# A Hitting Time Analysis for Stochastic Time-Varying Functions with Applications to Adversarial Attacks on Computation of Markov Decision Processes

Ali Yekkehkhany[1], Han Feng[1], Donghao Ying[1], Javad Lavaei[1]

[1] *Department of Industrial Engineering and Operations Research - University of California, Berkeley*
{*aliyek, han_feng, donghaoy, lavaei*}*@berkeley.edu*

**Abstract**

Stochastic time-varying optimization is an integral part of learning in which the shape of the function changes over time in a non-deterministic manner. This paper considers multiple models of stochastic time variation and analyzes the corresponding notion of hitting time for each model, i.e., the period after which optimizing the stochastic time-varying function reveals informative statistics on the optimization of the target function. The studied models of time variation are motivated by adversarial attacks on the computation of value iteration in Markov decision processes. In this application, the hitting time quantifies the extent that the computation is robust to adversarial disturbance. We develop upper bounds on the hitting time by analyzing the contraction-expansion transformation appeared in the time-variation models. We prove that the hitting time of the value function in the value iteration with a probabilistic contraction-expansion transformation is logarithmic in terms of the inverse of a desired precision. In addition, the hitting time is analyzed for optimization of unknown continuous or discrete time-varying functions whose noisy evaluations are revealed over time. The upper bound for a continuous function is super-quadratic (but sub-cubic) in terms of the inverse of a desired precision and the upper bound for a discrete function is logarithmic in terms of the cardinality of the function domain. Improved bounds for convex functions are obtained and we show that such functions are learned faster than non-convex functions. Finally, we study a time-varying linear model with additive noise, where hitting time is bounded with the notion of shape dominance.

***Keywords***— Stochastic time-varying functions, stochastic operators, hitting time, probabilistic contraction-expansion mapping, probabilistic Banach fixed-point theorem, adversarial Markov decision process

## 1 Introduction and Related Work

In many practical applications of optimization, such as those in the training of neural networks [1,2], online advertising [3], decision-making process of power systems [4,5], and the real-time state estimation of nonlinear systems [6], the parameters of the problem are often uncertain and change over time [7]. To put the time-varying and uncertainty of the systems into perspective in optimization problems, time-varying or online optimization aims to find the solution trajectories determined by

$$x_t^* = \underset{x\in\mathcal{X}}{\operatorname{argmin}} \left\{ f_t(x) = \mathbb{E}F_t(x,\xi) \right\}, \quad t \in \{1,2,\dots\}, \tag{1}$$

where the random variable $\xi$ models the uncertainty in the objective that comes from disturbance, inexactness of model, use of small batches, or injected noise, and where argmin denotes any global minimizer of the input function. Note that the expectation $\mathbb{E}$ over $\xi$ can only be evaluated approximately since the probability

distribution is unknown, and therefore the target function $f_t$ should be approximated by observed samples. The estimate of the target function may not capture the shape of the target function given a limited number of observed samples. However, there is a point of time, named *hitting time*, after which optimizing the estimated target function results in optimizing the target function up to some precision and confidence level. The hitting time captures the stochastic complexity of the time-varying problem in (1).

Table 1: Comparison of Selected Theorems in Sections II-III

| Theorem | Assumptions | Hitting Time Definition |
|---|---|---|
| 3 | Assumptions 3-4, bounded difference functions | (45) |
| 4 | Assumptions 3-6, convex bounded difference functions | (45) |
| 5 | Assumptions 3 and 7 | (65) |
| 6 | Assumptions 3 and 7, unimodal functions | (65) |
| 8 | linear dynamics and shape dominance | (84) |

## 1.1 Motivating Applications

In order to motivate the analysis of hitting time for time-varying probabilistic transformations, we first explain its applications in Markov Decision Process (MDP) and reinforcement learning (RL). Consider an MDP with the set of states (state space) $\mathcal{S}$, the set of actions (action space) $\mathcal{A}$, the time-invariant state transition $h$ such that $s_{k+1} = h(s_k, a_k, w_k)$, where $w_k$ for $k \in \{0, 1, \dots\}$ is a sequence of independent and identically distributed (i.i.d.) random variables, and the immediate reward $r(s_k, a_k, w_k)$ received after taking action $a_k$ in state $s_k$. A state-contingent decision policy is a mapping $\mu : \mathcal{S} \to \mathcal{A}$. Given a discount factor $0 < q < 1$ and a policy $\mu$, the value function $V^\mu : \mathcal{S} \to \mathcal{R}$ is defined as

$$V^\mu(s) = \mathbb{E}\left[\sum_{k=0}^{\infty} q^k \cdot r(s_k, \mu(s_k), w_k) \middle| s_0 = s\right], \tag{2}$$

where expectation is taken over $w_k$ for $k \geq 0$. Then, the optimal value function $V^*$ is defined by

$$V^*(s) = \max_{\mu} V^\mu(s). \tag{3}$$

For a finite action space, any policy $\mu^*$ given by

$$\mu^*(s) = \operatorname*{argmax}_{a \in \mathcal{A}} \ \mathbb{E}\big[r(s, a, w) + q \cdot V^*(h(s, a, w))\big] \tag{4}$$

is optimal in the sense that $V^*(s) = V^{\mu^*}(s)$, which gives rise to the Bellman equation

$$V^*(s) = \max_{a \in \mathcal{A}} \ \mathbb{E}\big[r(s, a, w) + q \cdot V^*(h(s, a, w))\big] \quad \forall s \in \mathcal{S}, \tag{5}$$

where $w$ is a random variable with the same distribution as $w_k$ for some $k$. Define the Bellman operator $\mathcal{T}$ as

$$(\mathcal{T}V)(s) = \max_{a \in \mathcal{A}} \ \mathbb{E}\big[r(s, a, w) + q \cdot V(h(s, a, w))\big] \tag{6}$$

Starting from an arbitrary $V_0$, the value iteration method constructs a sequence $\{V_0, V_1, V_2, \dots\}$ with $V_{t+1} = \mathcal{T}(V_t)$ for $t \in \{0, 1, \dots\}$. It is well known that the Bellman operator is a contraction mapping, which guarantees convergence to $V^*$. The optimal value function $V^*$ is unknown in MDP and RL applications. The value function $V_t$ is a time-varying function and may never be exactly equal to $V^*$. Moreover, $V_t$ is rarely computed exactly and is subject to adversarial attacks. We will introduce multiple models of attack and analyze the corresponding notion of hitting time for each model to be able to study the convergence of $V_t$.

## 1.2 Related Work

### 1.2.1 Approximate Dynamic Programming

The field approximate dynamic programming encompasses a wide range of techniques that overcomes the curse of dimensionality in the computation of Bellman operator. The adversarial attack model studied in this paper is motivated by the following approaches:

I. **Approximation in computing expectation:** There are different approaches to circumventing the costly computation of expectation in (6), e.g., a) assuming certainty equivalence by replacing stochastic quantities with deterministic ones to arrive at a deterministic optimization, b) using Monte Carlo tree search and adaptive simulation to determine which expectations associated with actions should be computed more accurately [8–12]. Both of these approaches introduce some errors in the expectation.

II. **Approximation in maximization:** The maximization in the Bellman operator in (6) can be over a large number of actions, possibly a continuous action space with an infinite number of actions. In addition to the discretization of the action space, nonlinear programming techniques are prone to errors especially when they are used in an online fashion.

III. **Approximation of value function:** Due to the large number of states in many recent applications of Markov decision processes and reinforcement learning, parametric feature-based approximation methods, such as neural network architectures, are used for value function representation [8, 13–16]. The parameterization of the value function is another source of error in value iteration that can cause expansion in value iteration [13, 14].

IV. **Adversarial value iteration:** The emergence of cloud, edge, and fog computing means that large-scale MDP and RL problems will likely be solved by distributed servers [17–19]. This swift shift to edge reinforcement learning brings a host of new adversarial attack challenges that can be catastrophic in critical applications of autonomous vehicles and Internet of Things (IoT) in general [20–22].

The first three causes have been studied extensively in the literature [23], while there is no mathematical analysis of adversarial attacks on the computation of the value functions.

### 1.2.2 Reinforcement Learning in Time-varying Environment

Consider a reinforcement learning framework in which the model is being learned or there is a time-varying environment whose state transition probabilities and rewards change over time [24]. An example of a time-varying environment is the changing environment at which autonomous vehicles interact with each other, human drivers, and pedestrians. In the context of reinforcement learning and Markov decision processes, this gradual change is translated into time-varying reward functions and transition probabilities. The relevance of time-varying functions to MDP and RL problems presented above is one of the many problems that can be described by time-varying functions whose hitting time analysis is of interest. Other applications of a time-varying framework, such as bandit optimization, model predictive control, and empirical risk minimization, are discussed in [25].

### 1.2.3 Scenario-based Approach for Optimization

Scenario-based approach for optimization [26–28] is concerned with decision making based on seen cases while having the ability to generalize to new situations. In this context, a bound on the violation probability captures the generalization of time-invariant decisions. The hitting time defined in this paper is related to the violation probability. Our work departs from this line of research in that we study a sequence of time-varying functions instead of a time-invariant function, which can potentially be corrupted by an adversary, and seeking to constantly adjusting our understanding of the optimal solution. The hitting time captures the time-varying aspect in our setting.

### 1.2.4 Dynamical Systems

Our work is also related to asynchronous dynamical systems [29], which have been extensively studied in the literature. Despite the mathematical resemblance, our work is different from this line of research since our focus is on analyzing the associated hitting times of different models and the dynamics considered in this work may not even be linear.

## 1.3 Contributions

We propose a probabilistic model of adversarial attacks, in which both expansion up to a constant and contraction occur with certain probabilities in iterates of the value iteration method. We then study the hitting time of such stochastic time-varying value functions in Section 2. We develop an upper bound on the hitting time under a time-varying contraction mapping with additive noise and develop an upper bound on the distance between the fixed point and the value function.

In the rest of this paper, different models of stochastic time variation for continuous and discrete functions are studied in Sections 2 and 3, respectively. In particular, probabilistic contraction-expansion mappings are studied in Section 2.1, time-varying probabilistic contraction-expansion mappings with additive noise are studied in Section 2.2, time-varying continuous functions with additive noise are studied in Section 2.3, and improved bounds for convex functions with additive noise are studied in Section 2.4. Time-varying discrete functions with additive noise are studied in Section 3.1, improved bounds for unimodal functions with additive noise are studied in Section 3.2, and a time-varying linear model with additive noise with the notion of shape dominance are studied in Section 3.3. We summarize the theorems and the associated assumptions as well as the hitting times definitions in Table 1. Finally, the simulation results are presented in Section 4 and the paper is concluded in Section 5 in which a discussion of opportunities for future work is presented as well.

# 2 The Hitting Time Analysis for Continuous Functions

In this section, three variants of stochastic time-varying models are studied and their hitting times are analyzed. In the first model, a probabilistic contraction-expansion mapping is analyzed, where the classical Banach fixed-point theorem cannot be applied to this model due to the probabilistic contraction-expansion nature of the problem. In the second model, a time-varying probabilistic contraction-expansion mapping with additive noise is investigated. The above two models are applicable to both continuous and discrete functions. In the last model, an unknown time-varying continuous function is observed with additive noise whose estimated function changes over time.

To motivate the three stochastic time-varying models, we revisit the motivating example in the previous section, where a sequence of value functions $V_0, V_1, \dots$ is generated by the Bellman operator $\mathcal{T}$ defined in (6). Note that the theoretical proof of convergence behind the value iteration method depends heavily on the contraction mapping parameter $q$ and the fact that $d\big(\mathcal{T}(V_{t+1}), \mathcal{T}(V_t)\big) \leq q \cdot d\big(V_{t+1}, V_t\big)$ deterministically, where $d(\cdot,\cdot)$ is a translation-invariant distance function induced by a norm. However, in an online implementation of the value iteration with large state or action spaces, the actual calculation in practice may result in the value iteration method not to satisfy the contraction condition $d\big(\mathcal{T}(V_{t+1}), \mathcal{T}(V_t)\big) \leq q \cdot d\big(V_{t+1}, V_t\big)$ in some iterations. Instead, the distance may expand up to a factor greater than one in some iterations of the value iteration, i.e., $d\big(\mathcal{T}(V_{t+1}), \mathcal{T}(V_t)\big) \leq Q \cdot d\big(V_{t+1}, V_t\big)$, where $Q \geq 1$. In this problem, the Bellman contraction mapping in value iteration may not be fixed anymore and could change over time. Hence, instead of applying the same transformation $\mathcal{T}$ in value iteration, a time-varying transformation $\mathcal{T}_t$ for $t \in \{0, 1, \dots\}$ may be applied to value iteration. Section 2.1 formalizes this observation.

## 2.1 Probabilistic Contraction-Expansion Mapping

Let $(X, \|\cdot\|)$ be a non-empty complete normed vector (linear) space, known as a Banach space, over the field $\mathbb{R}$ of real scalars, where $X$ is a vector space, e.g., a function space, together with a norm $\|\cdot\|$. The norm induces a translation invariant distance function, called canonical induced metric, as $d(f, g) = \|f - g\|$. Let $\|f\| = \langle f, f\rangle^{1/2}$, where the inner product of $f, g \in X$ in general is defined by $\langle f, g\rangle = \int f(x)g(x)dx$. Consider

a contraction mapping $\mathcal{T} : X \to X$ with the property that for all $f, g \in X$, there exists a scalar $q \in [0, 1)$ such that

$$d\big(\mathcal{T}(f), \mathcal{T}(g)\big) \leq q \cdot d(f, g). \tag{7}$$

In light of the Banach-Caccioppoli fixed-point theorem, this contraction mapping has its own unique fixed point, i.e., there exists $f^* \in X$ such that $\mathcal{T}(f^*) = f^*$. Furthermore, starting with an arbitrary function $f^0 \in X$, the sequence $\{f^n\}$ with $f^n = \mathcal{T}(f^{n-1})$ for $n \geq 1$ converges to $f^*$; in other words, $f^n \to f^*$, where $d\big(f^*, f^n\big) \leq \frac{q^n}{1-q} \cdot d(f^1, f^0)$. Note that in all iterations of the above value iteration, the mapping $\mathcal{T}$ operates as a contraction mapping according to (7) with probability one. However, in the rest of this subsection, we consider a probabilistic version of the Banach fixed-point theorem, where the mapping either contracts or expands the distance between any two points in a probabilistic manner.

Consider the time-varying function $f_t \in X$ for $t \in \{0, 1, 2, \dots\}$ evolving over time according to

$$f_{t+1} = \overline{\mathcal{T}}(f_t), \quad t \in \{0, 1, 2, \dots\}, \tag{8}$$

where $\overline{\mathcal{T}}$ is a probabilistic contraction-expansion mapping such that

$$d\big(\overline{\mathcal{T}}(f_{t+1}), \overline{\mathcal{T}}(f_t)\big) \leq \begin{cases} q \cdot d(f_{t+1}, f_t) & \text{w.p.} \quad p \\ Q \cdot d(f_{t+1}, f_t) & \text{otherwise} \end{cases}, \quad \forall t \in \mathbb{N}_0 \tag{9}$$

for some constants $q \in [0, 1)$, $Q \geq 1$, and $p \in (0, 1]$, where w.p. stands for "with probability" and $\mathbb{N}_0$ is natural numbers with zero. The expansion in (9) is caused by an adversary in an attempt to move the function sequence away from the fixed point. The contraction or expansion of $\overline{\mathcal{T}}$ is independent over time and $f^*$ is a fixed point of the mapping if $\overline{\mathcal{T}}(f^*) = f^*$. The shape of the function $f_t$ changes over time, but there can be a time, called hitting time $T$, at which $f_T$ reaches a neighborhood of $f^*$, as formally defined below.

