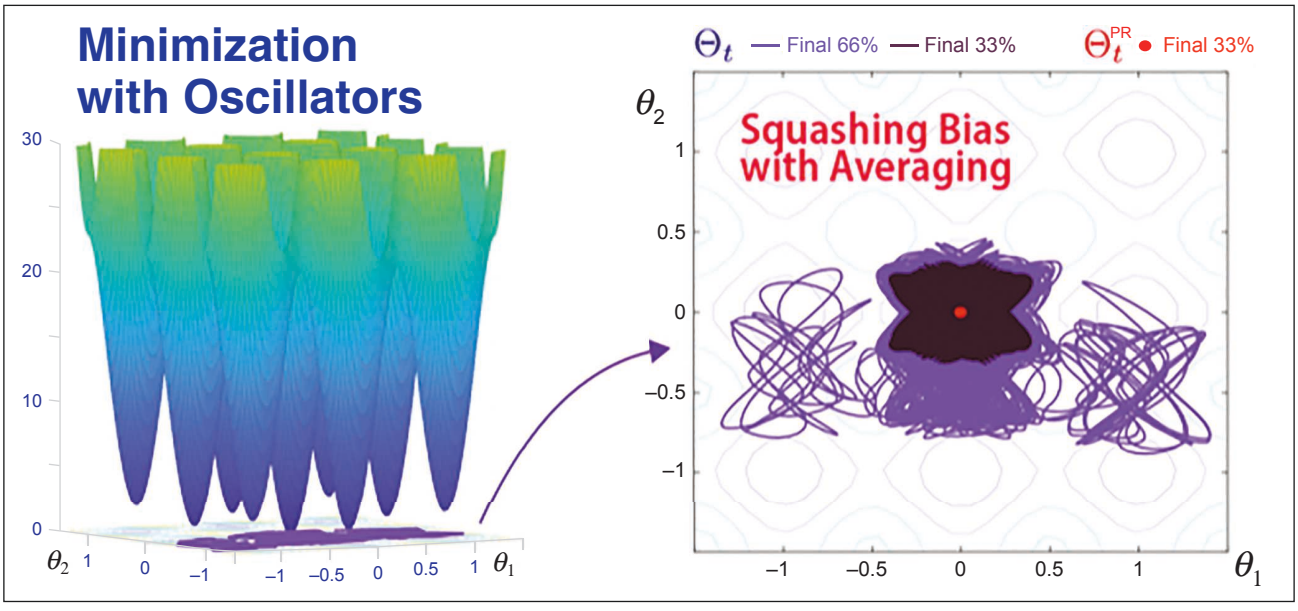
Quasi-Stochastic Approximation

DESIGN PRINCIPLES WITH APPLICATIONS TO EXTREMUM SEEKING CONTROL



CAIO KALIL LAUAND and SEAN MEYN

How can you optimize a function $\Gamma : \Re^d \to \Re$ based on evaluations of this function without access to its gradient? Kiefer and Wolfowitz proposed a solution in the early 1950s based on stochastic approximation (SA) [16], and in the 1920s, an engineer for the French railway system proposed an entirely deterministic approach that is now known as *extremum seeking control* (*ESC*) [27], [48]. Once you understand the ESC architecture, you will find that the ideas are very similar. A fundamental difference is that random noise is replaced with sinusoids for exploration.

The punchline: Techniques from the SA literature can be extended to the deterministic realm of quasi-stochastic

Digital Object Identifier 10.1109/MCS.2023.3291884
Date of current version: 18 September 2023

approximation (QSA), providing new techniques for algorithm design. When applied to ESC, we obtain algorithms that are globally stable and have astonishingly fast rates of convergence (see "Summary").

Before we can justify these claims, we require background, starting with the pioneering work of Robbins and Monro, who in [42] launched the field of SA. A summary can be found in "What Is Stochastic Approximation?" but a more concise explanation is presented here in the special case of minimization of a function Γ. An SA algorithm generates a sequence of estimates $\{\theta_n : n \geq 0\}$ of the minimizer θ^{opt} based on approximate gradient descent

$$\theta_{n+1} = \theta_n - \alpha_{n+1}\tilde{\nabla}_n\Gamma, \qquad n \geq 0 \tag{1}$$

in which $\{\alpha_{n+1} : n \geq 0\}$ is the step-size sequence and $\{\tilde{\nabla}_n\Gamma : n \geq 0\}$ is a random sequence, designed to approximate the respective gradients $\{\nabla\Gamma(\theta_n) : n \geq 0\}$. It is assumed that the approximation holds only in an average sense; the inherent filtering in (1) helps to reduce the impact of the noisy gradient estimates.

For gradient estimates that are asymptotically unbiased, theory establishing convergence is based on a proof of coupling with solutions of the gradient flow

$$\frac{d}{dt}\vartheta = -\nabla\Gamma(\vartheta). \tag{2}$$

See "What Is Stochastic Approximation?" to understand why we can expect solidarity between the gradient flow and stochastic gradient descent algorithms of the form (1).

Although there have been exciting advances in SA theory in recent decades, in many cases, this approach is not acceptable for the applications of interest in this article:

» *Constraints from physics*: Most versions of gradient-free optimization begin with the construction of $\{\tilde{\nabla}_n\Gamma : n \geq 0\}$ based on perturbed observations of the form $\mathcal{Y}_n = \Gamma(\theta_n + \varepsilon\xi_n)$, in which $\{\xi_n\}$ is a vector-valued probing sequence and $\varepsilon > 0$ is known as the *probing gain*. Within the realm of SA, the probing sequence is chosen to be independent and identically distributed (i.i.d.). Such high-frequency exploration

What Is Stochastic Approximation?

The goal of stochastic approximation (SA) is to solve the root-finding problem $\bar{f}(\theta^*) = 0$, in which $\bar{f} : \Re^d \to \Re^d$ is expressed as the expectation

$$\bar{f}(\theta) := \mathsf{E}[f(\theta, \xi)] \tag{S1}$$

with ξ a random vector taking values in $\Re^m$. In applications to optimization, the function f and distribution of ξ are selected so that $\bar{f}$ approximates a negative gradient. Theory has grown tremendously in the past few decades, driven in large part by applications to machine learning [S1], [S2], [S3] and reinforcement learning [S4], [S5] (see "Root Finding and Learning" to understand why).

The solution proposed in [42] is in essence the *ODE method*—a term coined by Ljung in [S6]. This consists of the following steps:

(i) By luck or design, ensure that the mean flow (5) is globally asymptotically stable,

(ii) Ensure that conditions are right, so that an Euler approximation of the mean flow is also globally convergent.

(iii) The basic SA algorithm is by definition the noisy Euler approximation

$$\theta_{n+1} = \theta_n + \alpha_{n+1}f(\theta_n, \xi_{n+1}), \qquad n \geq 0 \tag{S2}$$

where $\{\alpha_{n+1}\}$ is the nonnegative step-size sequence, and the sequence $\{\xi_{n+1}\}$ is random, with distribution converging to that of ξ as n tends to ∞.

The conditions ensuring convergence of $\{\theta_n\}$ to the desired value θ^* are not restrictive [S7].

While estimating bounds on the rate of convergence is far more challenging, there is now a well-developed theory based on the central limit theorem; the asymptotic covariance Σ_θ is the solution to a Lyapunov equation [S7], [S8]. Under stronger conditions (see [S9] and its references), its trace coincides with the scaled asymptotic mean square error

$$\lim_{n\to\infty}\frac{1}{\alpha_n}\mathsf{E}[\|\theta_n - \theta^*\|^2] = \text{trace}(\Sigma_\theta).$$

Lower bounds on the right-hand side are well known, along with algorithm design techniques to minimize this value.

REFERENCES

[S1] A. Fradkov and B. T. Polyak, "Adaptive and robust control in the USSR," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 1373–1378, Apr. 2021, doi: 10.1016/j.ifacol.2020.12.1882.

[S2] W. Mou, C. J. Li, M. J. Wainwright, P. L. Bartlett, and M. I. Jordan, "On linear stochastic approximation: Fine-grained Polyak-Ruppert and non-asymptotic concentration," in *Proc. Conf. Learn. Theory*, 2020, pp. 2947–2997.

[S3] E. Moulines and F. R. Bach, "Non-asymptotic analysis of stochastic approximation algorithms for machine learning," in *Proc. Adv. Neural Inf. Process. Syst. 24*, 2011, pp. 451–459.

[S4] D. P. Bertsekas, *Reinforcement Learning and Optimal Control*. Belmont, MA, USA: Athena Scientific, 2019.

[S5] S. Meyn, *Control Systems and Reinforcement Learning*. Cambridge, U.K.: Cambridge Univ. Press, 2022.

[S6] L. Ljung, "Analysis of recursive stochastic algorithms," *IEEE Trans. Autom. Control*, vol. AC-22, no. 4, pp. 551–575, Aug. 1977, doi: 10.1109/TAC.1977.1101561.

[S7] V. S. Borkar, *Stochastic Approximation: A Dynamical Systems Viewpoint*, 2nd ed. Delhi, India: Hindustan Book Agency, 2021.

[S8] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications*. New York, NY, USA: Springer-Verlag, 1997.

[S9] V. Borkar, S. Chen, A. Devraj, I. Kontoyiannis, and S. Meyn, "The ODE method for asymptotic statistics in stochastic approximation and reinforcement learning," 2021, *arXiv:2110.14427*.

may make no sense in truly online applications as the probing may be filtered out through inertia in the system or create stress on equipment.

» *Curse of variance*: In the majority of applications of SA, the mean-square error decays no faster than $O(1/n)$, as a consequence of the central limit theorem. *This slow convergence is often unacceptable.*

These constraints and curses are addressed through the flexibility we have in the applications of interest in this article: it is *we who design the exploration*. This is true in

Root Finding and Learning

In [S1], Polyak credits Tsypkin's 1971 monograph *Adaptation and Learning in Automatic Systems* [S10] for the realization that stochastic approximation (SA) is an invaluable ingredient in the creation of algorithms for learning. The following two classes of machine learning problems serve to justify Tsypkin's insight:

1) *Model-free optimization*: The goal is to approximate the minimizer of a function $\Gamma : \Re^d \to \Re$. We are free to choose the values $\{x_n\}$ to observe $y_n = \Gamma(x_n)$, but we may not have an analytical expression for the objective function or its gradient.

A close cousin is *gradient-free optimization*, whose theory began with the work of Kiefer and Wolfowitz [16] roughly two decades before [S10], with significant theoretical progress in the decades that followed.

The work of Spall stands out because of the elegant simplifications of the basic algorithms, along with analysis of convergence rates. Two versions of his simultaneous perturbation stochastic approximation (SPSA) algorithm can be expressed as SA in the form of (S2), differing only in the definition of f

$$\text{1SPSA:}\quad f(\theta,\xi) = -\frac{1}{\varepsilon}\xi\Gamma(\theta+\varepsilon\xi) \tag{S3a}$$

$$\text{2SPSA:}\quad f(\theta,\xi) = -\frac{1}{2\varepsilon}\xi[\Gamma(\theta+\varepsilon\xi) - \Gamma(\theta-\varepsilon\xi)]. \tag{S3b}$$

It will be seen that 1SPSA is a close cousin of extremum seeking control.

The 1SPSA recursion may be cast as an algorithm for model-free optimization: samples of $\Gamma(\theta_n + \varepsilon\xi_{n+1})$ may be collected from a physical system, without an analytical expression for the objective function Γ. The first-order difference approach, 2SPSA, will *not* be successful if there is substantial measurement noise.

1) *Reinforcement learning* (*RL*). It was observed in [S11] and [S12] that temporal difference methods (such as TD and Q-learning) may be regarded as SA approaches to solve a root-finding problem. Letting T denote the Bellman operator associated with the control problem of interest, and Q^θ be an approximation of the state-action value function, denote

$$\bar{f}(\theta) = \mathrm{E}[(TQ^\theta - Q^\theta)\zeta] \tag{S4}$$

in which the random vector ζ is a stationary realization of the *eligibility vector*. The root-finding problem $\bar{f}(\theta^*) = 0$ coincides with the *projected Bellman* equation [S4], [S5].

The definition of T depends on context. For the deterministic state-space model $x_{k+1} = F(x_k, u_k)$ and one-step cost function c, the total cost-value function is denoted as

$$Q^*(x,u) = \min \sum_{k=0}^{\infty} c(x_k, u_k), \qquad x_0 = x, u_0 = u$$

where the minimum is over all admissible inputs $\{u_1, u_2, \ldots\}$. The dynamic programming equation is expressed as $Q^* = TQ^*$, where for any function H, the function $H^+ = TH$ is defined by

$$H^+(x,u) := \min_{u_1}\{c(x,u) + H(F(x,u),u_1)\}.$$

Even in a fully deterministic setting, probabilistic tools are inevitable because *exploration* is a component of training algorithms for learning. Analysis is based on a steady-state realization of the input-state process. In the case of linear function approximation, $Q^\theta = \theta^\top\psi$ with $\psi(x,u) \in \Re^d$ for each state-input pair, a common choice of eligibility vector is $\zeta = \psi$, and (S5) becomes the steady-state mean

$$\bar{f}(\theta) = \mathrm{E}[([TQ^\theta](x_k,u_k) - Q^\theta(x_k,u_k))\psi(x_k,u_k)].$$

Actor-critic methods in RL may be regarded as an approach to model-free optimization in which the objective is average cost. The policy gradient theorem of [S13], a variant of Schweitzer's sensitivity formula [S14], leads to techniques to obtain unbiased estimates of the gradient of the objective function. One key ingredient is the approximation of a state-action value function, made possible through geometry revealed in the dissertations of Van Roy and Konda [S15], [S16], [S17], [S18].

REFERENCES

[S10] Y. Z. Tsypkin and Z. J. Nikolic, *Adaptation and Learning in Automatic Systems*. New York, NY, USA: Academic, 1971.

[S11] T. Jaakola, M. Jordan, and S. Singh, "On the convergence of stochastic iterative dynamic programming algorithms," *Neural Comput.*, vol. 6, pp. 1185–1201, Nov. 1994, doi: 10.1162/neco.1994.6.6.1185.

[S12] J. Tsitsiklis, "Asynchronous stochastic approximation and *Q*-learning," *Mach. Learn.*, vol. 16, pp. 185–202, 1994, doi: 10.1007/BF00993306.

[S13] R. S. Sutton, D. McAllester, S. Singh, and Y. Mansour, "Policy gradient methods for reinforcement learning with function approximation," in *Proc. Adv. Neural Inf. Process. Syst.*, 1999, vol. 12, pp. 1057–1063.

[S14] P. J. Schweitzer, "Perturbation theory and finite Markov chains," *J. Appl. Probability*, vol. 5, no. 2, pp. 401–403, Aug. 1968, doi: 10.1017/S0021900200110083.

[S15] V. Konda, "Actor-critic algorithms," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 2002.

[S16] V. R. Konda and J. N. Tsitsiklis, "On actor-critic algorithms," *SIAM J. Contr. Optim.*, vol. 42, no. 4, pp. 1143–1166, Apr. 2003, doi: 10.1137/S0363012901385691.

[S17] J. N. Tsitsiklis and B. Van Roy, "An analysis of temporal-difference learning with function approximation," *IEEE Trans. Autom. Control*, vol. 42, no. 5, pp. 674–690, May 1997, doi: 10.1109/9.580874.

[S18] B. Van Roy, "Learning and value function approximation in complex decision processes," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 1998.

applications to both optimization and reinforcement learning (RL) [S5], [47].

Probing signals can be designed so that mean-square error bounds are far smaller than $O(1/n)$. Without much effort, we obtain algorithms to achieve a mean-square convergence rate of order approaching $O(1/n^4)$.

However, such speedy rates of convergence are only possible through the use of a step-size sequence $\{\alpha_{n+1} : n \geq 0\}$ that is vanishing. If the ultimate goal is to track the evolving optimizer of a time-varying objective function, a vanishing step-size is not acceptable. In much of the article, we focus on algorithms similar to (1) in which α_{n+1} is independent of n.

Physical constraints require that we consider smooth probing. This is just one reason why we begin with a continuous time setting for algorithm construction and analysis.

What is QSA?

QSA is a deterministic analog of SA. In the fixed-gain setting that is the focus of this article, the *QSA* ordinary differential equation (*ODE*) is defined by the ordinary differential equation

$$\frac{d}{dt}\Theta_t = \alpha f(\Theta_t, \xi_t). \tag{3}$$

The gain $\alpha > 0$, m-dimensional probing signal $\boldsymbol{\xi}$, and vector field $f : \Re^d \times \Re^m \to \Re^d$ are design choices.

The mathematical objective is identical to SA: by design, the solution to the QSA ODE approximates the solution θ^* to the root-finding problem $\bar{f}(\theta^*) = 0$, with $\bar{f}$ defined by

$$\bar{f}(\theta) = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(\theta, \xi_t)\, dt. \tag{4}$$

We cannot expect convergence of $\{\Theta_t\}$ to θ^* when the gain is fixed. Instead, we obtain bounds on asymptotic bias of order $O(\alpha^2)$ and variance of order $O(\alpha^4)$. The theoretical development of QSA is also similar to SA, starting with comparison of solutions to the QSA ODE and solutions to the *mean flow*

$$\frac{d}{dt}\vartheta_t = \bar{f}(\vartheta_t). \tag{5}$$

Solidarity between the mean flow and the QSA ODE (3) may be addressed by following theory for its stochastic counterpart, or by recognizing that the constant-gain ODE may be analyzed through the *averaging principle* (see "The Averaging Principle" for a short history and references to a vast literature).

The Averaging Principle

The quasi-stochastic approximation (QSA) ODE with fixed gain (3) is not at all new to the dynamical systems community, for which solidarity of the QSA ODE and the mean flow is known as the *averaging principle*. Analysis of the larger state process $\boldsymbol{\Psi} = (\boldsymbol{\Theta}, \boldsymbol{\Phi})$ may also be cast in the setting of singular perturbation theory, in which $\boldsymbol{\Theta}$ is regarded as the slow variable.

The concepts are far older than SA, with heuristics applied in the 18th century to obtain models for coupled planetary systems. Firm theory emerged approximately one century ago [S19], which is several decades before Robbins and Monro introduced SA [42]. Averaging and singular perturbation theory grew within the control systems community beginning in the 1970s [S20] and became a foundation of adaptive control (a close cousin of RL) in the decades that followed. Any of the standard references will provide a fuller history, such as [S21], [S22], and [S23].

The academic fields of SA and singular perturbation theory are far from disjoint in terms of goals, and there has been a history of cross fertilization. The transfer of concepts from the deterministic to the stochastic domain includes the application of singular perturbation techniques in the analysis of two-timescale Markov chains [S22], or the more recent work [S24], which proposes improvements to 1SPSA that are inspired by extremum seeking control.

The main goal of this article is to transfer concepts in the reverse direction. Techniques from the SA literature have tremendous value in advancing the theory of averaging. Obviously, the most valuable is the disturbance decomposition introduced in the 1980s by Métivier and Priouret [S25], which is based on Poisson's equation for Markov chains. Multiple applications of this technique lead to the p-mean flow representation (7). We are not aware of any counterpart in the averaging literature.

