Batch-to-Batch Finite-Horizon LQ Control for Unknown Discrete-Time Linear Systems Via Stochastic Extremum Seeking

Shu-Jun Liu, Miroslav Krstic, and Tamer Başar

Abstract—We employ our recent discrete-time stochastic averaging theorems and stochastic extremum seeking to iteratively (batch-to-batch) optimize open-loop control sequences for unknown but reachable discrete-time linear systems with a scalar input and without known system dimension, for a cost that is quadratic in the measurable output and the input. First, for a multivariable discrete-time Newton-based stochastic extremum seeking algorithm we prove local exponential convergence to the optimal open-loop control sequence. Second, to remove the convergence rate’s dependence on the Hessian matrix of the cost function, which is unknown since the system’s model (the system matrices \( (A, B, C) \)) is unknown, we develop a multivariable discrete-time Newton-based stochastic extremum seeking method, design the Newton-based algorithm for the iteration of the input sequence, and prove local exponential convergence to the optimal input sequence. Finally, two simulation examples are given to illustrate the effectiveness of the two methods.

Index Terms—LQ control, stochastic averaging, stochastic extremum seeking.

I. INTRODUCTION

Extremum seeking (ES), a non-model-based method for real-time optimization, has inspired intensive research activity since the emergence of its first stability proof [17]. According to the type of probing signals, extremum seeking methods can be categorized as deterministic ES [2], [7], [34], [37], [28], [36] and stochastic ES [26], [22], [24]. Reference [37] provides non-local stability analysis of extremum seeking, whereas [28] proposes a framework for continuous-time extremum seeking control of systems with parameter uncertainties. [36] presents an extremum seeking based iterative learning model predictive control method under states and inputs constraints. Analysis of ES based on Lie brackets is presented in [9]. Although periodic perturbation signals dominate deterministic ES, stochastic measurement noise corrupting an ES algorithm that uses sinusoidal perturbations has been considered in a stochastic approximation-like framework in [34]. The first stochastic version of ES was developed in the discrete-time setting [26]. In [22] and [24], we have proposed continuous-time stochastic extremum seeking algorithms based on gradient and Newton algorithms, respectively. Examples of applications of ES can be found in [3], [40], [31], [35], [27], [23] and [33].

For the optimal control of partially known systems, there are many results based on learning and dynamic programming methods (e.g., [8], [19], [20], [21],[30], [38], [39]). Reference [21] gives a survey about the application of reinforcement learning and adaptive dynamic programming for feedback control. In [20], feedback policy iteration and value iteration algorithms are given for infinite-horizon optimal control of discrete-time linear systems of known order. Based on three neural networks, a near-optimal feedback controller is developed for unknown nonaffine nonlinear discrete-time systems with a discounting factor in the cost function in [38]. By online system identification and offline optimal control training via neural networks, an asymptotically optimal control law is designed for unknown affine nonlinear discrete-time systems in [39]. For continuous-time unknown systems, adaptive asymptotically optimal control algorithms based policy iteration are given in [30].

In contrast to these feedback regulation results with an infinite-horizon cost function and some known system information (e.g., known system order), we focus on batch-to-batch open-loop control for unknown but reachable discrete-time linear systems with only output measurement and corrupted by noise. Batch-to-batch means that we design an iterative learning algorithm where experiments are repeated with the same initial conditions ([13], [4]). We present two approaches (gradient-based and Newton-based) for iteratively solving a finite-horizon optimal control problem for a completely unknown discrete-time linear system with a scalar input by using a non-model based, stochastic extremum seeking algorithm. First, for the finite-horizon LQ control problem, when the cost function’s measurement or calculation is possibly contaminated with noise, we design an iteration algorithm via gradient-based stochastic extremum seeking to find the optimal control sequence and prove that the algorithm locally exponentially converges to the optimal open-loop control sequence. Then, to remove the convergence rate’s dependence on the unknown Hessian, we develop a discrete-time Newton-based stochastic extremum seeking approach, design a Newton-based iteration algorithm for the LQ problem, and prove its local exponential convergence. Finally, two simulation examples are given to illustrate the effectiveness of the two methods.