**Definition 1.** *Given $\epsilon > 0$ and $a \in (0, 1]$, the hitting time $T(\epsilon, a)$ for the stochastic function sequence introduced in (8) is defined as*

$$T(\epsilon, a) = \min\big\{T : \mathbb{P}\big\{d\big(f_t, f^*\big) < \epsilon\big\} \geq 1 - a, \ \forall t \geq T\big\}, \tag{10}$$

*where $f^*$ is a fixed point whose existence and uniqueness is proven in Theorem 1 and $\mathbb{P}\{\cdot\}$ takes the probability of the input event.*

As a result, the complexity of optimizing the functions $f_t$ for $t < T$ can be irrelevant to the optimization complexity of the functions $f_t$ for $t \geq T$. Consequently, the hitting time $T$ together with the optimization complexity of any function $f_t$ for $t \geq T$ captures the complexity of optimizing the time-varying sequence of functions $\{f_t\}$. In the following theorem, the limiting behavior of the function sequence $\{f_t\}$ is studied and an upper bound on the hitting time is derived.

**Theorem 1.** ***Probabilistic Banach Fixed-Point Theorem.*** *Let $(X, \|\cdot\|)$ be a non-empty complete normed vector space with a probabilistic contraction-expansion mapping $\overline{\mathcal{T}} : X \to X$ defined in (9) such that $q^2 \cdot p + Q^2 \cdot (1-p) < 1$. Starting with an arbitrary element $f_0 \in X$, the sequence $\{f_t\}$ defined in (8) converges to an element $f^* \in X$ with an associated confidence level $1 - a$, where $f^*$ is a unique fixed point for the mapping $\overline{\mathcal{T}}$. Furthermore, for every $0 < L < \frac{\epsilon}{d(f_1, f_0)}$, the hitting time $T(\epsilon, a)$ satisfies the inequality*

$$T(\epsilon, a) \leq \max\left\{\frac{\ln\left(\frac{a \cdot L^2 \cdot \big(1 - q \cdot p - Q \cdot (1-p)\big) \cdot \big(1 - q^2 \cdot p - Q^2 \cdot (1-p)\big)}{1 + q \cdot p + Q \cdot (1-p)}\right)}{\ln\big(q^2 \cdot p + Q^2 \cdot (1-p)\big)}, \frac{\ln\Big(\Big(\frac{\epsilon}{d(f_1, f_0)} - L\Big) \cdot \Big(1 - q \cdot p - Q \cdot (1-p)\Big)\Big)}{\ln\big(q \cdot p + Q \cdot (1-p)\big)}\right\}. \tag{11}$$

*Proof.* In order to find an upper bound on the hitting time $T(\epsilon, a)$ defined in Definition 1, we first need to study the convergence behavior of the function sequence $\{f_t\}$ in (8) under the probabilistic contraction-expansion mapping $\overline{\mathcal{T}}$. To this end, we prove that this function sequence is a Cauchy sequence with high

probability. Given arbitrary integer values $n$ and $m$ such that $n > m$, one can write

$$\begin{aligned} d\big(f_n, f_m\big) = d\big(\overline{\mathcal{T}}^n(f_0), \overline{\mathcal{T}}^m(f_0)\big) &\overset{(a)}{\leq} \sum_{i=1}^{n-m} d\big(\overline{\mathcal{T}}^{n-i+1}(f_0), \overline{\mathcal{T}}^{n-i}(f_0)\big) = \sum_{i=1}^{n-m} d\big(\overline{\mathcal{T}}^{n-i}(f_1), \overline{\mathcal{T}}^{n-i}(f_0)\big) \\ &\overset{(b)}{\leq} \sum_{i=1}^{n-m} \left(\prod_{j=1}^{n-i} B_j\right) \cdot d\big(f_1, f_0\big) = d\big(f_1, f_0\big) \cdot \sum_{i=1}^{n-m} \prod_{j=1}^{n-i} B_j, \end{aligned} \tag{12}$$

where triangular inequality is applied $n-m-1$ times in $(a)$ and the independent and identically distributed random variables $B_j$ for $j \in \{1, 2, \ldots, n-1\}$ used in $(b)$ have the distribution

$$B_j = \begin{cases} q & \text{w.p.} \quad p \\ Q & \text{otherwise} \end{cases}. \tag{13}$$

Next, we study the mean and variance of the random variable $S_{n,m} = \sum_{i=1}^{n-m} \prod_{j=1}^{n-i} B_j$ in (12). Using the independence of $B_j$ for $j \in \{1, 2, \ldots, n-1\}$, the mean can be upper-bounded as

$$\mathbb{E}[S_{n,m}] = \mathbb{E}\left[\sum_{i=1}^{n-m} \prod_{j=1}^{n-i} B_j\right] = \sum_{i=1}^{n-m} \prod_{j=1}^{n-i} \mathbb{E}\left[B_j\right] = \sum_{i=1}^{n-m} \big(q \cdot p + Q \cdot (1-p)\big)^{n-i} \leq \frac{\big(q \cdot p + Q \cdot (1-p)\big)^m}{1 - q \cdot p - Q \cdot (1-p)}. \tag{14}$$

On the other hand, $\mathrm{Var}\left(S_{n,m}\right) \leq \mathbb{E}\left[S_{n,m}^2\right]$, where $\mathrm{Var}\left(\cdot\right)$ takes the variance of the input random variable, and the second moment of $S_{n,m}$ will be upper-bounded next. Note that

$$S_{n,m} = B_1 \cdot B_2 \cdots B_m \cdot \big(1 + B_{m+1} + B_{m+1} \cdot B_{m+2} + \cdots + B_{m+1} \cdots B_{n-1}\big). \tag{15}$$

Let $\bar{S}_{n,m} = 1 + B_{m+1} + B_{m+1} \cdot B_{m+2} + \cdots + B_{m+1} \cdots B_{n-1}$, where $\bar{S}_{n,m}$ is a random variable independent of $B_j$ for $j \in \{1, 2, \ldots, m\}$, and $\bar{S} = \lim_{n\to\infty} \bar{S}_{n,m}$. We leave out the subscript $m$ since the limits $\lim_{n\to\infty} \bar{S}_{n,m}$ and $\lim_{n\to\infty} \bar{S}_{n,m'}$ are identically distributed for all $m, m' \geq 0$. This is because $\bar{S}$ is an infinite sum and $\{B_j\}$ are i.i.d. random variables. Since $\mathbb{E}[B_j] > 0$ for $j \geq 1$, we have $\mathbb{E}[\bar{S}_{n,m}^2] \leq \mathbb{E}[\bar{S}^2]$; hence, it follows from (15) that

$$\mathbb{E}\left[S_{n,m}^2\right] = \mathbb{E}\left[B_1^2\right] \cdots \mathbb{E}\left[B_m^2\right] \cdot \mathbb{E}\left[\bar{S}_{n,m}^2\right] \leq \mathbb{E}\left[B_1^2\right] \cdots \mathbb{E}\left[B_m^2\right] \cdot \mathbb{E}\left[\bar{S}^2\right]. \tag{16}$$

In order to find an upper bound on $\mathbb{E}\left[\bar{S}^2\right]$, we have

$$\bar{S} = 1 + B_{m+1} \cdot (1 + B_{m+2} + B_{m+2} \cdot B_{m+3} + B_{m+2} \cdot B_{m+3} \cdot B_{m+4} + \ldots) = 1 + B_{m+1} \cdot \tilde{S}, \tag{17}$$

where $\tilde{S}$ is independent of $B_{m+1}$, and the random variables $\bar{S}$ and $\tilde{S}$ are identically distributed but not independent of each other. By taking expectation on both sides of $\bar{S}^2 = (1 + B_{m+1} \cdot \tilde{S})^2$, and using the independence of $\tilde{S}$ and $B_{m+1}$ and the fact that $\mathbb{E}\big[\bar{S}^2\big] = \mathbb{E}\big[\tilde{S}^2\big]$, one can obtain

$$\mathbb{E}\left[\bar{S}^2\right] = 1 + \mathbb{E}\left[B_{m+1}^2\right] \cdot \mathbb{E}\left[\tilde{S}^2\right] + 2\mathbb{E}\left[B_{m+1}\right] \cdot \mathbb{E}\left[\tilde{S}\right] \Longrightarrow \mathbb{E}\left[\bar{S}^2\right] = \frac{1 + 2\mathbb{E}\left[B_{m+1}\right] \cdot \mathbb{E}\left[\tilde{S}\right]}{1 - \mathbb{E}\left[B_{m+1}^2\right]}. \tag{18}$$

In the same way as finding the mean of $S_{n,m}$ in (14), it is derived that $\mathbb{E}\left[\tilde{S}\right] = \frac{1}{1 - q \cdot p - Q \cdot (1-p)}$; furthermore, $\mathbb{E}\left[B_{m+1}\right] = q \cdot p + Q \cdot (1-p)$ and $\mathbb{E}\left[B_{m+1}^2\right] = q^2 \cdot p + Q^2 \cdot (1-p)$. As a result, if $q^2 \cdot p + Q^2 \cdot (1-p) < 1$, Equation (18) results in

$$\mathbb{E}\left[\bar{S}^2\right] = \frac{1 + q \cdot p + Q \cdot (1-p)}{\big(1 - q \cdot p - Q \cdot (1-p)\big) \cdot \big(1 - q^2 \cdot p - Q^2 \cdot (1-p)\big)}. \tag{19}$$

Using Equation (16), we have

$$\mathrm{Var}\left(S_{n,m}\right) \leq \mathbb{E}\left[S_{n,m}^2\right] \leq \big(q^2 \cdot p + Q^2 \cdot (1-p)\big)^m \times \frac{1 + q \cdot p + Q \cdot (1-p)}{\big(1 - q \cdot p - Q \cdot (1-p)\big) \cdot \big(1 - q^2 \cdot p - Q^2 \cdot (1-p)\big)}. \tag{20}$$

So far, it is shown that $d\big(\overline{\mathcal{T}}^n(f_0), \overline{\mathcal{T}}^m(f_0)\big) \leq S_{n,m} \cdot d\big(f_1, f_0\big)$, where $S_{n,m}$ is a random variable with its mean and variance upper-bounded in (14) and (20), respectively. Using Chebyshev's inequality, for any $L > 0$, we have

$$\begin{aligned} &\mathbb{P}\left\{|S_{n,m} - \mathbb{E}[S_{n,m}]| \leq L\right\} \geq 1 - \frac{\mathrm{Var}\left(S_{n,m}\right)}{L^2} \Longrightarrow \\ &\mathbb{P}\left\{S_{n,m} \leq \frac{\big(q \cdot p + Q \cdot (1-p)\big)^m}{1 - q \cdot p - Q \cdot (1-p)} + L\right\} \geq 1 - \frac{\big(q^2 \cdot p + Q^2 \cdot (1-p)\big)^m \cdot \big(1 + q \cdot p + Q \cdot (1-p)\big)}{L^2 \cdot \big(1 - q \cdot p - Q \cdot (1-p)\big) \cdot \big(1 - q^2 \cdot p - Q^2 \cdot (1-p)\big)}. \end{aligned} \tag{21}$$

As a result, for any $\epsilon > 0$ and $a \in (0, 1]$, we have $d(f_n, f_m) = d\big(\overline{\mathcal{T}}^n(f_0), \overline{\mathcal{T}}^m(f_0)\big) \leq \epsilon$ with the confidence level $1 - a$ if $m$ satisfies the two inequalities

$$\frac{\big(q^2 \cdot p + Q^2 \cdot (1-p)\big)^m \cdot \big(1 + q \cdot p + Q \cdot (1-p)\big)}{L^2 \cdot \big(1 - q \cdot p - Q \cdot (1-p)\big) \cdot \big(1 - q^2 \cdot p - Q^2 \cdot (1-p)\big)} \quad \leq a \tag{22a}$$

$$\left(\frac{\big(q \cdot p + Q \cdot (1-p)\big)^m}{1 - q \cdot p - Q \cdot (1-p)} + L\right) \cdot d\big(f_1, f_0\big) \quad \leq \epsilon. \tag{22b}$$

Assume that $d\big(f_1, f_0\big) \neq 0$; otherwise, $f_0$ is a fixed point by definition. Hence, for $0 < L < \frac{\epsilon}{d(f_1, f_0)}$, if $q \cdot p + Q \cdot (1-p) < 1$ and $q^2 \cdot p + Q^2 \cdot (1-p) < 1$, then the two inequalities in (22a) and (22b) are satisfied when

$$m \geq \max\left\{\frac{\ln\left(\frac{a \cdot L^2 \cdot \big(1 - q \cdot p - Q \cdot (1-p)\big) \cdot \big(1 - q^2 \cdot p - Q^2 \cdot (1-p)\big)}{1 + q \cdot p + Q \cdot (1-p)}\right)}{\ln\big(q^2 \cdot p + Q^2 \cdot (1-p)\big)}, \frac{\ln\left(\left(\frac{\epsilon}{d(f_1, f_0)} - L\right) \cdot \big(1 - q \cdot p - Q \cdot (1-p)\big)\right)}{\ln\big(q \cdot p + Q \cdot (1-p)\big)}\right\}. \tag{23}$$

Now, for every $\epsilon > 0$ and $a \in (0, 1]$, let $N_\epsilon$ be the constant on the right-hand side of (23). Then, with probability $1 - a$, it holds that $\lim_{n \to \infty} d(f_n, f_{N_\epsilon}) \leq \lim_{n \to \infty} S_{n,N_\epsilon} \cdot d\big(f_1, f_0\big) \leq \epsilon$. For all $n > m > N_\epsilon$, since $\{B_j\}$ are nonnegative, it holds that $S_{n,m} = B_1 \cdot B_2 \cdots B_{N_\epsilon} \cdot (B_{N_\epsilon+1} \cdots B_m + \cdots + B_{N_\epsilon+1} \cdots B_{n-1}) \leq B_1 \cdot B_2 \cdots B_{N_\epsilon} \cdot (1 + B_{N_\epsilon+1} + B_{N_\epsilon+1} B_{N_\epsilon+2} + \dots) = \lim_{n \to \infty} S_{n,N_\epsilon}$, which implies $d(f_n, f_m) \leq S_{n,m} \cdot d\big(f_1, f_0\big) \leq \epsilon$ as long as $\lim_{n \to \infty} S_{n,N_\epsilon} \cdot d\big(f_1, f_0\big) \leq \epsilon$. To conclude, the sequence $\{f_t\}$ is a Cauchy sequence with probability $1 - a$. Since the vector space $X$ is complete, the sequence $\{f_t\}$ converges to an element $f^*$ in the space with high probability. Moreover, $f^*$ is a fixed point of the mapping $\overline{\mathcal{T}}$ since with high probability we have

$$\overline{\mathcal{T}}(f^*) = \overline{\mathcal{T}}(\lim_{t \to \infty} f_t) \overset{(a)}{=} \lim_{t \to \infty} \overline{\mathcal{T}}(f_t) = \lim_{t \to \infty} f_{t+1} = f^*, \tag{24}$$

where $(a)$ is true as the mapping $\overline{\mathcal{T}}$ is continuous due to (9), which justifies bringing the limit outside the operator $\overline{\mathcal{T}}$. Lastly, there cannot be more than one fixed point for the mapping $\overline{\mathcal{T}}$, which can be proved by contradiction. Considering any pair of distinct fixed points $f_1^*$ and $f_2^*$, we have $d\big(\overline{\mathcal{T}}(f_1^*), \overline{\mathcal{T}}(f_2^*)\big) = d\big(f_1^*, f_2^*\big)$ with probability 1, which contradicts the fact that the distance between the mapped points contracts with a factor $q < 1$ with probability $p > 0$.