REFERENCES

[S19] D. R. Smith, *Singular-Perturbation Theory: An Introduction with Applications*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[S20] P. Kokotovic, R. O'Malley, and P. Sannuti, "Singular perturbations and order reduction in control theory — An overview," *Automatica*, vol. 12, no. 2, pp. 123–132, Mar. 1976, doi: 10.1016/0005-1098(76)90076-5.

[S21] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

[S22] P. Kokotović, H. K. Khalil, and J. O'Reilly, "Singular perturbation methods in control: analysis and design," in *Classics in Applied Mathematics*. Philadelphia, PA, USA: SIAM, 1999.

[S23] J. A. Sanders, F. Verhulst, and J. Murdock, *Averaging Methods in Nonlinear Dynamical Systems*, vol. 59. New York, NY, USA: Springer, 2007.

[S24] X. Chen, Y. Tang, and N. Li, "Improve single-point zeroth-order optimization using high-pass and low-pass filters," in *Proc. Int. Conf. Mach. Learn. (PMLR)*, 2022, pp. 3603–3620.

[S25] M. Métivier and P. Priouret, "Applications of a Kushner and Clark lemma to general classes of stochastic algorithms," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 2, pp. 140–151, Mar. 1984, doi: 10.1109/TIT.1984.1056894.

Techniques from the SA literature can be extended to the deterministic realm of quasi-stochastic approximation, providing new techniques for algorithm design.

The present survey is concerned primarily with translating SA techniques to the deterministic setting. Our starting point is the representation

$$\frac{d}{dt}\Theta_t = \alpha[\bar{f}(\Theta_t) + \tilde{\Xi}_t] \tag{6}$$

in which $\tilde{\Xi}_t = f(\Theta_t, \xi_t) - \bar{f}(\Theta_t)$ is called the *apparent noise*. This noise is called *additive* if $f(\theta, \xi) - \bar{f}(\theta)$ does not depend on θ, for any value ξ. Otherwise, we say there is *multiplicative noise*.

The next step in analysis is to obtain a representation of the apparent noise through multiple applications of Poisson's equation, borrowing from SA theory techniques. This brings us to the central equation on which design guidelines are built upon: the *perturbative mean flow* (or *p-mean flow*). Its justification requires assumptions that are explained in Theorem 1 in "Part 1: QSA."

P-mean flow: The solution to the QSA ODE admits the exact description

$$\begin{aligned}\frac{d}{dt}\Theta_t &= \alpha[\bar{f}(\Theta_t) - \alpha\overline{\Upsilon}_t + \mathcal{W}_t],\\ \mathcal{W}_t &= \alpha^2\mathcal{W}_t^0 + \alpha\frac{d}{dt}\mathcal{W}_t^1 + \frac{d^2}{dt^2}\mathcal{W}_t^2.\end{aligned} \tag{7}$$

The details are as follows:

- The deterministic processes $\{\mathcal{W}_t^i : i = 0, 1, 2\}$ have explicit representations, given in (29a)–(29c) as smooth functions of a larger state process.
- The function $\overline{\Upsilon}_t$ may be expressed as a static function of the parameter process

$$\overline{\Upsilon}_t = \overline{\Upsilon}(\Theta_t) \tag{8}$$

where $\overline{\Upsilon} : \Re^d \to \Re^d$ is continuous, which appears only when there is multiplicative noise. It can contribute significantly to the estimation error $\|\Theta_t - \theta^*\|$, resulting in a large bias and variance. Fortunately, it can be eliminated with careful design.

The implications of the p-mean flow representation to algorithm design is a focus of this article. There is one catch: although the representation holds in broad generality, we cannot use it to establish stability (in the sense of ultimate boundedness). Stability can be established through a separate Lyapunov function argument or based on the "ODE@∞" borrowed from the literature on SA. Both approaches are based entirely on consideration of the mean flow (5). Theorem 4 contains full details.

What is ESC?

The answer begins with an explanation of the appearance of $-M\tilde{\nabla}_t\Gamma$ in Figure 1.

A few simplifications will clarify the discussion. Although much of the recent ESC literature concerns tracking the minimizer of a time-varying objective (Γ depends on both the parameter θ and time t), we explain the main ideas in the context of global optimization of the static objective $\Gamma : \Re^d \to \Re$. Second, as will be made clear in "Part 2: ESC," it is often crucial to introduce a time-varying probing gain (the ϵ_t signal shown in Figure 1). Only here is it chosen fixed: $\epsilon_t \equiv \varepsilon$, independent of time.

Low-pass filters: We now explain Figure 1, subject to these simplifications. The low-pass filter with output $\{\Theta_t\}$

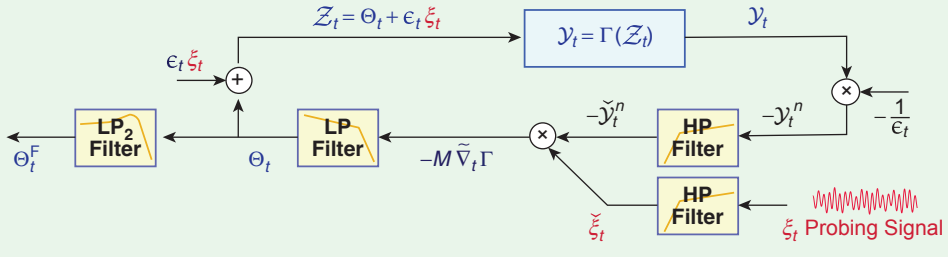


FIGURE 1 A typical architecture for ESC for gradient-free optimization. Observations of the objective Γ are perturbed by sinusoids and used as input to a combination of filters. The output Θ estimates θ^{opt} as time evolves. HP: high pass; LP: low pass.

Probing signals can be designed so that mean-square error bounds are far smaller than $O(1/n)$. Without much effort, we obtain algorithms to achieve a mean-square convergence rate of order approaching $O(1/n^4)$.

is designed so that the derivative of Θ_t is small enough in magnitude to justify a quasi-static analysis. An example is

$$\frac{d}{dt}\Theta_t = -\sigma[\Theta_t - \theta^{\text{ctr}}] + \alpha U_t, \qquad U_t = -M\tilde{\nabla}_t\Gamma \tag{9}$$

with parameters satisfying $0 < \sigma < \alpha$. The vector θ^{ctr} is an a-priori estimate of θ^{opt}.

In tracking applications, we cannot allow α to be too small, which means that the volatility of $\{\Theta_t\}$ will remain high. The second low-pass filter with output Θ_t^{F} is introduced to further reduce volatility. The p-mean flow motivates guidelines for design.

High-pass filters: For the high-pass filter, consider the two special cases:

1) *Pure differentiation*: The figure is interpreted as

$$M\tilde{\nabla}_t\Gamma = \left(\frac{d}{dt}\xi_t\right)\left(\frac{1}{\varepsilon}\frac{d}{dt}\Gamma(\Theta_t + \varepsilon\xi_t)\right). \tag{10}$$

Adopting the notation from the figure, with $\check{\xi}_t$ the derivative of ξ_t, we obtain via the chain rule

$$M\tilde{\nabla}_t\Gamma = \check{\xi}_t\check{\xi}_t^{\mathsf{T}}\nabla\Gamma(\Theta_t + \varepsilon\xi_t) + \mathcal{W}_t \tag{11}$$

where

$$\mathcal{W}_t = \check{\xi}_t\nabla^{\mathsf{T}}\Gamma(\Theta_t + \varepsilon\xi_t)\frac{d}{dt}\Theta_t$$

is small by design of the low-pass filter; consider (9), with $\alpha > 0$ small.

This justifies the diagram, with $M_t = \check{\xi}_t\check{\xi}_t^{\mathsf{T}}$ being time varying. Its time average $\Sigma_{\check{\xi}}$ is required to be full rank.

2) *All pass*: The high-pass filter is removed entirely

$$M\tilde{\nabla}_t\Gamma = \xi_t\frac{1}{\varepsilon}\Gamma(\Theta_t + \varepsilon\xi_t). \tag{12}$$

The analysis begins with an application of the fundamental theorem of calculus to obtain

$$M\tilde{\nabla}_t\Gamma = \frac{1}{\varepsilon}\xi_t\Gamma(\Theta_t) + \int_0^1 \xi_t\xi_t^{\mathsf{T}}\nabla\Gamma(\Theta_t + r\varepsilon\xi_t)\,dr. \tag{13}$$

This is interpreted as a "noisy" observation of $\Sigma_\xi\nabla\Gamma(\Theta_t)$, with Σ_ξ being the mean of $M_t = \xi_t\xi_t^{\mathsf{T}}$. The first term in (13) is small in an average sense, provided the probing signal has zero mean.

Is ESC QSA?

The answer is yes, provided we broaden our definitions as follows:

- For all pass, *yes*: The pair of equations, (9) and (12), is an example of the QSA ODE (3).
- The answer is also *yes* for pure differentiation, but only for purposes of analysis. For sufficiently small $\alpha > 0$, we may express the pair of equations, (9) and (10), as

$$\frac{d}{dt}\Theta = \alpha f(\Theta_t, \xi_t, \check{\xi}_t)$$

  where f inherits the smoothness properties of $\nabla\Gamma$. This is an instance of QSA with a *2d-dimensional* probing signal.
- For a general high-pass filter, the ESC ODE is an example of *two-timescale* QSA, which, in the setting of this article, is equivalently cast within the theory of singular perturbation theory [S21]. This theory justifies an *approximation* by the QSA ODE (3).

Even without approximation, the p-mean flow remains valid and useful for purposes of insight and design. Details are provided in "Part 2: ESC."

A very simple special case will receive special attention:

ESC-0: The QSA ODE (9) and (12) using $\sigma = 0$.

It is the most similar to a standard approach in the stochastic domain, known as *1SPSA* [see (S4a)], and will be a source of examples to illustrate the theory surveyed in "Part 1: QSA."

ESC-0 is an effective approach to gradient-free optimization if the probing signal is chosen with care, along with careful design of the second low-pass filter shown in Figure 1. It is highlighted here only because it is the simplest version available that is potentially successful.

High volatility can be expected when using ESC-0, based on a casual glance at (13): by design, we ensure that $\xi_t\Gamma(\Theta_t)$ is small *on average*, but nevertheless contributes greatly to volatility, especially when $|\Gamma(\theta^*)|$ is large. The remedy is found in the second low-pass filter shown in Figure 1.

A simple linear QSA example is introduced next to illustrate the value of filtering.

The role of filtering in QSA. A pair of scalar examples will serve to illustrate sources of estimation error and how they may be attenuated through a combination of filter design and design of the probing signal. The two QSA ODEs are linear, with multiplicative noise

$$\frac{d}{dt}\Theta_t = \alpha[A_t\Theta_t + \mathcal{U}_t^1], \qquad \mathcal{U}_t^1 = 2\sin(\omega t) + 1 \tag{14a}$$

$$\frac{d}{dt}\Theta_t = \alpha[A_t\Theta_t + \mathcal{U}_t^2], \qquad \mathcal{U}_t^2 = 2\cos(\omega t) + 1 \tag{14b}$$

and $A_t = -(1 + \sin(\omega t))$ with $\omega = 0.1$. They share the common mean vector field $\bar{f}(\theta) = -\theta + 1$, differing only by a phase shift in the input. Theory predicts that the solutions to either QSA ODE will eventually remain within $O(\alpha)$ of $\theta^* = 1$.

Results from the simulations are shown in Figure 2, which illustrate two points:

1) Figure 2(b) shows the sample paths obtained using $\alpha = 0.01$: the bias is 10% for (14b), while the bias observed using (14a) is far smaller [theory predicts it is $O(\alpha^2)$].
2) Figure 2(b) also shows that ESC-0 fails entirely with $\alpha = 0.01$, for either of the two QSA ODEs. It is only after filtering that acceptable results are obtained.

The impact of filtering is more fully illustrated in Figure 2(a), where we see that (14a) is the clear winner: volatility is of order $O(\alpha^2)$ for small α after filtering of parameter estimates.

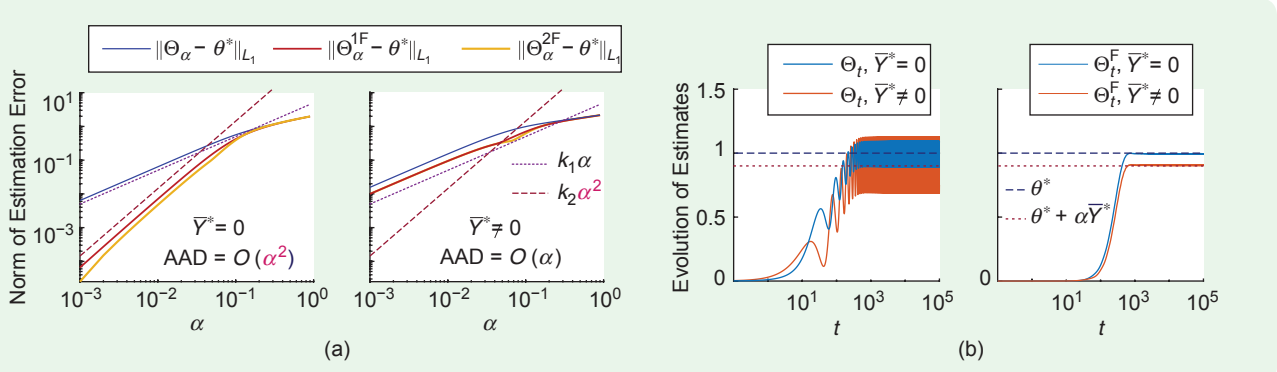


FIGURE 2 The impact of multiplicative noise on average absolute deviation (AAD) (the L_1-norm of estimation error) for QSA. (a) A plot for the L_1 error as a function of the gain α for different filtering techniques and (b) a plot of the impact of $\bar{Y}^* = -\bar{\Upsilon}(\theta^*)$ on estimation error with $\alpha = 0.01$. Filtering can dramatically reduce error when multiplicative noise is absent.

Summary

The goal of this article is twofold: survey the emerging theory of quasi-stochastic approximation (QSA) and its implication to design, and explain the intimate connection between QSA and extremum seeking control (ESC). The contributions go in two directions: ESC algorithm design can benefit by applying concepts from QSA theory, and the broader research community, with interest in gradient-free optimization, can benefit from the control theoretic approach inherent to ESC.

The following are surprising modes of analysis and outcomes:

- *Markovian analysis*: In SA with Markovian noise, the standard approach to variance analysis is to "whiten the noise" through a certain Poisson equation. A similar idea is used when the probing signal is defined as an analytic function of sinusoids. Three applications of this technique are required to obtain (7).
- Once stability has been established, the perturbative mean (p-mean) flow representation for QSA (7) provides insight into dynamic response, based on coupling with the mean flow. This also provides justification for the linearization of the QSA ODE (3)

$$\frac{d}{dt}\Theta_t = \alpha A^*[\Theta_t - \theta^* - \alpha\bar{Y}] + \alpha\mathcal{W}_t + O(\alpha^2) + o(1) \tag{S5}$$

where $A^* = \partial\bar{f}(\theta^*)$, $\bar{Y}^* = [A^*]^{-1}\bar{\Upsilon}^*(\theta^*)$ and $\{\mathcal{W}_t\}$ is a bounded process defined in (7).

- Techniques for establishing ultimate boundedness of the QSA ordinary differential equation (ODE) are obtained by adapting well-worn methods from the SA literature.

The implications to ESC are recent:

- The first are implications of the p-mean flow (see Theorem 9 for a summary).
- QSA stability theory relies strongly on Lipschitz continuity of all vector fields, which is typically *violated* for ESC. A remedy is introduced here for the first time, where a Lipschitz algorithm is obtained through the design of a parameter-dependent probing gain.

Global stability of the algorithm is easily established for the new class of ESC algorithms, under readily verifiable conditions. It is argued that the new design will also result in more efficient exploration.

Filtering cannot attenuate the estimation error for (14b): it remains of order $O(\alpha)$ with or without filtering.

These outcomes can be anticipated from the p-mean flow, along with Theorems 1–3 contained in "Part 1: QSA." This example will be revisited following exposition of QSA theory.

Design for tracking. Once we have confidence in design for the static optimization problem, these algorithms can be tested with an objective function that is time varying.

ESC-0 appears to be a poor choice as $\sigma = 0$ plays the role of a "forgetting factor," which is usually deemed crucial for tracking. There are, however, hidden dynamics that are partially revealed through the p-mean flow, which provide some degree of forgetting. For illustration, consider the problem of tracking a smooth two dimensional signal $\{\theta_t^{\text{opt}}\}$ based on the time-varying objective $\Gamma_t(\theta) = \Gamma(\theta - \theta_t^{\text{opt}})$. This may be posed as a gradient-free optimization problem if $\Gamma : \Re^2 \to \Re$ has global minimizer $\theta^{\text{opt}} = 0$. In this case, the observations driving the ESC-0 ODE are of the form

$$\mathcal{Y}_t^{\text{n}} = \frac{1}{\varepsilon}\Gamma_t(\Theta_t + \varepsilon\xi_t). \tag{15}$$

The plots that follow show results from the ESC-0 ODE in the following special case:

- Γ is the *Three-Hump Camel* [46], a standard benchmark used for testing optimization algorithms.
- The signal θ_t^{opt} is an epitrochoid curve.

A plot of $-\Gamma$ appears in Figure 3(a), showing two local maxima and a single global maxima attained at $\theta^{\text{opt}} = 0$. The signal $\{\theta_t^{\text{opt}}\}$ is indicated with the dashed curve shown in Figure 3.

With initialization at one of the nonoptimal extrema for Γ_0, it is seen in Figure 3 that the estimates $\{\Theta_t\}$ obtained from ESC-0 track a ball around $\{\theta_t^{\text{opt}}\}$ after a transient period, but the evolution is highly volatile. The filtered estimates $\{\Theta_t^{\text{F}}\}$ display much less variability while maintaining good tracking.

In conclusion, the cheapest ESC design works well, subject to constraints on the probing signal and additional filtering. However, it is worth repeating: *we are not advocating that the high-pass filters be abandoned*, and *we do not advocate setting* $\sigma = 0$ *in (9) in application to tracking*. Rather, we adopt the simplest instance of ESC to illustrate the application of general design principles.

The main content of the remainder of this article is divided into two parts, with the first on QSA fundamentals, and the second on implications to ESC. History and resources are included in the "History and Resources" section. All of the theory surveyed in parts 1 and 2 is taken from [15], [23], and [24], following [8] and [S5, Ch. 4].

Acknowledgment to an inspirational scientist. It was a sad day in February 2023 when Boris Polyak was taken from us.

