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

Metaheuristics-Based Approximation of Two-Dimensional Probability Distributions for Stochastic Systems Control

Kenta Osaki *, **Yohei Hosoe** *, and **Tomomichi Hagiwara** *

* *Department of Electrical Engineering, Kyoto University, Kyoto, Japan (e-mail: {osaki@jaguar., hosoe@, hagiwra@}kuee.kyoto-u.ac.jp)*

Abstract: This paper is concerned with approximation of probability distributions behind discrete-time systems with stochastic dynamics. The aim is to facilitate advances in studies for control of such stochastic systems. An approach to approximating two-dimensional probability distributions using metaheuristics is suggested. Then, its effectiveness is demonstrated with a numerical example of stabilization synthesis.

Keywords: Stochastic systems, Approximation of probability distributions, Metaheuristics, Kolmogorov distance, Discrete distribution, Linear matrix inequality (LMI)

1. INTRODUCTION

This paper studies metaheuristics-based approximation of two-dimensional probability distributions as a step toward applications in control problems for discrete-time systems with stochastic dynamics determined by an independent and identically distributed (i.i.d.) process. The concept of randomness is common in various fields, and developing a basis of taking it into account in control is an important issue. For example, randomness in packet interarrival times (Paxson and Floyd, 1995) could affect control performance of networked systems when the sampling time is affected by it and becomes aperiodic (Hetel et al., 2017). When we consider using information of such randomness in controller synthesis, we may have to search for controller parameters satisfying some inequality conditions involving random variables (representing the randomness). The search, however, becomes generally not simple since each random variable could take a value among infinitely many possible values; this issue will be revisited later for making our point clearer. To circumvent this issue, we consider approximating a given probability distribution underlying stochastic systems by a discrete distribution with finite support, which is theoretically and numerically tractable.

If the source of randomness is single and if we only have to deal with one-dimensional probability distribution, then such approximation could be relatively simple. For example, in the case of evaluating the accuracy of approximation by the Kolmogorov distance (Rachev et al., 2013), the corresponding optimal approximation can be analytically calculated without loss of generality (Hosoe et al., 2019). Instead of the Kolmogorov distance, other distances of probability distributions such as the Kantorovich distance (Rachev et al., 2013) can be also used in a simple fashion for the evaluation. In the case of multi-dimensional probability distributions, however, the situation is totally different, and the techniques used in the one-dimensional approximation do not work, in general. For example, even when each of two independent random variables is approximated by a discrete distribution with finite support (having a fixed number of possible values) optimally in the sense of the Kolmogorov distance, the joint distribution consisting of the two discrete distributions does not generally become optimal for the original two-dimensional distribution (i.e., about the two independent random variables) under the restricted support; this is true even in the simplest case of two uniform distributions. Hence, we need to find another way of approximation when the distributions are multi-dimensional. Our idea here is the use of metaheuristics for tackling the approximation problem viewed as an optimization problem in terms of the Kolmogorov distance. In particular, we deal with a genetic algorithm (GA) (Goldberg, 1989; Schwefel, 1995; Reeves, 2010) and simulated annealing (SA) (Van Laarhoven and Aarts, 1987; Schwefel, 1995; Nikolaev and Jacobson, 2010) for approximating two-dimensional probability distributions. Approximation of distributions is not an issue unique to the field of automatic control, and similar problems have been studied, e.g., in the field of finance (Pflug and Pichler, 2011). The contributions of this paper compared to such earlier studies are as follows: (i) to associate the present approximation problem (which is a part of the control problem) with GA and SA, and (ii) to show effectiveness of our approximation methods using GA and SA in a numerical example of stabilization synthesis.

To state the motivation of this study, and to confirm the effectiveness of the proposed approaches easily, this paper will basically deal only with linear systems with stochastic dynamics. However, the approximation itself is irrelevant to the system linearity, and the same approach can be used for nonlinear systems as long as the dynamics is determined by an i.i.d. process. In addition, there is no theoretical restriction on the dimension of probability distributions in our approximation approach, and higher-dimensional distributions are expected to be tractable in a similar fashion.

Copyright lies with the authors

This paper is organized as follows. Section 2 reviews existing results on the stabilization synthesis in the linear case and state the motivation of this study in detail. Then, Section 3 discusses our main results on metaheuristics-based methods of approximation for two-dimensional probability distributions through GA and SA. The effectiveness of such methods are demonstrated with a numerical example of stabilization synthesis in Section 4.