In this proof, both $q \cdot p + Q \cdot (1-p) < 1$ and $q^2 \cdot p + Q^2 \cdot (1-p) < 1$ must be satisfied to ensure that Equations (22a) and (22b) hold for a large enough $m$. However, $q^2 \cdot p + Q^2 \cdot (1-p) < 1$ implies $q \cdot p + Q \cdot (1-p) < 1$ since one can write

$$\begin{aligned} (1-p) \cdot (Q^2 - 2Q + 1) \geq 0 \Longrightarrow \quad & Q^2 \quad \cdot (1-p) - 2Q \cdot (1-p) + 1 - p \geq 0 \\ \overset{(a)}{\Longrightarrow} \quad & Q^2 \cdot (1-p)^2 - 2Q \cdot (1-p) + 1 \geq p \cdot \big(1 - (1-p) \cdot Q^2\big) \\ \overset{(b)}{\Longrightarrow} \quad & 1 - Q \cdot (1-p) \geq p \cdot \sqrt{\frac{1 - Q^2 \cdot (1-p)}{p}} \\ \overset{(c)}{\Longrightarrow} \quad & q \cdot p + Q \cdot (1-p) < 1, \end{aligned} \tag{25}$$

where $p - p \cdot (1-p) \cdot Q^2$ is added on both sides of inequality in $(a)$, the square root is taken from both sides in $(b)$, and $q^2 \cdot p + Q^2 \cdot (1-p) < 1$ is used in $(c)$ to draw the claimed conclusion. $\square$

Theorem 1 states that if contraction of an operator in the iterates of the value iteration is compromised by an adversary via expansions in the iterates of value iteration, the value function sequence can still converge to the fixed point of the operator with high probability. The standard Banach fixed-point theorem is a special case of Theorem 1 by setting $p = 1$ and $L = 0$. The analysis in the proof of this theorem suggests that the compromised operator being contractive on expectation is not enough for the convergence of the value function sequence with high probability since the introduced randomness to the operator by the adversary can lead to high variance in the elements of the value function sequence. Hence, the additional assumption $q^2 \cdot p + Q^2 \cdot (1-p) < 1$ is required to bound such a variance rooted from the expansion caused by the adversary. Furthermore, this theorem provides an upper bound on the number of rounds for value iteration to defeat the effect of the adversary that attempts to move the value function sequence away from the fixed point. If the adversary is not modeled, the user who expects a normal scenario may perform fewer iterations of the value iteration. This can lead to a highly inaccurate estimate of the fixed point in the presence of an adversary.

**Remark 1.** *The parameter* $L \in \left(0, \frac{\epsilon}{d(f_1,f_0)}\right)$ *serves as an auxiliary parameter used in* (21)*. We observe that the first term in the upper bound* (11) *is decreasing with respect to* $L$ *and the second term is increasing with respect to* $L$*. By minimizing the bound* (11) *over* $L$*, we have that* $T(\epsilon, a)$ *has the order* $\mathcal{O}\left(\frac{d(f_1,f_0)}{\epsilon}\right)$*.*

## 2.2 Time-Varying Probabilistic Contraction-Expansion Mapping with Additive Noise

Let $(X, \|\cdot\|)$ be the same complete normed vector space as in Section 2.1. Consider time-varying probabilistic contraction-expansion mappings $\overline{\mathcal{T}}_t(\cdot) : X \to X$ for $t \in \{0, 1, 2, \dots\}$ with parameters $p_t, q_t$, and $Q_t$, i.e.,

$$d\big(\overline{\mathcal{T}}_t(f), \overline{\mathcal{T}}_t(g)\big) \leq \begin{cases} q_t \cdot d(f,g) & \text{w.p.} \quad p_t \\ Q_t \cdot d(f,g) & \text{otherwise} \end{cases}, \quad \forall t \in \mathbb{N}_0. \tag{26}$$

By Theorem 1, starting with an arbitrary function $f^0 \in X$, the sequence $\{f^n\}$ with $f^n = \overline{\mathcal{T}}_t(f^{n-1})$ for $n \geq 1$, where the same probabilistic contraction-expansion mapping $\overline{\mathcal{T}}_t$ is applied repeatedly, converges to $f_t^*$ with high probability.

**Assumption 1.** *The fixed points of every two consecutive mappings are at most* $\epsilon_f > 0$ *away from each other, i.e.,* $d\big(f_t^*, f_{t-1}^*\big) \leq \epsilon_f$ *for all* $t \in \{1, 2, 3, \dots\}$*.*

It is worth mention that, even under Assumption 1, there can be non-consecutive mappings $\overline{\mathcal{T}}_t$ and $\overline{\mathcal{T}}_{t'}$ whose fixed points are arbitrarily far away from each other. Note that in all iterations of the probabilistic value iteration, the same probabilistic contraction-expansion mapping $\overline{\mathcal{T}}_t$ is applied to the function sequence $\{f^n\}$. However, in the remainder of this subsection, we consider a time-varying and noisy version of the probabilistic Banach fixed-point theorem, where the underlying mapping changes over time and noise functions are added to the outcome of the mapping in each iteration.

Consider the time-varying function $f_t \in X$ for $t \in \{0, 1, 2, \dots\}$ evolving over time according to

$$f_{t+1} = \widetilde{\mathcal{T}}_t(f_t) = \overline{\mathcal{T}}_t(f_t) + w_t, \quad t \in \{0, 1, 2, \dots\}, \tag{27}$$

where $w_t \in X$ is some additive noise.

**Assumption 2.** *The additive noise is uniformly upper-bounded by a constant* $\epsilon_w > 0$*, i.e.,* $\|w_t\| \leq \epsilon_w$ *for all* $t \in \{0, 1, 2, \dots\}$*.*

Note that the shape of the function $f_t$ can change over time and can be non-convex. However, the following theorem shows that an upper bound can be established for the distance between $f_t$ and the time-varying fixed point $f_t^*$.

**Theorem 2.** *Consider arbitrary time-varying probabilistic contraction-expansion mappings* $\mathcal{T}_t$ *with fixed points* $f_t^*$*, where* $\sup_t \left(q_t^2 \cdot p_t + Q_t^2 \cdot (1-p_t)\right) < 1$ *for* $t \in \{0, 1, 2, \dots\}$*. Let the time-varying function* $f_t$ *evolve*

*over time according to the time-varying noisy probabilistic transformation in* (27)*. Under Assumptions 1 and 2, it holds that*

$$d\big(f_t, f_t^*\big) \leq P_t \cdot d\big(f_0, f_0^*\big) + S_t \cdot (\epsilon_f + \epsilon_w), \tag{28}$$

*where* $P_t = \left(\prod_{i=0}^{t-1} B_i\right)$ *and* $S_t = \left(1 + \sum_{i=1}^{t-1} \prod_{j=1}^{t-i} B_j\right)$ *are random variables with independent random variables* $B_t$ *having the distribution*

$$B_t = \begin{cases} q_t & w.p. \quad p_t \\ Q_t & otherwise \end{cases}. \tag{29}$$

*The means and variances of* $P_t$ *and* $S_t$ *are upper-bounded as*

$$\begin{aligned} \mathbb{E}\left[P_t\right] &\leq \left(\sup_t \big(q_t \cdot p_t + Q_t \cdot (1 - p_t)\big)\right)^t \xrightarrow{t\to\infty} 0, \\ \mathrm{Var}\left(P_t\right) &\leq \left(\sup_t \big(q_t^2 \cdot p_t + Q_t^2 \cdot (1 - p_t)\big)\right)^t \xrightarrow{t\to\infty} 0, \end{aligned} \tag{30}$$

*and*

$$\begin{aligned} \mathbb{E}\left[S_t\right] &\leq \frac{1}{1 - \sup_t \big(q_t \cdot p_t + Q_t \cdot (1 - p_t)\big)}, \\ \mathrm{Var}\left(S_t\right) &\leq \frac{\big(\bar{q}^2 \cdot \bar{p} + \bar{Q}^2 \cdot (1 - \bar{p})\big) \cdot \big(1 + \bar{q} \cdot \bar{p} + \bar{Q} \cdot (1 - \bar{p})\big)}{\big(1 - \bar{q}^2 \cdot \bar{p} - \bar{Q}^2 \cdot (1 - \bar{p})\big) \cdot \big(1 - \bar{q} \cdot \bar{p} - \bar{Q} \cdot (1 - \bar{p})\big)}, \end{aligned} \tag{31}$$

*where* $\bar{q}$*,* $\bar{Q}$*, and* $\bar{p}$ *satisfy* $\bar{q} \cdot \bar{p} + \bar{Q} \cdot (1 - \bar{p}) \geq \sup_{t\geq 1} \mathbb{E}[B_t]$ *and* $\bar{q}^2 \cdot \bar{p} + \bar{Q}^2 \cdot (1 - \bar{p}) \geq \sup_{t\geq 1} \mathbb{E}[B_t^2]$.

*Proof.* Under the time-varying probabilistic contraction-expansion mappings with added noise functions introduced in (27), the distance between $f_t$ and $f_t^*$ can be upper-bounded as

$$\begin{aligned} d\big(f_t, f_t^*\big) &= d\big(\widetilde{\mathcal{T}}_{t-1} \circ \cdots \circ \widetilde{\mathcal{T}}_0(f_0), f_t^*\big) \\ &\overset{(a)}{=} d\big(\overline{\mathcal{T}}_{t-1}\big(\widetilde{\mathcal{T}}_{t-2} \circ \cdots \circ \widetilde{\mathcal{T}}_0(f_0)\big) + w_{t-1}, f_t^*\big) \\ &= \big\|\overline{\mathcal{T}}_{t-1}\big(\widetilde{\mathcal{T}}_{t-2} \circ \cdots \circ \widetilde{\mathcal{T}}_0(f_0)\big) + w_{t-1} - f_t^*\big\| \\ &\overset{(b)}{\leq} d\big(\overline{\mathcal{T}}_{t-1}\big(\widetilde{\mathcal{T}}_{t-2} \circ \cdots \circ \widetilde{\mathcal{T}}_0(f_0)\big), f_t^*\big) + \|w_{t-1}\| \\ &\overset{(c)}{\leq} d\big(\overline{\mathcal{T}}_{t-1}\big(\widetilde{\mathcal{T}}_{t-2} \circ \cdots \circ \widetilde{\mathcal{T}}_0(f_0)\big), f_{t-1}^*\big) + d\big(f_{t-1}^*, f_t^*\big) + \|w_{t-1}\| \\ &\overset{(d)}{\leq} B_{t-1} \cdot d\big(\widetilde{\mathcal{T}}_{t-2} \circ \cdots \circ \widetilde{\mathcal{T}}_0(f_0), f_{t-1}^*\big) + \epsilon_f + \epsilon_w, \end{aligned} \tag{32}$$

where $\circ$ denotes the composition of linear operators, the definition of the mapping $\widetilde{\mathcal{T}}_{t-1}$ in (27) is used in $(a)$, inequalities $(b)$ and $(c)$ are true by the triangular inequality, and $(d)$ follows from Assumptions 1 and 2 in addition to the probabilistic contraction-expansion property of the operator $\overline{\mathcal{T}}_{t-1}$ and the fact that $\overline{\mathcal{T}}_{t-1}(f_{t-1}^*) = f_{t-1}^*$. Furthermore, the independent random variables $B_t$ for $t \geq 0$ used in $(d)$ have the distribution as specified in (29). Taking similar steps as in (32), we have

$$\begin{aligned} d\big(f_t, f_t^*\big) &\leq B_{t-1} \cdot \Big(B_{t-2} \cdot d\big(\widetilde{\mathcal{T}}_{t-3} \circ \cdots \circ \widetilde{\mathcal{T}}_0(f_0), f_{t-2}^*\big) + \epsilon_f + \epsilon_w\Big) + \epsilon_f + \epsilon_w \\ &\leq B_{t-1} \cdot \Big(B_{t-2} \cdot \Big(B_{t-3} \cdot d\big(\widetilde{\mathcal{T}}_{t-4} \circ \cdots \circ \widetilde{\mathcal{T}}_0(f_0), f_{t-3}^*\big) + \epsilon_f + \epsilon_w\Big) + \epsilon_f + \epsilon_w\Big) + \epsilon_f + \epsilon_w \\ &\leq \left(\prod_{i=0}^{t-1} B_i\right) \cdot d\big(f_0, f_0^*\big) + \left(1 + \sum_{i=1}^{t-1} \prod_{j=1}^{t-i} B_j\right) \cdot (\epsilon_f + \epsilon_w) \\ &\leq P_t \cdot d\big(f_0, f_0^*\big) + S_t \cdot (\epsilon_f + \epsilon_w), \end{aligned} \tag{33}$$

where $P_t = \left(\prod_{i=0}^{t-1} B_i\right)$ and $S_t = \left(1 + \sum_{i=1}^{t-1} \prod_{j=1}^{t-i} B_j\right)$ are random variables whose means and variances will be calculated below. Using the independence of random variables $B_t$ for $t \geq 0$, we have

$$\mathbb{E}\left[P_t\right] = \mathbb{E}\left[\prod_{i=0}^{t-1} B_i\right] = \prod_{i=0}^{t-1} \mathbb{E}\left[B_i\right] = \prod_{i=0}^{t-1} \big(q_t \cdot p_t + Q_t \cdot (1 - p_t)\big) \leq \left(\sup_t \big(q_t \cdot p_t + Q_t \cdot (1 - p_t)\big)\right)^t \tag{34}$$

and

$$\begin{aligned}\operatorname{Var}(P_t) &= \mathbb{E}\left[P_t^2\right] - \left(\mathbb{E}\left[P_t\right]\right)^2 \\ &= \mathbb{E}\left[\prod_{i=0}^{t-1} B_i^2\right] - \prod_{i=0}^{t-1}\left(q_t \cdot p_t + Q_t \cdot (1-p_t)\right)^2 \\ &\leq \prod_{i=0}^{t-1}\left(q_t^2 \cdot p_t + Q_t^2 \cdot (1-p_t)\right) \\ &\leq \left(\sup_t\left(q_t^2 \cdot p_t + Q_t^2 \cdot (1-p_t)\right)\right)^t .\end{aligned} \tag{35}$$

Note that it is already shown in (25) that $q_t^2 \cdot p_t + Q_t^2 \cdot (1-p_t) < 1$ implies $q_t \cdot p_t + Q_t \cdot (1-p_t) < 1$, and therefore it suffices to assume that $\sup_t\left(q_t^2 \cdot p_t + Q_t^2 \cdot (1-p_t)\right) < 1$. Furthermore,

$$\begin{aligned}\mathbb{E}\left[S_t\right] &= \mathbb{E}\left[1 + \sum_{i=1}^{t-1}\prod_{j=1}^{t-i} B_j\right] \\ &= 1 + \sum_{i=1}^{t-1}\prod_{j=1}^{t-i}\mathbb{E}\left[B_j\right] \\ &= 1 + \sum_{i=1}^{t-1}\prod_{j=1}^{t-i}\left(q_j \cdot p_j + Q_j \cdot (1-p_j)\right) \\ &\leq 1 + \sum_{i=1}^{t-1}\left(\sup_j\left(q_j \cdot p_j + Q_j \cdot (1-p_j)\right)\right)^{t-i} \\ &\leq \frac{1}{1 - \sup_j\left(q_j \cdot p_j + Q_j \cdot (1-p_j)\right)}\end{aligned} \tag{36}$$

and

$$\operatorname{Var}(S_t) = \operatorname{Var}\left(1 + \sum_{i=1}^{t-1}\prod_{j=1}^{t-i} B_j\right) = \operatorname{Var}\left(\sum_{i=1}^{t-1}\prod_{j=1}^{t-i} B_j\right) \leq \mathbb{E}\left[\left(\sum_{i=1}^{t-1}\prod_{j=1}^{t-i} B_j\right)^2\right]. \tag{37}$$

