which for $\alpha = 2 \rightarrow$ Gn. The result for fGn is surprising in that some LRD can improve the convergence behavior. While the result from SαS is not surprising given the effect of filtering. All of the heavy tail behavior is lost due to the type of filter. Also the increase in performance when α approaches 0.5 results in more impulse behavior. This in turn increases the frequency and in the PESC as ω increases, so does the convergence speed. However, some systems may have bandwidth limitations. This makes the persistent behavior of fGn more desirable.

8. CONCLUSION

In this work we examine SESC with fractional dithering noises from SαS and fGn for a class of smooth convex functions. The stability depends on $f''(\theta^*) > 0$, $ka > 0$ and the second moment of η (i.e. choice of H or α). We found that:

- when parameter selection is based on the well width and the inverse of the maximum gradient of $f(\theta)$. The SESC will converge to the global extremum more consistently.

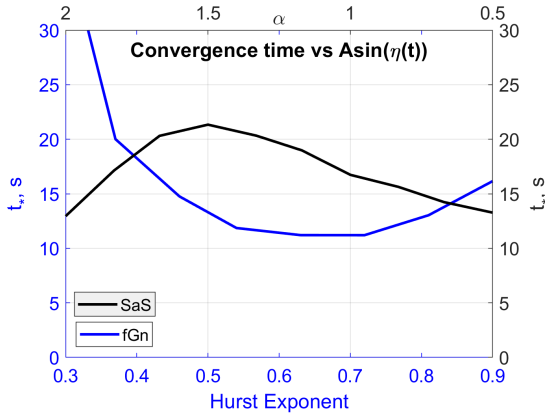

Fig. 8. The temporal convergence to the global minimum from the initial condition θ_0, denoted t^*, is shown as a function of fGn and SαS dithering.

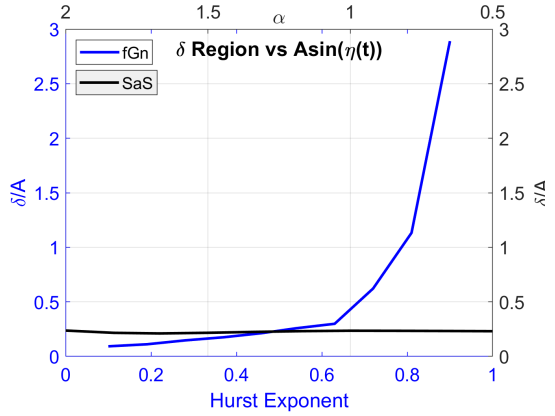

Fig. 9. The spatial deviation δ around the converged θ is shown as a function of fGn and SαS dithering.

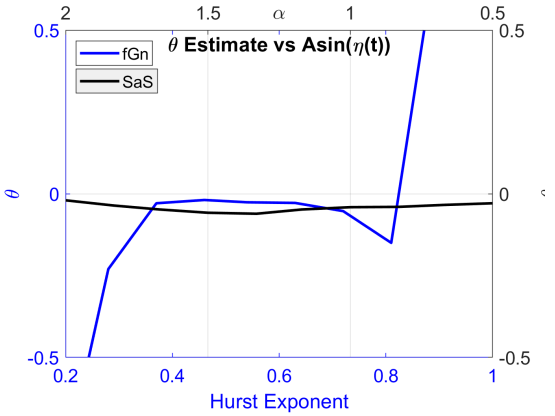

Fig. 10. The final converged θ is shown as function of fGn nd SαS dithering.

- the presence of fractional noise can improve convergence in the presence of local extrema.
- A more optimal SESC can be achieved with $H = 0.63$, in this example.

Future work will reconsider SαS and fGn under fractional order filtering and explore tempered Lévy distributions (Chen et al. (2018)) as well as examine non-static functions.

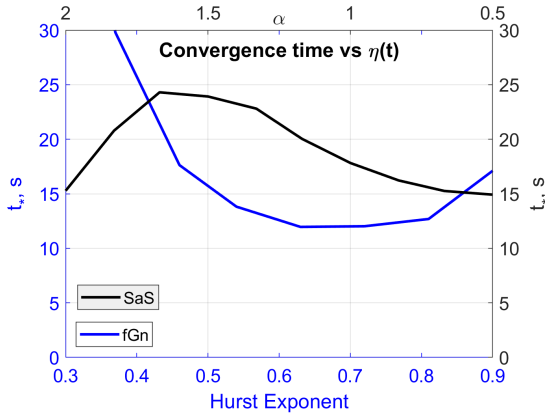

Fig. 11. The temporal convergence to the global minimum from the initial condition θ_0, denoted t^*, is shown as a function of fGn and SαS dithering.