The reader will find references to Prof. Polyak throughout this article. Among his wide-ranging scientific contributions is one that plays center stage in this article: a simple averaging technique to optimize the asymptotic covariance in SA, discovered contemporaneously with David Ruppert. Well before this breakthrough, he introduced in [36] what is now called the *heavy ball algorithm for optimization*. This is just one special case from a menu of acceleration techniques introduced in this article. These ideas led to momentum algorithms for accelerating gradient descent, introduced by his student, Yurii Nesterov. More recently, his survey *Adaptive and Robust Control in the USSR* describes exciting activity that is often missed in the West [S1]. Students of learning are advised to scour Polyak's bibliography to find mathematical gems that are not yet widely known.

PART 1: QSA

A full proof of the p-mean flow representation is provided here, along with its implications to design: Figure 4 provides a hint on the design of low-pass filters.

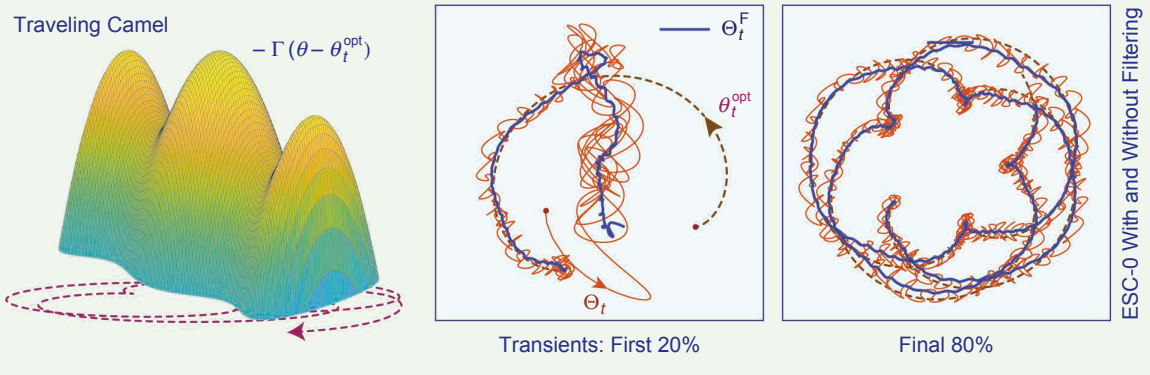


FIGURE 3 Tracking the moving maximizer for the Three-Hump Camel through ESC-0. The process $\{\Theta_t\}$ successfully tracks the moving target $\{\theta_t^{\text{opt}}\}$ after a transient period, but with high volatility. Filtering $\{\Theta_t\}$ to obtain $\{\Theta_t^{\text{F}}\}$ results in much lower volatility for tracking.

"Part 1: QSA" concludes with a brief overview of conclusions for QSA with vanishing gain; this is often the best option in static optimization and machine learning applications.

Markovian Foundations

In the theory of SA, the stochastic "probing" sequence $\{\xi_{n+1}\}$ appearing in (S3) is not always assumed to be i.i.d. Convergence holds under far weaker assumptions. If the probing sequence is a function of a Markov chain, it can be partially "whitened" through the technique of Métivier and Priouret [4, S25]. *It is by extension of this technique to QSA that we arrive at the p-mean flow representation.*

The probing signal is assumed to be a nonlinear function of sinusoids, $\xi_t = G_0(\xi_t^0)$, with

$$\xi_t^0 = [\cos(2\pi[\omega_1 t + \phi_1]), \ldots, \cos(2\pi[\omega_K t + \phi_K])]^\intercal \tag{16}$$

and $G_0 : \Re^K \to \Re^m$ smooth. The motivation for a nonlinearity may be to create rich probing signals from simple ones.

The probing signal $\boldsymbol{\xi}$ is a function of the K-dimensional clock process denoted as $\boldsymbol{\Phi}$, with entries

$$\Phi_t^i = \exp(2\pi j[\omega_i t + \phi_i]), \qquad t \ge 0 \tag{17}$$

which we regard as the underlying Markovian state process.

The notation $G(z) = G_0((z + 1/z)/2)$ is adopted throughout, where $1/z := (1/z_1, \ldots, 1/z_K)$ so that

$$\xi_t = G(\Phi_t). \tag{18}$$

The function G is analytic on $z \in \{C \setminus \{0\}\}^K$, provided G_0 is analytic on C^K. Properties of the clock process are summarized in "Ergodic Theory for the Clock Process." Crucial notation is summarized as follows:

- $\boldsymbol{\Phi}$ evolves on a compact set denoted $\Omega \subset C^K$. It has a unique invariant probability measure denoted as π, which is uniform on Ω.
- Its differential generator is denoted, for smooth h: $\Omega \to C$

$$\mathcal{D}h(z) = \nabla h(z) \cdot Wz, \qquad z \in \Omega \tag{19}$$

  with $W = 2\pi j \operatorname{diag}(\omega_i)$.
- The pair process $\boldsymbol{\Psi} = (\boldsymbol{\Theta}, \boldsymbol{\Phi})$ is also Markovian. For smooth functions h: $\Pi \to C$, its differential generator is

$$\mathcal{D}_{\text{QSA}} h(\theta, z) = \alpha[D^f h](\theta, z) + \partial_z h(\theta, z) \cdot [Wz] \tag{20a}$$

with $[D^f h](\theta, z) = \partial_\theta h(\theta, z) \cdot f(\theta, G(z))$. (20b)

- A crucial takeaway is the representation for the vector field for the mean flow: for $\theta \in \Re^d$

$$\bar{f}(\theta) = \mathsf{E}_\pi[f(\theta, G(\Phi))] := \int f(\theta, G(z))\, \pi(dz). \tag{21}$$

The differential generator for $\boldsymbol{\Phi}$ is used to define Poisson's equation for a Markov process. For functions $g, \hat{g} : \Omega \to \Re$, this is expressed as

$$\hat{g}(\Phi_0) = \hat{g}(\Phi_T) + \int_0^T \tilde{g}(\Phi_t)\, dt, \quad T \ge 0. \tag{22}$$

If a solution exists, then g is called the *forcing function* and $\hat{g}$ the *solution*. If $\hat{g}$ is continuously differentiable, then Poisson's equation is written in its differential form: $\mathcal{D}\hat{g} = -\tilde{g}$.

Three versions will be used in the following, one of which mirrors the use of Poisson's equation for SA with Markovian noise in [4] and [S25].

Functions g on the larger domain $\Pi = \Re^d \times \Omega$ are also considered through a slight abuse of notation: for a function on the joint state space $g : \Pi \to \Re$, for each θ, the function $\hat{g}(\theta, \cdot)$ is the solution to

$$\hat{g}(\theta, \Phi_0) = \hat{g}(\theta, \Phi_T) + \int_0^T \tilde{g}(\theta, \Phi_t)\, dt, \qquad T \ge 0. \tag{23}$$

That is, $\hat{g}(\theta, \cdot)$ solves Poisson's equation for $\boldsymbol{\Phi}$ for each θ, with forcing function $g(\theta, \cdot)$.

Please note:

- The solution $\hat{g}$ to (22) is not unique. It is always normalized so that $\mathsf{E}_\pi[\hat{g}(\Phi)] = 0$. A solution to (23) is assumed normalized so that $\mathsf{E}_\pi[\hat{g}(\theta, \Phi)] = 0$ for each θ.
- In most of the applications considered in this article, the function g depends on Φ_t only through ξ_t, but this is not generally true for $\hat{g}$.
- Finally, on notation: we often write $\hat{g}_t$ instead of $\hat{g}(\Theta_t, \Phi_t)$. When g is vector valued, $\hat{g}$ denotes the vector-valued function whose ith component solves Poisson's equation with forcing function g_i.

Assumptions

Some of the assumptions that follow are essential, and others are imposed only because of limitations in current theory.

The first assumption sets restrictions on frequencies.

(A0a) $\xi_t = G_0(\xi_t^0)$ for all t, with ξ_t^0 defined in (16). The function $G_0 : \Re^K \to \Re^m$ is assumed to be analytic, with the coefficients in the Taylor series expansion for $G_0(\xi_t^0)$ absolutely summable.

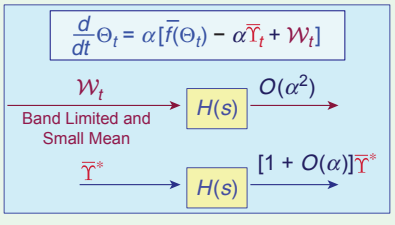


FIGURE 4 P-mean flow and its implications for design. Filtering attenuates the signal $\{\mathcal{W}_t\}$ to be of order $O(\alpha^2)$. However, if $\bar{\Upsilon}^* \neq 0$, a term of order $O(\alpha)$ remains, hindering estimation accuracy for QSA, even after filtering.

(Aob) The frequencies $\{\omega_1,\dots,\omega_K\}$ are chosen of the form

$$\omega_i = \log(a_i/b_i) > 0,\ 1 \le i \le K$$
$$\{\omega_i\}, \text{linearly independent over the rationals} \tag{24a}$$

and with $\{a_i, b_i\}$ positive integers.

(A1) The functions $\bar{f}$ and f are Lipschitz continuous: for a constant $\hat{L}_f < \infty$

$$\begin{aligned}\|\bar{f}(\theta') - \bar{f}(\theta)\| &\le \hat{L}_f \|\theta' - \theta\| \\ \|f(\theta', \xi) - f(\theta, \xi)\| &\le \hat{L}_f [\|\theta' - \theta\| + \|\xi' - \xi\|] \\ \|f(\theta, \xi') - f(\theta, \xi)\| &\le \hat{L}_f [\|\theta' - \theta\| + \|\xi' - \xi\|]\end{aligned} \tag{24b}$$

for all $\theta', \theta \in \Re^d$, $\xi, \xi' \in \Re^m$.

(A2) The vector fields f and $\bar{f}$ are each twice continuously differentiable, with derivatives denoted as

$$A(\theta, z) = \partial_\theta f(\theta, z), \qquad \bar{A}(\theta) = \partial_\theta \bar{f}(\theta). \tag{24c}$$

(A3) Solutions to Poisson's equation exist in the form (23) for the following three choices of $g : \Pi \to \Re^d$. In each case, the solution $\hat{g}$ is assumed normalized with $\mathsf{E}_\pi[\hat{g}(\theta, \Phi)] = 0$ for each θ, and $\hat{g} : \Pi \to \Re^d$ is assumed continuously differentiable.

Ergodic Theory for the Clock Process

Ergodicity of the clock process is well known to researchers in both dynamical systems and stochastic processes. This summary reviews notation and essential properties.

SUMMARY OF MARKOV TERMINOLOGY

The clock process evolves on a compact set, denoted as $\Omega \subset C^K$, and may be represented as the state process for a linear system:

$$\frac{d}{dt}\Phi_t = W\Phi_t, \quad \text{with } W = 2\pi j \operatorname{diag}(\omega_i), \quad \Phi_0 \in \Omega. \tag{S6}$$

It is a stationary Markov process when Φ_0 is chosen randomly, with $\Phi_0 \sim \pi$ (the uniform distribution on Ω).

The mean $\bar{g} := \int g(z)\pi(dz)$ is always finite when $g : \Omega \to \Re$ is continuous. The centered function is denoted as $\tilde{g}(z) = g(z) - \bar{g}$ for $z \in \Omega$.

THE LAW OF LARGE NUMBERS

The law of large numbers (LLN) tells us that for each initial condition Φ_0

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \tilde{g}(\Phi_t)\,dt = 0.$$

This is commonly used with $g(z) = h(G(z))$ so that $g(\Phi_t) = h(\xi_t)$ [recall (18)]. The probing signal $\boldsymbol{\xi}$ falls in the broader class of almost-periodic functions [S26], [S27].

If there is a continuous function $\hat{g} : \Omega \to \Re$ solving (22), then we have

$$\left|\frac{1}{T}\int_0^T \tilde{g}(\Phi_t)\,dt\right| \le 2\|\hat{g}\|_\infty \frac{1}{T}, \quad T > 0.$$

In the terminology of [S21, Ch. 8], we say that the LLN holds with *convergence function* $\kappa(T) = 1/T$.

The assumptions on G_0 in assumption (A0a) are imposed to ensure consistency of the two definitions of the mean flow vector field $\bar{f}$ in (4) and (21). The Lipschitz conditions in (A1) imply that convergence in the LLN is uniform in both time and parameter: for a constant $b_f > 0$,

$$\sup_{\theta,z} \frac{1}{1+\|\theta\|}\left|\frac{1}{T}\int_0^T [f(\theta, \xi_t) - \bar{f}(\theta)]\,dt\right| \le b_f \frac{1}{T} \tag{S7}$$

where the supremum is over $\theta \in \Re^d$ and $\Phi_0 = z \in \Omega$.

DIFFERENTIAL GENERATOR

The following two forms are required in analysis:

1) The differential generator for the clock process is defined in (19). If $h : C^k \to C$ is C^1 in a neighborhood of Ω, then the continuous function $g = \mathcal{D}h$ may be represented as

$$g(\Phi_t) = \frac{d}{dt}h(\Phi_t).$$

It follows that $h = \hat{g}$ is a solution to Poisson's equation, $\bar{g} = 0$, and the LLN holds for $\{g(\Phi_t) : t \ge 0\}$, with convergence function $\kappa(T) = 1/T$.

2) The pair process $\boldsymbol{\Psi} = (\boldsymbol{\Theta}, \boldsymbol{\Phi})$ is itself the state process for a time-homogeneous dynamical system on $\Pi = \Re^d \times \Omega$. It is also Markovian, with the differential generator defined in (20a), and the function $g = \mathcal{D}_{\text{QSA}}h$ may be represented as

$$g(\Psi_t) = \frac{d}{dt}h(\Psi_t)$$

for any function h that is continuously differentiable.

Suppose that the pair process $\{\Psi_t\}$ is a bounded function of time from some initial condition $\Psi_0 = (\theta, z)$. It follows from [32, Th. 12.1.2] that there exists an invariant probability measure for the joint process (the generalization to continuous time is only a change in notation). The LLN may not hold from each initial condition, however, for the function $g = \mathcal{D}_{\text{QSA}}h$, we have the familiar bound

$$\left|\frac{1}{T}\int_0^T g(\Psi_t)\,dt\right| \le b_h \frac{1}{T}, \quad T > 0 \tag{S8}$$

with $b_h = \sup_t |h(\Psi_t) - h(\Psi_0)|$.

REFERENCES

[S26] L. Amerio and G. Prouse, *Almost-Periodic Functions and Functional Equations*. New York, NY, USA: Springer Science & Business Media, 2013.

[S27] H. Bohr, *Almost Periodic Functions* (Reprint of the 1947 Edition). New York, NY, USA: Chelsea, 2018.

1) The solution $\hat{f}$ with forcing function f equal to the QSA vector field. Its Jacobian with respect to θ is denoted

$$\hat{A}(\theta, z) \ := \partial_\theta \hat{f}(\theta, z). \tag{24d}$$

2) The solution $\hat{\hat{f}}$ with forcing function $\hat{f}$.
3) The solution $\hat{\Upsilon}$ with forcing function Υ, where

$$\Upsilon(\theta, z) = -[D^f \hat{f}](\theta, z) = -\hat{A}(\theta, z) f(\theta, G(z)) \tag{24e}$$

with D^f defined in (20b). That is, for all $0 \le t_0 \le t_1$,

$$\begin{aligned}
\hat{f}(\theta, \Phi_{t_0}) &= \int_{t_0}^{t_1} [f(\theta, \Phi_t) - \bar{f}(\theta)]\, dt + \hat{f}(\theta, \Phi_{t_1}) \\
\hat{\hat{f}}(\theta, \Phi_{t_0}) &= \int_{t_0}^{t_1} \hat{f}(\theta, \Phi_t)\, dt + \hat{\hat{f}}(\theta, \Phi_{t_1}) \\
\hat{\Upsilon}(\theta, \Phi_{t_0}) &= \int_{t_0}^{t_1} [\Upsilon(\theta, \Phi_t) - \overline{\Upsilon}(\theta)]\, dt + \hat{\Upsilon}(\theta, \Phi_{t_1})
\end{aligned}$$

with $\bar{f}$ defined in (21), and

$$\overline{\Upsilon}(\theta) = \mathsf{E}[\Upsilon(\theta, \Phi)] = -\int_\Omega \hat{A}(\theta, z) f(\theta, G(z))\, \pi(dz). \tag{24f}$$

Rationale. Assumption (A0) is imposed for two reasons. Subject to the assumption that $f(\theta, G(z))$ is an analytic function of (θ, z) on an appropriate domain, assumption (A0) has two important consequences:

1) Assumption (A3) holds.
2) $\overline{\Upsilon}(\theta) \equiv 0$.

Assumptions (A1)–(A3) and further assumptions are required to bound bias and variance, and Lipschitz continuity is also crucial in establishing criteria for ultimate boundedness of $\boldsymbol{\Psi}$.

Three Steps to the P-Mean Flow

The three steps in the derivation of (7) are based on the three solutions to Poisson's equation in (A3).

The differential generator $\mathcal{D}_{\text{QSA}}$ defined in (20a) plays a role, even though we never consider Poisson's equation for the full generator. Rather, suppose that $g\colon \Pi \to \Re$ is a smooth function on Π, and there exists a smooth function $\hat{g}$ solving (23) for each θ and Φ_0. The following identity then follows from the chain rule, using the notation (20b):

$$\frac{d}{dt}\hat{g}_t = \mathcal{D}_{\text{QSA}}\hat{g}(\Theta_t, \Phi_t) = \alpha[D^f \hat{h}](\Theta_t, \Phi_t) - [g_t - \bar{g}_t] \tag{25}$$

where $\hat{g}_t \equiv \hat{g}(\Theta_t, \Phi_t)$, and a similar compact notation is used for the remaining terms on the right-hand side.

We now proceed through the three steps, starting with representation (6). Understanding (7) is equivalent to determining the functions $\{\mathcal{W}^i\}$ in the representation

$$\tilde{\Xi}_t = -\alpha\overline{\Upsilon}_t + \alpha^2 \mathcal{W}_t^0 + \alpha\frac{d}{dt}\mathcal{W}_t^1 + \frac{d^2}{dt^2}\mathcal{W}_t^2. \tag{26}$$

Step 1: Apply (25) with $h = \hat{f}$

$$\frac{d}{dt}\hat{f}(\Theta_t, \Phi_t) = \partial_\theta \hat{f}(\Theta_t, \Phi_t)\frac{d}{dt}\Theta_t - [f(\Theta_t, \xi_t) - \bar{f}(\Theta_t)].$$

This gives the first transformation of the apparent noise

$$\tilde{\Xi}_t = \underbrace{-\frac{d}{dt}\hat{f}(\Theta_t, \Phi_t)}_{\text{High pass}} + \underbrace{\alpha\partial_\theta \hat{f}(\Theta_t, \Phi_t) f(\Theta_t, \xi_t)}_{\text{Attenuation}}.$$

Recalling (24e) gives, in shorthand notation,

$$\tilde{\Xi}_t = -\frac{d}{dt}\hat{f}_t - \alpha\Upsilon_t. \tag{27}$$

Step 2: The arguments in step 1 are repeated, using $\hat{f}$, to get

$$\frac{d}{dt}\hat{f}_t = \alpha\frac{d}{dt}[D^f\hat{f}](\Theta_t, \Phi_t) - \frac{d^2}{dt^2}\hat{\hat{f}}_t.$$

Step 3: Repeat with Υ, to achieve

$$\Upsilon_t = \overline{\Upsilon}_t + \alpha[D^f\hat{\Upsilon}](\Theta_t, \Phi_t) - \frac{d}{dt}\hat{\Upsilon}_t.$$

Steps 2 and 3, combined with (27), lead to the p-mean flow representation.