Consider the sequence of independent and identically distributed random variables $\bar{B}_t$ for $t \in \{1, 2, \dots\}$ that have the distribution

$$\bar{B}_t = \begin{cases} \bar{q} & \text{w.p. } \bar{p} \\ \bar{Q} & \text{otherwise} \end{cases} \tag{38}$$

such that $\mathbb{E}[\bar{B}_t] \geq \sup_{i\geq 1}\mathbb{E}[B_i]$ and $\mathbb{E}[\bar{B}_t^2] \geq \sup_{i\geq 1}\mathbb{E}[B_i^2]$. Proceeding with (37), one can write

$$\operatorname{Var}(S_t) \leq \mathbb{E}\left[\left(\sum_{i=1}^{t-1}\prod_{j=1}^{t-i} B_j\right)^2\right] \leq \mathbb{E}\left[\left(\sum_{i=1}^{t-1}\prod_{j=1}^{t-i} \bar{B}_j\right)^2\right] \leq \mathbb{E}\left[\left(\sum_{i=1}^{\infty}\prod_{j=1}^{i} \bar{B}_j\right)^2\right] = \mathbb{E}\left[\bar{S}^2\right], \tag{39}$$

where $\bar{S} = \sum_{i=1}^{\infty}\prod_{j=1}^{i}\bar{B}_j$. We have $\mathbb{E}[\bar{S}] = \frac{\bar{q}\cdot\bar{p}+\bar{Q}\cdot(1-\bar{p})}{1-\bar{q}\cdot\bar{p}-\bar{Q}\cdot(1-\bar{p})}$ and $\bar{S} = \bar{B}_1 \cdot (1 + \bar{B}_2 + \bar{B}_2 \cdot \bar{B}_3 + \cdots) = \bar{B}_1 \cdot (1 + \tilde{S})$, where $\tilde{S}$ is independent of $B_1$, and the random variables $\bar{S}$ and $\tilde{S}$ are identically distributed but not independent of each other. Taking expectation on both sides of $\bar{S}^2 = \bar{B}_1^2 \cdot (1+\tilde{S})^2$, and using the independence of $\tilde{S}$ and $B_1$ and the fact that $\mathbb{E}[\bar{S}^2] = \mathbb{E}[\tilde{S}^2]$, we have

$$\begin{aligned}\mathbb{E}[\bar{S}^2] = \mathbb{E}[\bar{B}_1^2]\cdot\mathbb{E}[1 + 2\tilde{S} + \tilde{S}^2] &= \left(\bar{q}^2\cdot\bar{p} + \bar{Q}^2\cdot(1-\bar{p})\right) \times \left(1 + \frac{2\left(\bar{q}\cdot\bar{p} + \bar{Q}\cdot(1-\bar{p})\right)}{1-\bar{q}\cdot\bar{p}-\bar{Q}\cdot(1-\bar{p})} + \mathbb{E}[\tilde{S}^2]\right) \\ \implies \mathbb{E}[\bar{S}^2] &= \frac{\left(\bar{q}^2\cdot\bar{p} + \bar{Q}^2\cdot(1-\bar{p})\right)\cdot\left(1 + \bar{q}\cdot\bar{p} + \bar{Q}\cdot(1-\bar{p})\right)}{\left(1-\bar{q}^2\cdot\bar{p} - \bar{Q}^2\cdot(1-\bar{p})\right)\cdot\left(1-\bar{q}\cdot\bar{p}-\bar{Q}\cdot(1-\bar{p})\right)}.\end{aligned} \tag{40}$$

Putting (39) and (40) together, it can be concluded that $\mathrm{Var}\,(S_t) \leq \frac{\left(\bar{q}^2\cdot\bar{p}+\bar{Q}^2\cdot(1-\bar{p})\right)\cdot\left(1+\bar{q}\cdot\bar{p}+\bar{Q}\cdot(1-\bar{p})\right)}{\left(1-\bar{q}^2\cdot\bar{p}-\bar{Q}^2\cdot(1-\bar{p})\right)\cdot\left(1-\bar{q}\cdot\bar{p}-\bar{Q}\cdot(1-\bar{p})\right)}$, which completes the proof. □

In the absence of the adversary, the probabilistic contraction-expansion mapping $\overline{\mathcal{T}}_t$ is purely a contraction with the rate $q_t$. We obtain the following corollary as a direct consequence of Theorem 2.

**Corollary 1.** *Consider arbitrary time-varying contraction mappings $\overline{\mathcal{T}}_t$ with the contraction constants $q_t$ and fixed points $f_t^*$. Suppose that $q = \sup_t q_t < 1$ and that Assumption 1 holds. Let the time-varying function $f_t$ evolve over time according to (27). For $\epsilon > 0$, we define the hitting time as $T(\epsilon) = \min\left\{T : d\big(f_t, f_t^*\big) < \epsilon,\ \forall t \geq T\right\}$. If $\epsilon \in (\frac{1}{1-q}\cdot(\epsilon_f+\epsilon_w), \frac{1}{1-q}\cdot(\epsilon_f+\epsilon_w)+D]$, then*

$$T(\epsilon) \leq 1 + \ln\left(\left(\epsilon - \frac{1}{1-q}\cdot(\epsilon_f+\epsilon_w)\right)\Big/ D\right)\Big/ \ln(q), \tag{41}$$

*where $\epsilon_w$ is an upper bound on the norm of each noise function and $D > 0$ is an upper bound on $d\big(f_0^*, f_0\big)$.*

*Proof.* When the time-varying mappings $\{\mathcal{T}_t\}$ are only contraction mappings, the random variable $B_t$ is equal to $q_t$ with probability 1 in (29). As a result, Equation (33) has the following form:

$$d\big(f_t, f_t^*\big) \leq q^t \cdot d\big(f_0, f_0^*\big) + \frac{1}{1-q}\cdot(\epsilon_f+\epsilon_w), \tag{42}$$

where we use $q = \sup_t q_t$. Since the right-hand side of (42) is decreasing in $t$, the hitting time $T(\epsilon)$ is upper-bounded by the minimum value of $t$ that satisfies $q^t\cdot d\big(f_0, f_0^*\big) + \frac{1}{1-q}\cdot(\epsilon_f+\epsilon_w) \leq \epsilon$. The proof is completed by noticing that $d\big(f_0^*, f_0\big)$ is upper-bounded by a constant $D > 0$ and $\frac{1}{1-q}\cdot(\epsilon_f+\epsilon_w) \leq \epsilon \leq \frac{1}{1-q}\cdot(\epsilon_f+\epsilon_w)+D$. □

Corollary 1 formalizes how many iterations are required in the value iteration with additive noise and a time-varying contraction operator – that can be caused by a time-varying environment – to guarantee that the ultimate function value is in an $\epsilon$-neighborhood of the fixed point.

**Remark 2.** *Tighter bounds on the hitting time for Theorems 1 and 2 may be obtained by applying concentration inequalities involving higher moments instead of Chebyshev's inequality. However, since our bounds already have logarithmic dependence on the relevant parameters p, Q, L, $\epsilon$, and $d(f_1, f_0)$, they are sufficient for most practical purposes as long as those parameters do not scale exponentially with the problem size.*

## 2.3 Optimization of Time-Varying Functions with Additive Noise

Consider the unknown time-varying continuous function $f_t : \mathcal{D} \to \mathcal{R}$ with the known bounded Lipschitz constant $K_t$, over the discrete-time horizon $t \in \{1, 2, \dots\}$, where $\mathcal{D} \subset \mathbb{R}^d$ is a compact set and $\mathcal{R} \subset \mathbb{R}$. The goal is to $\epsilon$-optimize the unknown time-varying function $f_t$, i.e., to find a possibly time-varying point $\widehat{x}_t^*$ such that $|f_t(\widehat{x}_t^*) - f_t(x_t^*)| \leq \epsilon$ for $\epsilon > 0$, where $x_t^* = \mathrm{argmin}_{x\in\mathcal{D}} f_t(x)$. Although the function $f_t$ is unknown, inquiries of the function values at given input points can be made in consecutive rounds, which are evaluated with added noise. More precisely, at round $t \in \{1, 2, \dots\}$, we consider querying the function $f_t$ on the set of input points $\mathcal{P} = \{x_1, \dots, x_n\} \subset \mathcal{D}$, and the revealed values are

$$\widetilde{f}_t(x_i) = f_t(x_i) + N_t(x_i), \tag{43}$$

where $N_t(x_i)$ is some noise satisfying the following assumption.

**Assumption 3.** *The noise parameters $N_t(x_i)$ are bounded i.i.d. random variables with zero mean, i.e., $\mathbb{E}[N_t(x_i)] = 0$, for which there exists $L_N > 0$ such that $[\sup\{N_t(x_i)\} - \inf\{N_t(x_i)\}] < L_N$ for all $t \in \{1, 2, \dots\}$ and $x_i \in \mathcal{P}$.*

If the noise is disruptive enough, a single set of observed noisy function values $f_t(x_i)$ for all $x_i \in \mathcal{P}$ may not represent the unknown target function accurately, making it impossible to $\epsilon$-optimize the function with a few number of observations. Furthermore, since the function changes over time, old observations may not

be useful in $\epsilon$-optimizing the time-varying function as $t$ increases. Putting these two facts into perspective, the estimate of the target function $f_t$ at round $t-1$, namely $\widehat{f}_{t-1}$, may need to be updated with the new observation at round $t$, while discarding inaccurate old observations. We propose the following formula for estimating $f_t$:

$$\widehat{f}_t(x_i) = \frac{\min\{t, T+1\}-1}{\min\{t,T\}} \cdot \widehat{f}_{t-1}(x_i) + \frac{1}{\min\{t,T\}} \cdot \widetilde{f}_t(x_i) - \frac{1}{T} \cdot \widetilde{f}_{t-T}(x_i) \cdot \mathbb{1}\{t > T\}, \tag{44}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function. The parameter $T$, whose value to be specified, should be chosen such that old data is discarded due to the time-varying nature of the function while not harming accurate estimation of the function value in the presence of noise. The computational cost of (44) is on the same order of that of the moving average update in reinforcement learning, but in (44) there is a need for storing the previous $T$ observations in order to have access to $\widetilde{f}_{t-T}(x_i)$.

The estimation function $\widehat{f}_t(x_i)$ changes over time and may not represent the target function for small values of $t$. However, there may exist a hitting time $T$ that is used in (44) after which optimizing the estimated function $\widehat{f}_t$ $\epsilon$-optimizes the target function $f_t$ with an associated confidence level $1-a$, where $0 < a \leq 1$. As a result, the complexity of $\epsilon$-optimizing the unknown time-varying target function $f_t$ in long-run is irrelevant to the complexity of optimizing function $\widehat{f}_t$ up to the hitting time $T$. Consequently, the hitting time $T$ as well as the optimization complexity of $\widehat{f}_t$ for $t \geq T$ captures the difficulty of $\epsilon$-optimizing the target function $f_t$ rather than the cumulative optimization complexities of functions $\widehat{f}_t$ for $t < T$. Formally speaking, the hitting time $T(\epsilon, a)$ is defined below.

**Definition 2.** *Given $\epsilon > 0$ and $a \in (0,1]$, the hitting time $T(\epsilon, a)$ is defined as*

$$T(\epsilon, a) = \min\Big\{T : \mathbb{P}\big(|f_t(\widehat{x}_t^*) - f_t(x_t^*)| \leq \epsilon\big) \geq 1-a,\ \forall t \geq T\Big\}, \tag{45}$$

*where $\widehat{x}_t^* = \operatorname{argmin}_{x\in\mathcal{P}} \widehat{f}_t(x)$ and $x_t^* = \operatorname{argmin}_{x\in\mathcal{D}} f_t(x)$.*

To make the time-varying problem amenable to optimization, we also make the following assumption about the set of input points $\mathcal{P}$.

**Assumption 4.** *For a given $\epsilon > 0$, the set of input points $\mathcal{P} = \{x_1, x_2, \ldots, x_n\}$ is a $\delta$-uniform grid of the function domain $\mathcal{D}$ such that $\delta < \frac{2\epsilon}{7\sqrt{d}K}$, where $K = \sup_{t\geq 1} K_t$ with $K_t$ being the Lipschitz constant of function $f_t$.*

Recall that being a $\delta$-uniform grid means that $\mathcal{P}$ satisfies two properties: (i) $\{x_i + \delta e_j, x_i - \delta e_j\} \cap \mathcal{D} \in \mathcal{P}$ for all $i \in \{1,\ldots,n\}$ and $j \in \{1,\ldots,d\}$, where $e_1, \ldots, e_d$ are the standard basis of $\mathbb{R}^d$, and (ii) for every $x \in \mathcal{D}$ there exists $x_i \in \mathcal{P}$ such that $\|x_i - x\| \leq \sqrt{d}\delta/2$. The fine granularity assumption, i.e., $\delta < \frac{2\epsilon}{7\sqrt{d}K}$, assures that there exists a grid point whose unknown function value at time $t$ is at least $\frac{\epsilon}{7}$ close to the minimum of function $f_t$. Denote such points of the grid $\mathcal{P}$ by $\mathcal{N}_t(\frac{\epsilon}{7}) = \{x_i \in \mathcal{P} : f_t(x_i) - f_t(x_t^*) \leq \frac{\epsilon}{7}\}$ and let $\overline{\mathcal{N}}_t(\epsilon) = \{x_i \in \mathcal{P} : f_t(x_i) - f_t(x_t^*) > \epsilon\}$. Without loss of generality, we assume that $\overline{\mathcal{N}}_t(\epsilon) \neq \emptyset$; otherwise, any point in $\mathcal{P}$ $\epsilon$-optimizes function $f_t$. The following theorem presents an upper bound on the hitting time.

**Theorem 3.** *Consider the unknown time-varying function $f_t$ with the property $|f_t(x) - f_{t-1}(x)| \leq \frac{\epsilon^3}{43L_N^2 \cdot \ln(\frac{n}{a})}$, for all $t \geq 1$ and $x \in \mathcal{D}$. Given $\epsilon > 0$ and $a \in (0,1]$, let Assumptions 3 and 4 hold. Then, the hitting time $T(\epsilon, a)$ satisfies the inequality*

$$T(\epsilon, a) \leq \frac{49L_N^2}{8\epsilon^2} \cdot \ln\left(\frac{n}{a}\right) + 1. \tag{46}$$

*Proof.* In order to find an upper bound on the hitting time $T(\epsilon, a)$, it is reasonable to assume that the function variation over time is upper-bounded; otherwise, there may not be enough time for learning the rapidly changing functions $\{f_t\}$. Assume that the time-variation of the unknown time-varying target function $f_t$ is upper-bounded by

$$|f_t(x) - f_{t-1}(x)| \leq \frac{\epsilon}{7T}, \quad \forall t \geq 1, \forall x \in \mathcal{D}. \tag{47}$$

Then, under Assumption 4, the hitting event defined in (45) satisfies the following condition

$$\begin{aligned}&\left\{\exists x_i \in \mathcal{N}_t(\frac{\epsilon}{7}) \text{ such that } \frac{1}{T}\cdot\sum_{s=t-T+1}^{t} N_s(x_i) \le \frac{2\epsilon}{7} \textbf{ and } \frac{1}{T}\cdot\sum_{s=t-T+1}^{t} N_s(x_i) \ge -\frac{2\epsilon}{7}, \forall x_i \in \overline{\mathcal{N}}_t(\epsilon)\right\} \\ &\subseteq \left\{\left|f_t(\widehat{x}_t^*) - f_t(x_t^*)\right| \le \epsilon\right\}, \quad \forall t \ge T.\end{aligned} \tag{48}$$

The above equation holds true because (43) and (44) result in $\widehat{f}_t(x_i) = \frac{1}{T}\cdot\sum_{s=t-T+1}^{t} f_s(x_i) + \frac{1}{T}\cdot\sum_{s=t-T+1}^{t} N_s(x_i)$ for $t \ge T$, and by (47), one can write

$$\begin{aligned}\widehat{f}_t(x_i) &\le f_t(x_i) + \frac{\epsilon}{7} + \frac{1}{T}\cdot\sum_{s=t-T+1}^{t} N_s(x_i), \quad \forall x_i \in \mathcal{N}_t(\frac{\epsilon}{7}), \\ \widehat{f}_t(\overline{x}_j) &\ge f_t(\overline{x}_j) - \frac{\epsilon}{7} + \frac{1}{T}\cdot\sum_{s=t-T+1}^{t} N_s(\overline{x}_j), \quad \forall \overline{x}_j \in \overline{\mathcal{N}}_t(\epsilon).\end{aligned} \tag{49}$$

Furthermore, $f_t(\overline{x}_j) - f_t(x_i) > \frac{6\epsilon}{7}$ for all $\overline{x}_j \in \overline{\mathcal{N}}_t(\epsilon)$ and $x_i \in \mathcal{N}_t(\frac{\epsilon}{7})$. Taking the difference of the two inequalities in (49) yields that $\widehat{f}_t(\overline{x}_j) - \widehat{f}_t(x_i) > \frac{4\epsilon}{7} + \sum_{s=t-T+1}^{t} N_s(\overline{x}_j) - \sum_{s=t-T+1}^{t} N_s(x_i)$. If the event on the left-hand side of (48) is true, then $\widehat{f}_t(\overline{x}_j) - \widehat{f}_t(x_i) > 0$, which means that there exists $\widetilde{x}_t^* \in \mathcal{N}_t(\frac{\epsilon}{7})$ whose estimated function value is less than the estimated function value at all points $\overline{x}_j \in \overline{\mathcal{N}}_t(\epsilon)$. Note that the estimated function value at a point $\overline{x}_t^* \in \mathcal{P} \setminus \left(\mathcal{N}_t(\frac{\epsilon}{7}) \cup \overline{\mathcal{N}}_t(\epsilon)\right)$ can be less than $\widehat{f}_t(\widetilde{x}_t^*)$, but such a point also $\epsilon$-optimizes the function $f_t$. Hence, $\widehat{x}_t^* = \operatorname{argmin}_{x\in\mathcal{P}} \widehat{f}_t(x)$ $\epsilon$-optimizes the function $f_t$, which means that the event on right-hand side of (48) is true.