We extend the method of [10] to the stochastic case by using our recently developed discrete-time stochastic extremum seeking [25]. The contributions of this technical note and advantages of the stochastic approach are: (1) the Newton algorithm offers considerable advantages for LQ problems with longer horizons; (2) incorporation of the measurement noise for the cost function; (3) independent stochastic perturbation signals in stochastic ES are easier to choose than “sufficiently spaced” perturbation frequencies in deterministic ES.
The differences with the previous work on optimal control-based learning and adaptive dynamic programming are: (i) we study open-loop control with finite-horizon cost functions when the model structure \((A, B, C)\) and the system’s dimension are unknown; (ii) we prove the convergence of the control sequence without making the persistent excitation assumption (e.g., [5]) or discounting the cost function (e.g., [20], [39], [38]); (iii) we simultaneously consider the output measurement noise and the dither signal as stochastic perturbations, thanks to our new discrete-time stochastic averaging theory [25].

Due to the batch-to-batch/iterative nature of the present work, it may be confused with iterative learning control (ILC). We explain the differences next in (a)–(d).

a) The goal of ILC is to track a desired reference trajectory ([16]), while ES is based on a cost function to be optimized. In [11], [18], a quadratic cost function of the tracking error and the distance of two
quent to our article, deal with ES-based ILC (the latter one deals with
continuous-time nonlinear plants with periodic sampled-data inputs,
without focusing on a specific method of gradient estimation).

b) Most optimization-based ILC rely on model knowledge, while ES
is non-model based. In [12] and [32], the updating control laws as well
as convergence of the ILC methods depend on the precise knowledge
of the nominal model.

c) ILC can be approached with ES, but the converse is not necessarily
the case. In particular, [15], [16], both developed and submitted subse-
quent to our article, deal with ES-based ILC (the latter one deals with
continuous-time nonlinear plants with periodic sampled-data inputs,
without focusing on a specific method of gradient estimation).

d) Our convergence analysis is different than those for ILC, which
are generally dependent on the system (e.g., transfer function or control
coefficient, see [29], [6]) and in the case of noisy measurement, it is hard
to give the convergence analysis. In [11], which is an ILC algorithm
without requiring detailed knowledge of the system that is the closest
to our work, no convergence is analyzed for the LQ problem. This is
the biggest difference between [11] and our work. Besides, the ILC
algorithm in [11] limits the control input to be in the space spanned by
a smaller number of basis functions (or vectors), needs training data
to identify the conjugate basis vectors, and then computes the control
input, while our method is to directly design the control input sequence
based on output measurement without training data, and especially, we
guarantee convergence by design parameters of the ES algorithm.

Organization of the Paper: We state the problem in Section II, in-
roduce a gradient-based stochastic algorithm in Section III, introduce
a Newton-based algorithm in Section IV, and give simulation examples
in Section V, closing with conclusions in Section VI.

II. PROBLEM STATEMENT

Consider the single-input linear discrete-time system

\[
x_{k+1} = A_k x_k + B_k u_k, \tag{1}
\]

\[
y_k = C_k x_k, \tag{2}
\]

where \(x_k \in \mathbb{R}^n, u_k \in \mathbb{R}, \) and \(y_k \in \mathbb{R}^p\), denote the state, control input, and output, respectively, and \(A_k, B_k, C_k\) denote unknown matrices of appropriate dimensions at discrete time \(k, k = 0, 1, 2, \ldots\). The state dimension may be unknown and the only available information is the output \(y_k\) and the scalar input \(u_k\).

Our objective is to find the optimal, open-loop control sequence \(\{u_k\}_{k=0}^{N-1}\) that minimizes the cost function

\[
J(u) = \frac{1}{2} y_k^T \hat{Q} y_N + \frac{1}{2} \sum_{k=0}^{N-1} \left( y_k^T \hat{Q}_k y_k + R_k u_k^2 \right), \tag{3}
\]

where \(u = [u_0, \ldots, u_{N-1}]^T, \) and \(Q_k, R_k \geq 0, R_k > 0 \) for all \(k \in \{0, \ldots, N-1\}.\) Namely, we want to solve the discrete-time, finite-horizon LQ optimal control problem,

\[
\min_{u \in \mathbb{R}} J(u), \text{ subject to } (1) - (2) \tag{4}
\]

with initial condition \(x_0 \in \mathbb{R}^n.\) We seek an open-loop solution instead of an optimal state feedback policy since the system is unknown and state information is not available. The cost function (3) can be written in terms of the state \(x_k\) as