We use the following notation in this paper. The sets of real numbers, non-negative integers and natural numbers are denoted by $\mathbf{R}$, $\mathbf{N}_0$ and $\mathbf{N}$, respectively. The set of n-dimensional real column vectors and that of $m \times n$ real matrices are denoted by $\mathbf{R}^n$ and $\mathbf{R}^{m\times n}$, respectively. The set of $n \times n$ symmetric matrices and that of $n \times n$ positive definite matrices are denoted by $\mathbf{S}^{n\times n}$ and $\mathbf{S}_+^{n\times n}$, respectively. The Euclidean norm is denoted by $\|\cdot\|$. The expectation (i.e., the expected value) of a random variable is denoted by $E[\cdot]$; this notation is also used for the expectation of a random matrix. If s is a random variable obeying the distribution D, then we represent it as $s \sim D$.

2. DISCRETE-TIME LINEAR SYSTEMS WITH STOCHASTIC DYNAMICS AND CONTROLLER SYNTHESIS

2.1 Discrete-Time Linear Systems with Stochastic Dynamics and Stability

To make the motivation of this study clearer, this section first revisits stability conditions for discrete-time linear systems with dynamics determined by an i.i.d. process. Let us consider the Z-dimensional discrete-time stochastic process $\xi = (\xi_k)_{k\in\mathbf{N}_0}$ satisfying the following assumption.

Assumption 1. ξ_k is independent and identically distributed (i.i.d.) with respect to the discrete time $k \in \mathbf{N}_0$.

This assumption naturally makes ξ stationary and ergodic (Klenke, 2014). For this stochastic process ξ, we denote the cumulative distribution function of ξ_k and the corresponding support by $\mathcal{F}(\cdot)$ and $\boldsymbol{\varXi}$, respectively.

Let us further consider the discrete-time linear system

$$x_{k+1} = A(\xi_k)x_k, \tag{1}$$

where $A : \boldsymbol{\varXi} \to \mathbf{R}^{n\times n}$, and the initial state x_0 is assumed to be deterministic. Since $A(\xi_k)$ is a random matrix, the dynamics of the above system is stochastic.

To define second-moment stability for system (1), we introduce the following assumption.

Assumption 2. The squares of elements of $A(\xi_k)$ are all Lebesgue integrable, i.e.,

$$E\left[A_{ij}(\xi_k)^2\right] < \infty \qquad (\forall i, j = 1, \ldots, n), \tag{2}$$

where $A_{ij}(\xi_k)$ denotes the (i, j)-entry of $A(\xi_k)$.

Then, second-moment exponential stability (Kozin, 1969) can be defined as follows.

Definition 3. The system (1) satisfying Assumptions 1 and 2 is said to be exponentially stable in the second moment if there exist $a > 0$ and $\lambda \in (0, 1)$ such that

$$\sqrt{E\left[\|x_k\|^2\right]} \le a\|x_0\|\lambda^k \quad (\forall k \in \mathbf{N}_0,\ \forall x_0 \in \mathbf{R}^n). \tag{3}$$

This stability notion is known to be characterized by a Lyapunov inequality as follows (Hosoe and Hagiwara, 2019).

Theorem 4. Suppose the system (1) satisfies Assumptions 1 and 2. The following two conditions are equivalent.

1. The system (1) is exponentially stable in the second moment.
2. There exist $P \in \mathbf{S}_+^{n\times n}$ and $\lambda \in (0, 1)$ such that

$$E\left[\lambda^2 P - A(\xi_0)^T P A(\xi_0)\right] \ge 0. \tag{4}$$

2.2 Intractable Issues in Controller Synthesis for General Distributions

Let us next consider the Z-dimensional process ξ satisfying Assumption 1, and the associated system

$$x_{k+1} = A_{\rm op}(\xi_k)x_k + B_{\rm op}(\xi_k)u_k \tag{5}$$

with input u, where $A_{\rm op} : \boldsymbol{\varXi} \to \mathbf{R}^{n\times n}$, $B_{\rm op} : \boldsymbol{\varXi} \to \mathbf{R}^{n\times m}$, and the initial state x_0 is assumed to be deterministic. Then, the closed-loop system consisting of this system and the state feedback

$$u_k = Fx_k \tag{6}$$

with the static time-invariant gain $F \in \mathbf{R}^{m\times n}$ can be described as follows.

$$x_{k+1} = A_{\rm cl}(\xi_k)x_k, \quad A_{\rm cl}(\xi_k) = A_{\rm op}(\xi_k) + B_{\rm op}(\xi_k)F \tag{7}$$

The Lyapunov inequality for this closed-loop system is given by (4) with $A(\xi_0)$ replaced by $A_{\rm cl}(\xi_0)$, i.e.,

$$E\left[\lambda^2 P - A_{\rm cl}(\xi_0)^T P A_{\rm cl}(\xi_0)\right] \ge 0. \tag{8}$$

Hence, for stabilization synthesis, we need to search for P and F satisfying the Lyapunov inequality (8) (under $\lambda < 1$). This inequality condition involves random variables contained in the expectation operation, and a direct search for such P and F is generally difficult. In addition, the nonlinearity of the inequality with respect to P and F is also an obstacle. Although our standpoint here is to only consider stabilization synthesis as an example, these issues are common in other controller synthesis problems.