Denote the event on the left-hand side of (48) as $E_t$, whose probability can be lower-bounded as

$$\begin{aligned}\mathbb{P}\{E_t\} &\overset{(a)}{\ge} \mathbb{P}\left\{\frac{1}{T}\cdot\sum_{s=t-T+1}^{t} N_s(x_i) \le \frac{2\epsilon}{7}, x_i \in \mathcal{N}_t(\frac{\epsilon}{7})\right\} \times \prod_{x_i\in\overline{\mathcal{N}}_t(\epsilon)} \mathbb{P}\left\{\frac{1}{T}\cdot\sum_{s=t-T+1}^{t} N_s(x_i) \ge -\frac{2\epsilon}{7}\right\} \\ &\overset{(b)}{\ge} \prod_{x_i\in\mathcal{P}} \left(1 - \exp\left(-\frac{8T\epsilon^2}{49L_N^2}\right)\right) \\ &> 1 - n\cdot\exp\left(-\frac{8T\epsilon^2}{49L_N^2}\right),\end{aligned} \tag{50}$$

where $(a)$ is true as the added noise signals are independent of each other and $(b)$ follows from Hoeffding's inequality and possibly multiplying by positive terms that are less than one. Putting (48) and (50) together, we have

$$\mathbb{P}\left\{\left|f_t(\widehat{x}_t^*) - f_t(x_t^*)\right| \le \epsilon\right\} \ge 1 - n\cdot\exp\left(-\frac{8T\epsilon^2}{49L_N^2}\right), \forall t \ge T. \tag{51}$$

If $1 - n\cdot\exp\left(-\frac{8T\epsilon^2}{49L_N^2}\right) \ge 1 - a$ or equivalently $T \ge \frac{49L_N^2}{8\epsilon^2}\cdot\ln\left(\frac{n}{a}\right)$, we have

$$\mathbb{P}\left\{\left|f_t(\widehat{x}_t^*) - f_t(x_t^*)\right| \le \epsilon\right\} \ge 1 - a, \quad \forall t \ge T. \tag{52}$$

As a result, an upper bound on the hitting time $T(\epsilon, a)$ defined in (45) is provided as

$$T(\epsilon, a) \le \frac{49L_N^2}{8\epsilon^2}\cdot\ln\left(\frac{n}{a}\right) + 1. \tag{53}$$

We substitute the upper bound on $T(\epsilon, a)$ into (47). It follows that the above analysis is valid if

$$|f_t(x) - f_{t-1}(x)| \le \frac{8\epsilon^3}{343L_N^2\cdot\ln(\frac{n}{a})}, \quad \forall t \ge 1, \forall x \in \mathcal{D}. \tag{54}$$

This completes the proof. □

**Remark 3.** *Note that the cardinality of the $\delta$-grid with $\delta < \frac{2\epsilon}{7\sqrt{d}K}$ used in Theorem 3, namely $n = |\mathcal{P}|$, depends on $\epsilon$. As an example, if $\mathcal{D}$ can be written as the Cartesian product of $d$ intervals of length at most $M$ as $\mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2 \times \cdots \times \mathcal{D}_d$, then the cardinality of the $\delta$-grid would be $n = \mathcal{O}\left(\left(\frac{\sqrt{d}KM}{\epsilon}\right)^d\right)$, and therefore the upper bound on the hitting time in Theorem 3 is given by $T(\epsilon, a) \leq \mathcal{O}\left(\frac{dL_N^2}{\epsilon^2} \cdot \ln\left(\frac{\sqrt{d}KM}{\sqrt[d]{a}\epsilon}\right)\right)$.*

Theorem 3 determines how fast the unknown function $f_t$ is allowed to change over time such that one can still learn the estimation function $\widehat{f}_t$ which is used to $\epsilon$-optimize the target function $f_t$ with a confidence level. The parameter $T$ in (44) can be set to the upper bound provided in Theorem 3 so that old inaccurate observations are discarded and at the same time enough observations are used for an accurate estimation of $f_t$.

## 2.4 Improved Bounds for Convex Functions

Consider the same framework as in Section 2.3 under additional assumptions to be stated here. Let $f_t$ be a convex function for all $t \geq 1$. Denote the lower contour set of the convex function $f_t$ by $C_t(c) = \{x \in \mathcal{D} : f_t(x) - f_t(x_t^*) \leq c\}$ and the level set of the convex function $f_t$ by $L_t(c) = \{x \in \mathcal{D} : f_t(x) - f_t(x_t^*) = c\}$ for $c > 0$. Define $\overline{C}_t(c_1, c_2) = \{x \in \mathcal{D} : c_1 < f_t(x) - f_t(x_t^*) \leq c_2\}$ when $c_2 > c_1$. Let $\mathcal{M}_t(c) = \{x_i \in \mathcal{P} : x_i \in C_t(c)\}$ and $\overline{\mathcal{M}}_t(c_1, c_2) = \{x_i \in \mathcal{P} : x_i \in \overline{C}_t(c_1, c_2)\}$.

**Assumption 5.** *There exists $M > 0$ such that $L_t(M)$ is homeomorphic to a $d$-dimensional sphere and is inside $\mathcal{D}$ for all $t \geq 1$.*

If $d = 1$ or $d = 2$, a sphere is defined as two distinctive points or a circle, respectively. Note that a lower bound on $M$ can be estimated up to a precision with high probability, but $M$ is assumed to be known to simplify the proof concepts.

**Assumption 6.** *There exists $k > 0$ such that $\|\nabla f_t(x)\| \geq k$, for all $t \geq 1$ and $x \in \mathcal{D} \setminus C_t(\epsilon)$.*

Intuitively, Assumption 6 requires every convex function $f_t$ have enough curvature inside its lower contour set $C_t(\epsilon)$, so that $\|\nabla f_t(x)\|$ can be uniformly lower-bounded by a positive constant $k$ in $\mathcal{D} \setminus C_t(\epsilon)$ for all $t \geq 1$.

Leveraging the new assumptions on the time-varying functions $\{f_t\}$, the following theorem presents a tighter upper bound on the hitting time compared to Theorem 3.

**Theorem 4.** *Consider the unknown time-varying convex function $f_t$ with the property $|f_t(x) - f_{t-1}(x)| \leq \frac{\epsilon^3}{43L_N^2 \cdot \ln(\frac{n}{a})}$, for all $t \geq 1$ and $x \in \mathcal{D}$. Given $\epsilon > 0$ and $a \in (0, 1]$, suppose that Assumptions 3-6 hold. Then, the hitting time $T(\epsilon, a)$ is upper-bounded by the minimum $T$ satisfying the inequality*

$$\sum_{l=0}^{l_m} n_l \cdot \exp\Big(-\frac{2T\big(l+\frac{2}{7}\big)^2\epsilon^2}{L_N^2}\Big) \leq a, \tag{55}$$

*where $\sum_{l=0}^{l_m} n_l = n$ and $l_m \leq \lfloor \frac{M}{\epsilon} \rfloor - 3$ such that $n_l = \frac{m_l}{1+m_l} \cdot n + 1$ for $l \in \{0, 1, \ldots, l_m - 1\}$ with $m_l = \frac{2^{d+1} \cdot K \cdot \epsilon}{k \cdot \big(M - (l+4)\epsilon\big)}$.*

*Proof.* Following the same logic as in (48) and leveraging the convexity of $\{f_t\}$, we obtain that the the hitting event in (45) satisfies the condition

$$\begin{aligned}
&\Big\{\exists x_i \in \mathcal{M}_t(\frac{\epsilon}{7}) \text{ such that } \frac{1}{T} \cdot \sum_{s=t-T+1}^{t} N_s(x_i) \leq \frac{2\epsilon}{7} \textbf{ and } \frac{1}{T} \cdot \sum_{s=t-T+1}^{t} N_s(x_i) \geq -\frac{2\epsilon}{7}, \forall x_i \in \overline{\mathcal{M}}_t\Big(\epsilon, 2\epsilon\Big) \textbf{ and} \\
&\quad \frac{1}{T} \cdot \sum_{s=t-T+1}^{t} N_s(x_i) \geq -\Big(l + \frac{2}{7}\Big)\epsilon, \forall x_i \in \overline{\mathcal{M}}_t\Big((l+1)\epsilon, (l+2)\epsilon\Big), \forall 1 \leq l \leq \Big\lfloor \frac{M}{\epsilon} \Big\rfloor \Big\} \\
&\subseteq \big\{\big|f_t(\widehat{x}_t^*) - f_t(x_t^*)\big| \leq \epsilon\big\}, \quad \forall t \geq T.
\end{aligned} \tag{56}$$

Denote the event on the left-hand side of (56) as $E_t$, whose probability can be lower-bounded as

$$\begin{aligned}\mathbb{P}\{E_t\} &\overset{(a)}{\geq} \mathbb{P}\left\{\frac{1}{T}\cdot\sum_{s=t-T+1}^{t} N_s(x_i) \leq \frac{2\epsilon}{7}, x_i \in \mathcal{M}_t(\tfrac{\epsilon}{7})\right\} \times \prod_{x_i\in\overline{\mathcal{M}}_t\left(\epsilon,2\epsilon\right)} \mathbb{P}\left\{\frac{1}{T}\cdot\sum_{s=t-T+1}^{t} N_s(x_i) \geq -\frac{2\epsilon}{7}\right\} \\ &\quad\times \prod_{l=1}^{\lfloor\frac{M}{\epsilon}\rfloor} \prod_{x_i\in\overline{\mathcal{M}}_t\left((l+1)\epsilon,(l+2)\epsilon\right)} \mathbb{P}\left\{\frac{1}{T}\cdot\sum_{s=t-T+1}^{t} N_s(x_i) \geq -\big(l+\frac{2}{7}\big)\epsilon\right\} \\ &\overset{(b)}{\geq} \left[1-\exp\left(-\frac{8T\epsilon^2}{49L_N^2}\right)\right]^{\overline{n}_0+1} \times \prod_{l=1}^{l_m}\left[1-\exp\left(-\frac{2T\big(l+\frac{2}{7}\big)^2\epsilon^2}{L_N^2}\right)\right]^{n_l} \\ &\geq 1-\sum_{l=0}^{l_m} n_l\cdot\exp\left(-\frac{2T\big(l+\frac{2}{7}\big)^2\epsilon^2}{L_N^2}\right)\end{aligned} \tag{57}$$

where $(a)$ is true as the added noise signals are independent of each other and $(b)$ follows from Hoeffding's inequality, $\overline{n}_0$ is an upper bound on the number of grid points in the set $\overline{\mathcal{M}}_t\big(\epsilon,2\epsilon\big)$ and $n_0=\overline{n}_0+1$, and $n_l$ is an upper bound on the number of grid points in the set $\overline{\mathcal{M}}_t\big((l+1)\epsilon,(l+2)\epsilon\big)$, where $l_m$ satisfies $\sum_{l=0}^{l_m} n_l = n$ and $l_m \leq \lfloor\frac{M}{\epsilon}\rfloor - 3$. Note that the last nonzero $n_l$ is not a free parameter since the sum of all $n_l$ should be $n$. Putting (56) and (57) together, we have $\mathbb{P}\left\{\left|f_t(\widehat{x}_t^*)-f_t(x_t^*)\right|\leq\epsilon\right\}\geq 1-a$ for all $t\geq T$ provided that

$$\sum_{l=0}^{l_m} n_l\cdot\exp\Big(-\frac{2T\big(l+\frac{2}{7}\big)^2\epsilon^2}{L_N^2}\Big) \leq a, \tag{58}$$

which provides an upper bound on the hitting time $T(\epsilon,a)$ defined in (45). As stated earlier in (47), the above analysis is true if $|f_t(x)-f_{t-1}(x)|\leq\frac{\epsilon}{7T(\epsilon,a)}$ for all $t\geq 1$ and $x\in\mathcal{D}$. Using the general upper bound on the hitting time provided in Theorem 3, the analysis holds if $|f_t(x)-f_{t-1}(x)|\leq\frac{\epsilon^3}{43L_N^2\cdot\ln(\frac{n}{a})}$ for all $t\geq 1$ and $x\in\mathcal{D}$.

In the rest of the proof, the values of $n_l$ for $0\leq l\leq l_m$ are computed. The key ideas behind finding these upper bounds are that the level sets $\overline{L}_t\big((l+1)\epsilon\big)$ for $0\leq l\leq l_m+2$ are nested surfaces that are homeomorphic to a $d$-dimensional sphere inside the function domain and that the minimum distance between any point of a level set from any of the other level set is controlled by $K$ and $k$. Let $Vol(\cdot)$ denote the volume of an input $d$-dimensional set and $A(\cdot)$ denote the area of an input $(d-1)$-dimensional surface. By convention, the area of a $d$-dimensional sphere for $d=1$ and $d=2$ is equal to 2 and the length of the sphere, respectively. For every $l\in\{0,1,\ldots,l_m\}$, one can write

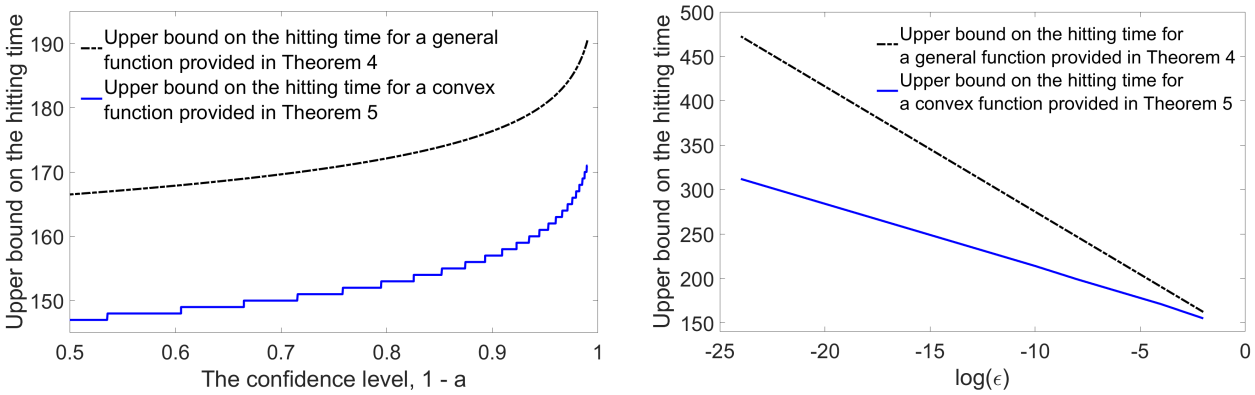
$$\begin{aligned} n_l - 1 &\leq \frac{2^d\cdot Vol\left(C_t\big((l+1)\epsilon,(l+3)\epsilon\big)\right)}{\delta^d} \leq \frac{2^d\cdot\frac{2\epsilon}{k}\cdot A\left(P_t\big((l+1)\epsilon,(l+3)\epsilon\big)\right)}{\delta^d}, \\ \sum_{\bar{l}=l+1}^{l_m} n_{\bar{l}} &\geq \frac{Vol\left(C_t\big((l+3)\epsilon,M-\epsilon\big)\right)}{\delta^d} \geq \frac{\frac{M-(l+4)\epsilon}{K}\cdot A\left(P_t\big((l+3)\epsilon,M-\epsilon\big)\right)}{\delta^d}, \end{aligned} \tag{59}$$

where the term $2^d$ comes from the facts that each $d$-dimensional cube has at most $2^d$ endpoints and $P_t\big((l+1)\epsilon,(l+3)\epsilon\big)\subset C_t\big((l+1)\epsilon,(l+3)\epsilon\big)$ and $P_t\big((l+3)\epsilon,M-\epsilon\big)\subset C_t\big((l+3)\epsilon,M-\epsilon\big)$ are two $(d-1)$-dimensional planes such that $A\left(P_t\big((l+1)\epsilon,(l+3)\epsilon\big)\right)\leq A\left(L_t\big((l+3)\epsilon\big)\right)\leq A\left(P_t\big((l+3)\epsilon,M-\epsilon\big)\right)$. Then,

$$\frac{n_l-1}{n-n_l}\leq\frac{n_l-1}{\sum_{\bar{l}=l+1}^{l_m} n_{\bar{l}}}\leq\frac{2^{d+1}\cdot K\cdot\epsilon}{k\cdot\big(M-(l+4)\epsilon\big)}=m_l \implies n_l\leq\frac{m_l}{1+m_l}\cdot n+1, \tag{60}$$

which completes the proof. □

**Remark 4.** *We note that, since the left-hand side of* (55) *is monotone decreasing in* $T$*, a number* $T$ *satisfying* (55) *always exists. By substituting the bound in* (46) *into* (55)*, it can be verified that Theorem 4 provides a better bound than Theorem 3 since some properties of convex functions are leveraged. A comparison of the results of Theorems 3 and 4 along with the simulation details is depicted in Figure 1.*




(a) $\epsilon = L = 10^{-4}, n = 3.2 \times 10^{11}$, and $a$ varies.

