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

Ali Yekkehkhany, Han Feng, and Javad Lavaei
Department of Industrial Engineering and Operations Research
University of California, Berkeley
{aliyek, han_feng, lavaei}@berkeley.edu

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. In order to throw light on this framework, multiple models of stochastic time variation are studied for both discrete and continuous functions in this paper. The main goal is to develop upper bounds on the hitting time after which optimizing the stochastic time-varying function reveals informative statistics on the optimization of the target function. In particular, time-varying probabilistic contraction-expansion transformations are studied with applications to adversarial attacks on the computation of value iteration in Markov decision processes and reinforcement learning. In this application, the contraction of value iteration is reversed to an expansion up to a constant by an adversary in a probabilistic manner, violating the contraction criterion of the Banach fixed-point theorem. To address the problem, we establish a probabilistic Banach fixed-point theorem in which the convergence of the value iteration method with a probabilistic contraction-expansion transformation is proved with an associated confidence level. 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. In addition, the hitting time is analyzed 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 provide improved bounds for convex functions and show that such functions are learned faster than non-convex functions. Finally, the hitting time is studied for a time-varying linear model with additive noise under the notion of shape dominance.

Key words: Stochastic time-varying functions, 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 (Gu et al. 2020, Sun 2019), online advertising (Bottou et al. 2013), decision-making process of power systems (Mulvaney-Kemp, Fattahi, and Lavaei 2020, Park

et al. 2020), and the real-time state estimation of nonlinear systems (Rao, Rawlings, and Mayne 2003), the parameters of the problem are often uncertain and change over time. 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

$$x_t^* = \arg\min_{x\in\mathcal{X}} \{f_t(x) = \mathsf{E}F_t(x,\xi)\}, \quad t \in \{1,2,\dots\}, \tag{1}$$

where the random variable ξ models the uncertainty in the objective that comes from disturbance, inexactness of model, use of small batches, or injected noise, and where arg min denotes any global minimizer of the input function. Note that the expectation E over ξ can only be evaluated approximately since the nature of the noise is often 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 Equation (1), which is studied for multiple models in this work. The focus of this paper is on the stochastic complexity of stochastic time-varying functions, captured by the hitting time notion. For an optimization-based study of time-varying problems, the reader is referred to the works by Ding, Lavaei, and Arcak (2019), Fattahi et al. (2019), Simonetto et al. (2016), Tang et al. (2018) and the references therein.

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 μ, the value function $V^\mu : \mathcal{S} \to \mathcal{R}$ is defined as

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

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

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

A preliminary treatment of the problem discussed in this paper has been introduced in our earlier work that is submitted to the American Control Conference 2021 (Feng, Yekkehkhany, and Lavaei 2020). Compared with the conference version whose focus is on discrete time-varying functions, this paper includes multiple new models of time variation in continuous domains. The new studies in this paper include the analysis of adversarial attacks on the computation of Markov decision processes, which also distinguishes this work from our conference paper.

It is well known in the context of dynamic programming that for a finite action space, any policy μ^* given by

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

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

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

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

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

It is well known that the Bellman operator is a contraction mapping associated with the parameter q on a complete metric space that has a unique fixed point $V^* = \mathcal{T}(V^*)$, which is discussed in detail in Section 2.1. Using the value iteration method, starting from any arbitrary V_0, the sequence $\{V_0, V_1, V_2, \dots\}$ with $V_{t+1} = \mathcal{T}(V_t)$ for $t\in\{0,1,\dots\}$ converges to V^*, where V^* is mostly unknown in MDP and RL applications. The value function V_t is a time-varying function and may never be exactly equal to V^*, but by defining some hitting-time notion, it can be analyzed how many iterations should be performed so that the maximum value of the argument function in Equation (4) is approximated by maximizing the time-varying function $\mathsf{E}\big[r(s,a,w)+q\cdot V_t(h(s,a,w))\big]$ with a desired precision. Note that the theoretical proof of convergence behind the value iteration method and the hitting time depends heavily on the contraction mapping parameter q and the fact that $d\big(\mathcal{T}(V_{t+1}),\mathcal{T}(V_t)\big) \le q\cdot d\big(V_{t+1},V_t\big)$ 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 and action spaces, there are sticking points from the practical point of view that may result in the value iteration method not to satisfy the contraction condition $d\big(\mathcal{T}(V_{t+1}),\mathcal{T}(V_t)\big) \le q\cdot d\big(V_{t+1},V_t\big)$ in some iterations. Instead, the distance may expand up to a factor greater than one in some iterations of the value iteration, i.e., $d\big(\mathcal{T}(V_{t+1}),\mathcal{T}(V_t)\big) \le Q\cdot d\big(V_{t+1},V_t\big)$, where $Q\ge 1$. There are multiple reasons for the contraction not to hold in value iteration for a realistic setting that are discussed below:

I. **Approximation errors in computing expectation:** Computing the expectation in the Bellman operator in Equation (6) can be costly, given that it should be computed for every action in order to obtain the action that gives the highest reward. There are different approaches to circumvent this issue (Dimitri 2017), e.g., a) assuming certainty equivalence by replacing stochastic quantities with deterministic ones to arrive at a deterministic optimization, which can possibly degrade the accuracy significantly, b) using Monte Carlo tree search and adaptive simulation to determine which expectations associated with actions should be computed more accurately (Browne

et al. 2012, Chang et al. 2013, Coulom 2006, Fu 2017). In both of the above approaches, the value of the expectation could be subject to errors. There is another line of research on robust dynamic programming that addresses the uncertainty on the model parameters (Iyengar 2005, Nilim and El Ghaoui 2005).

II. **Approximation errors in maximization:** The maximization in the Bellman operator in Equation (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, such as gradient methods, Newton's method, and quadratic programming for deterministic problems with continuous action spaces, and stochastic programming for stochastic problems, are used to handle the large number of actions over which the maximization is performed (Dimitri 2017). These methods are prone to errors especially when they are used in an online fashion. Hence, the approximation error in maximization is another source of error.

III. **Approximation errors of value function:** Due to the large number of states in many recent applications of Markov decision processes and reinforcement learning, it may not be practical to have a memory associated with every state. Instead, parametric feature-based approximation methods, such as neural network architectures, are used for value function representation (Busoniu et al. 2010, Dimitri 2017, Tsitsiklis and Van Roy 1996, Van Roy 1998, 2006). The parameterization of the value function is another source of error in value iteration that can cause expansion in value iteration (Tsitsiklis and Van Roy 1996, Van Roy 1998). This problem is resolved by taking further assumptions on parameterization methods to ensure contraction in the iterates of value iteration (Bertsekas 2011, De Farias and Van Roy 2000, Roy 2006). However, under the next cause of error, adversarial attacks, the contraction in all iterates of value iteration may not hold anymore whose effect is studied in this paper.

IV. **Adversarial value iteration:** By the emergence of cloud, edge, and fog computing, the computations associated with large-scale MDP and RL problems are moved away from agents to distributed servers (Li, Ota, and Dong 2018, Mach and Becvar 2017, Satyanarayanan 2017). 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 (Ansari et al. 2020, Isakov et al. 2019, Xiao et al. 2019). A natural example of such adversarial attacks is to contaminate the computation of value iteration so that the contraction of the Bellman operator would not hold in all iterations. As a result, the contraction condition of the Banach fixed-point theorem would not be satisfied, resulting in miscalculation of the value function that is supposed to maximize the expected received reward in a possibly critical application.

The effect of the above-mentioned sources of errors and disturbances in value iteration can be modeled in different ways. The first three causes have been studied extensively in the literature

(Powell 2009), while there is no mathematical analysis of adversarial attacks on the computation of the value functions. 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 refer to the transformation in this probabilistic approach as a probabilistic contraction-expansion mapping, where the expansion can result from adversarial attacks disturbing the computation of value iteration on edge computing.

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. 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. An example of a time-varying environment is the changing environment at which autonomous vehicles interact with each other, human drivers, and pedestrians. Due to the emergence of autonomous vehicles, the slim human-robot interaction has been shifted toward a robot-robot interaction among autonomous vehicles and a different human-robot interaction among vehicles, pedestrians, and human drivers as people adapt their actions to this new technology. In the context of reinforcement learning and Markov decision processes, this gradual change is translated into time-varying reward functions and transition probabilities that enforce the idea of a time-varying Bellman mapping discussed above. In this work, 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 at an iterate of the value iteration under a time-varying probabilistic contraction-expansion mapping with additive noise.

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 Feng, Yekkehkhany, and Lavaei (2020). 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 contraction mappings with additive noise are studied in Section 2.2, time-varying probabilistic contraction-expansion mappings with additive noise are studied in Section 2.3, time-varying continuous functions with additive noise are studied in Section 2.4, and improved bounds for convex functions with additive noise are studied in Section 2.5 where we prove that convex functions are learned faster in general. 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. 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, four variants of time-varying stochastic functions are studied. 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 but deterministic contraction mapping with additive noise is studied. In the third model, a time-varying probabilistic contraction-expansion mapping with additive noise is investigated. All of the above 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.