Theorem 1 (P-Mean Flow)
Subject to (A3),
1) the pre-p-mean flow representation holds

$$\begin{aligned}
\frac{d}{dt}Y_t &= \alpha[\bar{f}(Y_t) - \alpha(B_t\hat{f}_t + \Upsilon_t)] \\
\Theta_t &= Y_t - \alpha\hat{f}_t, \qquad Y_0 = \Theta_0 + \alpha\hat{f}_0
\end{aligned} \tag{28}$$

with $B_t = \int_0^1 \bar{A}(Y_t - r\alpha\hat{f}_t)\, dr$.
2) The p-mean flow representation (7) holds with

$$\mathcal{W}_t^0 = \mathcal{W}^0(\Theta_t, \Phi_t) := -[D^f\hat{\Upsilon}](\Theta_t, \Phi_t) \tag{29a}$$

$$\mathcal{W}_t^1 = \mathcal{W}^1(\Theta_t, \Phi_t) \ := -[D^f\hat{\hat{f}}](\Theta_t, \Phi_t) + \hat{\Upsilon}(\Theta_t, \Phi_t) \tag{29b}$$

$$\mathcal{W}_t^2 = \mathcal{W}^2(\Theta_t, \Phi_t) := \hat{\hat{f}}(\Theta_t, \Phi_t). \tag{29c}$$

In the remainder of this part of the article, it is assumed that $\bar{f}$ has a unique root $\theta^* \in \Re^d$. The goal is to explain how the p-mean flow can provide insight into how to design QSA ODEs that provide good estimates of θ^* after a short transient.

"Measuring Algorithmic Performance" summarizes metrics for assessing performance of an algorithm. Three receive focus in this article: bias, variance, and average absolute deviation (AAD) (the L_1-norm of estimation error). Bounds on these quantities will follow from absolute bounds on the estimation error $\|\Theta_t - \theta^*\|$, which in part follow from bounds on the *target bias* (S12).

Bias, variance, and target bias may be related through the following simple approximation. Recall the

definition $A^* = \bar{A}(\theta^*)$ following (S1). The target bias $\beta_{\bar{f}}$ is defined below.

Lemma 1 (Bias and Variance)

Suppose that $\theta^* \in \Re^d$ is the unique solution to $\bar{f}(\theta) = 0$. Suppose moreover that assumptions (A1) and (A2) hold, and denote $\tilde{\theta} = \theta - \theta^*$. Then,

1) There is a function $\mathcal{E}_A : \Re^d \to \Re^d$ satisfying

$$\bar{f}(\theta) = A^*\tilde{\theta} + \mathcal{E}_A(\theta), \qquad \theta \in \Re^d. \tag{30a}$$

The error term is Lipschitz continuous and admits the quadratic bound $\mathcal{E}_A(\theta) \le L_A \| \tilde{\theta} \|^2$.

2) If A^* is invertible

$$\Theta_t - \theta^* = [A^*]^{-1}[\bar{f}(\Theta_t) - \mathcal{E}_A(\Theta_t)]. \tag{30b}$$

And provided the target bias and variance are finite

$$\beta_\Theta \le \| [A^*]^{-1} \|_F [\beta_{\bar{f}} + L_A \sigma_\Theta^2] \tag{30c}$$

where the subscript F indicates the Frobenius norm.

Measuring Algorithmic Performance

How can we assess algorithmic performance? Standard performance metrics from statistics are adopted here, along with a nonstandard statistic, known as *target bias*.

BIAS AND VARIANCE

The usual statistical definitions of bias and covariance are

$$b_\theta = \bar{\theta} - \theta^*, \qquad \Sigma_\Theta = \mathsf{E}_\varpi[\Theta\Theta^\intercal] - \bar{\theta}\bar{\theta}^\intercal$$

with $\bar{\theta} = \mathsf{E}_\varpi[\Theta]$ and $\varpi \sim (\Theta, \Phi)$ where ϖ is a unique invariant measure. On denoting $\sigma_\Theta^2 = \text{trace}(\Sigma_\Theta)$ and $\beta_\Theta = |b_\theta|$

$$\| \Theta - \theta^* \|_{L_2}^2 := E_\varpi[\| \Theta - \theta^* \|^2] = \sigma_\theta^2 + \beta_\Theta^2.$$

The existence of an invariant measure is guaranteed for quasi-stochastic approximation (QSA) whenever the sample path $\boldsymbol{\Theta}$ is bounded from at least one initial condition. This follows from the fact that $\boldsymbol{\Psi} = (\boldsymbol{\Theta}, \boldsymbol{\Phi})$ is a Feller–Markov process. We do not know whether ϖ is unique, so expectations are replaced with sample-path averages

$$\begin{aligned} \beta_\Theta &= \limsup_{T\to\infty} \left\| \frac{1}{T}\int_0^T [\Theta_t - \theta^*] dt \right\| \\ \sigma_\Theta^2 &= \left(\limsup_{T\to\infty} \frac{1}{T}\int_0^T \| \Theta_t - \theta^* \|^2 dt \right) - \| \beta_\Theta \|^2. \end{aligned} \tag{S9}$$

L_P ERROR AND AVERAGE ABSOLUTE DEVIATION

The standard L_p-norms will also be considered in their sample-path forms:

$$\begin{aligned} \| \Theta - \theta^* \|_{L_1} &= \limsup_{T\to\infty} \frac{1}{T}\int_0^T \| \Theta_t - \theta^* \| dt \\ \| \Theta - \theta^* \|_{L_2} &= \sqrt{\limsup_{T\to\infty} \frac{1}{T}\int_0^T \| \Theta_t - \theta^* \|^2 dt}\,. \end{aligned} \tag{S10}$$

The L_1-norm is also referred to as the *average absolute deviation* (*AAD*). These quantities are related via

$$\| \Theta - \theta^* \|_{L_1} \le \| \Theta - \theta^* \|_{L_2} = \sqrt{\sigma_\theta^2 + \beta_\Theta^2}\,. \tag{S11}$$

TARGET BIAS

The goal of SA is to estimate θ^* such that $\bar{f}(\theta^*) = 0$, so we regard $0 \in \Re^d$ as the *target*. The *target bias* is defined as another sample-path average

$$b_{\bar{f}} := \lim_{T\to\infty} \frac{1}{T}\int_0^T \bar{f}(\Theta_t) dt \tag{S12}$$

provided the limit exists, and $\beta_{\bar{f}} := \| b_{\bar{f}} \|$.

ESTIMATING STATISTICS

Two approaches are adopted for estimating bias and other quantities. Given data up to time T, estimates of bias, variance, and AAD are denoted as $\hat{b}_T, \hat{\sigma}_T^2$, and $\widehat{\text{AAD}}_T$, respectively.

- *Single-path estimates*: Based on observations of $\{\Theta_t : 0 \le t \le T\}$ from a single initial condition, the estimates are determined by

$$\begin{aligned} \hat{b}_T &= \bar{\Theta}_T - \theta^*, \quad \bar{\Theta}_T = \frac{1}{T - T_0}\int_{T_0}^T \Theta_\tau d\tau \\ \hat{\sigma}_T^2 &= \frac{1}{T - T_0}\int_{T_0}^T \| \Theta_\tau - \theta^* \|^2 d\tau - \| \bar{\Theta}_T \|^2 \\ \widehat{\text{AAD}}_T &= \frac{1}{T - T_0}\int_{T_0}^T \| \Theta_\tau - \theta^* \| d\tau \end{aligned} \tag{S13}$$

where $T_0 \in [0, T)$ is introduced to reduce the impact of transients.

- *Batch mean methods*: The potential problem with estimates from a single sample path is that the sample path may be special, yielding misleading results. Consider application of a gradient-free optimization algorithm to an objective with multiple local extrema; how would you know whether or not your estimates are reaching the global minimum?

  The batch means method involves computation of M solutions to the QSA ODE, distinguished by distinct initial conditions $\Theta_0^m, 1 \le m \le M$. These should be spaced widely apart to ensure that the impact of each initial condition is not ignored entirely; this may also be interpreted as a form of exploration. Based on these data, only a single time point T is used to estimate bias, variance, and AAD as follows:

$$\begin{aligned} \hat{b}_T &= \bar{\Theta}_T - \theta^*, \qquad \bar{\Theta}_T = \frac{1}{M}\sum_{i=1}^M \Theta_T^i \\ \hat{\sigma}_T^2 &= \frac{1}{M}\sum_{i=1}^M \| \Theta_T^i \|^2 - \| \bar{\Theta}_T \|^2 \\ \widehat{\text{AAD}}_T &= \sum_i \| \Theta_T^i - \theta^* \|. \end{aligned} \tag{S14}$$

The representation (30a) is an instance of the mean value theorem, and the remaining conclusions are immediate from the definitions.

The value of (30c) comes from the fact that bounds on target bias are easily obtained through the p-mean flow representation. The proof of the following is obtained by combining part 2 of Theorem 1 with (S8):

Lemma 2 (Target Bias Representation)

Suppose that the limit (S12) exists for a given initial condition. Then, for the same initial condition

$$b_{\bar{f}} = \lim_{T\to\infty} \frac{1}{T}\int_0^T [\alpha \overline{\Upsilon}_t - \alpha^2 \mathcal{W}_t^0]\, dt.$$

These two lemmas suggest that bias bounds of the form $\beta_\Theta = O(\alpha)$ can be expected. Much better bounds on bias and variance are obtained through 1) additional filtering, and 2) elimination of the function $\overline{\Upsilon}_t$ appearing in the p-mean flow representation. The signal $\overline{\Upsilon}_t$ is addressed next.

Source of Poor Performance and Its Elimination

The two terms, $\{\mathcal{W}_t^i : i = 1,2\}$, are easily attenuated via filtering, and the first term, $\mathcal{W}_t^0$, is scaled by α^2 in (7), so it does not contribute significantly to bias or variance. The problem is $\overline{\Upsilon}_t$, which may contain significant dc content, and hence cannot be filtered away. Rather, this signal will be eliminated through design of the probing signal.

This is possible through the geometry illustrated in Figure 5. The green region indicates all functions g: $C^K \to C$ that are analytic in a neighborhood of Ω. The set S denotes analytic functions of the form $g(z) = h(G(z))$, where G appears in (18); that is, $g(\Phi_t) = h(\xi_t)$ for each t. The second function class $\hat{S}$ denotes all functions $\hat{g}$ that solve Poisson's equation for some $g \in S$.

Theorem 2 (Bounds on Target Bias)

Suppose (A0a), (A1), and (A3) hold for the QSA ODE, but with arbitrary choices of frequencies $\{\omega_i\}$. Then,

1) The target bias admits the bound

$$\beta_{\bar{f}} := \| b_{\bar{f}} \| = O(\alpha). \tag{31a}$$

2) If, in addition, (A0) holds, then $\overline{\Upsilon}(\theta) = 0$ for each $\theta \in \Re^d$, and the p-mean flow representation (7) reduces to

$$\frac{d}{dt}\Theta_t = \alpha\left[\bar{f}(\Theta_t) + \alpha^2 \mathcal{W}_t^0 + \alpha \frac{d}{dt}\mathcal{W}_t^1 + \frac{d^2}{dt^2}\mathcal{W}_t^2\right].$$

In this case, the bias bound is improved:

$$\beta_{\bar{f}} = O(\alpha^2). \tag{31b}$$

Proof Overview

The proof of 1) follows from (27). For 2), the function classes S and $\hat{S}$ are *orthogonal*: for $g = h \circ G \in S$ and $\ell \in \hat{S}$, we must have

$$\int h(G(z))\, \ell(z)\, \pi(dz) = 0. \tag{32}$$

In view of (24e), the ith entry of $\Upsilon(\theta)$ may be expressed as

$$\Upsilon_i = -\sum_{j=1}^d \hat{A}_{i,j} g_j$$

with $g_j(\theta, z) = f_j(\theta, G(z))$ so that $g_j \in S$. For each $\theta \in \Re^d$, we have $\hat{A}_{i,j}(\theta, \cdot) \in \hat{S}$, so the result follows from (32).

The conclusion that the target bias is of order $O(\alpha^2)$ follows from Lemma 2. Theorem 1 combined with Lemma 1 imply similar bounds for the parameter estimation bias β_Θ defined in (30c). Theorem 3 contains a much stronger conclusion in terms of bounds on both bias and AAD.

Filtering and Acceleration

With $\overline{\Upsilon}$ eliminated, it is time to attenuate $\{\mathcal{W}_t^i : i = 1, 2\}$ using a low-pass filter.

The first requirement of a filter is that it has unity dc gain. To reduce AAD to $O(\alpha^2)$ then requires consideration of (7): the bound on the input

$$\alpha \frac{d}{dt}\mathcal{W}_t^1$$

can be reduced to $O(\alpha^2)$ using a first-order low-pass filter with bandwidth $O(\alpha)$. The second derivative term has no scaling, so a first-order filter will not do, but a second-order filter will suffice. The filter to be considered is expressed as a second-order transfer function with relative degree two, or the equivalent time domain representation

$$\frac{d^2}{dt^2}\Theta_t^{\mathsf{F}} + 2\gamma\zeta \frac{d}{dt}\Theta_t^{\mathsf{F}} + \gamma^2 \Theta_t^{\mathsf{F}} = \gamma^2 \Theta_t. \tag{33}$$

This is subject to the constraint $\gamma = O(\alpha)$ as the natural frequency γ determines the bandwidth of the filter.

Uniform stability and uniform bounds on performance. To make precise statements regarding bias and variance as functions of α requires consideration of a family of

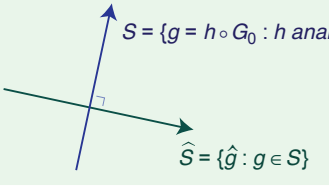


FIGURE 5 Orthogonality of functions of the probing signal and corresponding solutions to Poisson's equation. If the frequencies of the probing signal respect (A0), the two function classes are orthogonal. Orthogonality leads to the conclusion that $\bar{\Upsilon} = 0$.

QSA ODEs over a range of α, along with a uniform notion of stability.

For a given $\alpha^0 > 0$, the family of QSA ODEs (3) is called α^0-*ultimately bounded* if there is a fixed constant B such that for each $\alpha \in (0, \alpha^0]$ and initial condition $(\Theta_0, \Phi_0) = (\theta, z)$, there is a finite time $t_0 = t_0(\theta, z, \alpha)$ such that the solution to (3) satisfies

$$\|\Theta_t\| \le B, \qquad t \ge t_0 \tag{34}$$

with t_0 continuous on its domain.

Criteria for α^0-ultimate boundedness are discussed in the next section. The strong conclusions regarding bias and variance require this assumption and something more.

(A4) The family of QSA ODE models is α^0-ultimately bounded, and the mean flow satisfies the two conditions:

1) The ODE

$$\frac{d}{dt}\vartheta_t = \bar{f}(\vartheta_t)$$

is globally asymptotically stable with unique equilibrium θ^*.

2) The matrix $A^* = \bar{A}(\theta^*)$ is Hurwitz.

The filter must be designed based on the gain α. Specifications are provided in Theorem 3.

Theorem 3 (Error Attenuation)

Suppose (A1)–(A4) hold, and the second-order high-pass filter is chosen subject to the following constraints: the damping ratio $\zeta \in (0, 1)$ is independent of α, and a constant $\eta > 0$ is also fixed to define the natural frequency, $\gamma = \eta\alpha$ for each α.

Then, for $0 < \alpha \le \alpha_0$ and large t, the estimates admit the following approximations:

$$\Theta_t = \theta^* + O(\alpha) + o(1) \tag{35a}$$

$$\Theta_t^{\mathsf{F}} = \theta^* + \alpha\overline{Y}^* + O(\alpha^2) + o(1) \tag{35b}$$

where $o(1) \to 0$ as $t \to \infty$, $\overline{\Upsilon}^* = \overline{\Upsilon}(\theta^*)$ and $\overline{Y}^* = [A^*]^{-1}\overline{\Upsilon}^*$.

The approximations (35) imply bounds on the absolute deviation of parameter estimates, and hence the AAD. Bounds on bias and variance also follow as corollaries to Theorem 3.

Corollary 1 (Bias and Variance)

Under the assumptions of Theorem 3,

1) The asymptotic bias and variance (S9) admit the bounds

$$\beta_\Theta = O(\alpha), \quad \sigma^2_{\tilde{\Theta}} = O(\alpha^2) \tag{36a}$$

$$\beta_{\Theta^{\mathsf{F}}} = O(\alpha), \quad \sigma^2_{\tilde{\Theta}^{\mathsf{F}}} = O(\alpha^2). \tag{36b}$$

2) If, in addition, (A0) holds, then

$$\beta_\Theta = O(\alpha^2), \quad \sigma^2_{\tilde{\Theta}} = O(\alpha^2) \tag{36c}$$

$$\beta_{\Theta^{\mathsf{F}}} = O(\alpha^2), \quad \sigma^2_{\tilde{\Theta}^{\mathsf{F}}} = O(\alpha^4). \tag{36d}$$

Assumption (A0) has the largest impact on bias and variance. Equation (36c) tells us that the variance is of order $O(\alpha^2)$, subject to this restriction on frequencies, which is remarkable when compared with standard results from SA theory [see (63) and the discussion that follows]. Filtering brings the variance down to $O(\alpha^4)$: a restatement of the second bound in (36d).

Proof Overview of Theorem 3

The main ideas are surveyed here only to illustrate application of the p-mean flow representation.

It is assumed that the initial condition is selected so that $\Theta_t \in \mathcal{R}$ for all $t \ge 0$, with $\mathcal{R} = \{\theta : \|\theta\| \le B\}$. This is without loss of generality as every solution eventually remains within this region under the assumptions of the theorem.

The mean flow is locally exponentially asymptotically stable under the given assumptions, with a region of exponential asymptotic stability, including the region $\mathcal{R}$. It follows that there is a function $V : \Re^d \to \Re_+$, with the Lipschitz gradient satisfying for some $\delta_V > 0$

$$\delta_V \|x - \theta^*\|^2 \le V(x) \le \delta_V^{-1}\|x - \theta^*\|^2, \qquad x \in \mathcal{R}. \tag{37}$$

This Lyapunov function is then applied to the representation (28). This, combined with (37), implies (35a).

The proof of (35b) begins with an application of (35a) to justify a linearization of the p-mean flow (7) around θ^* so that bounds are obtained based on the linear approximation (S1). The approximation (35b) also follows from (S1). See "Frequency Domain Design for Quasi-Stochastic Approximation" for further details.

QSA Theory and Practice

The examples that follow illustrate application of the theory presented thus far.