\[
J(u) = \frac{1}{2} x_N^T Q_N x_N + \frac{1}{2} \sum_{k=0}^{N-1} \left( x_k^T Q_k x_k + R_k u_k^2 \right), \tag{5}
\]

where \(Q_k = C_k^T \hat{Q}_k C_k, k \in \{0, \ldots, N\}, \) which is the standard cost function for the LQ optimal control problem with a positive weight on control at each stage. Different from the deterministic method [10], we consider here the situation when the cost function \(J(u)\) in (3) is disturbed by some noise, which may be the result of (i) the output measurement being corrupted by noise or (ii) unknown initial conditions. For this LQ optimal control problem with the disturbed cost function, we will use the stochastic extremum seeking method developed in [25] to find the optimal control sequence.

By solving for the system’s state trajectory from (1), we have

\[
x_k = \Phi_{10} x_0 + \sum_{j=0}^{k-1} \Phi_{k,j+1} B_j u_j, \tag{6}
\]

where

\[
\Phi_{ij} = \begin{cases} 1, & i > j \\ A_{i-1} A_{i-2} \cdots A_j, & i = j, \end{cases}
\]

Thus the cost function (5) can be written in terms of only the initial state \(x_0\) and the control \(u\), namely,

\[
J(u) = \frac{1}{2} x_0^T Q x_0 + x_0^T G u + \frac{1}{2} u^T H u, \tag{7}
\]

where \(Q = Q_k + \Phi^T Q \Phi, G = \Phi^T Q A, H = A^T Q A + R, \Phi = [\Phi_{10}, \ldots, \Phi_{N0}]^T, Q = \text{diag}[Q_1, \ldots, Q_N], R = \text{diag}[R_0, \ldots, R_{N-1}],\)

\[
\Lambda = \begin{bmatrix} B_0 & 0 & 0 & \cdots & 0 \\ \Phi_{21} B_0 & B_1 & 0 & \cdots & 0 \\ \Phi_{31} B_0 & \Phi_{32} B_1 & B_2 & \cdots & 0 \\ & \vdots & \ddots & \ddots & \ddots \\ \Phi_{N1} B_0 & \Phi_{N2} B_1 & \Phi_{N3} B_2 & \cdots & B_{N-1} \end{bmatrix},
\]

and \(\text{diag}[\cdot]\) denotes a block diagonal matrix. The Hessian \(H\) is clearly positive definite symmetric being the sum of \(\Lambda^T Q A \Lambda \geq 0\) and \(R > 0.\) If the system model were known and there were no noise in the measurement of the cost function, the optimal open-loop control values could be found directly by solving the corresponding quadratic program.

III. OPTIMAL CONTROL VIA STOCHASTIC EXTREMUM SEEKING

As mentioned earlier, we allow for the possibility that the measurement or calculation of the cost function may be disturbed by noise or error. The stochastic extremum seeking controller is given in vector
where \( \hat{u} = [\hat{u}_0, \ldots, \hat{u}_N] \) is the optimal control sequence that minimizes the cost function \( J(u) \). Substituting (14) and (15) into (8)–(9) yields the error system as

\[
\begin{align*}
\hat{u}(l + 1) - \hat{u}(l) &= -\varepsilon KM(l + 1) (\hat{J}(l) - \hat{\xi}(l)), \\
\hat{\xi}(l + 1) - \hat{\xi}(l) &= -\varepsilon h \hat{\xi}(l) + \varepsilon h \hat{J}(l), \\
\hat{u}(l) &= \hat{u}(l) + S(l + 1),
\end{align*}
\]

for \( k = 0, 1, \ldots, N - 1 \), where for \( \varepsilon > 0 \) is a small parameter, \( K \) is a positive diagonal matrix, \( h > 0 \) is a design parameter, \( \hat{J}(l) \) is the \( l \)th step measured cost function with additive noise \( W(l + 1) \), which is caused from the measurement noise \( \nu(l + 1) \), the stochastic excitation signals \( M(\cdot) \) and \( S(\cdot) \) are given by

\[
M(l) = \begin{bmatrix}
\frac{1}{a_0 G_0} \sin (v_0(l)), \ldots, \\
\frac{1}{a_{N-1} G_{N-1}} \sin (v_{N-1}(l))
\end{bmatrix}^T,
\]

\[
S(l) = [a_0 \sin (v_0(l)), \ldots, a_{N-1} \sin (v_{N-1}(l))]^T,
\]