(b) $a = 0.01$ and $\epsilon = L$ varies.

Figure 1: A comparison of the upper bounds in Theorems 3 and 4 when $M = K = 16, k = 2 \times 10^{-2}$, and $d = 2$. In Figure 1b, the value of $n$ depends on $\epsilon$, which is taken into account for drawing the plots.

# 3 The Hitting Time Analysis for Discrete Functions

In this section, two variants of stochastic time-varying models are studied for discrete functions. In the first model, an unknown discrete function is observed with additive noise whose estimation function changes over time due to the presence of noise. In the second model, a time-varying linear model with additive noise is studied.

## 3.1 Optimization of Functions with Additive Noise

Consider an unknown discrete function $f : \mathcal{X} \to \mathcal{R}$, where $\mathcal{X} \subset \mathbb{Z}^d$ is a bounded subset of $d$ integer tuples and $\mathcal{R} \subset \mathbb{R}$ is a subset of real numbers ($\mathbb{Z}$ denotes the set of integer numbers). Denote the strict local minima and maxima, known collectively as strict local extrema, of the unknown function $f$ by $\mathcal{X}^*$ defined as

$$\mathcal{X}^* = \{x^* \in \mathcal{X} : f(x^*) < f(x), \forall x \in \mathcal{B}(x^*)\} \cup \{x^* \in \mathcal{X} : f(x^*) > f(x), \forall x \in \mathcal{B}(x^*)\} \tag{61}$$

where $\mathcal{B}(x^*) = \cup_{j=1}^{d} \{x^* + h_j, x^* - h_j\} \cap \mathcal{X}$ with $h_1, \dots, h_d$ being the standard basis of $\mathbb{Z}^d$. The goal is to find $\mathcal{X}^*$, the set of strict local extrema of the unknown function $f$. Although the function $f$ is unknown, inquiries of the function values at points in the domain can be made in consecutive rounds, which are evaluated with added noise signals that are mean zero, independent and identically distributed over time and over $\mathcal{X}$. Formally speaking, the revealed values of the target function $f$ at round $t \in \{1, 2, \dots\}$ are

$$f_t(x) = f(x) + N_t(x), \quad \forall x \in \mathcal{X}, \tag{62}$$

where $N_t(x)$ are noise signals satisfying Assumption 3. Note that if the noise is disruptive enough, a single set of observed noisy function values $f_t(x)$ for all $x \in \mathcal{X}$ may not represent the unknown target function accurately, making it impossible to find local extrema of the function. To address this issue, we estimate the target function $f$ at round $t - 1$ by leveraging the new observations at round $t \in \{2, 3, \dots\}$ as

$$\widehat{f}_t(x) = \frac{t-1}{t} \cdot \widehat{f}_{t-1}(x) + \frac{1}{t} \cdot f_t(x), \quad \forall x \in \mathcal{X}. \tag{63}$$

Note that the estimation function $\widehat{f}_t(x)$ changes over time and may not represent the shape of the unknown target function $f$ when $t$ is small. However, there may exist a hitting time $T$ after which the estimation function $\widehat{f}_t$ shares the same set of local extrema as the target function $f$ with an associated confidence level $1 - a$, where $0 < a \leq 1$. As a result, the complexity of finding the local extrema of the target function $f$ may be irrelevant to the complexity of finding the local extrema of function $\widehat{f}_t$ before the hitting time $T$.

Consequently, the complexity of finding the local extrema of the unknown target function $f$ is related to the hitting time $T$ as well as the computational complexity of optimizing function $\widehat{f}_T$. Denote the set of strict local extrema of $\widehat{f}_t$ by $\widehat{\mathcal{X}}_t^*$, defined as

$$\widehat{\mathcal{X}}_t^* = \left\{\widehat{x}^* \in \mathcal{X} : \widehat{f}_t(\widehat{x}^*) < \widehat{f}_t(x), \forall x \in \mathcal{B}(\widehat{x}^*)\right\} \cup \left\{\widehat{x}^* \in \mathcal{X} : \widehat{f}_t(\widehat{x}^*) > \widehat{f}_t(x), \forall x \in \mathcal{B}(\widehat{x}^*)\right\}. \tag{64}$$

**Definition 3.** *Given $a \in (0, 1]$, the hitting time $T(a)$ for an unknown discrete function $f$ is defined as*

$$T(a) = \min\left\{T : \mathbb{P}\left(\widehat{\mathcal{X}}_t^* = \mathcal{X}^*\right) \geq 1 - a,\ \forall t \geq T\right\}, \tag{65}$$

*where $\mathcal{X}^*$ and $\widehat{\mathcal{X}}_t^*$ are defined in* (61) *and* (64)*, respectively.*

The hitting time $T(a)$ depends on the minimum distance of the function values of $f$ at point $x \in \mathcal{X}$ from the function values at its neighbor points. This distance, denoted by $\delta(x)$, is defined as

$$\delta(x) = \min_{x' \in \mathcal{B}(x)} |f(x) - f(x')|. \tag{66}$$

In order to simply the analysis, we make the following assumption about the target function $f$.

**Assumption 7.** *The minimum distance $\delta(x)$ of function $f$ is uniformly lower-bounded by a positive number for all $x \in \mathcal{X}$, i.e., $\delta_m = \min_{x \in \mathcal{X}} \delta(x) > 0$.*

Intuitively, Assumption 7 ensures that function values of $f$ at adjacent points are different, so that their noisy values become distinguishable after enough observations. The following theorem presents an upper bound on the hitting time $T(a)$.

**Theorem 5.** *Consider the time-varying function $\widehat{f}_t$ in* (63)*. Under Assumptions 3 and 7, given $a \in (0, 1]$, the associated hitting time $T(a)$ defined in* (65)*, satisfies the inequality*

$$T(a) \leq \frac{2L_N^2}{\delta_m^2} \cdot \ln\left(\frac{2|\mathcal{X}|}{a}\right), \tag{67}$$

*where $|\mathcal{X}|$ denotes the number of elements in the set $\mathcal{X}$.*

*Proof.* In order to find an upper bound on the hitting time $T(a)$, note that the hitting event used in (65) satisfies the condition

$$\left\{\frac{1}{T} \cdot \left\|\sum_{t=1}^{T} N_t(x)\right\| < \frac{\delta(x)}{2},\ \forall x \in \mathcal{X}\right\} \subseteq \left\{\widehat{\mathcal{X}}_T^* = \mathcal{X}^*\right\}. \tag{68}$$

The above equation holds because (62) and (63) result in $\widehat{f}_T(x) = f(x) + \frac{1}{T} \cdot \sum_{t=1}^{T} N_t(x)$, and if the magnitude of the noise added to the true value of function $f$ at point $x$ is less than $\delta(x)/2$ for all $x \in \mathcal{X}$, then the set of local extrema of the function $\widehat{f}_T$ coincides with the set $\mathcal{X}^*$, the local extrema of function $f$. The probability of the event on the left-hand side of (68) can be lower-bounded as

$$\begin{aligned}
\mathbb{P}\left\{\frac{1}{T} \cdot \left\|\sum_{t=1}^{T} N_t(x)\right\| < \frac{\delta(x)}{2},\ \forall x \in \mathcal{X}\right\} &\overset{(a)}{=} \prod_{i=1}^{|\mathcal{X}|} \mathbb{P}\left\{\frac{1}{T} \cdot \left\|\sum_{t=1}^{T} N_t(x)\right\| < \frac{\delta(x)}{2}\right\} \\
&\overset{(b)}{\geq} \prod_{i=1}^{|\mathcal{X}|} \left(1 - 2\exp\left(-\frac{T\delta(x)^2}{2L_N^2}\right)\right) \\
&> 1 - 2\sum_{i=1}^{|\mathcal{X}|} \exp\left(-\frac{T\delta(x)^2}{2L_N^2}\right) \\
&\geq 1 - 2|\mathcal{X}| \cdot \exp\left(-\frac{T\delta_m^2}{2L_N^2}\right),
\end{aligned} \tag{69}$$

where $(a)$ holds because the added noise signals are independent from each other and $(b)$ follows from Hoeffding's inequality. Putting (68) and (69) together, we have

$$\mathbb{P}\left\{\widehat{\mathcal{X}}_T^* = \mathcal{X}^*\right\} > 1 - 2|\mathcal{X}| \cdot \exp\left(-\frac{T\delta_m^2}{2L_N^2}\right). \tag{70}$$

If $1 - 2|\mathcal{X}| \cdot \exp\left(-\frac{T\delta_m^2}{2L_N^2}\right) \geq 1 - a$ or equivalently $T \geq \frac{2L_N^2}{\delta_m^2} \cdot \ln\left(\frac{2|\mathcal{X}|}{a}\right)$, we have $\mathbb{P}\left\{\widehat{\mathcal{X}}_T^* = \mathcal{X}^*\right\} > 1 - a$, from which the upper bound in (65) follows. □

## 3.2 A Special Case for Unimodal Functions

A function $f$ over a bounded set $\mathcal{X} \subset \mathbb{Z}$ is called unimodal if it has only one global minimum $x^* \in \mathcal{X}$ and $f(i) > f(j)$ for all $i < j \leq x^*$, $i, j \in \mathcal{X}$, while $f(i) < f(j)$ for all $x^* \leq i < j$. Assume that the unknown target function $f$ is unimodal over $\mathcal{X}$, which implies it has a single global minimum. As mentioned earlier, the time-varying function $\widehat{f}_t$ may not even be unimodal for small values of $t$ under disruptive noise, and therefore it could have multiple local extrema. However, the single global minimum of the function $f$ becomes known after the hitting time with an associated confidence level. In this section, a new notion of hitting time is proposed for unimodal functions that captures the complexity of finding the global minimum of the function and does not take the local extrema of the estimated function $\widehat{f}_t$ into account.

Without loss of generality, we additionally assume that the noise signals $N_t(x)$ are continuous random variables. This implies that the estimation function $\widehat{f}_t$ has a single global minimum with probability 1. Let $\widehat{x}_t^* = \operatorname{argmin}_{x \in \mathcal{X}} \ \widehat{f}_t(x)$ denote the global minimum. The hitting time for a unimodal function $f$ is defined below.

**Definition 4.** *Given $a \in (0, 1]$, the hitting time $T_u(a)$ for a unimodal function $f$ with its global minimum at $x^* = \operatorname{argmin}_{x \in \mathcal{X}} \ f(x)$ and its estimated global minimum $\widehat{x}_t^* = \operatorname{argmin}_{x \in \mathcal{X}} \ \widehat{f}_t(x)$ is defined as*

$$T_u(a) = \min\left\{T : \mathbb{P}\big(\widehat{x}_t^* = x^*\big) \geq 1 - a, \ \forall t \geq T\right\}. \tag{71}$$

The distance of the function value at point $x \in \mathcal{X}$ from the minimum function value is denoted by $\Delta(x)$, which is defined as

$$\Delta(x) = \begin{cases} f(x) - f(x^*), & \text{if } x \in \mathcal{X} \setminus \{x^*\}, \\ \min\{f(x^* - 1) - f(x^*), f(x^* + 1) - f(x^*)\}, & \text{if } x = x^*. \end{cases} \tag{72}$$

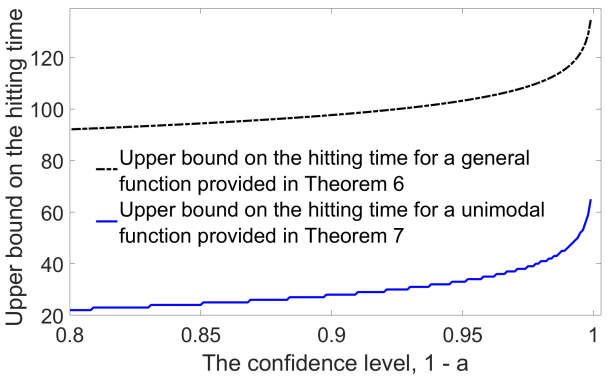
The following theorem presents an upper bound on the hitting time for a unimodal function.

**Theorem 6.** *Consider the time-varying function $\widehat{f}_t$ defined in (63) with $f$ being a unimodal function. Suppose that Assumptions 3 and 7 hold. Given $a \in (0, 1]$, the associated hitting time $T_u(a)$ satisfies the inequality $T_u(a) \leq T$, where $T$ is the smallest number such that*

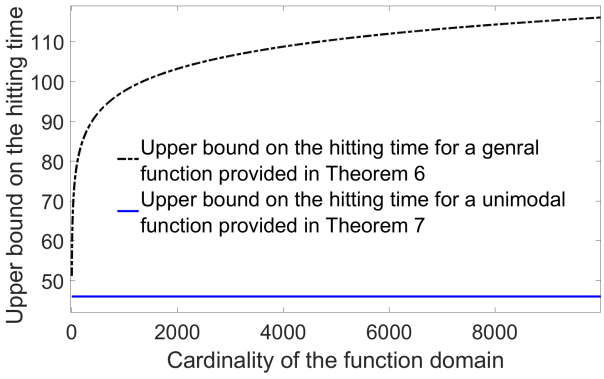
$$\exp\left(-\frac{\delta_m^2 T}{2L_N^2}\right) + 2 \sum_{i \in [\lfloor |\mathcal{X}|/2 \rfloor]} \exp\left(-\frac{i^2 \delta_m^2 T}{2L_N^2}\right) \leq a. \tag{73}$$

*Proof.* By construction, we have $\Delta(x) > 0$ for all $x \in \mathcal{X}$. In order to find an upper bound on the hitting time $T_u(a)$, note that the hitting event used in (71) satisfies the condition

$$\left\{\frac{1}{T} \cdot \sum_{t=1}^{T} N_t(x) > -\frac{\Delta(x)}{2}, \forall x \in \mathcal{X} \setminus \{x^*\} \ \textbf{and} \ \frac{1}{T} \cdot \sum_{t=1}^{T} N_t(x^*) < \frac{\Delta(x^*)}{2}\right\} \subseteq \left\{\widehat{x}_T^* = x^*\right\}. \tag{74}$$




(a) $|\mathcal{D}| = 10000$ and $a$ varies.




(b) $a = 0.01$ and $|\mathcal{D}|$ varies.

Figure 2: A comparison of the upper bounds in Theorems 5 and 6 when $L_N = 0.02$ and $\delta_m = 0.01$.