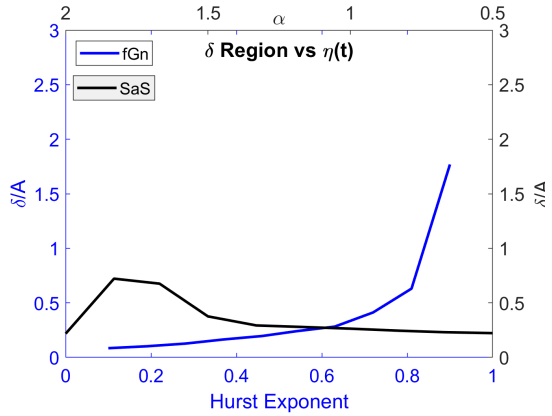

Fig. 12. The spatial deviation δ around the converged θ is shown as a function of fGn and SαS dithering.

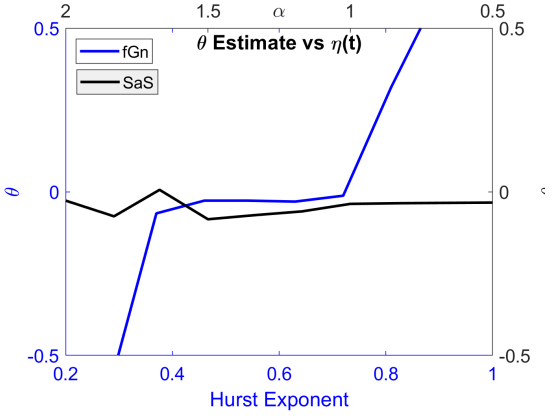

Fig. 13. The final converged θ is shown as function of fGn nd SαS dithering.

REFERENCES

Chen, Y., Hollenbeck, D., Wang, Y., and Chen, Y. (2018). On optimal tempered Lévy flight foraging. *Frontiers in Physics*, 6, 111.

Coito, F., Lemos, J., and Alves, S. (2005). Stochastic extremum seeking in the presence of constraints. *IFAC Proceedings Volumes*, 38(1), 276–281.

Dietrich, C.R. and Newsam, G.N. (1997). Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. *SIAM Journal on Scientific Computing*, 18(4), 1088–1107.

Khalil, H.K. (2015). *Nonlinear control*. Pearson New York.

Krstic, M. and Wang, H.H. (1997). Design and stability analysis of extremum seeking feedback for general nonlinear systems. In *Proceedings of the 36th IEEE Conference on Decision and Control*, volume 2, 1743–1748. IEEE.

Leblanc, M. (1922). Sur l'electrification des chemins de fer au moyen de courants alternatifs de frequence elevee. *Revue générale de l'électricité*, 12(8), 275–277.

Liu, S.J. and Krstic, M. (2012). *Stochastic averaging and stochastic extremum seeking*. Springer Science & Business Media.

Liu, S.J. and Krstić, M. (2014). Discrete-time stochastic extremum seeking. *IFAC Proceedings Volumes*, 47(3), 3274–3279.

Manzie, C. and Krstic, M. (2009). Extremum seeking with stochastic perturbations. *IEEE Transactions on Automatic Control*, 54(3), 580–585.

Matthes, J., Groll, L., and Keller, H.B. (2005). Source localization by spatially distributed electronic noses for advection and diffusion. *IEEE Transactions on Signal Processing*, 53(5), 1711–1719.

Mesquita, A.R., Hespanha, J.P., and Åström, K. (2008). Optimotaxis: A stochastic multi-agent optimization procedure with point measurements. In *International workshop on hybrid systems: Computation and control*, 358–371. Springer.

Nehorai, A., Porat, B., and Paldi, E. (1995). Detection and localization of vapor-emitting sources. *IEEE Transactions on Signal Processing*, 43(1), 243–253.

Nikias, C.L. and Shao, M. (1995). *Signal processing with alpha-stable distributions and applications*. Wiley-Interscience.

Shao, M. and Nikias, C.L. (1993). Signal processing with fractional lower order moments: stable processes and their applications. *Proceedings of the IEEE*, 81(7), 986–1010.

Zarzhitsky, D., Spears, D.F., Spears, W.M., and Thayer, D.R. (2004). A fluid dynamics approach to multi-robot chemical plume tracing. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 3*, 1476–1477. IEEE Computer Society.

Zhang, C. and Ordóñez, R. (2011). *Extremum-seeking control and applications: a numerical optimization-based approach*. Springer Science & Business Media.

Appendix A. GRONWALL-BELLMAN LEMMA

A Let $\lambda : [a,b] \to R$ be continuous and $\mu : [a,b] \to R$ be continuous and nonnegative. If a continuous function $y : [a,b] \to R$ satisfies

$$y(t) \leq \lambda(t) + \int_a^t \mu(s)y(s)ds$$

for $a \leq t \leq$, then on the same interval

$$y(t) \leq \lambda(t) + \int_a^t \lambda(s)\mu(s)exp[\int_s^t \mu(\tau)d\tau]ds$$

In particular, if $\lambda(t) \equiv \lambda$ is a constant, then

$$y(t) \leq \lambda(s)exp[\int_s^t \mu(\tau)d\tau]$$

If, in addition, $\mu(t) \equiv \mu \geq 0$ is a constant, then

$$y(t) \leq \lambda exp[\mu(t-a)]$$