Regarding stabilization synthesis, it is shown in Hosoe and Hagiwara (2019) that (8) can be rewritten as a standard matrix inequality with deterministic coefficients calculated only with $A_{\rm op}(\xi_0)$ and $B_{\rm op}(\xi_0)$. The resulting matrix inequality can also be linearized without loss of generality, and hence, the issues stated above can be resolved. However, this approach is not ensured to be available even for other controller synthesis problems. To facilitate advances in such problems for systems with stochastic dynamics, this paper focuses on another approach to dealing with inequality conditions involving random variables (for details of the other approach, see Hosoe and Hagiwara (2019)). The approach is to separate the expectation-based matrix inequality such as (8) into the expectation part and the matrix inequality part by introducing an auxiliary mapping. With this approach, together with the conventional techniques about linear matrix inequality (LMI) (Boyd et al., 1994), we can show the following theorem (Hosoe et al., 2019).

Theorem 5. Suppose ξ satisfies Assumption 1, and the squares of elements of $A_{\rm op}(\xi_k)$ and $B_{\rm op}(\xi_k)$ are all Lebesgue integrable (so that Assumption 2 is satisfied for the closed-loop system). There exists a state feedback gain F that stabilizes the closed-loop system (7) if and only if there exist $X \in \mathbf{S}_{+}^{n\times n}$, $Y \in \mathbf{R}^{n\times m}$, a mapping $S : \varXi \to \mathbf{S}^{n\times n}$ and $\lambda \in (0,1)$ satisfying

$$E\left[S(\xi_0)\right] \le 0, \tag{9}$$

$$\begin{bmatrix} \lambda^2 X + S(\xi_\star) & XA_{\mathrm{op}}(\xi_\star)^T + YB_{\mathrm{op}}(\xi_\star)^T \\ * & X \end{bmatrix} \ge 0 \quad (\forall \xi_\star \in \varXi). \tag{10}$$

In particular, $F = Y^T X^{-1}$ is one such stabilization gain.

In the above theorem, S is the introduced auxiliary mapping. The inequality condition (9) and (10) can be obtained from (8). As we can see, (10) is linear in the decision variables for each fixed λ. Hence, the remaining issue is only the treatment of ξ_0 and $\varXi$ in the condition, associated with the introduced mapping S. If the support $\varXi$ is not finite, we have to search for a mapping S satisfying an infinite-dimensional LMI condition, which is infeasible. To circumvent this issue, this paper studies approximation of the probability distribution of ξ_0 (i.e., the cumulative distribution $\mathcal{F}$) by a distribution with finite support.

2.3 Equivalent Simplified Condition for Discrete Distributions

As mentioned above, solving the inequality condition (9) and (10) is generally difficult. However, if the distribution of ξ_0, which we call the population distribution in the following, is a discrete distribution with finite support, the inequality condition reduces to a standard LMI, which can be numerically easily solved. To see this, let us consider as the population distribution the discrete distribution $D(q,p)$ with parameters $q = [q_1 \ \cdots \ q_Q] \in \mathbf{R}^{Z\times Q}$ and $p = [p_1 \ \cdots \ p_Q] \in \mathbf{R}^{1\times Q}$ such that ξ_0 obeying this distribution takes the value q_i with probability p_i for $i = 1,\ldots,Q$. Then, (9) and (10) immediately reduce to

$$\sum_{i=1}^{Q} p_i S_i \le 0, \tag{11}$$

$$\begin{bmatrix} \lambda^2 X + S_i & XA_{\mathrm{op}}(q_i)^T + YB_{\mathrm{op}}(q_i)^T \\ * & X \end{bmatrix} \ge 0 \quad (i = 1,\ldots,Q), \tag{12}$$

where $S_i \in \mathbf{S}^{n\times n}$ is a decision variable corresponding to $S(q_i)$. This inequality condition consists of standard finite-dimensional LMIs with decision variables S_i ($i = 1,\ldots,Q$), X and Y for each fixed λ. Hence, with a bisection with respect to λ^2, its solution leading to a minimal λ can be easily obtained, which is optimal in the sense of the convergence rate about exponential stability (see Definition 3).

The above arguments imply that if the population distribution can be approximated by $D(q,p)$ with high accuracy in some sense, the original inequality condition (9) and (10) can be solved approximately, even when the support is not finite. To make such synthesis possible, the following section discusses approximation of two-dimensional distributions using metaheuristics, as a first step toward multi-dimensional approximation.