and we note that \( \nabla J(u^*) \) is zero. Denote the error relative to the optimal control sequence as

\[
\tilde{u}(l) = u(l) - S(l + 1) - u^*;
\]

and substitute into (13) to obtain

\[
\begin{align*}
J(l) &= J(\tilde{u}(l)) = J(u^*) + \frac{1}{2} (\tilde{u}(l))^T H \tilde{u}(l) + S(l + 1)^T H S(l + 1) \\
&+ \frac{1}{2} (\tilde{u}(l))^T H S(l + 1) + \frac{1}{2} S(l + 1)^T H S(l + 1) + W(l + 1) - \xi(l)),
\end{align*}
\]

To calculate the average system of the error dynamics (16)–(17), we assume that the following noise condition holds:

**Assumption 1:** The measurement noise process \( W(\cdot) \) is a bounded ergodic sequence with invariant distribution \( \nu \) with zero mean, which is independent of stochastic sequence \( v_k(\cdot) \) for any \( k = 0, 1, \ldots, N - 1 \). Since

\[
\text{Ave} [\sin (v_k(\cdot))] \triangleq \int_R \sin (y) \mu_k (dy) = 0,
\]

\[
\text{Ave} [\sin^3 (v_k(\cdot))] \triangleq \int_R \sin^3 (y) \mu_k (dy) = 0,
\]

\[
\text{Ave} [\sin^2 (v_k(\cdot))] \triangleq \int_R \sin^2 (y) \mu_k (dy) = 0,
\]

\[
\text{Ave} [\sin (v_i(\cdot)) \sin (v_j(\cdot))] \triangleq \int_R \sin (y) \sin (z) \mu_i (dy) \times \mu_j (dz) = 0, \quad i \neq j,
\]

\[
\text{Ave} [\sin^2 (v_i(\cdot)) \sin (v_j(\cdot))] \triangleq \int_R \sin^2 (y) \sin (z) \mu_i (dy) \times \mu_j (dz) = 0, \quad i \neq j,
\]

\[
\text{Ave} [\sin (v_i(\cdot)) \sin (v_j(\cdot)) \sin (v_k(\cdot))] \triangleq \int_R \sin (x) \sin (y) \sin (z) \mu_i (dx) \times \mu_j (dy) \times \mu_k (dz) = 0, \quad i \neq j, k \neq j, k \neq j,
\]

\[
\text{Ave} [\sin (v_i(\cdot)) W(\cdot)] \triangleq \int_R \sin (y) \mu_i (dy) \int_R z \nu (dz) = 0,
\]

\[
\text{Ave} [W(\cdot)] = \int_R \nu (dy) = 0,
\]

we have

\[
\text{Ave} [M(l + 1)(J(u^*) + \frac{1}{2} (\tilde{u}(l))^T H \tilde{u}(l) + \frac{1}{2} S(l + 1)^T H S(l + 1) + W(l + 1) - \xi(l))] = 0,
\]

\[
\text{Ave} [M(l + 1) S(l + 1)^T H S(l + 1)] = I,
\]

\[
\text{Ave} [S(l + 1)^T H \tilde{u}(l)] = 0,
\]

\[
\text{Ave} [\frac{1}{2} S(l + 1)^T H S(l + 1)] = \frac{1}{2} \sum_{k=0}^{N-1} a_k^2 G_k (\sigma_k) H_{k+k},
\]

where for \( k = 0, \ldots, N - 1 \), \( H_{k+k} \) is the \( k \)th diagonal element of the matrix \( H \). Thus we obtain the average system of the error system (16)–(17) as

\[
\begin{align*}
\tilde{u}_{av}(l + 1) - \tilde{u}_{av}(l) &= -\varepsilon K H \tilde{u}_{av}(l), \\
\xi_{av}(l + 1) - \xi_{av}(l) &= -eh \xi_{av}(l) + eh (J(u^*)) \\
&\quad + \frac{1}{2} (\tilde{u}_{av}(l))^T H \tilde{u}_{av} + \frac{1}{2} \sum_{k=0}^{N-1} a_k^2 G_k (\sigma_k) H_{k+k},
\end{align*}
\]
whose equilibrium is $(\hat{u}^{ave}, \xi^{ave}) = (0, J(u^*)) + \frac{1}{2} \sum_{k=0}^{N-1} \alpha_k^2 G_k(\sigma_k) H_{kk}$). Its Jacobian matrix is $Jac^{ave} = \left[ 1 - \epsilon \hat{K} H \right]$. For sufficiently small $\epsilon$, i.e., $\epsilon \in (0, \frac{1}{2} \gamma \frac{1}{\max(\|H_{kk}\|, 1)} )$, we have that the equilibrium of the average system is locally exponentially stable. By the averaging theorem (Theorem 9 in [25]), we have the following result.