Revisiting the Linear Example

The results obtained for the two linear models (14) are no surprise when viewed through the lens of the p-mean flow, along with the details provided in Theorems 1–3.

The QSA ODE (14a) respects the constraints on frequencies imposed in (A0). Theorem 2 implies that $\overline{\Upsilon}(\theta) = 0$, independent of θ.

Theorem 2 cannot be applied in analysis of (14b) because assumption (A0) is *violated*. An appeal to Theorem 1 leads to a calculation of the major contribution to bias: the definition (24e) along with elementary calculations gives $\overline{\Upsilon}(\theta) = 1/\omega$, independent of θ. The p-mean flow tells us

that the $O(\alpha)$ contribution to both bias and AAD is precisely $\alpha\overline{Y}^* := \alpha[A^*]^{-1}\overline{\Upsilon}(\theta^*) = -\alpha/\omega$.

These results illustrate the importance of maintaining distinct frequencies: the inclusion of a phase shift in the input might appear harmless. In fact, this small change results in significant estimation bias: 10% in this example when $\alpha = 0.01$ and $\omega = 0.1$.

Filtering was performed following the assumptions of Theorem 3 to obtain a second-order filter, and a first-order filter was also constructed

$$\Theta^{1F}(s) = H_1(s)\Theta(s), \qquad \Theta^{2F}(s) = H_2(s)\Theta(s)$$
$$\text{where } H_1(s) = \frac{\gamma}{s+\gamma}, \qquad H_2(s) = \frac{\gamma^2}{s^2 + 2\zeta\gamma s + \gamma^2} \tag{38}$$

with $\zeta = 0.8$ and $\gamma = \eta\alpha$ using $\eta = 1$.

As suggested by (S13), an approximation of AAD is obtained using a sample-path average over the final 20% of the run, denoted as

$$\|\tilde{\Theta}_\alpha\|_{L_1} := \frac{1}{T - T_0}\int_{T_0}^{T} |\Theta_\tau - \theta^*|\, d\tau, \quad T_0 = 0.8T \tag{39}$$

where $\tilde{\Theta}_\alpha = \Theta_\alpha - \theta^*$.

This is repeated to obtain $\|\tilde{\Theta}_\alpha^{1F}\|_{L_1}$ and $\|\tilde{\Theta}_\alpha^{2F}\|_{L_1}$.

Figure 2(a) shows plots of the approximate AAD as a function of α, along with polynomials $r_1(\alpha) = k_1\alpha$, $r_2(\alpha) = k_2\alpha^2$; the constants k_1, k_2 were chosen to ease comparison. Figure 2 shows what is expected: $\alpha\overline{Y}^*$ dominates AAD when $\overline{Y}^* \neq 0$. In this case, filtering has no improvement on reducing AAD below $O(\alpha)$.

Figure 2 shows that both filtering choices reduce AAD to $O(\alpha^2)$ when $\overline{Y}^* = 0$, and $\alpha < 0.1$. The reason for the success of a first-order filter is explained in "Frequency Domain Design for Quasi-Stochastic Approximation."

Control of Volatility in Tracking

The filter used to obtain the smooth tracking in Figure 3 was chosen based on the criterion of Theorem 3, using $\gamma = \eta\alpha$ with $\eta = 5$. The larger bandwidth was needed to avoid excessive lag. This value of η was found to be useful through trial and error: the best value depends of course on properties of the target signal $\{\theta_t^{\text{opt}}\}$.

Consider a signal defined over a time horizon $[0, T]$, continuous on $[0, T/2]$ with components equal to triangle waves, and with components equal to square waves on the following subinterval $[T/2, T]$. ESC-0 works well for both first- and second-order filters of the form (38) for a range of η, but the best filter on the first subinterval will be very different from the best choice for the second.

This is illustrated in Figure 6, showing the evolution of $\{\Gamma(\Theta_t), \Gamma(\Theta_t^{1F}), \Gamma(\Theta_t^{2F})\}$ as functions of time using filter H_i to obtain Θ_t^{iF}. The first row shows results obtained using $\eta = 5$, and the second using $\eta = 15$. A first-order filter outperforms a second-order filter on the subinterval $[0, T/2]$, for which the target is consistently varying. In fact, in this case, the cost as a function of time without filtering appears to be the most successful. The second-order filter results in significant improvement in performance on the second subinterval (ignoring brief transients following each discontinuity of the target). As the theory anticipates,

Frequency Domain Design for Quasi-Stochastic Approximation

The linearization of the p-mean flow (S1) invites the application of Laplace transform techniques for design and analysis.

Consider the linear system approximating the quasi-stochastic approximation ODE, motivated by the representation (S1)

$$\frac{d}{dt}x_t = \alpha A^* x_t + \alpha\mathcal{W}_t.$$

The definition of $\mathcal{W}_t$ remains the same: a function of (Θ_t, Φ_t). Once we establish that $\|\Theta_t - x_t\| = O(\alpha^2)$ for common initial conditions $\Theta_0 = x_0$ within a bounded region $\mathcal{R}$, justification of (35b) can be conducted entirely in the frequency domain. This viewpoint leads to refinements of the second-order filter proposed in Theorem 9 and much greater insight.

Let $X(s)$, $W(s)$ denote the respective Laplace transforms of the state and input for this linear system, and $W^i(s)$ the transforms of the components of $\mathcal{W}_t$ shown in Theorem 1. Taking Laplace transforms of each side gives

$$X(s) = \alpha[Is - \alpha A^*]^{-1}W(s)$$
$$= \alpha[Is - \alpha A^*]^{-1}[\alpha^2 W^0(s) + \alpha s W^1(s) + s^2 W^2(s)].$$

Also, using a superscript "F" for the filtered signals

$$X^F(s) = \alpha[Is - \alpha A^*]^{-1}W^F(s)$$
$$= \alpha[Is - \alpha A^*]^{-1}[\alpha^2 W^{0F}(s) + \alpha s W^{1F}(s) + s^2 W^{2F}(s)].$$

The filter H is designed so that the inverse Laplace transforms of $s^2 W^{2F}(s)$ and $sW^{1F}(s)$ are each of order $O(\alpha^2)$. The induced operator norm of $\alpha[Is - \alpha A^*]^{-1}$, viewed as a mapping on L_∞, is uniformly bounded over $0 < \alpha \leq 1$. These arguments constitute the proof of (35b).

An important conclusion from the final representation of X^F is that *additional filtering comes from system dynamics*. The matrix-valued transfer function $[Is - \alpha A^*]^{-1}$ may attenuate some of these signals if α is small and the signals are band limited. This is the case for the linear examples (14) and explains the success of the first-order filter, as illustrated in Figure 2(a). In this example, $[Is - \alpha A^*]^{-1} = 1/(s + \alpha)$, and the spectrum of $\mathcal{W}_t$ is discrete.

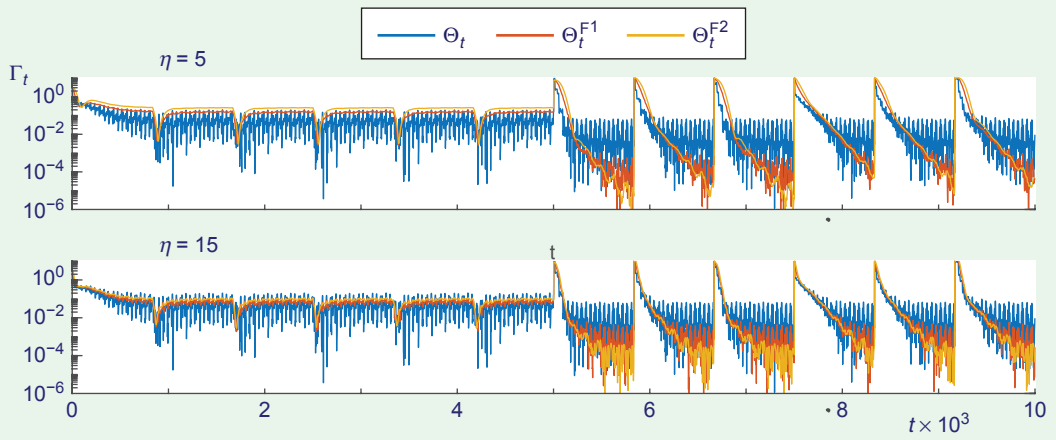


FIGURE 6 The impact of filtering on estimation error for tracking. Second-order filtering can dramatically reduce the norm of the error when the objective moves slowly. As the rate of change of $\{\theta_t^{\text{opt}}\}$ increases, performance of filtering is degraded.

a second-order filter is preferable when the rate of change of $\{\theta_t^{\text{opt}}\}$ is small.

Stability

There are two common approaches to establishing stability in SA that lend themselves to establishing α^0-ultimately boundedness for QSA:

1) Lyapunov criteria, similar to what was discussed in the proof overview of Theorem 3.
2) Stability of a mean flow with a scaled vector field, known as the *ODE@∞*.

Lipschitz Lyapunov function. This is the standard criterion used to establish ultimate boundedness of state-space models [S21, Ch. 4]. The Lyapunov function $V : \Re^d \to \Re_+$ is assumed C^1, and together with a constant $\delta_0 > 0$, satisfies $\nabla V(x) \cdot \bar{f}(x) \le -\delta_0 V(x)$ when $\|x\| \ge \delta_0^{-1}$. In the time domain

$$\frac{d}{dt} V(\vartheta_t) \le -\delta_0 V(\vartheta_t), \quad \text{when } \|\vartheta_t\| > \delta_0^{-1}. \tag{40}$$

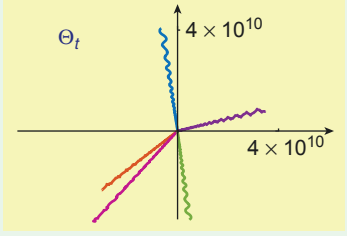


FIGURE 7 Trajectories of ESC-0 for the Rastrigin objective from large initial conditions. The stable behavior of the five trajectories shown is consistent with approximate coupling of solutions to QSA and the ODE@∞.

The application of V to the QSA ODE is successful if the function V is globally Lipschitz continuous. This fails for the standard quadratic option $V_1(x) = x^\top P x$ for a $d \times d$, matrix P with $P > 0$. However, if $\nabla V_1(x) \cdot \bar{f}(x) \le -\delta_1 V_1(x)$ for $\|x\| \ge \delta_1^{-1}$, then the chain rule gives the desired bound for the Lipschitz function $V = \sqrt{1 + V_1}$ and $\delta_0 \in (\delta_1/2, 1)$. (See [51] and [52] for recent Lyapunov theory for SA.)

Stability of the ODE@∞. This criterion is motivated by considering the mean flow starting from a large initial condition and examining the dynamics following scaling.

For fixed $r > 0$, consider the scaled vector field $\bar{f}^r(\theta) = r^{-1}\bar{f}(r\theta), \theta \in \Re^d$. If ϑ_t is a solution to the mean flow with initial condition of magnitude $r = \|\vartheta_0\|$, then the scaled process $\vartheta_t^r := r^{-1}\vartheta_t$ is a solution to the ODE with scaled vector field

$$\frac{d}{dt}\vartheta_t^r = \bar{f}^r(\vartheta_t^r), \qquad \|\vartheta_0^r\| = 1.$$

It is often the case that the scaled vector field is convergent as $r \to \infty$ to obtain

$$\bar{f}_\infty(\theta) := \lim_{r\to\infty} \bar{f}^r(\theta), \quad \theta \in \Re^d. \tag{41}$$

The ODE@∞ is then defined by

$$\frac{d}{dt}\vartheta_t^\infty = \bar{f}_\infty(\vartheta_t^\infty).$$

In several applications, such as in Q-learning, the scaled vector field $\bar{f}_\infty$ is much simpler than $\bar{f}$ [6].

Figure 7 shows the evolution of the solutions to the ESC-0 ODE from a very large initial condition, of order 10^{10}, applied to the Rastrigin objective [46]

$$\Gamma(\theta) = \|\theta\|^2 + 20 - 10[\cos(2\pi\theta_1) + \cos(2\pi\theta_2)]. \tag{42}$$

See Figure 8(a) for a plot of this function. Ultimate boundedness is apparent, and there is also coupling with the ODE@∞.

Theorem 4 (Criteria for α^0-Ultimately Boundedness)

Suppose that assumption (A1) holds, and that either of the following conditions hold:

1) There is a pair V, δ_0 satisfying (40). In addition, V is globally Lipschitz continuous, and $V(x) \geq \delta_0 \|x\|$ for $\|x\| > \delta_0^{-1}$.
2) The ODE@∞ is locally asymptotically stable.

Then, there is $\alpha^0 > 0$ and positive constants b and δ such that the following bounds hold for any $\alpha \in (0, \alpha^0]$ and any initial condition Θ_0, Φ_0:

$$\|\Theta_t\| \leq b\|\Theta_0\| \exp(-\alpha\delta t) \quad \text{for } t \leq T_1$$
$$\text{where } T_1 = \min\{t : \|\Theta_t\| \leq \delta^{-1}\}. \tag{43}$$

Consequently, the family of QSA ODE models is α^0-ultimately bounded.

Proof Overview

The proof of (43) under the Lyapunov criterion is similar to the proof of (35a) in Theorem 3.

Analysis under the second criterion begins with the following two observations:

1) The convergence in (41) is uniform on compact subsets of $\Re^d$.
2) If $\bar{f}_\infty$ is locally asymptotically stable, then it must be globally *exponentially* asymptotically stable, with the origin being the unique stationary point.

This leads to a string of conclusions, ending with a Lipschitz Lyapunov function for the ODE@∞, and then the mean flow. This suffices to obtain the uniform bounds in (43).

Vanishing Gain

There is a parallel theory for QSA with vanishing gain

$$\frac{d}{dt}\Theta_t = a_t f(\Theta_t, \xi_t). \tag{44}$$

The development is similar, leading to familiar choices in design:

- Assumption (A0) is imposed.
- Filtering is performed to obtain the final estimates $\{\Theta_t^{\mathrm{F}}\}$.

A significant difference is the intended goal: convergence of the estimates to θ^* is guaranteed under mild assumptions, which means that both probing design and filtering are performed to improve the *rate of convergence*.

The vanishing gain is chosen of the form

$$a_t = \alpha(1 + t/t_e)^{-\rho} \tag{45}$$

in which $\alpha > 0$ and $t_e > 0$ are arbitrary. Theory requires $\rho \in (1/2, 1)$.

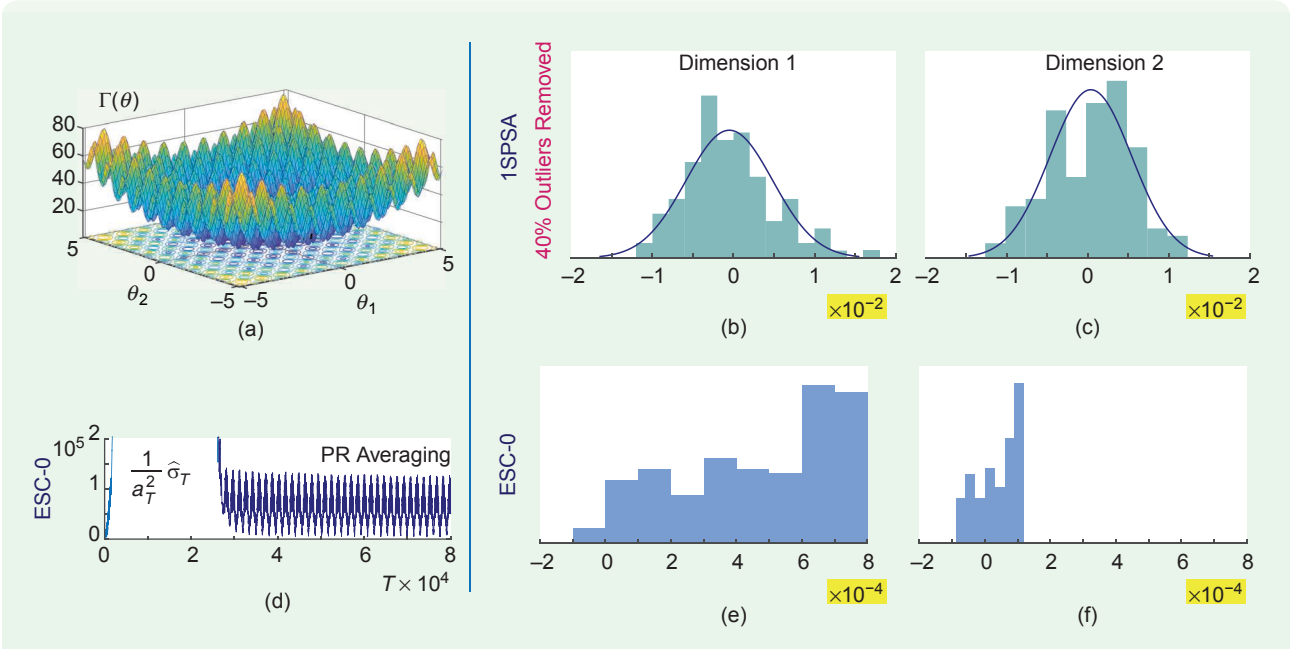


FIGURE 8 A comparison between quasi-stochastic and stochastic algorithms for minimization of the Rastrigin objective. (a) A plot of the objective, (d) evolution of the scaled empirical variance, (b) and (c) histograms of estimation error for 1SPSA with Polyak-Ruppert (PR) averaging, and (e) and (f) histograms of estimation error for ESC-0 with PR averaging. The deterministic algorithm achieves convergence rates arbitrarily close to $O(T^{-4})$, while presenting less variability in estimating θ^{opt}.

Theorem 1 admits an exact extension to the vanishing gain setting, beginning with the notation in terms of apparent noise

$$\frac{d}{dt}\Theta_t = a_t[\tilde{f}(\Theta_t) + \tilde{\Xi}_t]$$
$$\tilde{\Xi}_t = -a_t\overline{\Upsilon}_t + \mathcal{W}_t. \tag{46}$$

Theorem 5 (P-Mean Flow)

Subject to (A3), the p-mean flow representation (46) holds with

$$\mathcal{W}_t = \sum_{i=0}^{2} a_t^{2-i}\frac{d^i}{dt^i}\mathcal{W}_t^i$$

where the first term is modified

$$\mathcal{W}_t^0 = -[D^f\hat{\Upsilon}]_t + \frac{r_t}{a_t}[D^f\hat{\hat{f}}]_t$$
$$\text{with}\quad r_t = -\frac{d}{dt}\log(a_t) = \frac{\rho}{t_e + t}.$$

The remaining terms are unchanged: $\Upsilon_t = \Upsilon(\Theta_t, \Phi_t)$ with Υ defined in (24e) and

$$\mathcal{W}_t^1 = -[D^f\hat{\hat{f}}]_t + \hat{\Upsilon}_t,\quad \mathcal{W}_t^2 = \hat{\hat{f}}_t.$$

This leads to convergence of $\boldsymbol{\Theta}$ to θ^* with rate $O(a_t)$, which is improved to $O(a_t^2)$ with filtering. This is an astonishing conclusion: the rate can be arbitrarily close to $O(t^{-2})$ by choosing ρ close to unity.