3. APPROXIMATION OF TWO-DIMENSIONAL PROBABILITY DISTRIBUTIONS USING METAHEURISTICS

In this section, we discuss specific methods of approximating two-dimensional population distributions with infinite support. The index of accuracy of approximation we use is the Kolmogorov distance. Since it is difficult to simply extend the conventional methods of approximating one-dimensional distributions as stated in Section 1, we here take another way in which a genetic algorithm (GA) and simulated annealing (SA) are employed. GA is known to be good at global search, while SA is at systematic search in the neighborhood of an initial state (Nikolaev and Jacobson, 2010; Reeves, 2010). Hence, a combinational use of these two algorithms may lead us to more accurate approximation than using only one of them. After stating the definition of the Kolmogorov distance as well as the treatment of the cumulative distribution function of a discrete distribution, we discuss GA-based and GA&SA-based methods of approximation.

We use the following notation in this section. We denote the cumulative distribution function of $D(q,p)$ and the vector corresponding to ξ_0 respectively by $\mathcal{F}_D(\cdot)$ and $\mathcal{U}$, where $\mathcal{U} \in \varXi \subset \mathbf{R}^2$ and $\mathcal{U} =: [\mathcal{X} \ \mathcal{Y}]^T$. In addition, we abbreviate $\mathcal{F}([\mathcal{X} \ \mathcal{Y}]^T)$ as $\mathcal{F}(\mathcal{X},\mathcal{Y})$ for simplicity.

3.1 Kolmogorov Distance and Cumulative Distribution Function of Discrete Distribution

The Kolmogorov distance for Z-dimensional distributions is defined as follows (Rachev et al., 2013).

Definition 6. For given cumulative distribution functions $\mathcal{F}_i : \mathbf{R}^Z \to [0,1]$ ($i = 1,2$), the Kolmogorov distance $\rho_Z(\mathcal{F}_1,\mathcal{F}_2)$ between $\mathcal{F}_1$ and $\mathcal{F}_2$ is defined by

$$\rho_Z(\mathcal{F}_1,\mathcal{F}_2) := \sup_{\mathcal{U}\in\mathbf{R}^Z} |\mathcal{F}_1(\mathcal{U}) - \mathcal{F}_2(\mathcal{U})|. \tag{13}$$

The purpose of our approximation is to construct $D(q,p)$ (i.e., $\mathcal{F}_D$) minimizing the above distance from the (two-dimensional) population distribution (i.e., $\mathcal{F}$).

The cost function for the approximation problem viewed as an optimization problem is given by

$$\rho_2(\mathcal{F},\mathcal{F}_D) = \sup_{\mathcal{U}\in\varXi} |\mathcal{F}(\mathcal{U}) - \mathcal{F}_D(\mathcal{U})|, \tag{14}$$

which is denoted by d_{K} for simplicity. Since $\mathcal{F}_D$ is the cumulative distribution of a discrete distribution with finite support, this d_{K} can be obtained by just calculating $|\mathcal{F}(\mathcal{U}) - \mathcal{F}_D(\mathcal{U})|$ at a finite number of $\mathcal{U} \in \varXi$ (Justel et al., 1997). Hence, the distance computation is numerically tractable when $\mathcal{F}_D$ is fixed.

Let us next consider the situation where Q points $(\mathcal{X}_i,\mathcal{Y}_i)$ ($i = 1,\ldots,Q$) are on the $\mathcal{X}\mathcal{Y}$ plane, and $\mathcal{F}_D(\mathcal{U}) = \mathcal{F}_D(\mathcal{X},\mathcal{Y})$ is described by

$$\mathcal{F}_D(\mathcal{X},\mathcal{Y}) := \sum_{i=1}^{Q} \Delta h_i \cdot \mathbf{1}_{(-\infty,\mathcal{X}]\times(-\infty,\mathcal{Y}]}(\mathcal{X}_i,\mathcal{Y}_i), \tag{15}$$

where Δh_i and $\mathbf{1}_{\mathcal{A}}(\mathcal{U})$ respectively denote the increment of $\mathcal{F}_D$ at the point $(\mathcal{X}_i,\mathcal{Y}_i)$ and the indicator function

$$\mathbf{1}_{\mathcal{A}}(\mathcal{U}) := \begin{cases} 1 & (\mathcal{U} \in \mathcal{A}) \\ 0 & (\mathcal{U} \notin \mathcal{A}). \end{cases} \tag{16}$$

By definition of the cumulative distribution,

$$\sum_{i=1}^{Q} \Delta h_i = 1 \tag{17}$$

should be satisfied. Then, a random variable obeying the corresponding discrete distribution $D(q,p)$ takes the value $q_i = [\mathcal{X}_i \ \mathcal{Y}_i]^T$ with probability $p_i = \Delta h_i$ for each $i = 1,\ldots,Q$.