**Theorem 1:** Consider the finite-horizon optimal control problem (4) under the extremum seeking controller (8)–(9). If the measurement noise process $W(\cdot)$ satisfies Assumption 1, then the error system temporally exponentially converges to a neighborhood of the equilibrium, i.e., there exist constants $r > 0$, $c_r > 0$, and $0 < \gamma_e < 1$ such that for any initial condition $|\Lambda_i(0)| < r$ and any $\gamma$, we have

$$\lim_{\gamma_r \to 0} \inf \{ t \in N : |\Lambda_i(t)| > c_r |\Lambda_i(0)|^{(q-r)} \} = +\infty \text{ a.s.},$$

and

$$\lim_{\gamma_r \to 0} P \{ |\Lambda_i(t)| \leq c_r |\Lambda_i(0)|^{(q-r)} \} = 1,$$

where $\Lambda_i(0) = (\hat{u}^{ave}, \xi^{ave}) - J(u^*) + \frac{1}{2} \sum_{k=0}^{N-1} \alpha_k^2 G_k(\sigma_k) H_{kk}$, and $N_0$ is any natural number.

Here the convergence properties are a kind of weak exponential convergence in the almost sure sense and in probability. The term “weak” because the properties involve $\lim_{\gamma_r \to 0}$ and given through the first exist time from a set. These exponential convergences imply that the norm of the estimation error vector $\hat{u}(t)$ exponentially converges, both almost surely and in probability, to below an arbitrarily small residual value $\gamma$, over an arbitrarily long time interval, which tends to infinity as $\gamma$ goes to zero. Thus the limit of the sequence $\{\hat{u}(t)\}$ is the approximation of the optimal control solution we sought.

The estimate of the region of attraction $r$ can be conservatively taken as independent of the $\gamma$, for $\gamma$ chosen sufficiently small. This fact can be seen by going through the proof of the exponential stability theorem (Theorem 4.13 of [14]), from which discrete-time exponential stability can be derived easily for discrete-time nonlinear systems. The convergence rate $\gamma_e$ is determined by the small parameter $\gamma$, the gain matrix $K$, the Hessian matrix $H$ of the cost function, and the filter parameter $h$.

**IV. NEWTON-METHOD BASED CONTROLLER DESIGN**

In the gradient-based method, we have seen that the convergence rate depends on the Hessian matrix $H$. In this section, to remove this dependence, we present a Newton-based ES algorithm shown in Fig. 2.

The Newton-based algorithm is constructed as follows:

$$\hat{u}(l + 1) - \hat{u}(l) = -\epsilon K \hat{G}(l) H \hat{u}(l) - \epsilon K \hat{u}_{ave}(l),$$

$$\hat{G}(l + 1) - \hat{G}(l) = -\epsilon \beta \hat{G}(l) H \hat{u}(l) - \epsilon \beta \hat{u}_{ave}(l),$$

$$\hat{\xi}(l + 1) - \hat{\xi}(l) = -\epsilon h \hat{\xi}(l) + \epsilon h \hat{u}(l),$$

$$u(l) = \hat{u}(l) + S(l + 1),$$

where $M(\cdot)$ and $S(\cdot)$ are given as (11) and (12), the excitation process $v(l)$, and the measurement noise process $W(\cdot)$ are the same as in the gradient-based method, $\beta > 0$ is a constant, the $N \times N$ matrix $\Xi$ is chosen as $\Xi_i(l) = \frac{1}{\alpha_i^2} G_i(2\sigma_i) \left[ \frac{\sin^2(v(l)) - G_i(\sigma_i)}{\sin(v(l)) \sin(v(l))} \right]$, $i \neq j$, $i = 0, \ldots, N - 1$, and $G_i(\sigma_i) = \frac{1}{2} (1 - e^{-2\pi \sigma_i^2})$.