Denote the event on the left-hand side of (74) as $E_t$, whose probability can be lower-bounded as

$$\begin{aligned}
\mathbb{P}\{E_t\} &\overset{(a)}{=} \mathbb{P}\left\{\frac{1}{T}\cdot\sum_{t=1}^{T} N_t(x^*) < \frac{\Delta(x^*)}{2}\right\} \times \prod_{x\in\mathcal{X}\setminus\{x^*\}} \mathbb{P}\left\{\frac{1}{T}\cdot\sum_{t=1}^{T} N_t(x) > -\frac{\Delta(x)}{2}\right\} \\
&\overset{(b)}{\geq} \left(1-\exp\left(-\frac{T\Delta(x^*)^2}{2L_N^2}\right)\right) \times \prod_{x\in\mathcal{X}\setminus\{x^*\}} \left(1-\exp\left(-\frac{T\Delta(x)^2}{2L_N^2}\right)\right) \\
&> 1-\exp\left(-\frac{T\Delta(x^*)^2}{2L_N^2}\right) - \sum_{x\in\mathcal{X}\setminus\{x^*\}} \exp\left(-\frac{T\Delta(x)^2}{2L_N^2}\right) \\
&\overset{(c)}{\geq} 1-\exp\left(-\frac{T\delta_m^2}{2L_N^2}\right) - \sum_{x\in\mathcal{X}\setminus\{x^*\}} \exp\left(-\frac{T(x-x^*)^2\delta_m^2}{2L_N^2}\right) \\
&\overset{(d)}{\geq} 1-\exp\left(-\frac{T\delta_m^2}{2L_N^2}\right) - 2\sum_{i\in[\lfloor|\mathcal{X}|/2\rfloor]} \exp\left(-\frac{Ti^2\delta_m^2}{2L_N^2}\right)
\end{aligned} \tag{75}$$

where $(a)$ holds true by the independence property of the added noise signals, $(b)$ is due to Hoeffding's inequality, $(c)$ is true because function $f$ is unimodal, $\Delta(x^*) \geq \delta_m$, and $\Delta(x) \geq (x-x^*)\delta_m$, and $(d)$ results from minimizing the equation with respect to all possible values of $x^*$, which gives rise to $x^* = \lceil|\mathcal{X}|/2\rceil$ (taking the ceiling corresponding to the summation through $\lfloor|\mathcal{X}|/2\rfloor$). Putting (74) and (75) together concludes the proof. □

**Remark 5.** *A number $T$ that satisfies* (73) *must exists because the left-hand side of* (73) *approaches* 0 *when $T \to \infty$. Also, by substituting the bound in* (67) *into* (73)*, it can be verified that Theorem 6 provides a better bound than Theorem 5 as the properties of unimodal functions are leveraged. A comparison of the results of Theorems 5 and 6 along with the details of the simulation model is depicted in Figure 2.*

### 3.3 Time-Varying Linear Model with Additive Noise

In this section, we study a linear model of time-variation and analyze the hitting time under shape-dominant operators. Consider the Hilbert space $L^2(\mathcal{X})$, where the inner product of $f$ and $g \in L^2(\mathcal{X})$ is defined by $\langle f, g\rangle = \int_{\mathcal{X}} f(x)g(x)dx$. We use the same inner product notation when the domain $\mathcal{X}$ is a discrete set. For any nonzero functions $f, g \in L^2$, there exists a bounded linear transformation $\mathcal{T} : L^2(\mathcal{X}) \to L^2(\mathcal{X})$ such that $\mathcal{T}f = g$. In fact, one such transformation is given by $\mathcal{T}h = \frac{\langle f,h\rangle}{\langle f,f\rangle} g$. Since the zero function is trivial to optimize, the restriction to linear transformation is a general framework that captures the varying nature of nonlinear functions.

We further note that for any scalar $\lambda > 0$, the functions $f$ and $\lambda f$ share the same set of local minima. Rescaling by a positive number does not affect the complexity of the optimization problem. Hence, restricting the linear operators $\mathcal{T}$ to have norm 1 incurs no loss of generality.

In practice, the functions to be minimized are often not specified exactly, due to the rounding error of numerical computation or the inexact nature of the model. We model this limitation by the random perturbation $w$ sampled from some distribution. Given a sequence of linear operators $\{\mathcal{A}_t\}$ such that $\|\mathcal{A}_t\| = \sup_{f\neq 0} \frac{\|\mathcal{A}_t f\|}{\|f\|} = 1$ together with the perturbations $\{w_t\}$, consider the following model of linear time variation:

$$f_{t+1} = \mathcal{T}_t f_t = \mathcal{A}_t f_t + w_t, \quad \text{for } t \in \{0, 1, \dots\}. \tag{76}$$

*What properties the operators $\{\mathcal{T}_t\}$ should satisfy in order for $f_t$ to almost reach a target function $f^*$ at time $t = T$?* We will provide an answer using the notion of shape dominant operator. To understand the importance of this problem, suppose that at time $t = 0$, we optimize $f_0$ around a poor local minimum $x_0^*$. If at $t = T$, the function $f_T$ becomes convex with a unique global minimum $x_T^*$, then no matter how optimization is carried out for $f_1$ through $f_{T-1}$, minimizing $f_T$ will yield the same solution $x_T^*$, which is globally optimal. The effect of minimizing $f_T$ cancels out the sub-optimality at time $t = 0$. Moreover, under some technical conditions, the global solution at time $T$ can be used to find global solutions at future times using tracking methods [30–32]. In other words, the shape of $f_T$ affects the complexity of online optimization in the long run.

Now, we introduce the notion of shape dominant operator. Consider time-varying functions $\{f_t\}$ defined on a finite discrete set $\mathcal{X} = \{x_1, \dots, x_n\} \subset \mathbb{Z}^d$. Equivalently, $f_t$ can be viewed as a vector in $\mathbb{R}^n$. For the noisy linear operator $\mathcal{T}_t$ defined in (76), let $A_t$ denote the associated matrix of the linear operator $\mathcal{A}_t$ represented under the standard basis, for $t \in \{1, 2, \dots\}$. Let $P(A_t, w_t)$ denote the joint distribution of $A_t$ and $w_t$.

**Definition 5.** *The joint distribution $P(A, w)$ is said to be $(\delta, \sigma, f^*, \phi^*)$ shape dominant if following conditions hold with probability 1: 1) the unit vector $f^*$ is the eigenvector of $A$ associated with eigenvalue 1; 2) the unit vector $\phi^*$ is the eigenvector of $A^\top$ associated with eigenvalue 1; 3) $\langle f^*, \phi^* \rangle \neq 0$; 4) all other eigenvalues of $A$ have absolute values less than $1 - \delta$; 5) conditioned on $A$, the noise $w$ has zero mean and is sub-Gaussian with parameter $\sigma^2$ in the sense that for all $u \in \mathbb{R}^n$ with $\|u\| \leq 1$, it holds that $\mathbb{E}[\exp(su^\top w)] \leq \exp\left(\frac{\sigma^2 s^2}{2}\right)$.*

**Theorem 7.** *For the time-varying operator $\mathcal{T}_t$ defined in (76), suppose that $P(A_t, w_t)$ is $(\delta, \sigma_t, f^*, \phi^*)$ shape dominant and independent for all $t \in \{0, 1, \dots, T-1\}$, then,*

$$f_T = \frac{\langle \phi^*, f_0 + \sum_{t=0}^{T-1} w_t \rangle}{\langle \phi^*, f^* \rangle} f^* + v + w, \tag{77}$$

*where $\|v\| \leq (1-\delta)^T \left( \|f_0\| + \frac{\langle \phi^*, f_0 \rangle}{\langle \phi^*, f^* \rangle} \right)$ and $w$ is sub-Gaussian with parameter $\sigma^2 = \left(1 + \frac{1}{\langle \phi^*, f^* \rangle^2}\right) \sum_{t=0}^{T-1} (1 - \delta)^{2(T-t)} \sigma_t^2$.*

*Proof.* Consider the subspace $\mathcal{G} = \{g \in \mathbb{R}^n, \langle \phi^*, g \rangle = 0\}$. Since $\langle \phi^*, f^* \rangle \neq 0$, we have $f^* \notin \mathcal{G}$. Since $\phi^*$ is the eigenvector of $A_t^\top$, the following holds for all $g \in \mathcal{G}$

$$\langle \phi^*, A_t g \rangle = \langle {A_t}^\top \phi^*, g \rangle = \langle \phi^*, g \rangle = 0. \tag{78}$$

Therefore, $A_t g \in \mathcal{G}$, and $\mathcal{G}$ is an invariant subspace of $A_t$ in $\mathbb{R}^n$ for $t \in \{0, 1, \dots, T-1\}$. Let a basis of $\mathcal{G}$ be given by $\{g_1, \dots, g_{n-1}\}$. Then, $B = \{f^*, g_1, \dots, g_{n-1}\}$ is a basis of $\mathbb{R}^n$, under which the linear operator $A_t$ takes the form

$$A_t = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & A_t' & \\ 0 & & & \end{bmatrix}, \tag{79}$$

where $A_t'$ is a random matrix in $\mathbb{R}^{(n-1)\times(n-1)}$. With a slight abuse of notation, we regard $A_t'$ as a linear transformation from $\mathcal{G}$ to $\mathcal{G}$. Note that $\|A_t'\| \leq 1-\delta$ because all other eigenvalues of $A_t$ have norm less than $1-\delta$. Under the basis $B$, $f_0$ has the representation $f_0 = \frac{\langle \phi^*, f_0\rangle}{\langle \phi^*, f^*\rangle} f^* + g$, where $g \in \mathcal{G}$. As a result,

$$\begin{aligned} f_T &= \mathcal{T}_{T-1} \circ \cdots \circ \mathcal{T}_0 f_0 \\ &= A_{T-1}\cdots A_0 f_0 + \sum_{t=0}^{T-1} A_{T-1}\cdots A_{t+1} w_t \\ &= \frac{\langle \phi^*, f_0\rangle}{\langle \phi^*, f^*\rangle} f^* + A_{T-1}'\ldots A_1' g + \sum_{t=0}^{T-1} A_{T-1}\cdots A_{t+1} w_t. \end{aligned} \tag{80}$$

The norm estimate gives rise to

$$\left\|A_{T-1}'\ldots A_1' g\right\| \leq (1-\delta)^T \cdot \|g\| \leq (1-\delta)^T \cdot \left(\|f_0\| + \left|\frac{\langle \phi^*, f_0\rangle}{\langle \phi^*, f^*\rangle}\right|\right), \tag{81}$$

where the triangle inequality is used. Similarly, one can write $w_t = \frac{\langle \phi^*, w_t\rangle}{\langle \phi^*, f^*\rangle} f^* + h_t$, where $h_t \in \mathcal{G}$. We have

$$A_{T-1}\cdots A_{t+1} w_t = \frac{\langle \phi^*, w_t\rangle}{\langle \phi^*, f^*\rangle} f^* + A_{T-1}'\cdots A_{t+1}' h_t. \tag{82}$$

For all $u \in \mathbb{R}^n$ with $\|u\| \leq 1$, it holds that

$$\begin{aligned} &\mathbb{E}\left[\exp\left(s\left\langle u, A_{T-1}'\cdots A_{t+1}' h_t\right\rangle\right)\right] \\ &= \mathbb{E}\left[\exp\left(s\left\langle A_{t+1}'^{\top}\cdots A_{T-1}'^{\top} u, h_t\right\rangle\right)\right] \\ &= \mathbb{E}\left[\exp\left(s\left\langle A_{t+1}'^{\top}\cdots A_{T-1}'^{\top} u, w_t - \frac{\langle \phi^*, w_t\rangle}{\langle \phi^*, f^*\rangle} f^*\right\rangle\right)\right] \\ &= \mathbb{E}\left[\exp\left(s\left\langle A_{t+1}'^{\top}\cdots A_{T-1}'^{\top} u, w_t\right\rangle\right) \times \exp\left(s\left\langle -\frac{\langle A_{t+1}'^{\top}\cdots A_{T-1}'^{\top} u, f^*\rangle}{\langle \phi^*, f^*\rangle}\phi^*, w_t\right\rangle\right)\right] \\ &\leq \exp\left(\frac{\sigma_t^2 s^2 \left\|A_{t+1}'^{\top}\cdots A_{T-1}'^{\top} u\right\|^2}{2}\right) \times \exp\left(\frac{\sigma_t^2 s^2}{2}\left(\frac{\langle A_{t+1}'^{\top}\cdots A_{T-1}'^{\top} u, f^*\rangle}{\langle \phi^*, f^*\rangle}\right)^2\right) \\ &\leq \exp\left(\frac{\sigma_t^2 s^2 (1-\delta)^{2(T-t)}\left(1+\frac{1}{\langle \phi^*, f^*\rangle^2}\right)}{2}\right), \end{aligned} \tag{83}$$

which implies that $A_{T-1}'\cdots A_{t+1}' h_t$ is sub-Gaussian with parameter $\sigma_t^2(1-\delta)^{2(T-t)}\left(1+\frac{1}{\langle \phi^*, f^*\rangle^2}\right)$, and thereby, $\sum_{t=0}^{T-1} A_{T-1}'\cdots A_{t+1}' h_t$ is sub-Gaussian with parameter $\sigma^2 = \left(1+\frac{1}{\langle \phi^*, f^*\rangle^2}\right)\sum_{t=0}^{T-1}(1-\delta)^{2(T-t)}\sigma_t^2$. This completes the proof. □

Theorem 7 states that if the time-varying model is given by shape dominant operators, the function $f_T$ decomposes into the sum of dominating shape $f^*$, a bias term $v$ that gradually fades away, and a cumulating noise term that discounts noise in previous iterations. We provide a bound on the hitting time below.

**Theorem 8.** *Under the same assumptions made in Theorem 7, for a given $\epsilon > 0$, define the associated hitting time $T(\epsilon)$ as*

$$T(\epsilon) = \min\left\{T : \exists \lambda \in \mathbb{R} \text{ s.t. } \|f_T - \lambda f^*\| < \epsilon\right\}. \tag{84}$$

*Then, for all $T > \frac{\log 2\left(\|f_0\| + \left|\frac{\langle \phi^*, f_0\rangle}{\langle \phi^*, f^*\rangle}\right|\right) - \log \epsilon}{\log \frac{1}{1-\delta}}$, it holds that*

$$\mathbb{P}(T(\epsilon) \geq T) \leq C_n \exp\left(-\frac{\epsilon^2}{32\left(1+\frac{1}{\langle \phi^*, f^*\rangle^2}\right)\sum_{t=0}^{T-1}(1-\delta)^{2(T-t)}\sigma_t^2}\right), \tag{85}$$

*where $C_n$ is a universal constant depending only on $n$.*

*Proof.* By Theorem 7, for a fixed number $T$, we have the following decomposition for $f_T$:

$$f_T = \frac{\langle \phi^*, f_0 + \sum_{t=0}^{T-1} w_t \rangle}{\langle \phi^*, f^* \rangle} f^* + v^{(T)} + w^{(T)}, \tag{86}$$

where $\left\|v^{(T)}\right\| < (1-\delta)^T \left( \|f_0\| + \left| \frac{\langle \phi^*, f_0 \rangle}{\langle \phi^*, f^* \rangle} \right| \right)$ and $w^{(T)} = \sum_{t=0}^{T-1} A'_{T-1} \cdots A'_{t+1} h_t$ is sub-Gaussian with parameter $\sigma^2 = \left(1 + \frac{1}{\langle \phi^*, f^* \rangle^2}\right) \sum_{t=0}^{T-1} (1-\delta)^{2(T-t)} \sigma_t^2$. From the definition of the hitting time $T(\epsilon)$ in (84), we have

$$\mathbb{P}(T(\epsilon) < T) \geq \mathbb{P}\left( \left\|v^{(T)}\right\| < \epsilon/2, \left\|w^{(T)}\right\| < \epsilon/2 \right). \tag{87}$$

When $T > \frac{\log 2\left(\|f_0\| + \left|\frac{\langle \phi^*, f_0 \rangle}{\langle \phi^*, f^* \rangle}\right|\right) - \log \epsilon}{\log \frac{1}{1-\delta}}$, the bound $\left\|v^{(T)}\right\| < \epsilon/2$ is satisfied. Since $w^{(T)}$ is sub-Gaussian with parameter $\sigma^2$, the tail-bound for $w^{(T)}$ yields

$$\mathbb{P}\left( \left\|w^{(T)}\right\| < \epsilon/2 \right) = 1 - \mathbb{P}\left( \left\|w^{(T)}\right\| > \epsilon/2 \right) \geq 1 - C_n \exp\left( -\frac{\epsilon^2}{32\sigma^2} \right), \tag{88}$$

where $C_n$ is a universal constant depending only on $n$. This completes the proof. □

To understand the above bound, consider a fixed time $T$. When $\sigma_t$ decreases, the bound becomes smaller. As a result, with a smaller random perturbation, it is more likely to reach the target function faster. When $\epsilon$ increases, the bound also becomes smaller, which matches the intuition that a larger neighborhood is easier to reach than a smaller one.