Here, to make the problem simple, we introduce the constraint that the increment Δh_i of $\mathcal{F}_D$ is uniform in $i = 1,\ldots,Q$. That is, $p_1 = \cdots = p_Q = 1/Q$ (recall $\Delta h_i = p_i$), which naturally makes (17) satisfied, and $\mathcal{F}_D$ in (15) is given by

$$\mathcal{F}_D(\mathcal{X},\mathcal{Y}) = \frac{1}{Q}\sum_{i=1}^{Q} \mathbf{1}_{(-\infty,\mathcal{X}]\times(-\infty,\mathcal{Y}]}(\mathcal{X}_i,\mathcal{Y}_i). \tag{18}$$

The problem of our approximation is to determine the Q points $(\mathcal{X}_i,\mathcal{Y}_i)$ $(i = 1,\ldots,Q)$ of this $\mathcal{F}_D$ that minimizes d_{K} for given $\mathcal{F}$. Thus, our optimization problem can be described as follows.

$$\min_{q} \quad d_{\mathrm{K}} \tag{19}$$

$$\text{s.t.} \quad q_i \in \varXi \qquad (i = 1,\ldots,Q) \tag{20}$$

$$p_i = \frac{1}{Q} \qquad (i = 1,\ldots,Q) \tag{21}$$

3.2 Approximation Using GA

This subsection discusses a GA-based method of approximation, which produces a discrete distribution from a population distribution that tries to minimize d_{K} as much as possible. To associate the present approximation problem with a GA, we first make a grid on the search area from which $(\mathcal{X}_i,\mathcal{Y}_i)$ $(i = 1,\ldots,Q)$ in the $\mathcal{XY}$ plane are taken. The search area here is given by $[a_{\mathcal{X}}, b_{\mathcal{X}}] \times [a_{\mathcal{Y}}, b_{\mathcal{Y}}] \in \varXi$ with $a_{\mathcal{X}}, b_{\mathcal{X}}, a_{\mathcal{Y}}, b_{\mathcal{Y}} \in \mathbf{R}$ appropriately selected by using information of the population distribution. The grid fineness is characterized by an integer r (≥ 2). That is,

$$q_i = \begin{bmatrix} \mathcal{X}_i \\ \mathcal{Y}_i \end{bmatrix} \in \Theta \qquad (i = 1,\ldots,Q) \tag{22}$$

is assumed to hold for the set

$$\Theta := \left\{ \begin{bmatrix} a_{\mathcal{X}} + \dfrac{b_{\mathcal{X}} - a_{\mathcal{X}}}{r-1}\cdot v \\ a_{\mathcal{Y}} + \dfrac{b_{\mathcal{Y}} - a_{\mathcal{Y}}}{r-1}\cdot w \end{bmatrix} \in \varXi \;\middle|\; \begin{array}{l} v = 0,\ldots,r-1; \\ w = 0,\ldots,r-1 \end{array} \right\} \tag{23}$$

consisting of the grid points in our GA-based approximation; this restriction is used also in GA&SA-based approximation discussed later. With this restriction, the Q selection points to be determined can be dealt with in GA. Thus, (20) in our optimization problem is replaced by (22) and (23).

We next develop a specific GA-based method of approximating the population distribution. The basic concept of GAs is shown in Goldberg (1989); Schwefel (1995); Reeves (2010). Our specific method follows the basic concept, and has some steps: Initialization, Selection, Crossover, and Mutation. An additional idea that we use is elitism (Goldberg, 1989), which passes the best individual (called an elite individual) of each generation to the next generation. This will ensure d_{K} to be monotonically non-increasing throughout the generation changes in our method (the details will be clearer soon). Our method starts from Initialization with $g = 0$, where g denotes the generation number. After Mutation, if $g < g_{\mathrm{end}}$, g is incremented and we return to Selection, where g_{end} denotes the last generation number. If $g = g_{\mathrm{end}}$, our method ends. Each step of our GA-based approximation method is shown in the following.