Define the error variables as $\hat{u}(l) = u(l) - S(l + 1) - u^*, \hat{G}(l) = G(l) - H^{-1}, \hat{\xi}(l) = \xi(l) - J(u^*)$. Then by (15), we have the error dynamics as

$$\begin{align*}
\hat{u}(l + 1) - \hat{u}(l) & = -\epsilon K \hat{G}(l) H \hat{u}(l) - \epsilon K \hat{u}_{ave}(l), \\
\hat{G}(l + 1) - \hat{G}(l) & = -\epsilon \beta \hat{G}(l) H \hat{u}(l) - \epsilon \beta \hat{u}_{ave}(l), \\
\hat{\xi}(l + 1) - \hat{\xi}(l) & = -\epsilon h \hat{\xi}(l) + \epsilon h \hat{u}(l).
\end{align*}$$

By (18)–(25), we have the average error system as follows

$$\begin{align*}
\hat{u}_{ave}(l + 1) - \hat{u}_{ave}(l) & = -\epsilon K \hat{G}_{ave}(l) H \hat{u}_{ave}(l) - \epsilon K \hat{u}_{ave}(l), \\
\hat{G}_{ave}(l + 1) - \hat{G}_{ave}(l) & = -\epsilon \beta \hat{G}_{ave}(l) H \hat{u}_{ave}(l) - \epsilon \beta \hat{u}_{ave}(l), \\
\hat{\xi}_{ave}(l + 1) - \hat{\xi}_{ave}(l) & = -\epsilon h \hat{\xi}_{ave}(l) + \epsilon h \left( \frac{1}{2} \hat{u}_{ave}(l) \right),
\end{align*}$$

for which $(\hat{u}_{ave}, \hat{G}_{ave}, \hat{\xi}_{ave}) = (0, 0, \frac{1}{2} \sum_{k=0}^{N-1} \alpha_k^2 G_k(\sigma_k) H_{kk})$ is an equilibrium. Then the Jacobian matrix at the equilibrium $(0, 0, \frac{1}{2} \sum_{k=0}^{N-1} \alpha_k^2 G_k(\sigma_k) H_{kk})$ is
conditions are guaranteed by conservatively taking of attraction hard to manifest the relationship between the estimate of the region where 
\[ h \]
which implies that the equilibrium \( (0,0) \) is locally exponentially stable. Thus by the averaging theorem (Theorem 9 in [25]), we have the following result.

**Theorem 2:** Consider the finite-horizon optimal control problem (4) under the Newton-based extremum seeking scheme (30)–(32). If the measurement noise process \( W(\cdot) \) satisfies Assumption 1, then there exist constants \( r > 0, c_\varepsilon > 0, 0 < \gamma_\varepsilon < 1 \) such that for any initial condition \( |A_\varepsilon^r(0)| < r \) and any \( \delta > 0 \),

\[
\lim_{\varepsilon \to 0} \inf \left\{ l \in N : |A_\varepsilon^r(l)| > c_\varepsilon |A_\varepsilon^r(0)(\gamma_\varepsilon)^l + \delta \right\} = \infty, \text{ a.s.}
\]

and

\[
\lim_{\varepsilon \to 0} P \left\{ |A_\varepsilon^r(l)| \leq c_\varepsilon |A_\varepsilon^r(0)|(|\gamma_\varepsilon|^l + \delta), \forall l = 0, 1, \ldots, \lfloor N_0/\varepsilon \rfloor \} = 1,
\]

where \( A_\varepsilon^r(l) \triangleq \langle a^T(l), \mathcal{C}(\Gamma(l)), \xi - \frac{1}{2} \sum_{k=0}^{N-1} a_k^T G_k(\sigma_k) H_k k \rangle, N_0 \) is any natural number, \( \mathcal{C}(A) \triangleq \langle A_1^T, \ldots, A_{N_0}^T \rangle \), and \( A_i, i = 1, \ldots, l \), denotes the column vector of any matrix \( A \in \mathbb{R}^{n \times n} \).

From the Jacobi matrix II and averaging theorem (Theorem 9 in [25]), we can see that the convergence rate \( \gamma_\varepsilon \) is determined solely by the design parameters: small parameter \( \varepsilon \), the gain matrix \( K \), estimate gain \( \beta \) of the inverse of the Hessian matrix and the filter parameter \( h \). Thus it is independent of the unknown Hessian matrix \( H \). It is hard to manifest the relationship between the estimate of the region of attraction \( r \) and the design parameter according to the exponential stability theorem (Theorem 4.13 of [14]), but the existence can be guaranteed by conservatively taking \( r \) sufficiently small.