The second-order filter is abandoned and replaced by a simple time average, known as *Polyak–Ruppert averaging*

$$\Theta_T^{\mathsf{PR}} := \frac{1}{T - T_0}\int_{T_0}^{T}\Theta_t dt. \tag{47}$$

The interval $[0, T_0]$ is known as the *burn-in period*; estimates from this period are abandoned to reduce the impact of transients in early stages of the run.

Theorem 6 (Acceleration With Vanishing Gain)

Suppose that assumptions (A1)–(A3) and assumption (A4) hold with one modification: α^0-ultimate boundedness for the family of QSA ODEs (3) is not assumed, but the QSA ODE (44) is assumed to have bounded solutions from each initial condition.

Suppose that $\rho \in (1/2, 1)$ and $t_e > 0, \alpha > 0$. Suppose, moreover, that T_0 is selected to solve $1/(T - T_0) = \kappa/T$ with $\kappa > 1$. Then, the following approximations hold for (44) and the averaged estimates:

$$\Theta_t = \theta^* - a_t\hat{f}_t^* + O(a_t\|\overline{Y}^*\|) + o(a_t) \tag{48a}$$
$$\Theta_T^{\mathsf{PR}} = \theta^* + a_T[c(\kappa, \rho) + o(1)]\overline{Y}^* + O(T^{-2\rho}) \tag{48b}$$

where $c(\kappa, \rho) > 0$, $\overline{Y}^* = [A^*]^{-1}\overline{\Upsilon}^*$, and $\hat{f}_t^* = \hat{f}(\theta^*, \Phi_t)$.

Consequently, Θ_T^{PR} converges to θ^* with rate bounded by $O(T^{-2\rho})$ if and only if $\overline{Y}^* = 0$.

Gain Selection for Static Optimization

The focus on fixed-gain algorithms was motivated entirely by applications to tracking. In the static root-finding problems found in optimization and RL, Theorem 6 suggests that a vanishing gain algorithm may prove to be far more efficient and not very sensitive to the coefficients in the gain process (45). The results from experiments using ESC-0 will make this point clear.

Vanishing or fixed gain? Vanishing gain algorithms provide extra degrees of freedom: a single scalar α cannot balance transient response and asymptotic performance. The next set of experiments are designed to illustrate this conflict.

Recall the Rastrigin objective defined in (42), for which a plot is shown in Figure 8. Optimization is challenging because of the infinite number of local extrema and saddle points. Three choices of a_t in (44) are considered in the ESC-0 ODE:

1) $a_t = 0.1(t+1)^{-0.65}$
2) $a_t \equiv \alpha_b = 3 \times 10^{-3}$
3) $a_t \equiv \alpha_s = 7 \times 10^{-4}$.

The top row of Figure 9 shows the evolution of $\boldsymbol{\Theta}$ for each choice of gain and several initial conditions. The bottom row shows the evolution of $\{\Gamma(\Theta_T^{\mathsf{PR}}), \Gamma(\Theta_T^{\mathsf{1F}}), \Gamma(\Theta_T^{\mathsf{2F}}), \Gamma(\Theta_T)\}$ for the single path yielding the best performance for each gain choice across all runs.

The following takeaways are noted:

- For $a_t = 0.1(t+1)^{-0.65}$ (case 1), Figure 9 illustrates the advantage of vanishing gain algorithms: the algorithm explores much more in the beginning of the run, and the objective remains very small after a brief transient period. The parameter estimates converge to $\theta^{\mathsf{opt}} = 0$ in each experiment.
- For the runs that used α_b (case 2), a good amount of exploration is observed, but the steady-state behavior is poor. Case 3, using the smaller value of α, often yielded better results in steady state, but in several cases, the trajectory remains trapped near a nonoptimal local minimum.
- Figure 9 shows the benefit of bias reduction from a second-order filter as opposed to a first-order filter, based on runs that used α_s. As opposed to the results in Figure 2, this example shows that a first-order filter is not always sufficient to obtain AAD of order $O(\alpha^2)$.

  When the trajectory is not trapped near a nonoptimal local minima, the final estimates obtained using the second-order filter are comparable to what is obtained in the vanishing gain experiments, in terms of quality of the approximation of θ^{opt}.

In conclusion, a constant gain QSA ODE can be fine-tuned to obtain good results, but the vanishing gain algorithm is far more reliable in these experiments.

Impact of Dimension. According to Theorem 6, the convergence rate of $\{\Theta_T^{\text{PR}}\}$ to θ^* is of order $O(T^{-2\rho})$, so that the empirical variance (S14) vanishes at a rate bounded by $O(T^{-4\rho})$. There is no theory available that indicates how the constants in these bounds are impacted by dimension, so we explore the impact through another application of ESC-0, this time for the Ackley objective [46].

Figure 10 shows the evolution of $T^{2\rho}\hat{\sigma}_T$ for $d = 2$ and for $d = 30$ [see (S14) for the definition of the empirical variance $\hat{\sigma}_T^2$]. Simulations confirm that the variance is bounded by $O(a_T^4)$ but grows with dimension.

Figure 10 also shows that performance is not unacceptable: the averaged sample paths $\{\Theta_T^{\text{PR}}\}$ approach the optimizer $\theta^{\text{opt}} = 0$.

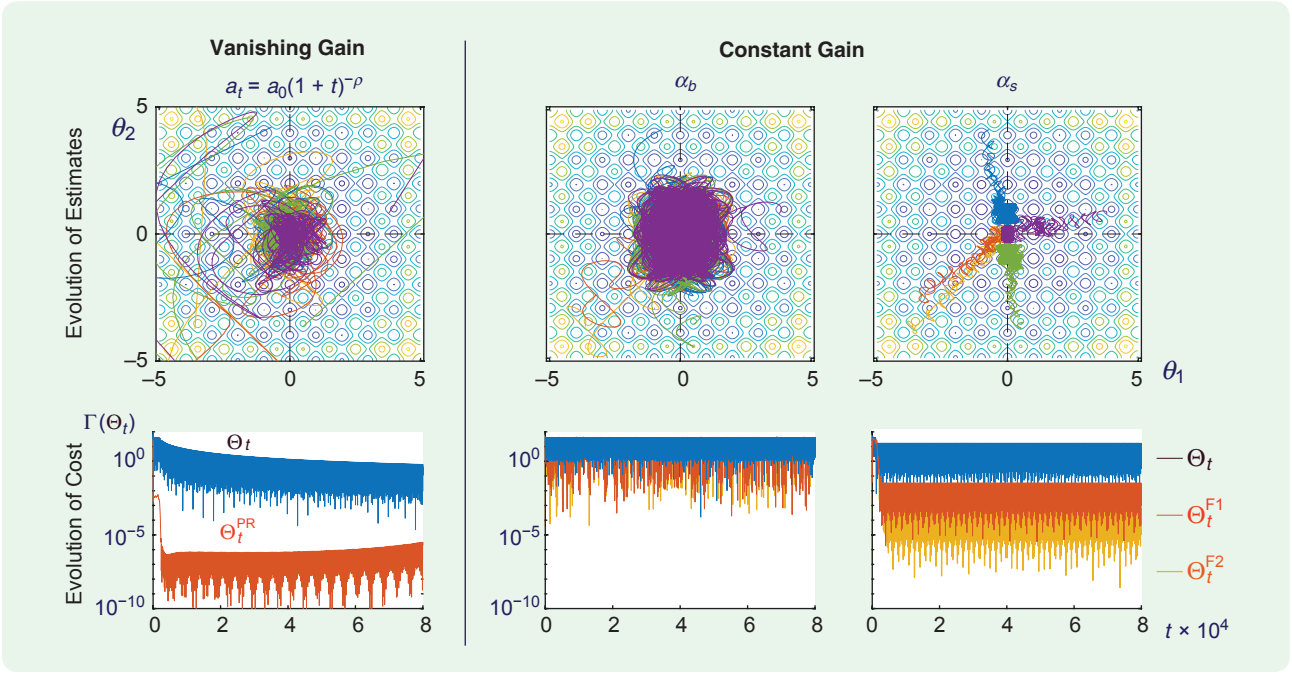


FIGURE 9 A comparison between vanishing and constant gain ESC-0 for optimization of the Rastrigin objective. The vanishing gain algorithm has a lot of exploration power and approaches θ^{opt} quickly. For the case with fixed gain, the steady state is poor when the gain is large. As the fixed gain decreases, the algorithm loses its exploration power.

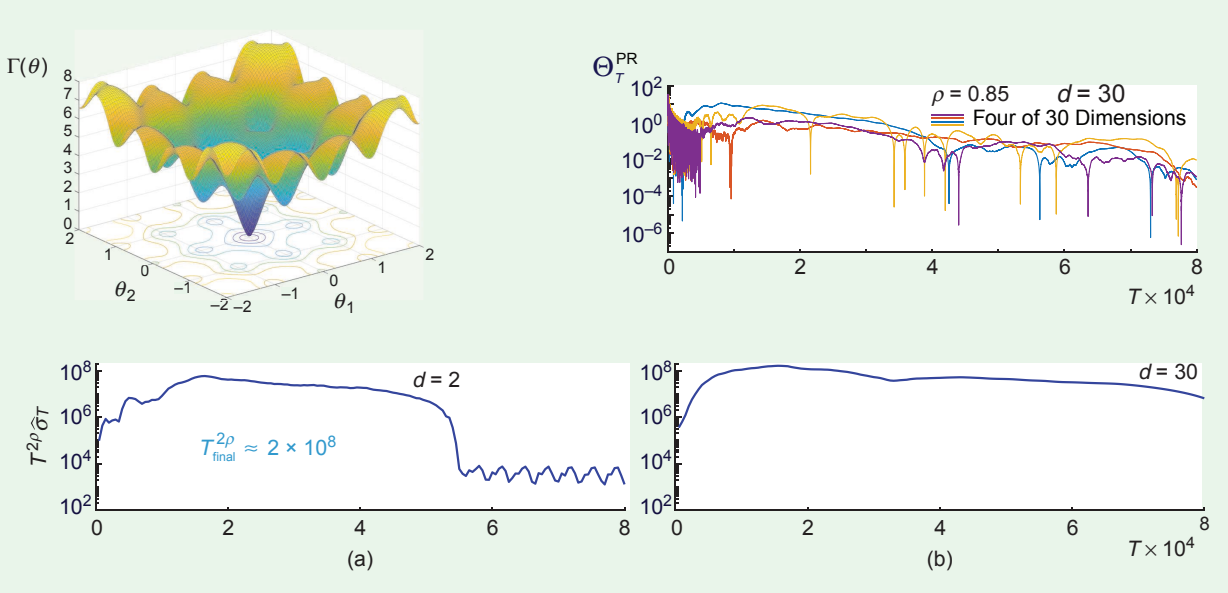


FIGURE 10 The impact of dimension. (a) The Ackley objective for dimension $d = 2$. (b) The evolution of sample paths of four dimensions of $\{\Theta_T^{\text{PR}}\}$ for the objective with $d = 30$. The two plots on the bottom show evolution of the scaled empirical variance for (a) $d = 2$ and (b) $d = 30$.

Random or smooth exploration? The use of i.i.d. exploration has great appeal because of its simplicity and the many tools for analysis. This is why it is a standard approach to exploration in many areas of machine learning. Also, it might be assumed that this approach to exploration is efficient in applications to gradient-free optimization due to the high-frequency content in the probing signal. This is far from the truth (as can be seen from theory) but is best made clear through illustration.

The 1SPSA algorithm (S4a) is the stochastic counterpart to ESC-0 and is implemented along with its deterministic version here to illustrate the benefits of carefully designing exploration signals.

Results from the application of multiple instances of 1SPSA and ESC-0 for minimization of the Rastrigin objective are shown in Figure 8. The following can be seen:

- Figure 8 shows histograms for the estimation error $\Theta_T^{\mathsf{PR}} - \theta^{\mathsf{opt}}$ for each experiment. The variance of the estimation error for the deterministic algorithm is much smaller than for its stochastic counterpart: the reduction is roughly two orders of magnitude. Roughly 40% of the estimates were considered outliers for the stochastic algorithm, while none were observed for its deterministic counterpart.
- Figure 8 shows the evolution of the scaled empirical variance (S14) across all instances of the deterministic algorithm. This process is bounded as expected for a convergence rate of order $O(T^{-2\rho})$.

Summary of Design Principles

By now, it is clear that QSA theory leads to a toolbox for design. "Part 1: QSA" is concluded with a brief summary:

- Ensure by luck or design that $\bar{f}$ and f are globally Lipschitz continuous, and that the mean flow is globally asymptotically stable.
- The probing signal is a smooth function of sinusoids, but of a special form. Frequencies must be distinct and respect (A0) to ensure that $\overline{\Upsilon} = 0$. Bias may be significant if this constraint is ignored.
- *Perform filtering*: a second-order low-pass filter can reduce estimation bias and variance dramatically.
- *Test your algorithm*: perform repeated trials to estimate variance and outliers.

In some applications, it may not be possible to ensure Lipschitz continuity. In such cases, a projection of estimates is required to ensure boundedness. If it is known that if f violates the Lipschitz bounds, then projection alone is not sufficient: the larger the domain of projection, the smaller the choice of α in (3).

PART 2: ESC

This second part is devoted to explaining how QSA theory applies to ESC for the purposes of

- Stability verification
- Bounds on transient behavior
- Bounds on asymptotic bias.

Precise statements on each point are provided for static optimization, but only empirical results in the case of tracking.

Probing is assumed to be a true mixture of sinusoids, which is obtained when G_0 in (A0a) is linear

$$\xi_t = \sum_{i=1}^{K} v^i \cos(2\pi[\omega_i t + \phi_i]) \tag{49}$$

with $v^i \in \Re^d$ for each i, $K \geq d$ and the K frequencies are positive and *distinct*. The covariance matrix is thus

$$\Sigma_\xi = \mathsf{E}_\pi[G(\Phi)G(\Phi)^\intercal] = \frac{1}{2}VV^\intercal \tag{50}$$

with V the $d \times K$ matrix with columns equal to the v^i appearing in (49).

This structure is imposed to avoid unnecessary abstractions and because the bandwidth of the apparent noise is controlled when the probing gain is small [recall $\tilde{\Xi}$ defined in (6)].

Theorem 10 provides a QSA representation for ESC in broad generality, not just the special case of ESC-0.

Approximations for each of the terms in the p-mean flow representation are available, subject to assumptions on the objective function. The following assumptions are listed in order, paralleling assumptions (A0)–(A4). Note that there is no assumption (E3) because (E2) will justify both (A2) and (A3).

(E0): The probing signal is of the form (49), with frequencies satisfying (A0) and $\Sigma_\xi > 0$.

(E1): The objective Γ is C^2 and has a Lipschitz continuous gradient.

(E2): The objective is analytic.

(E4): The objective satisfies

- $\|\nabla\Gamma(\theta)\| \geq \delta\|\theta\|$ for some $\delta > 0$ and all $\|\theta\| \geq \delta^{-1}$.
- It has a unique minimizer θ^{opt}, and it is the only solution to $\nabla\Gamma(\theta) = 0$.
- $P = \nabla^2\Gamma(\theta^{\mathsf{opt}})$ is positive definite.

Just as (A0) and (A1) were valuable in QSA theory, so are (E0) and (E1) here. It will be seen that (A1) follows from (E1), and (E4) implies (A4), subject to (A0) and (A1).

However, none of these implications are valid without a small change in the definition of the ESC observations.

QSA Theory Requires Lipschitz Continuity

Recall the early warning in the introduction: ESC ODEs are not Lipschitz continuous unless the observations $\{\mathcal{Y}_t^n\}$ defined in (15) are Lipschitz continuous as functions of Θ_t. This is rarely the case in practice, so a first step is to modify the algorithm so that assumption (A1) is satisfied.

A state-dependent probing gain ϵ_t is adopted for two important reasons:

1) Assumption (A1) will follow from (E1).
2) If the observed cost $\Gamma(\mathcal{Z}_t)$ is large, then it makes sense to increase the exploration gain to move quickly to a more desirable region of the parameter space.

The state-dependent probing gain ensures that the Lipschitz condition (A1) is satisfied, which is essential to global stability theory.

Two choices for $\epsilon_t \equiv \epsilon(\Theta_t)$ are proposed here:

$$\epsilon(\theta) = \varepsilon\sqrt{1+\Gamma(\theta)-\Gamma^-} \tag{51a}$$

$$\epsilon(\theta) = \varepsilon\sqrt{\frac{1+\|\theta-\theta^{\text{ctr}}\|^2}{\sigma_p^2}} \tag{51b}$$

where in (51a), the constant Γ^- is chosen so that $\Gamma(\theta)\ge\Gamma^-$ for all θ. In the second option, θ^{ctr} is interpreted as an a-priori estimate of θ^{opt} as in (9), and σ_p plays the role of standard deviation around this prior.

The first is the most intuitive as it directly addresses 2): the exploration gain ϵ_t is large when $\Gamma(\Theta_t)$ is far from its optimal value. However, it does not lead to an online algorithm because $\Gamma(\Theta_t)$ is not observed. In a discrete-time implementation, an online version is adopted:

$$\epsilon_{t_n} = \varepsilon\sqrt{1+\mathcal{Y}_{t_{n-1}}-\Gamma^-}$$

with $\mathcal{Y}_t := \Gamma(\Theta_t+\epsilon_t\xi_t)$ for $t=t_n$, $n\ge 0$ (compare Figure 1).

In cases (51a) or (51b), we adopt the new definition

$$\mathcal{Y}^{\text{n}}(\theta,\xi) = \frac{1}{\epsilon}\Gamma(\theta+\epsilon\xi) \tag{52}$$

with the understanding that $\epsilon=\epsilon(\theta)$. The signal $\mathcal{Y}_t^{\text{n}} = \mathcal{Y}^{\text{n}}(\Theta_t,\xi_t)$ is an important part of the feedback loop in any interpretation of Figure 1.

Theorem 7 (Lipschitz Observations for ESC)

The function $\mathcal{Y}^n$ defined in (52) is uniformly Lipschitz continuous in θ, subject to (E1) and either of the following:

1) ϵ is defined by (51b).
2) ϵ is defined by (51a), and (E4a) holds.

Moreover, under either 1) or 2), the following approximation holds:

$$\mathcal{Y}^{\text{n}}(\theta,\xi) = \frac{1}{\epsilon(\theta)}\Gamma(\theta)+\xi^{\intercal}\nabla\Gamma(\theta)+O(\epsilon) \tag{53}$$

where the error term $O(\epsilon)$ is bounded by a fixed constant times $\epsilon(\theta)$.

The plots shown in Figure 7 were obtained using ESC-0 with probing gain (51b) and normalized "observations" (52). The state-dependent probing gain ensures that the Lipschitz condition (A1) is satisfied, which is essential to global stability theory. An example of divergence using a fixed probing gain is contained in "Finite Escape Time for Extremum Seeking Control."