Initialization

By taking into account the use of elitism, the number of individuals M is assumed to be odd. In addition, to improve the efficiency of search, the grid fineness r is assumed to be a power of 2. Then, we first give a rule that associates a candidate of the solution of the problem (i.e., the coordinates of the Q selection points in Θ, which we call the set of selection points with Q pairs of binary strings that each of the individuals has (which we call a binary representation of an individual). By (22) and (23), the coordinates of $(\mathcal{X}_i,\mathcal{Y}_i)$ $(i = 1,\ldots,Q)$ in the $\mathcal{XY}$ plane are determined by v and w. Hence, by considering the set of (v_i, w_i), $i = 1,\ldots,Q$ that each of the individuals is supposed to represent, a candidate of the discrete distribution $\mathcal{F}_D$ can be represented (recall (18)) by each of the individuals. Here, v_i and w_i are both representable by $\log_2 r$-digit binary numbers. Since binary numbers are compatible with GA, we regard them as the binary representation of an individual. That is, we consider the correspondence relationship exemplified as follows ($Q = 3$ and $r = 2^7 = 128$ in this example).

$$\text{set of selection points} \begin{bmatrix} 53 & 19 & 79 \\ 28 & 63 & 94 \end{bmatrix}$$

$$\updownarrow \text{correspond}$$

$$\text{binary representation} \begin{bmatrix} 0110101 & 0010011 & 1001111 \\ 0011100 & 0111111 & 1011110 \end{bmatrix}$$

The first row of the set of selection points represents v_i and the second row does w_i for $i = 1,\ldots,Q$. With these settings, we randomly generate an initial population consisting of M individuals, each of which has the binary representation for the set of Q selection points.

Selection

The index of accuracy of approximation is the Kolmogorov distance between $\mathcal{F}$ and $\mathcal{F}_D$. We assign a number from 1 to M to each individual and define the corresponding fitness f_j as

$$f_j = \frac{1}{d_{\mathrm{K},j} - \dfrac{1}{2Q}} \qquad (j = 1,\ldots,M), \tag{24}$$

where $d_{\mathrm{K},j}$ is the Kolmogorov distance for individual j. $1/(2Q)$ is the Kolmogorov distance when optimal approximation is achieved in the case of one-dimensional approximation (Hosoe et al., 2019). We use this fitness because $1/(2Q)$ is a lower bound of $d_{\mathrm{K},j}$ under the current problem settings.

The individual having the highest fitness in each generation is called an elite individual (in the generation). To keep the number of individuals unchanged over generations, we repeat the operations of selecting two individuals as a parent pair from M individuals including an elite individual using a roulette wheel (Goldberg, 1989), and generating two individuals as offspring (i.e., next generation individuals) $(M-1)/2$ times. Since $M-1$ offspring are generated by this operation, we add an elite individual of the parent generation after mutation. In the selection using a roulette wheel, the probability that the individual j is selected is given by $f_j/\sum_{j=1}^{M} f_j$, and thus, the individual with higher fitness is selected as a parent with higher probability.

Crossover and **Mutation**

We introduce simple crossover (Goldberg, 1989) as the rule for our crossover. Simple crossover is a method in which a crossover point is randomly selected for binary strings of the binary representation, and binary bits after the crossover point of the binary representation of one parent is swapped for that of the other parent by the crossover probability $p_{\rm c} \in [0,1]$.

After crossover, mutation inverts all the binary bits of the binary representation of offspring generated by crossover with probability $p_{\rm m} \in [0,1]$ ($p_{\rm m}$ is called the mutation probability).

3.3 Approximation Using GA and SA

This subsection further discusses a method of improving the approximation accuracy using SA after obtaining discrete distributions by our GA, which we call the GA&SA-based method of approximation. In the GA&SA-based method, we store not only the best individual of the last generation but also those of (some of) other generations in the GA part, and then, use each of them as the initial state of the SA. This is considered to be helpful for finding relatively good local optimal solutions in the SA part.

The basic concept of SA is shown in Schwefel (1995); Nikolaev and Jacobson (2010). In our SA, the set of selection points and the Kolmogorov distance $d_{\rm K}$ are regarded as the state and the energy, respectively. Then, for a state, the family of the set of selection points that is obtained by moving only one among the Q points represented by the state to an adjacent grid point is defined as the neighborhood. We select a perturbed set of selection points randomly. Then, we use

$$T(t) := \alpha^t T_{\rm init} \qquad (T_{\rm init} > 0) \tag{25}$$

as the temperature of SA, where α and $T_{\rm init}$ denote the cooling factor ($\alpha \in (0,1)$) and the start temperature, respectively. The parameters α and $T_{\rm init}$ are adjusted by trial and error through observing the output of SA. The end time of SA (i.e., the number of iterations) is denoted by $t_{\rm end}$. We use this part of method after the GA part in GA&SA-based approximation.