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 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 $d\big(f^*, f^n\big) \leq \frac{q^n}{1-q} \cdot d(f^1, f^0)$. If the goal is to optimize the function f^*, the optimization of the functions in the sequence $\{f^n\}$ may not be reliable up until a hitting time in which the function sequence is close enough to f^* and does not change dramatically. In the current and next subsections, different variants of mappings are studied and their hitting times are analyzed. Note that in all iterations of the above value iteration, the same contraction mapping $\mathcal{T}$ is applied to the function sequence that operates as a contraction mapping according to Equation (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

$$d\big(\overline{\mathcal{T}}(f_{t+1}), \overline{\mathcal{T}}(f_t)\big) \leq \begin{cases} q \cdot d(f_{t+1}, f_t) & \text{w.p.} \quad p \\ Q \cdot d(f_{t+1}, f_t) & \text{otherwise} \end{cases}, \qquad \forall t \in \{0,1,2,\dots\} \tag{9}$$

for some constants $q \in [0,1)$, $Q \geq 1$, and $p \in (0,1]$, where w.p. stands for "with probability". 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 expansion irregularity in Equation (9) is caused by an adversary in an attempt to move the function sequence away from the fixed point. The shape of the function f_t possibly changes over time; however, there can be a time, called hitting time T, after which f_T reaches a neighborhood of some desirable function, as explained in more details below. 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 $\{f_t\}$. The focus of the rest of this section is to find an upper bound on the hitting time for the function sequence $\{f_t\}$, where hitting time is 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 Equation (8) is defined as

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

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

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 Equation (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 Equation (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, the hitting time $T(\epsilon, a)$ satisfies the inequality*

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

for any $0 < L < \frac{\epsilon}{d(f_1,f_0)}$.

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 function sequence $\{f_t\}$ in Equation (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

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

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 Equation (12). Using the independence of B_j for $j \in \{1, 2, \ldots, n-1\}$, the mean can be upper bounded as

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

On the other hand, $\mathrm{Var}\left(S_{n,m}\right) \leq \mathsf{E}\left[S_{n,m}^2\right]$, 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 the random variable $\bar{S}_{n,m}$ is independent of B_j for $j \in \{1, 2, \ldots, m\}$, and $\bar{S} = \lim_{n\to\infty} \bar{S}_{n,m}$ (the dependence on m is dropped after taking the limit as $\bar{S}$ is an infinite sum and B_j are i.i.d. random variables). Since $\mathsf{E}[B_j] > 0$ for $j \geq 1$, we have $\mathsf{E}[\bar{S}_{n,m}^2] \leq \mathsf{E}[\bar{S}^2]$; hence, it follows from Equation (15) that

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

In order to find an upper bound on $\mathsf{E}\left[\bar{S}^2\right]$, 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 $\mathsf{E}\left[\bar{S}^2\right] = \mathsf{E}\left[\tilde{S}^2\right]$, one can obtain

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

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

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

Using Equation (16), we have

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

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 Eqautions (14) and (20), respectively. Using Chebyshev's inequality, for any $L > 0$, we have

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

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

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

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

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 Equations (22a) and (22b) are satisfied when

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

As a result, for a constant N_ϵ that is greater than the term on the right-hand side of Equation (23), for all $n>m>N_\epsilon$ it holds that $d\big(\overline{\mathcal{T}}^n(f_0),\overline{\mathcal{T}}^m(f_0)\big)\leq S_{n,m}\cdot d\big(f_1,f_0\big)=B_1\cdot B_2\cdots B_{N_\epsilon}\cdot\big(B_{N_\epsilon+1}\cdots B_m+\cdots+B_{N_\epsilon+1}\cdots B_{n-1}\big)\cdot d\big(f_1,f_0\big)\leq\epsilon$ with high probability since $\big(B_{N_\epsilon+1}\cdots B_m+\cdots+B_{N_\epsilon+1}\cdots B_{n-1}\big)$ is deterministically increasing as n goes to infinity and it is already proven above that $S_{n,m}\cdot d\big(f_1,f_0\big)\leq\epsilon$ with high probability as n goes to infinity. To conclude, the sequence $\{f_t\}$ is a Cauchy sequence with high probability. 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 Equation (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

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

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.

Theorem 1 states that if contraction of an operator in the iterates of the value iteration method 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 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 $q^2 \cdot p + Q^2 \cdot (1-p) < 1$ on the contraction and expansion of the operator 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 may perform a fewer number of iterates in the value iteration method by using the hitting time for the perfect scenario. This can lead to a highly inaccurate estimate of the fixed point that is used to maximize the expected reward in a possibly critical application.

REMARK 1. By minimizing the upper bound on the hitting time in Equation (11) over the parameter $0 < L < \frac{\epsilon}{d(f_1,f_0)}$, we have that the upper bound is logarithmic in terms of $\frac{d(f_1,f_0)}{\epsilon}$.