Finite Escape Time for Extremum Seeking Control

The Lipschitz conditions in (A1) cannot be relaxed in the global stability theory for quasi-stochastic approximation. This is why establishing global stability of extremum seeking control (ESC) is challenging when Γ is not Lipschitz continuous.

The ESC-0 ODE is recalled here:

$$\dot{\Theta}_t = -\alpha\frac{1}{\varepsilon}\xi_t\Gamma(Y_t),\quad Y_t=\Theta_t+\varepsilon\xi_t.$$

Consider the scalar ODE with quadratic objective $\Gamma(\theta)=\theta^2$ and probing signal $\xi_t=\cos(\omega_0 t)$. For this simple example, we obtain

$$\begin{aligned}\frac{d}{dt}Y_t &= -\frac{\alpha}{\varepsilon}\xi_t\Gamma(Y_t)+\varepsilon\frac{d}{dt}\xi_t\\ &= -\frac{\alpha}{\varepsilon}\cos(\omega_0 t)Y_t^2-\varepsilon\omega_0\sin(\omega_0 t).\end{aligned}$$

This ODE has finite escape time when $Y_0<0$ and $|Y_0|$ is sufficiently large.

To justify this claim, we bound $\{Y_t : 0\le t<t_\diamond\}$ with

$$t_\diamond = 2\varepsilon\frac{1}{\alpha}\frac{1}{|Y_0|}.$$

Assume that $\varepsilon|Y_0|^{-1}$ is sufficiently small so that $\cos(\omega_0 t)\ge(1/2)$ for $0\le t\le t_\diamond$. This implies the lower bound

$$-\frac{d}{dt}Y_t \ge \alpha\frac{Y_t^2}{2\varepsilon}$$

and hence

$$\frac{d}{dt}\frac{1}{Y_t} = -\frac{1}{Y_t^2}\Big(\frac{d}{dt}Y_t\Big)\ge\frac{\alpha}{2\varepsilon},\qquad \text{for } t<t_\diamond.$$

Integrating both sides from zero to any value $T<t_\diamond$ gives

$$\frac{1}{Y_T}-\frac{1}{Y_0}\ge\frac{\alpha}{2\varepsilon}T\ \Rightarrow\ Y_T\le\Big(\frac{1}{Y_0}+\frac{\alpha}{2\varepsilon}T\Big)^{-1}.$$

In conclusion, for a value $t_\bullet\in(0,t_\diamond)$, the solution $\{Y_t : 0\le t<t_\bullet\}$ is continuous and decreasing, with

$$\lim_{T\uparrow t_\bullet} Y_T = -\infty.$$

Global stability is ensured if the probing gain is state dependent; either of the choices in (51) ensure success.

P-Mean Flow for ESC-0

We begin with ESC-0 because of the simple approximations for both the mean flow and the p-mean flow representation, starting with the ESC-0 vector field

$$f(\theta,\xi) = -\mathcal{Y}^{\mathrm{n}}(\theta,\xi)\xi.$$

The following approximations hold under (E0) and (E2), through an application of Theorem 7:

$$f(\theta,\xi) = -\frac{1}{\epsilon(\theta)}\Gamma(\theta)\xi - \xi\xi^{\intercal}\nabla\Gamma(\theta) + O(\epsilon) \tag{54a}$$

$$\bar{f}(\theta) = -\Sigma_{\xi}\nabla\Gamma(\theta) + O(\epsilon^2) \tag{54b}$$

$$\hat{f}(\theta,\Phi) = -\frac{1}{\epsilon(\theta)}\Gamma(\theta)\hat{G}(\Phi) - \hat{\Sigma}(\Phi)\nabla\Gamma(\theta) + O(\epsilon). \tag{54c}$$

In (54a) and (54b), the error terms $O(\epsilon)$ and $O(\epsilon^2)$ represent a uniform bound over $\Re^d$. In (54c), the approximation is uniform on compact subsets of $\Re^d$. The function $\hat{G}$ is the solution to Poisson's equation with forcing function G defined in (18), so that $G(\Phi_t) = \xi_t$. The function $\hat{\Sigma}(\Phi)$, is a matrix-valued solution to Poisson's equation: the forcing function for entry (i, j) is G_iG_j.

The term $\hat{f}_t = \hat{f}(\Theta_t, \Phi_t)$ appears in the pre-p-mean flow equation (28). Although zero mean, we can expect the division by $\epsilon(\Theta_t)$ to induce high volatility.

Only the approximation (54b) is required for verifying stability, which means that (E2) may be relaxed in the following.

Theorem 8 (Stability Criteria for ESC-0)

If (E0), (E1), and (E4a) hold, then ESC-0 is α^0-ultimately bounded.

All of the approximations in (54) are imposed to approximate the terms in the p-mean flow.

Theorem 9 (QSA Theory for ESC-0)

The p-mean flow representation holds under (E2):

$$\frac{d}{dt}\Theta_t = \alpha[\bar{f}(\Theta_t) - \alpha\overline{\Upsilon}_t + \mathcal{W}_t].$$

If, in addition, (E0) holds, then $\overline{\Upsilon}_t \equiv 0$.

If (E0)–(E4) hold, then there is $\varepsilon^0 > 0$ such that the following uniform bounds hold for $0 < \alpha \le \alpha^0$, $0 < \varepsilon \le \varepsilon^0$, and $t \ge t_0(\theta, z, \alpha)$:

» *Approximate gradient descent*:

$$\frac{d}{dt}\Theta_t = -\alpha[\Sigma_{\xi}\nabla\Gamma(\Theta_t) + \mathcal{W}_t + O(\varepsilon^2)].$$

» *Approximate linear dynamics*:

$$\frac{d}{dt}\Theta_t = \alpha[-\Sigma_{\xi}P[\Theta_t - \theta^{\mathrm{opt}}] + \mathcal{W}_t + O(\alpha^2+\varepsilon^2) + o(1)]. \tag{55a}$$

» *Approximate consistency*:

1) $\|\Theta_t - \theta^{\mathrm{opt}}\| \le O(\alpha+\varepsilon^2) + o(1)$.
2) $\|\Theta_t^{\mathrm{F}} - \theta^{\mathrm{opt}}\| \le O(\alpha^2+\varepsilon^2) + o(1)$, with a filtered estimate obtained using the criteria of Theorem 3.

The proof of Theorem 9 follows from Theorem 4 using $V = \sqrt{\Gamma(\theta) - \Gamma^{-}}$, where Γ^{-} is chosen so that V takes on positive values [recall (51a)].

Models and Approximations for General ESC

To simplify the discussion, it is best to maintain the first-order low-pass filter

$$\frac{d}{dt}\Theta_t = -\sigma\Theta_t - \alpha\check{\nabla}_t\Gamma, \qquad \check{\nabla}_t\Gamma = \check{\xi}_t\check{\mathcal{Y}}_t^{\mathrm{n}}. \tag{56}$$

If the high-pass filter is taken to be all pass (a scalar gain), then it is a simple task to generalize Theorems 8 and 9 to $\sigma > 0$. Modeling for genuine high-pass filters within the framework of QSA requires more effort.

Consider a high-pass filter with state-space realization of dimension $q \ge 1$

$$\frac{d}{dt}\mathrm{Z}_t = \mathrm{F}\mathrm{Z}_t + \mathrm{G}u_t \tag{57a}$$

$$y_t = \mathrm{H}^{\intercal}\mathrm{Z}_t + \mathrm{J}u_t \tag{57b}$$

with (F, G, H, J) of compatible dimension. In this equation, u_t is the scalar input, y_t the scalar output, and Z_t the q-dimensional state process.

The $(d+q)$-dimensional state process for ESC has the form $X_t = (\Theta_t; Z_t)$, in which Z_t is (57a) with input $u_t = \mathcal{Y}_t^{\mathrm{n}}$. Its evolution is described by the controlled nonlinear state-space model

$$\frac{d}{dt}X_t = \alpha\begin{bmatrix} -\frac{\sigma}{\alpha}I & -\check{\xi}_t\mathrm{H}^{\intercal} \\ 0 & \frac{1}{\alpha}\mathrm{F}\end{bmatrix}X_t + \alpha\begin{bmatrix}-\mathrm{J}\check{\xi}_t \\ \frac{1}{\alpha}\mathrm{G}\end{bmatrix}\mathcal{Y}_t^{\mathrm{n}} \tag{58}$$

driven by the *2d-dimensional* input $(\xi_t, \check{\xi}_t)$.

To match the architecture shown in Figure 1, the high-pass filter is used for $d+1$ different choices of input: in addition to $u_t = \mathcal{Y}_t^{\mathrm{n}}$, giving $y_t = \check{\mathcal{Y}}_t^{\mathrm{n}}$, the input $u_t = \xi_t^i$ gives $y_t = \check{\xi}_t^i$ for each i.

P-mean flow representation. We can freely apply Theorem 1 to the state-space representation (58) because the theorem makes no assumptions on the magnitude of α, or even the stability of the QSA ODE.

Remember that α is a fixed constant, so the fact that f depends on this gain is irrelevant in the definition for the QSA vector field

$$f(x,\check{\xi},\xi) = \begin{bmatrix} -\frac{\sigma}{\alpha}I & -\check{\xi}\mathrm{H}^{\intercal} \\ 0 & \frac{1}{\alpha}\mathrm{F}\end{bmatrix}x + \begin{bmatrix}-\mathrm{J}\check{\xi} \\ \frac{1}{\alpha}\mathrm{G}\end{bmatrix}\mathcal{Y}^{\mathrm{n}}(\theta,\xi) \tag{59}$$

where $x = (\theta;\varsigma) \in \Re^{d+q}$ denotes an arbitrary value for X_t.

Three solutions to Poisson's equation are required to write down the p-mean flow:

3) The solution $\hat{\mathcal{Y}}^{\mathsf{n}}$ with forcing function $\mathcal{Y}^{\mathsf{n}}$.
4) $\hat{\check{\xi}}$ with forcing function $\check{\xi}$ [similar to $\hat{G}$ in (54c)].
5) $\hat{Q}$ with forcing function $Q(\theta,\Phi) = -\mathrm{J}\check{\xi}\mathcal{Y}^{\mathsf{n}}(\theta,\xi)$.

Theorem 10 (P-Mean Flow for ESC)

The p-mean flow representation holds under (E2)

$$\frac{d}{dt}X_t = \alpha[\bar{f}(X_t) - \alpha\overline{\Upsilon}_t + \mathcal{W}_t] \tag{60a}$$

in which for any $x = (\theta;\varsigma)$ and $z \in \Omega$

$$\bar{f}(x) = \begin{bmatrix} -\frac{\sigma}{\alpha}I & 0 \\ 0 & \frac{1}{\alpha}\mathrm{F} \end{bmatrix} x + \begin{bmatrix} -\mathrm{J}\mathsf{E}[\check{\xi}(\Phi)\mathcal{Y}^{\mathsf{n}}(\theta,\xi)] \\ \frac{1}{\alpha}\mathrm{G}\mathsf{E}[\mathcal{Y}^{\mathsf{n}}(\theta,\xi)] \end{bmatrix} \tag{60b}$$

with expectations in steady state. The functions $\hat{f}$ and $\overline{\Upsilon}$ admit the representations

$$\hat{f}(x,z) = \begin{bmatrix} 0 & -\hat{\check{\xi}}(z)\mathrm{H}^{\top} \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} -\mathrm{J}\hat{Q}(\theta,z) \\ \frac{1}{\alpha}\mathrm{G}\hat{\mathcal{Y}}^{\mathsf{n}}(\theta,z) \end{bmatrix} \tag{60c}$$

$$\Upsilon(x,z) = \frac{1}{\alpha}\begin{bmatrix} \hat{\check{\xi}}(z)\mathrm{H}^{\top}\{\mathrm{F}x + \mathrm{G}\mathcal{Y}^{\mathsf{n}}(\theta,\xi(z))\} \\ 0 \end{bmatrix}. \tag{60d}$$

Proof

The expression (60c) follows directly from (59). There is simplification because terms not involving ξ or $\check{\xi}$ vanish. The formula (60d) then follows from the definition $\gamma(x,z) = -\partial_x\hat{f}(x,z)f(x,z)$. □

Interpretation of the p-mean flow representation is entirely different here, because Υ is no longer a nuisance term but a critical part of the dynamics. Application to design remains a topic for future research.

ESC as two time-scale QSA. The state-space model (58) is an instance of two-timescale QSA, provided the low-pass filter gain scales with α, so that $\sigma = O(\alpha)$. The pair (Z_t,Φ_t) represents the fast state variables, and as always, Θ_t is the slow variable. See [S7, Ch. 8] for a survey of the rich theory of two-timescale SA.

In this deterministic setting, with constant gain α, theory of two-timescale QSA is a subset of singular perturbation theory. The objective is model reduction, which in this case amounts to approximating (58) by the d-dimensional instance of QSA

$$\begin{aligned} \frac{d}{dt}\Theta_t^{\circ} &= -\sigma\Theta_t^{\circ} - \alpha M_t\nabla\Gamma(\Theta_t^{\circ}) - \alpha\check{\xi}_t\frac{1}{\epsilon_t}(h_0+\mathrm{J})\Gamma(\Theta_t^{\circ}) + O(\alpha\epsilon) \\ &\text{with}\quad M_t = \check{\xi}_t[\check{\xi}_t + J\xi_t]^{\top}, \qquad \Theta_0^{\circ} = \Theta_0 \end{aligned} \tag{61a}$$

where $h_0 = -\mathrm{H}^{\top}\mathrm{F}^{-1}\mathrm{G}$ is the dc gain of the high-pass filter. Its mean flow is easily identified:

$$\frac{d}{dt}\vartheta_t^{\circ} = -\sigma\vartheta_t^{\circ} - \alpha M\nabla\Gamma(\vartheta_t^{\circ}) + O(\alpha\epsilon) \tag{61b}$$

with $M = \mathsf{E}_{\pi}[M_t]$. An analysis of this ODE is far more tractable than the original ESC ODE. In particular, the mean flow (61b) is stable, provided the high-pass filter is *passive*, such as a lead compensator. Passivity combined with positivity of Σ_{ξ} implies that $M + M^{\top} > 0$.

The approximation is based on *freezing* the slow variable Θ_t in the fast dynamics to obtain an approximation for $\boldsymbol{Z}$. For a given time t, let $\{\overline{Z}_r : r \ge t\}$ denote the solution to the state-space model defining $\boldsymbol{Z}$ with $\Theta_r \equiv \theta$ for all $-\infty < r < \infty$:

$$\overline{Z}_r = \int_{-\infty}^{r} e^{\mathrm{F}(r-\tau)}G\mathcal{Y}^{\mathsf{n}}(\theta,\xi_{\tau})\,d\tau.$$

On substituting $\xi_{\tau} = G(\exp([\tau - r]W\Phi_r)$, it follows that $\overline{Z}_r = \overline{Z}(\theta,\Phi_r)$ for some function $\overline{Z}$ and each r and θ. The next step is to substitute the solution to obtain the approximate dynamics

$$\frac{d}{dt}\Theta_t \approx -\sigma\Theta_t - \alpha[\check{\xi}_t H^{\top}\overline{Z}(\Theta_t,\Phi_t) + \mathrm{J}\check{\xi}_t\mathcal{Y}_t^{\mathsf{n}}]. \tag{62}$$

Defining $\mathcal{Y}^{\circ}(\theta,\xi) := \frac{1}{\epsilon(\theta)}\Gamma(\theta) + \xi^{\top}\nabla\Gamma(\theta)$ and applying Theorem 7 gives

$$\begin{aligned} \mathcal{Y}_t^{\mathsf{n}} &= \mathcal{Y}^{\circ}(\Theta_t,\xi_t) + O(\epsilon) \\ \mathrm{H}^{\top}\overline{Z}(\theta,\Phi_t) &= \mathrm{H}^{\top}\int_{-\infty}^{t} e^{\mathrm{F}(t-\tau)}\mathrm{G}\mathcal{Y}^{\circ}(\theta,\xi_{\tau})\,d\tau + O(\epsilon) \\ &= h_0\frac{1}{\epsilon(\theta)}\Gamma(\theta) + \check{\xi}_t^{\top}\nabla\Gamma(\theta) \end{aligned}$$

and substitution into (62) justifies the claim that (61a) is an approximation of (58):

$$\frac{d}{dt}\Theta_t \approx -\sigma\Theta_t - \alpha M_t\nabla\Gamma(\Theta_t) - \alpha\check{\xi}_t\frac{1}{\epsilon_t}(h_0+\mathrm{J})\Gamma(\Theta_t) + O(\alpha\epsilon).$$

CONCLUSIONS AND OUTLOOK

The perturbative mean flow (p-mean flow) representation opens many doors for analysis of algorithms and provides a clear path to obtain both transient and steady-state performance bounds.

There remains much more to unveil:

Δ The use of filtering for acceleration of algorithms is not at all new. It will be exciting to investigate the implications of the acceleration techniques pioneered by Polyak and Nesterov for nonlinear optimization, particularly in their modern form (see [26] and [33] and the references therein).

The integration of these two disciplines may provide insight into how to design the high-pass filters shown in Figure 1 or suggest entirely new architectures.

Δ The introduction of normalization into the observations in the general form (52) was crucial to obtain global stability of ESC ODEs. There are many improvements to

The perturbative mean flow (p-mean flow) representation opens many doors for analysis of algorithms and provides a clear path to obtain both transient and steady-state performance bounds.

consider. First, on considering the Taylor series approximation (53), performance is most likely improved via a second normalization

$$\mathcal{Y}_t^{\mathrm{n}} = \frac{1}{\epsilon_t}[\Gamma(\Theta_t + \epsilon_t\xi_t) - \Gamma_t^*]$$

in which $\{\Gamma_t^*\}$ are estimates of the minimum of the objective. These might be obtained by passing $\{\mathcal{Y}_r := \Gamma(\Theta_r + \epsilon_r\xi_r)\}$ through a low-pass filter.

Δ Far better performance might be obtained through an observation process inspired by 2SPSA. Consider first a potential improvement of 2SPSA: a state-dependent exploration gain is introduced so that (S4b) becomes

$$\theta_{n+1} = \theta_n - \alpha_{n+1}\frac{1}{2\varepsilon}\xi_{n+1}[\Gamma(\theta_n + \epsilon_n\xi_{n+1}) - \Gamma(\theta_n - \epsilon_n\xi_{n+1})]$$

with $\epsilon_n = \epsilon(\theta_n)$. The division by 2ε (independent of state) remains as 2SPSA in its original form satisfies the required Lipschitz conditions for SA, provided $\nabla\Gamma$ is Lipschitz continuous.