4. DEMONSTRATION OF EFFECTIVENESS OF METAHEURISTICS-BASED APPROXIMATION

This section numerically demonstrates the effectiveness of our proposed approximation. Let us consider the two-dimensional stochastic process ξ that satisfies Assumption 1 and is given by the sequence of $\xi_k = [\xi_{1k}\ \xi_{2k}]^T \sim N(\mu, \Sigma)$, where $N(\mu,\Sigma)$ is the (two-dimensional) normal distribution with mean μ and covariance Σ, which corresponds to the population distribution. In this paper, we use

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 0.7^2 & 0.2 \\ 0.2 & 0.6^2 \end{bmatrix}. \tag{26}$$

Let us further consider the system (7) with

$$A_{\rm op}(\xi_k) = \begin{bmatrix} -0.2+\xi_{1k} & -0.4 & -0.7 \\ 0.5 & -0.7 & -0.3+\xi_{2k} \\ -1.0 & -0.9+c(\xi_k) & -0.2 \end{bmatrix},$$
$$B_{\rm op}(\xi_k) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad c(\xi_k) = \frac{1}{1+\xi_{1k}^2+\xi_{2k}^2}. \tag{27}$$

For this system, we compare the following three approximation methods under $Q = 10$: random sampling of ξ_0, GA-based approximation, and GA&SA-based approximation.

First, we construct $\mathcal{F}_D$ from the population distribution by random sampling. We generated 20,000 sets of samples for $\mathcal{F}_D$ (each set consists of $Q = 10$ samples of ξ_0, and each sample point is in $\mathbf{R}^2$), and compared them with respect to the Kolmogorov distance $d_{\rm K}$. Then, we obtained the discrete distribution $\mathcal{F}_D$ shown in blue in Fig. 1 as the best among them, where the population distribution $\mathcal{F}$ is shown in green. Table 1 shows the Kolmogorov distance $d_{\rm K}$ and the execution time of the program of this simple method; all the programs in this section were executed with MATLAB on a PC equipped with 32 GB RAM and Intel(R) Core(TM) i7-4790 CPU @ 3.60GHz.

Next, we state the result of GA-based approximation. In our computation, we set $r = 2^7$, $a_{\mathcal{X}} = -2$, $a_{\mathcal{Y}} = -1$, $b_{\mathcal{X}} = 2$, $b_{\mathcal{Y}} = 1$, $M = 31$, $g_{\rm end} = 400$, $p_{\rm c} = 0.75$, $p_{\rm m} = 2.2 \times 10^{-3}$. Then, the best $\mathcal{F}_D$ with respect to $d_{\rm K}$ for this approach was as shown in blue in Fig. 2. The Kolmogorov distance $d_{\rm K}$ and the execution time are also shown in Table 1.

Finally, we state the result of GA&SA-based approximation. We used top three individuals of each of the generations No. 100, 200, 300 and 400 (hence 12 individuals in total) obtained through the above GA-based approximation as the initial state of SA. The parameters in the SA part were set as $t_{\rm end} = 2000$, $\alpha = 0.99$, $T_{\rm init} = 1$. For each of 12 initial states, we executed SA and obtained 12 discrete distributions. Then, the best with respect to $d_{\rm K}$ for this approach was as shown in blue in Fig. 3. The Kolmogorov distance $d_{\rm K}$ and the execution time are in Table 1.

According to the above results, at least in this example, GA&SA-based approximation gave a better result than the others, and GA-based approximation did than random sampling with respect to $d_{\rm K}$. The purpose of this study is to reduce the influence of approximation in controller synthesis, and hence, we next compare the results in stabilization synthesis for the system given above. We searched for a solution that minimizes λ with respect to (11) and (12) with each of the three discrete distributions obtained above. Then, we obtained $\lambda_{\rm app}$ in Table 2 as the minimal λ for each method. Since these $\lambda_{\rm app}$ are obtained under the approximate distributions, and since we can check the true minimal λ (denoted by $\lambda_{\rm min}$) with the designed F for the present stability problem (Hosoe and Hagiwara, 2019), we performed such post-synthesis analysis for each of the three cases. Then, we obtained the result in Ta-

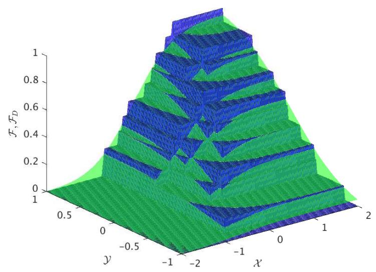

Fig. 1. Approximation based on random sampling.

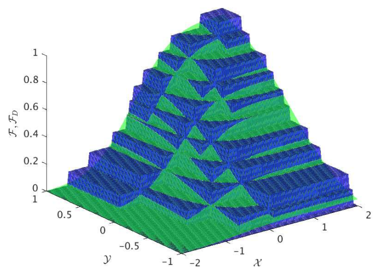

Fig. 2. GA-based approximation.