**Remark 6.** *The analysis in this section can be generalized to continuous functions by working through eigenfunctions as opposed to eigenvectors. We briefly discuss this in the special case where $L^2(\mathcal{X})$ has a finite number of bases. Let the inner product be $\langle f, g \rangle = \int_{\mathcal{X}} f(x) \cdot g(x) dx$ and the function space to have an orthonormal basis given by the set of functions $\{u_1, u_2, \ldots, u_n\}$ such that*

$$\langle u_i, u_j \rangle = \int_{\mathcal{X}} u_i(x) \cdot u_j(x) dx = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}. \tag{89}$$

*Note that any function can be decomposed into a linear combination of the basis functions, i.e., $f(x) = \sum_{j=1}^{n} a_j \cdot u_j(x)$, where the coefficients can be stacked into a column vector $a = [a_1, a_2, \ldots, a_n]^T$. Define the matrix $A$ representing the linear operator $\mathcal{T}$ with the elements*

$$A_{ij} = \langle u_i, \mathcal{T}(u_j) \rangle = \int_{\mathcal{X}} u_i(x) \cdot \mathcal{T}\big(u_j(x)\big) dx. \tag{90}$$

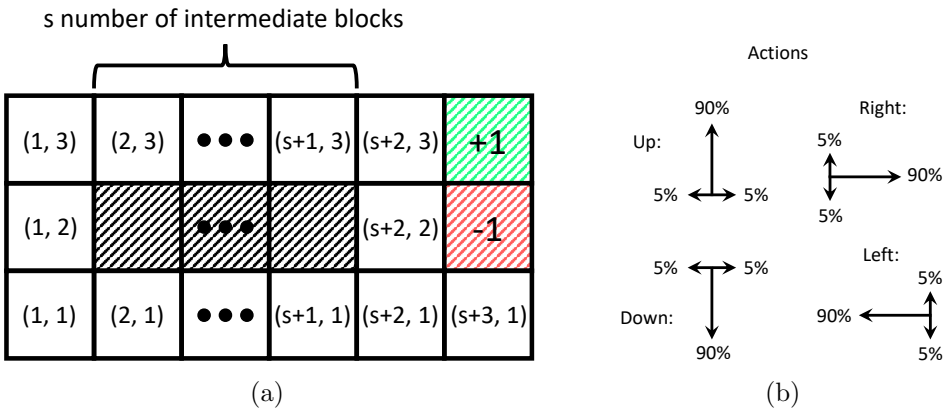
*There exists a vector $b = [b_1, b_2, \ldots, b_n]^T$ such that applying the operator $\mathcal{T}$ on the decomposed form of $f(x)$ yields*

$$\mathcal{T}\big(f(x)\big) = \sum_{j=1}^{n} a_j \cdot \mathcal{T}\big(u_j(x)\big) = \sum_{j=1}^{n} b_j \cdot u_j(x). \tag{91}$$

*Taking the inner product of both sides of the above equation with an arbitrary basis function $u_i$ leads to*

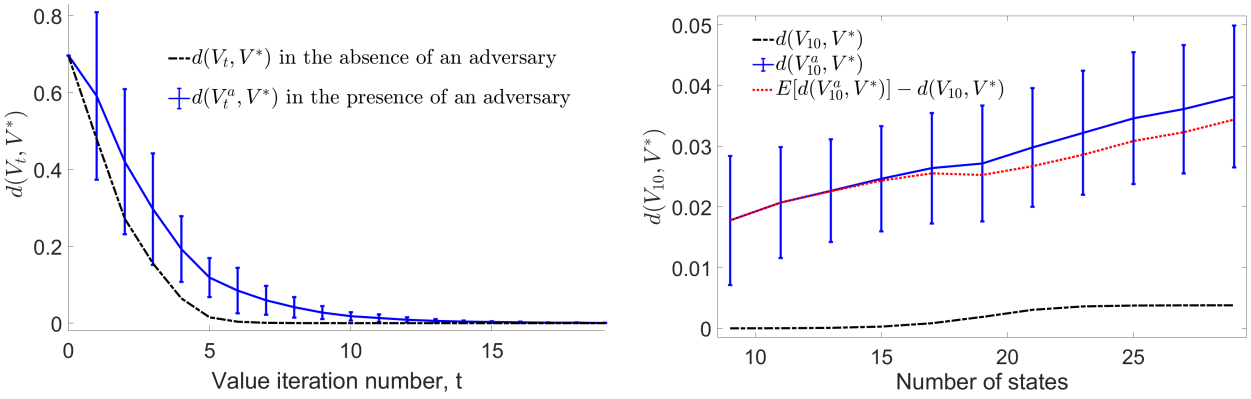
$$\sum_{j=1}^{n} a_j \cdot \big\langle u_i, \mathcal{T}\big(u_j\big) \big\rangle = \sum_{j=1}^{n} b_j \cdot \langle u_i, u_j \rangle \Rightarrow \sum_{j=1}^{n} a_j \cdot A_{ij} = b_i. \tag{92}$$

*The above equation is the matrix multiplication $Aa = b$, which is the matrix associated with $\mathcal{T}$ acting upon the function $f(x)$ expressed in the orthonormal basis. If $f(x)$ is an eigenfunction of transformation $\mathcal{T}$ with eigenvalue $\lambda$, we have $Aa = \lambda a$. Hence, the results of Theorem 8 can be applied to continuous functions in a function space with a finite number of bases. The extension to the case with an infinite, but countable, number of bases is similar under some technical assumptions.*




(a) (b)

Figure 3: (a) the agent interacts with an environment, (b) the agent has a set of four actions in each state.




(a) A comparison of value iteration convergence in the absence and presence of an adversary.

(b) The effect of an adversary versus the number of states.

Figure 4: The effect of an adversary on the convergence of value iteration.

## 4 Simulation Results

In this section, the adversarial attack on the computation of value iteration is simulated for an agent interacting with an environment depicted in Figure 3. The agent can take any of the four actions Up, Down, Right, and Left in each of the non-terminal states. By taking an action, the agent moves one block toward the desired action 90% of the time, or moves one block to the right or left of the desired taken action uniformly at random 10% of the time. The agent bounces back to its original state before taking an action if movement in the direction described above is not possible due to the walls marked with diagonal strips or exiting the environment. The agent is incurred a cost of 0.02 by each move and there are two terminal states in which the agent receives an immediate reward of +1 and -1 as shown in Figure 3. In order to determine the optimal path for the agent starting from any of the states, the value function is calculated using synchronous value iteration. In our simulated example, an adversary contaminates the value function by expanding up to $Q = 1.8$ in a random direction, withholding the contraction, 20% of the time. As a result, the distance of the time-varying value function from the true value function based on the $L^2$-norm is affected negatively as depicted in Figure 4a, where the starting function is the all-zero function in our simulations and the average and standard deviations are estimated by 1000 rounds of independent runs of the value iteration. Furthermore, the negative effect of the adversary is worsened by increasing the cardinality of the state space in the studied example. In order to show this, the number of intermediate blocks in Figure 3 is changed from 1 to 10, i.e., the number of states is changed from 9 to 27, and the distance between the

value function at the tenth iterate and the true value function is depicted in Figure 4b. As shown in Figure 4b, $\mathbb{E}\big[d(V_{10}^a, V^*)\big] - d(V_{10}, V^*)$ has an increasing trend as the number of states increases, where $V_{10}^a$ is value function at the tenth iterate in the presence of an adversary and $V_{10}$ is the corresponding function in the absence of an adversary, and the dependence of value function on the number of states is eliminated to keep the notations simple.

# 5 Conclusion and Future work

Multiple models of stochastic time variation along with their corresponding notions of hitting time are studied in this paper. In particular, we develop a probabilistic Banach fixed-point theorem that proves the convergence of the value iteration method with a probabilistic contraction-expansion transformation with an associated confidence level, which finds applications to adversarial attacks on computation of the value iteration method. We prove that the hitting time of the value function in the value iteration method with a probabilistic contraction-expansion transformation is logarithmic in terms of the inverse of a desired precision. Furthermore, we develop upper bounds on the hitting time for optimization of unknown discrete and continuous time-varying functions whose noisy evaluations are revealed over time. The upper bound for a discrete function is logarithmic in terms of the cardinality of the function domain and the upper bound for a continuous function is super-quadratic (but sub-cubic) in terms of the inverse of a desired precision. In this framework, we show that convex functions are learned faster than non-convex functions. Finally, an upper bound on the hitting time is developed for a time-varying linear model with additive noise under the notion of shape dominance for discrete functions. Future research directions include: studying how an environment with time-varying parameters modeled by transition probabilities and rewards affects the Bellman transformation and its fixed point, obtaining upper bounds on the rate of change of the time-varying parameters such that the time-varying fixed points are achievable after a hitting time, and studying the effect of an adversary in applications of reinforcement learning whose computations are performed via edge computing.

# References


[1] Ruoyu Sun. Optimization for deep learning: theory and algorithms. *arXiv preprint arXiv:1912.08957*, 2019.

[2] Fangda Gu, Heng Chang, Wenwu Zhu, Somayeh Sojoudi, and Laurent El Ghaoui. Implicit graph neural networks. *Advances in Neural Information Processing Systems*, 33, 2020.

[3] Léon Bottou, Jonas Peters, Joaquin Quiñonero-Candela, Denis X Charles, D Max Chickering, Elon Portugaly, Dipankar Ray, Patrice Simard, and Ed Snelson. Counterfactual reasoning and learning systems: The example of computational advertising. *The Journal of Machine Learning Research*, 14(1):3207–3260, 2013.

[4] Julie Mulvaney-Kemp, Salar Fattahi, and Javad Lavaei. Load variation enables escaping poor solutions of time-varying optimal power flow. In *2020 IEEE Power & Energy Society General Meeting (PESGM)*, pages 1–5. IEEE, 2020.

[5] SangWoo Park, Elizabeth Glista, Javad Lavaei, and Somayeh Sojoudi. Homotopy method for finding the global solution of post-contingency optimal power flow. In *2020 American Control Conference (ACC)*, pages 3126–3133. IEEE, 2020.

[6] Christopher V Rao, James B Rawlings, and David Q Mayne. Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations. *IEEE transactions on automatic control*, 48(2):246–258, 2003.

[7] Amirhossein Ajalloeian, Andrea Simonetto, and Emiliano Dall'Anese. Inexact online proximal-gradient method for time-varying convex optimization. In *2020 American Control Conference (ACC)*, pages 2850–2857. IEEE, 2020.

[8] P Bertsekas Dimitri. *Dynamic programming and optimal control.* Athena Scientific, 2017.

[9] Hyeong Soo Chang, Jiaqiao Hu, Michael C Fu, and Steven I Marcus. *Simulation-based algorithms for Markov decision processes*. Springer Science & Business Media, 2013.

[10] Rémi Coulom. Efficient selectivity and backup operators in monte-carlo tree search. In *International conference on computers and games*, pages 72–83. Springer, 2006.

[11] Cameron B Browne, Edward Powley, Daniel Whitehouse, Simon M Lucas, Peter I Cowling, Philipp Rohlfshagen, Stephen Tavener, Diego Perez, Spyridon Samothrakis, and Simon Colton. A survey of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in games*, 4(1):1–43, 2012.

[12] Michael C Fu. Markov decision processes, alphago, and monte carlo tree search: Back to the future. In *Leading Developments from INFORMS Communities*, pages 68–88. INFORMS, 2017.

[13] Benjamin Van Roy. *Learning and value function approximation in complex decision processes*. PhD thesis, Massachusetts Institute of Technology, 1998.

[14] John N Tsitsiklis and Benjamin Van Roy. Feature-based methods for large scale dynamic programming. *Machine Learning*, 22(1-3):59–94, 1996.

[15] Benjamin Van Roy. Performance loss bounds for approximate value iteration with state aggregation. *Mathematics of Operations Research*, 31(2):234–244, 2006.

[16] Lucian Busoniu, Robert Babuska, Bart De Schutter, and Damien Ernst. *Reinforcement learning and dynamic programming using function approximators*, volume 39. CRC press, 2010.

[17] Mahadev Satyanarayanan. The emergence of edge computing. *Computer*, 50(1):30–39, 2017.

[18] He Li, Kaoru Ota, and Mianxiong Dong. Learning iot in edge: Deep learning for the internet of things with edge computing. *IEEE network*, 32(1):96–101, 2018.

[19] Pavel Mach and Zdenek Becvar. Mobile edge computing: A survey on architecture and computation offloading. *IEEE Communications Surveys & Tutorials*, 19(3):1628–1656, 2017.

[20] Mihailo Isakov, Vijay Gadepally, Karen M Gettings, and Michel A Kinsy. Survey of attacks and defenses on edge-deployed neural networks. In *2019 IEEE High Performance Extreme Computing Conference (HPEC)*, pages 1–8. IEEE, 2019.

[21] Mohammad S Ansari, Saeed H Alsamhi, Yuansong Qiao, Yuhang Ye, and Brian Lee. Security of distributed intelligence in edge computing: Threats and countermeasures. In *The Cloud-to-Thing Continuum*, pages 95–122. Palgrave Macmillan, Cham, 2020.

[22] Yinhao Xiao, Yizhen Jia, Chunchi Liu, Xiuzhen Cheng, Jiguo Yu, and Weifeng Lv. Edge computing security: State of the art and challenges. *Proceedings of the IEEE*, 107(8):1608–1631, 2019.

[23] Warren B Powell. What you should know about approximate dynamic programming. *Naval Research Logistics (NRL)*, 56(3):239–249, 2009.

[24] Lantao Liu and Gaurav S Sukhatme. A solution to time-varying markov decision processes. *IEEE Robotics and Automation Letters*, 3(3):1631–1638, 2018.

[25] Han Feng, Ali Yekkehkhany, and Javad Lavaei. A hitting time analysis of non-convex optimization with time-varying revelations. *https://lavaei.ieor.berkeley.edu/Online_opt_2020_1.pdf*, 2020.

[26] Giuseppe Calafiore and Marco C Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46, 2005.

[27] Marco C Campi and Simone Garatti. The exact feasibility of randomized solutions of uncertain convex programs. *SIAM Journal on Optimization*, 19(3):1211–1230, 2008.

[28] Marco Claudio Campi, Simone Garatti, and Federico Alessandro Ramponi. A general scenario theory for nonconvex optimization and decision making. *IEEE Transactions on Automatic Control*, 63(12):4067–4078, 2018.

[29] Arash Hassibi, Stephen P Boyd, and Jonathan P How. Control of asynchronous dynamical systems with rate constraints on events. In *Proceedings of the 38th IEEE Conference on Decision and Control (Cat. No. 99CH36304)*, volume 2, pages 1345–1351. IEEE, 1999.

[30] Yuhao Ding, Javad Lavaei, and Murat Arcak. Escaping spurious local minimum trajectories in online time-varying nonconvex optimization. In *2021 American Control Conference (ACC)*, pages 454–461. IEEE, 2021.

[31] Salar Fattahi, Cedric Josz, Reza Mohammadi, Javad Lavaei, and Somayeh Sojoudi. Absence of spurious local trajectories in time-varying optimization: A control-theoretic perspective. In *2020 IEEE Conference on Control Technology and Applications (CCTA)*, pages 140–147. IEEE, 2020.

[32] Olivier Massicot and Jakub Marecek. On-line non-convex constrained optimization. *arXiv preprint arXiv:1909.07492*, 2019.