**V. SIMULATION**

**A. Scalar System**

We consider the system \( x_{k+1} = 2 \sin(k) x_k + u_k \), where \( x_k \in \mathbb{R} \) and \( u_k \in \mathbb{R} \). The linear quadratic cost function is chosen as \( J(u) = \frac{1}{2} x_k^2 + \frac{1}{2} \sum_{k=0}^{-1} \langle x_k^2 + u_k^2 \rangle \).

Fig. 3 depicts some simulation results for discrete-time stochastic ES where the measurement noise \( W(\cdot) \) is assumed as the sequence of independent Gaussian random variables with identical distribution \( N(0, 0.02^2) \). The design parameters are \( h = 1, \sigma_k \equiv 2, \varepsilon = 0.2, a = [0.5, 0.5, 0.5]^T, K_0 = \text{diag}[1/1500, 1/1000, 1/500], \) and the initial conditions are \( \hat{w}_0(0) = [0, 0, 0]^T, \xi(0) = 1, \) and \( x_0 = 1 \).

For a comparison with the gradient-based algorithm (a) of Fig. 3, the same parameters except the gain matrix are used in the Newton-based algorithm (b) of Fig. 3. The selection of \( K_0 \) and \( \Gamma(0) \) for (b) of Fig. 3 is made such that the initial gain \( K_0, \Gamma(0) \) is the same as the gain \( K_0 \) used in the gradient-based design. We choose parameters \( K_0 = \text{diag}[2, 2, 2] \times 10^{-2}, \beta = 0.02 \) and the initial condition \( \Gamma(0) = \text{diag}[1/30, 1/20, 1/10] \) in the Newton-based algorithm.

It can be seen from Fig. 3 that the iteration sequence \( \hat{u}_k(l) \) can move to a small neighborhood of the optimal control \( (u_0^*, u_1^*, u_2^*) = (-1.23, -0.42, -0.06)^T \) for both algorithms. In Fig. 3, (a)-1 and (a)-2 are the same simulation over different numbers of iterations, from which we can see that the iteration sequence in the Newton-based algorithm converges after about 2000 iterations, while in the gradient-based algorithm, it needs about 15000 iterations, which is caused by the selection of design parameters in the Newton-based algorithm.

Fig. 4 shows the evolution of the Hessian matrix estimator \( \left( \Gamma(l) \right)^{-1} \) in the Newton-based algorithm. It can be seen that the components of \( \left( \Gamma(l) \right)^{-1} \) can approach a small neighborhood of the corresponding components of the Hessian matrix \( H \), and components of \( \left( \Gamma(l) \right)^{-1} \) (after about 1000 iterations) converge more quickly than the iteration sequence \( \hat{u}_k(l) \) (after about 2000 iterations).

Fig. 5 shows the convergence properties of the corresponding iteration sequences in the two algorithms. It is shown that the
shown that the cost function sequence can approach the neighborhood. Fig. 6 shows the evolution of the cost function for the two cases. It is because the initial condition $u_0(0) = 0$ is too close to the optimal value $u_* = -0.06$ and the stochastic probing signal with persistent excitation makes the trajectory fluctuation near the optimal value.

Fig. 6 shows the cost function of stochastic ES algorithms for the scalar system: (a) gradient-based, and (b) Newton-based.

Newton-based algorithm can move the iteration sequence $\hat{u}_k(l)$ to the optimal value faster than the gradient-based algorithm, in which the iteration sequence converges over a larger number of iteration steps (see Fig. 3, (a)-1). The larger fluctuation in $\hat{u}_k$ of Fig. 5 is because the initial condition $\hat{u}_0(0) = 0$ is too close to the optimal value $\hat{u}_* = -0.06$ and the stochastic probing signal with persistent excitation makes the trajectory fluctuation near the optimal value.