There are surely many ways to obtain an online version based on QSA. One approach is through sampling: denote $T_n = nT$ for a given sampling interval $T > 0$ and take $\mathcal{Y}_t^{\mathrm{n}}$ constant on each interval $[T_n, T_{n+1})$, designed to mimic 2SPSA. One option is the simple average

$$\mathcal{Y}_{T_{n+1}}^{\mathrm{n}} := \frac{2}{T}\frac{1}{2\varepsilon}\int_{T_n}^{T_n+T/2}\xi_t[\Gamma(\theta_n + \epsilon_t\xi_t) - \Gamma(\theta_n - \epsilon_t\xi_t)]\,dt$$

with $\theta_n = \Theta_{T_n}$. This can be computed in real time, based on two sets of observations:

$$\begin{aligned}&\Gamma(\theta_n + \epsilon_t\xi_t), && T_n \le t \le T_n + T/2\\ &\Gamma(\theta_n - \epsilon_n\xi_{t-T/2}), && T_n + T/2 \le t \le T_{n+1}.\end{aligned}$$

Δ The implications for RL deserve much greater attention. The applications of QSA and ESC in [17], [30], and [S5] are only the beginning.

Δ It may be straightforward to extend the p-mean flow representation (7) to tracking problems. This requires theory for time-inhomogeneous QSA of the form

$$\frac{d}{dt}\Theta_t = \alpha f(\Theta_t, \xi_t; t).$$

Analysis would require consideration of solutions to Poisson's equation, such as $\hat{f}(\theta, \cdot\,; t)$ for each $\theta \in \Re^d$ and $t \in \Re$. The representation will be more complex than (7) but will likely lead to sharper bounds than are presently available.

HISTORY AND RESOURCES

Sources for Main Results

Many of the main results presented here are taken from recent publications. The p-mean flow representation (7) first appeared in the preprint [24], along with the general QSA theory contained in Theorems 1–4, and implications to ESC contained in Theorems 7–10. These results are based on a parallel theory for QSA with vanishing gain [8], [15], [23], [S5]; the convergence rates in Theorem 6 for QSA with vanishing gain are taken from [15] and [23].

QSA

Recall from "The Averaging Principle" that the QSA ODE (3) with fixed gain $\alpha > 0$ has a long history within the theory of averaging theory. The discussion that follows concerns QSA with vanishing gain, which is the typical setting of SA theory.

QSA was proposed in [19] and [21] for applications to finance and applied in [30] for application to Q-learning (one approach to RL). QSA and ESC are also applied to actor-only RL in [S5, Ch. 4] and [17]. Something similar to QSA appears in [5], with applications to gradient-free optimization.

The first convergence rate results for QSA were obtained for quasi-periodic linear systems in [44], which was extended to the nonlinear setting in [7], [8], and [S5]. The appearance of Υ and its implication to rates of convergence in QSA is one topic of [S5, Sec. 4.9]. In all of this previous work, it was assumed that a convergence rate of $O(1/T)$ would be the best possible. Theorem 6, taken from [15], demonstrates that this assumption is a fallacy.

Gradient-Free Optimization

The field is far too vast to survey in this article. Instead, we provide a few leads for the reader, beginning with a warning regarding terminology: the terms *zeroth order* and *gradient-free optimization* refer to identical goals and similar approaches. The goal of ESC is not exactly the same, but the methodology is closely related.

Kiefer and Wolfowitz introduced gradient-free methods for optimization in [16], shortly after SA was introduced in [42]. "What Is Stochastic Approximation?" describes a simplification of the original approach due to Spall, and his monograph [45] contains further history and many more insights on algorithm design.

Much of this literature focuses on design for convergence of the estimates $\{\theta_n\}$ to the global optimizer θ^{opt}, which in general requires a vanishing probing gain. For example, for 1SPSA, this amounts to

$$\theta_{n+1} = \theta_n - \alpha_{n+1}\left[\frac{1}{\varepsilon_{n+1}}\xi_{n+1}\Gamma(\theta_n + \varepsilon_{n+1}\xi_{n+1})\right]$$

in which both $\{\alpha_n\}$ and $\{\varepsilon_n\}$ are vanishing nonnegative sequences, and $\{\xi_n\}$ is assumed to be i.i.d.

Polyak was a major contributor to the theory of convergence rates for algorithms that are asymptotically unbiased: it was established in [39] that the best possible convergence rate for the mean square error is $O(n^{-\beta})$ with $\beta = (p-1)/2p$, provided the objective function is p-fold differentiable at θ^*. Upper bounds on convergence rates appeared much earlier in the work of Fabian [14]. See [10], [11], [20], and [35], along with [45] for more recent history.

Extremum seeking control is said to be the oldest approach to gradient-free optimization, with a 1922 patent the alleged starting point [25], [48]. Success stories on the application of ESC to various problems have been shared over the 20th century, for example, in [12], [29], [34], [40], and [41]. Theory has lagged behind practice: the first Lyapunov stability analysis for ESC algorithms appeared in the 1970s for a very special case [28].

Bounds on bias and variance for ESC were established 30 years later in [1] and [18]. Global stability results were not obtained due to the absence of Lipschitz continuity, although parameters can be chosen to achieve an arbitrarily large region of initial conditions for which the solution is bounded [49], [50]. See [2] and [27] for further history.

Convergence Rates for SA

Theory has largely focused on the vanishing gain setting. Most relevant to the current article are the remarkable averaging techniques of Polyak and Ruppert [37], [38], [43] (see [9], [S7], [S8], and [S9] for recent theory and a more complete history).

Poisson's equation appears in many domains in stochastic processes. In addition to SA, versions of this equation appear in the theory of simulation of Markov processes and average-cost optimal control [3], [4], [31], [32], [S5].

There is an equally long history of analysis for algorithms with constant step-size. The most recent literature on constant gain SA for applications to tracking is contained in [S7, Sec. 9.3].

It was first shown in [6] that stability of the ODE@∞ implies a strong form of geometric ergodicity when the probing signal is i.i.d., and based on this, bounds were obtained on the L_2 error of the form

$$\begin{aligned}\mathsf{E}[\|\theta_n - \theta^*\|^2] &= E_\pi[\|\theta_\infty - \theta^*\|^2] + O(\|\theta_0 - \theta^*\|\varrho^n)\\ \mathsf{E}_\pi[\|\theta_\infty - \theta^*\|^2] &= O(\alpha)\end{aligned} \tag{63}$$

where $\varrho < 1$, and $\alpha > 0$ denotes the (fixed) step-size.

In recent work, it is shown that averaging can eliminate variance [13], [S2], provided the apparent noise is a martingale difference sequence. These optimistic conclusions cannot be expected in general [22].

ACKNOWLEDGMENT

The authors thank the Army Research Office (ARO) and the National Science Foundation (NSF) for their financial support for this work. This work was supported by Award W911NF2010055 from the ARO and Award EPCN 1935389 from the NSF.

AUTHOR INFORMATION

Caio Kalil Lauand received the B.S.E.E. degree from the University of North Florida. He is a Ph.D. student at the University of Florida, Gainesville, FL 32611 USA, under the supervision of Prof. Sean Meyn. His research interests include stochastic approximation and applications such as optimization and reinforcement learning. He is a Student Member of IEEE.

Sean Meyn (meyn@ece.ufl.edu) received the B.A. degree in mathematics from the University of California, Los Angeles and the Ph.D. degree in electrical and computer engineering (ECE) from McGill University under the supervision of Prof. Peter Caines. He is an IEEE Control Systems Society Distinguished Lecturer on topics related to reinforcement learning, stochastic processes, and energy systems. Following 20 years as a professor of ECE at the University of Illinois, he joined the University of Florida, Gainesville, FL 32611 USA, where he is a professor and holds the Robert C. Pittman Eminent Scholar Chair. He is a Fellow of IEEE.

REFERENCES

[1] K. B. Ariyur and M. Krstić, "Analysis and design of multivariable extremum seeking," in *Proc. IEEE Amer. Contr. Conf.*, 2002, vol. 4, pp. 2903–2908, doi: 10.1109/ACC.2002.1025231.

[2] K. B. Ariyur and M. Krstić, *Real Time Optimization by Extremum Seeking Control*. New York, NY, USA: Wiley, 2003.

[3] S. Asmussen and P. W. Glynn, *Stochastic Simulation: Algorithms and Analysis*, vol. 57. New York, NY, USA: Springer-Verlag, 2007.

[4] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*, vol. 22. Berlin, Germany: Springer Science & Business Media, 2012.

[5] S. Bhatnagar, M. C. Fu, S. I. Marcus, and I.-J. Wang, "Two-timescale simultaneous perturbation stochastic approximation using deterministic perturbation sequences," *ACM Trans. Model. Comput. Simul.*, vol. 13, no. 2, pp. 180–209, Apr. 2003, doi: 10.1145/858481.858486.

[6] V. S. Borkar and S. P. Meyn, "The O.D.E. method for convergence of stochastic approximation and reinforcement learning," *SIAM J. Contr. Optim.*, vol. 38, no. 2, pp. 447–469, 2000, doi: 10.1137/S0363012997331639.
[7] S. Chen, A. Devraj, A. Bernstein, and S. Meyn, "Accelerating optimization and reinforcement learning with quasi stochastic approximation," in *Proc. Amer. Contr. Conf.*, May 2021, pp. 1965–1972, doi: 10.23919/ACC50511.2021.9482825.
[8] S. Chen, A. Devraj, A. Bernstein, and S. Meyn, "Revisiting the ODE method for recursive algorithms: Fast convergence using quasi stochastic approximation," *J. Syst. Sci. Complexity*, vol. 34, no. 5, pp. 1681–1702, Oct. 2021, doi: 10.1007/s11424-021-1251-5.
[9] A. M. Devraj, A. Bušić, and S. Meyn, "Fundamental design principles for reinforcement learning algorithms," in *Handbook on Reinforcement Learning and Control, Studies in Systems, Decision and Control Series*, vol. 325, K. G. Vamvoudakis, Y. Wan, F. L. Lewis, and D. Cansever, Eds., Cham, Switzerland: Springer, 2021, pp. 75–137.
[10] J. Dippon, "Accelerated randomized stochastic optimization," *Ann. Statist.*, vol. 31, no. 4, pp. 1260–1281, Aug. 2003, doi: 10.1214/aos/1059655913.
[11] J. Dippon and J. Renz, "Weighted means in stochastic approximation of minima," *SIAM J. Contr. Optim.*, vol. 35, no. 5, pp. 1811–1827, 1997, doi: 10.1137/S0363012995283789.
[12] C. S. Draper and Y. T. Li, *Principles of Optimalizing Control Systems and an Application to the Internal Combustion Engine*. New York, NY, USA: American Society of Mechanical Engineers, 1951.
[13] A. Durmus, E. Moulines, A. Naumov, S. Samsonov, K. Scaman, and H.-T. Wai, "Tight high probability bounds for linear stochastic approximation with fixed step-size," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, vol. 34, pp. 30,063–30,074.
[14] V. Fabian, "On the choice of design in stochastic approximation methods," *Ann. Math. Statist.*, vo. 39, no. 2, pp. 457–465, Apr. 1968, doi: 10.1214/aoms/1177698409.
[15] C. Kalil Lauand and S. Meyn, "Approaching quartic convergence rates for quasi-stochastic approximation with application to gradient-free optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds. Red Hook, NY, USA: Curran Associates, 2022, vol. 35, pp. 15,743–15,756.
[16] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *Ann. Math. Statist.*, vol. 23, no. 3, pp. 462–466, Sep. 1952, doi: 10.1214/aoms/1177729392.
[17] N. J. Killingsworth and M. Krstic, "PID tuning using extremum seeking: Online, model-free performance optimization," *IEEE Control Syst. Mag.*, vol. 26, no. 1, pp. 70–79, Feb. 2006, doi: 10.1109/MCS.2006.1580155.
[18] M. Krstić and H.-H. Wang, "Stability of extremum seeking feedback for general nonlinear dynamic systems," *Automatica*, vol. 36, no. 4, pp. 595–601, Apr. 2000, doi: 10.1016/S0005-1098(99)00183-1.
[19] B. Lapeybe, G. Pages, and K. Sab, "Sequences with low discrepancy generalisation and application to Bobbins-Monbo algorithm," *Statistics*, vol. 21, no. 2, pp. 251–272, 1990, doi: 10.1080/02331889008802246.
[20] J. Larson, M. Menickelly, and S. M. Wild, "Derivative-free optimization methods," *Acta Numer.*, vol. 28, pp. 287–404, Jun. 2019, doi: 10.1017/S0962492919000060.
[21] S. Laruelle and G. Pagès, "Stochastic approximation with averaging innovation applied to finance," *Monte Carlo Methods Appl.*, vol. 18, no. 1, pp. 1–51, Feb. 2012, doi: 10.1515/mcma-2011-0018.
[22] C. K. Lauand and S. Meyn, "Bias in stochastic approximation cannot be eliminated with averaging," in *Proc. Allerton Conf. Commun., Contr., Comput.*, Sep. 2022, pp. 1–4.
[23] C. K. Lauand and S. Meyn, "Extremely fast convergence rates for extremum seeking control with Polyak-Ruppert averaging," 2022, *arXiv:2206.00814*.
[24] C. K. Lauand and S. Meyn, "Markovian foundations for quasi stochastic approximation with applications to extremum seeking control," 2022, *arXiv:2207.06371*.
[25] M. Le Blanc, "Sur l'electrification des chemins de fer au moyen de courants alternatifs de frequence elevee [On the Electrification of Railways by Means of Alternating Currents of High Frequency]," *Revue Generale de l'Electricite*, vol. 12, no. 8, pp. 275–277, 1922.
[26] L. Lessard, "The analysis of optimization algorithms: A dissipativity approach," *IEEE Control Syst. Mag.*, vol. 42, no. 3, pp. 58–72, Jun. 2022, doi: 10.1109/MCS.2022.3157115.
[27] S. Liu and M. Krstic, "Introduction to extremum seeking," in *Stochastic Averaging and Stochastic Extremum Seeking, Communications and Control Engineering*. London, U.K.: Springer, 2012.
[28] J. C. Luxat and L. H. Lees, "Stability of peak-holding control systems," *IEEE Trans. Ind. Electron. Contr. Instrum.*, vol. IECI-18, no. 1, pp. 11–15, Feb. 1971, doi: 10.1109/TIECI.1971.230455.
[29] S. M. Meerkov, "Asymptotic methods for investigating a class of forced states in extremal systems," *Autom. Remote Contr.*, vol. 28, no. 12, pp. 1916–1920, 1967.
[30] P. G. Mehta and S. P. Meyn, "Q-learning and Pontryagin's minimum principle," in *Proc. Conf. Decis. Contr.*, Dec. 2009, pp. 3598–3605, doi: 10.1109/CDC.2009.5399753.
[31] M. Metivier and P. Priouret, "Théorèmes de convergence presque sure pour une classe d'algorithmes stochastiques à pas décroissant," *Probl. Theory Related Fields*, vol. 74, pp. 403–428, Sep. 1987, doi: 10.1007/BF00699098.
[32] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge, U.K: Cambridge Univ. Press, 2009.
[33] H. Mohammadi, M. Razaviyayn, and M. R. Jovanović, "Robustness of accelerated first-order algorithms for strongly convex optimization problems," *IEEE Trans. Autom. Control*, vol. 66, no. 6, pp. 2480–2495, Jun. 2021, doi: 10.1109/TAC.2020.3008297.
[34] V. Obabkov, "Theory of multichannel extremal control systems with sinusoidal probe signals," *Automat. Remote Contr.*, vol. 28, pp. 48–54, 1967.
[35] R. Pasupathy and S. Ghosh, "Simulation optimization: A concise overview and implementation guide," in *Proc. Theory Driven Influential Appl.*, 2013, pp. 122–150, doi: 10.1287/educ.2013.0118.
[36] B. T. Polyak, "Some methods of speeding up the convergence of iteration methods," *USSR Comput. Math. Math. Phys.*, vol. 4, no. 5, pp. 1–17, 1964, doi: 10.1016/0041-5553(64)90137-5.
[37] B. T. Polyak, "A new method of stochastic approximation type," *Avtomatika i telemekhanika* (in Russian) (Transl.: *Automat. Remote Contr.*), vol. 51, 1991, pp. 98–107.
[38] B. T. Polyak and A. B. Juditsky, "Acceleration of stochastic approximation by averaging," *SIAM J. Contr. Optim.*, vol. 30, no. 4, pp. 838–855, 1992, doi: 10.1137/0330046.
[39] B. T. Polyak and A. B. Tsybakov, "Optimal orders of accuracy for search algorithms of stochastic optimization," *Probl. Inf. Transmiss.*, vol. 26, no. 2, pp. 45–53, Oct. 1990.
[40] L. Rastrigin, "Extremum control by means of random scan," *Avtomatika i Telemekhanika*, vol. 21, no. 9, pp. 1264–1271, 1960.
[41] L. A. Rastrigin, "Random search in problems of optimization, identification and training of control systems," *J. Cybern.*, vol. 3, no. 3, pp. 93–103, 1973, doi: 10.1080/01969727308546050.
[42] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, vol. 22, pp. 400–407, 1951, doi: 10.1214/aoms/1177729586.
[43] D. Ruppert, "Efficient estimators from a slowly convergent Robbins-Monro processes," School Oper. Res. Ind. Eng., Cornell Univ., Ithaca, NY, USA, Tech. Rep. No. 781, 1988.
[44] S. Shirodkar and S. Meyn, "Quasi stochastic approximation," in *Proc. Amer. Contr. Conf.*, Jul. 2011, pp. 2429–2435, doi: 10.1109/ACC.2011.5991485.
[45] J. C. Spall, "Stochastic optimization," in *Handbook of Computational Statistics*, J. Gentle, W. Härdle, and Y. Mori, Eds. Heidelberg, Germany: Springer-Verlag, 2012, pp. 173–201.
[46] S. Surjanovic and D. Bingham. "Virtual library of simulation experiments: Test functions and datasets." Accessed: May 16, 2022. [Online]. Available: http://www.sfu.ca/~ssurjano/optimization.html
[47] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.
[48] Y. Tan, W. H. Moase, C. Manzie, D. Nešić, and I. Mareels, "Extremum seeking from 1922 to 2010," in *Proc. 29th IEEE Chin. Contr. Conf.*, 2010, pp. 14–26.
[49] Y. Tan, D. Nešić, and I. Mareels, "On non-local stability properties of extremum seeking control," *Automatica*, vol. 42, no. 6, pp. 889–903, Jun. 2006, doi: 10.1016/j.automatica.2006.01.014.
[50] A. Teel and D. Popovic, "Solving smooth and nonsmooth multivariable extremum seeking problems by the methods of nonlinear programming," in *Proc. Amer. Contr. Conf.*, Jun. 2001, vol. 3, pp. 2394–2399, doi: 10.1109/ACC.2001.946111.
[51] M. Vidyasagar, "A new converse Lyapunov theorem for global exponential stability and applications to stochastic approximation," in *Proc. IEEE 61st Conf. Decis. Contr. (CDC)*, 2022, pp. 2319–2321, doi: 10.1109/CDC51059.2022.9992831.
[52] M. Vidyasagar, "Convergence of stochastic approximation via martingale and converse Lyapunov methods," *Math. Contr., Signals, Syst.*, vol. 35, pp. 351–374, Jan. 2023, doi: 10.1007/s00498-023-00342-9.