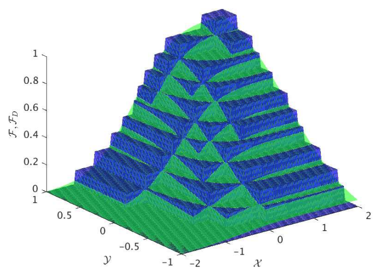

Fig. 3. GA&SA-based approximation.

Table 1. The Kolmogorov distance d_{K} and the execution time for each method.

	random sampling	GA	GA & SA
d_{K}	0.1640	0.1333	0.1098
execution time	174.1 sec	40.9 sec	119.2 sec

Table 2. Results of approximate synthesis and post-synthesis (strict) analysis.

	random sampling	GA	GA & SA
λ_{app}	0.6940	0.8783	0.8334
λ_{min}	0.7963	0.7891	0.7866
ϵ_λ	0.1023	0.0892	0.0468

ble 2, in which the gap $\epsilon_\lambda := |\lambda_{\mathrm{app}} - \lambda_{\mathrm{min}}|$ is also shown. According to this result, we can confirm that minimizing d_{K} in approximation contributed to reducing ϵ_λ, and the latter further contributed to reducing (not λ_{app} but) λ_{min}. This suggests the effectiveness of our minimization of d_{K} in approximation of probability distributions through the use of metaheuristics.

5. CONCLUSION

In this paper, we discussed approximation of two-dimensional probability distributions using metaheuristics toward future control applications. A similar idea is considered to be effective also for higher-dimensional cases of probability distribution. In addition, there would be varieties of methods in terms of evaluation functions and metaheuristics in the present approximation. Further investigations from these viewpoints will be a future work.

REFERENCES

Arnold, L. (1998). *Random Dynamical Systems*. Springer-Verlag.

Boyd, S., Ghaoui, L.E., Feron, E., and Balakrishman, V. (1994). *Linear Matrix Inequalities in System and Control Theory*. SIAM.

Goldberg, D.E. (1989). *Genetic Algorithm in Search, Optimization, and Machine Learning*. Addison-Wesley Publishing Company, Inc.

Hetel, L., Fiter, C., Omran, H., Seuret, A., Fridman, E., Richard, J.P., and Niculescu, S.I. (2017). Recent developments on the stability of systems with aperiodic sampling: An overview. *Automatica*, Vol. 76, No. 2, pp. 309–335.

Hosoe, Y., Okuno, M., and Hagiwara, T. (2019). Approximation of cumulative distribution functions based on L_∞ and L_1 norms for stochastic systems control (in Japanese). *Transactions of the Institute of Systems, Control and Information Engineers*, Vol. 32, No. 4, pp. 145–153.

Hosoe, Y. and Hagiwara, T. (2019). Equivalent stability notions, Lyapunov inequality, and its application in discrete-time linear systems with stochastic dynamics determined by an i.i.d. process. *IEEE Transactions on Automatic Control*, Vol. 64, No. 11, pp. 4764–4771.

Justel, A., Peña, D., and Zamar, R. (1997). A multivariate Kolmogorov-Smirnov test of goodness of fit. *Statistics & Probability Letters*, Vol. 35, No. 3, pp. 251–259.

Klenke, A. (2014). *Probability Theory: A Comprehensive Course 2nd edition*. Springer-Verlag.

Kozin, F. (1969). A survey of stability of stochastic systems. *Automatica*. Vol. 5, No. 1, pp. 95–112.

Nikolaev, A.G. and Jacobson, S.H. (2010). Simulated annealing. In M. Gendreau and J.-Y. Potvin (eds.), *Handbook of Metaheuristics*, pp. 1–40. Springer, 2nd edition.

Paxson, V. and Floyd, S. (1995). Wide area traffic: The failure of Poisson modeling. *IEEE/ACM Transactions on Networking*. Vol. 3, No. 3, pp. 226–244.

Pflug, G.C. and Pichler, A. (2011). Approximation for probability distributions and stochastic optimization problems. In M. Bertocchi, G. Consigli, and M.A.H. Dempster (eds.), *Stochastic Optimization Methods in Finance and Energy*, pp. 343–387. Springer.

Rachev, S.T., Klebanov, L.B., Stoyanov, S.V., and Fabozzi, F.J. (2013). *The Methods of Distances in the Theory of Probability and Statistics*. Springer.

Reeves, C.R. (2010). Genetic algorithm. In M. Gendreau and J.-Y. Potvin (eds.), *Handbook of Metaheuristics*, pp. 109–140. Springer, 2nd edition.

Schwefel, H.-P. (1995). *Evolution and Optimum Seeking*. John Wiley & Sons, Inc.

Van Laarhoven, P.J.M. and Aarts, E.H.L. (1987). *Simulated Annealing: Theory and Applications*. D. Reidel Publishing Company.