Fig. 7 displays simulation results for this third-order system. For the gradient-based and Newton-based algorithms, the common design parameter values are $h = 0.5$, $\sigma_k \equiv 2$, $\varepsilon = 0.134$, $a = [0.5, 0.5, 0.5, 0.5, 0.5]^T$, and the initial conditions are $\hat{u}(0) = [0, 0, 0, 0, 0]^T$ and $\xi(0) = 1$. The parameters that are different for the two algorithms are $K = \text{diag}[1/2500, 1/2000, 1/1500, 1/1000, 1/500]$ for the gradient-based algorithm, and $K = \text{diag}[0.02, 0.02, 0.02, 0.02, 0.02]$ for the Newton-based algorithm.

From Fig. 7, we can see that the input sequence $\hat{u}_k(l)$ in both algorithms enters a small neighborhood of the optimal control $(u_0^*, u_1^*, u_2^*, u_3^*)^T = (1.206, 0.757, 0.414, 0.197, 0.026)^T$. It can be seen from Fig. 8 that the components of the Hessian matrix estimator $(\Gamma(l))^{-1}$ go to a small neighborhood of the corresponding components of the Hessian matrix, and as in the scalar system, we can see that the components of $(\Gamma(l))^{-1}$ converge faster than the iteration sequence $\hat{u}_k(l)$. The oscillation from the optimal control is related to trade-offs in the choices of parameters. In the practical applications, the tradeoff between the convergence rate and the convergence precision should be considered by the concrete optimization demands.

From Fig. 9, we can see that with the Newton-based algorithm $\hat{u}_k(l)$, $i = 0, \ldots, 3$ converges to the optimal value much faster than the gradient-based algorithm. Since the convergence rate of the Newton-based algorithm can be regulated by more parameters $\beta$, initial condition $\Gamma(0)$, and the variance $\sigma_k$ of stochastic excitation signal, the Newton-based algorithm can achieve better convergence than the gradient-based algorithm. Generally, if the initial condition $\Gamma^{-1}(0)$ is chosen near the corresponding components of $H$, the input sequence can converge with fewer iterations, and if $\Gamma^{-1}(0)$ is taken far from the corresponding components of $H$, the convergence can still be obtained but requires a large number of iterations. The bigger variance $\sigma_k$ of stochastic excitation signal can make the convergence quicker because of a stronger excitation, but makes the fluctuation of the trajectory bigger. Moreover, the bigger fluctuation of $\hat{u}_k/l$, $i = 3, 4$ in Fig. 9 is because the initial condition $\hat{u}_k(0), \hat{u}_k(0) = (0, 0)$ is too close to the optimal value $(u_0^*, u_1^*) = (0.197, 0.026)$ and the stochastic probing signal makes the trajectory fluctuate near the optimal value.

Fig. 10 shows the evolution of the cost function for the two cases. It is shown that the cost function sequence can approach the neighborhood of the optimal value $J^* = 11.32$. 

B. Third-Order System

Consider the following system

$$
x_{k+1} = \begin{bmatrix} 0.5 & 0.5 & 0.3 \\ 0.6 & 0.5 & 0.2 \\ 0.5 & 0.5 & 0.75 \end{bmatrix} x_k + \begin{bmatrix} -0.15 \\ 0.5 \\ 0.75 \end{bmatrix} u_k,
$$

$$
y_k = \begin{bmatrix} 1 & -0.15 & 0 \\ -1.5 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} x_k, \text{ with initial condition } x_0 = [-1, -0.5, 0.1]^T.
$$

The system’s eigenvalues are $1.2207, -0.1062, 0.3856$. The cost function is $J = \frac{1}{2} x_N y_N^T y_N + \sum_{k=0}^{N-1} (r_k y_k^T y_k + 5u_k^2)$ with $N = 5, r_1 = 3, r_2 = 5, r_3 = 7, r_4 = 9, r_5 = 11, r_6 = 1.$

Fig. 7 displays simulation results for this third-order system. For the gradient-based and Newton-based algorithms, the common
VI. CONCLUSION

We have presented two optimal open-loop controller design algorithms (gradient-based and Newton-based) for unknown discrete-time linear systems via the stochastic extremum seeking method. With our recently developed discrete-time averaging theory, we have proven the local exponential convergence of both algorithms. Since currently only local convergence can be proved, the choice of the initial condition is still a challenge. The convergence in the current framework of stochastic extremum seeking algorithm is in the “weak” sense given through the first exit time from a set. In practice, convergence with probability one is more attractive. Thus in our future work, we plan to investigate other types of convergence for the stochastic extremum seeking algorithm.

REFERENCES