REMARK 2. The Banach fixed-point theorem is a special case of Theorem 1 by setting $p = 1$. In that case, the first term inside the maximum in Equation (11) can be ignored since it is related to the randomness of the problem. Since there is no randomness in this case, we can choose $L = 0$ to reduce the second term to $\frac{\ln\left(\frac{(1-q)\epsilon}{d(f_1,f_0)}\right)}{\ln(q)}$, which is compatible with the results from the Banach fixed-point theorem.

2.2. Time-Varying Contraction Mapping with Additive Noise Let $(X, \|\cdot\|)$ be the same complete normed vector space as in Section 2.1. Consider time-varying transformations $\mathcal{T}_t(\cdot) : X \to X$ for $t \in \{0,1,2,\dots\}$ that are contraction mappings with the contraction factors $q_t \in [0,1)$, meaning that

$$d\big(\mathcal{T}_t(f), \mathcal{T}_t(g)\big) \leq q_t \cdot d(f,g), \quad \forall t \in \{0,1,2,\dots\}$$

for all $f, g \in X$. As a result, using the Banach-Caccioppoli fixed-point theorem, each of the contraction mappings has its own unique fixed point, i.e., there exists $f_t^* \in X$ such that $\mathcal{T}_t(f_t^*) = f_t^*$ for all $t \in \{0,1,2,\dots\}$. Furthermore, starting with an arbitrary function $f^0 \in X$, the sequence $\{f^n\}$ with $f^n = \mathcal{T}_t(f^{n-1})$ for $n \geq 1$, where the same contraction mapping $\mathcal{T}_t$ is applied repeatedly, converges to f_t^*; in other words, $f^n \to f_t^*$, where $d\big(f_t^*, f^n\big) \leq \frac{q_t^n}{1-q_t} \cdot d(f^1, f^0)$. Assume that the fixed points of any two consecutive transformations 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 $t \in \{1,2,3,\dots\}$ and the time-varying function $f_t \in X$ for $t \in \{0,1,2,\dots\}$ evolves over time according to

$$f_{t+1} = \widehat{\mathcal{T}}_t(f_t) = \mathcal{T}_t(f_t) + w_t, \quad t \in \{0,1,2,\dots\}, \tag{26}$$

where $w_t \in X$ is some additive noise with the property that $\|w_t\| \leq \epsilon_w$ for all $t \in \{0,1,2,\dots\}$ for some constant $\epsilon_w > 0$. Note that the shape of function f_t possibly changes over time and can be non-convex. However, there may exist a hitting time T after which f_T reaches a neighborhood of some desirable time-varying function as formalized below.

DEFINITION 2. Given $\epsilon > 0$, the hitting time $T(\epsilon)$ for the function sequence introduced in Equation (26) is defined as

$$T(\epsilon) = \min\left\{T : d\big(f_t, f_t^*\big) < \epsilon,\ \forall t \geq T\right\}. \tag{27}$$

The next theorem presents an upper bound on the hitting time $T(\epsilon)$ for the sequence of functions introduced in Equation (26) for noisy time-varying contraction mappings $\widehat{\mathcal{T}}_t$. Interestingly, we observe that the noise values added at previous rounds become discounted at an exponential rate, and therefore the overall impact of added noise functions is automatically controlled by the effect of contraction mappings.

THEOREM 2. *Consider arbitrary time-varying contraction mappings $\mathcal{T}_t$ with the contraction constants q_t and fixed points f_t^*, where $q = \sup_t q_t < 1$ and $d\big(f_t^*, f_{t+1}^*\big) < \epsilon_f$ for $t \in \{0,1,2,\dots\}$. Let the time-varying function f_t evolve over time according to the time-varying noisy transformation in Equation (26). If $\epsilon \in [\frac{1}{1-q}\cdot(\epsilon_f+\epsilon_w), \frac{1}{1-q}\cdot(\epsilon_f+\epsilon_w) + D]$, then the hitting time $T(\epsilon)$ satisfies the inequality*

$$T(\epsilon) \leq 1 + \frac{\ln\left(\frac{\left(\epsilon - \frac{1}{1-q}\cdot(\epsilon_f+\epsilon_w)\right)}{D}\right)}{\ln(q)}, \tag{28}$$

where ϵ_w is an upper bound on the norm of each noise function and D is an upper bound on the distance between f_0 and f_0^*.

Proof: In order to find an upper bound on the hitting time $T(\epsilon)$, we first study how time-varying transformations and added noise functions affect the distance between f_t and f_t^*. The distance can be computed as follows:

$$\begin{aligned}
d\big(f_t, f_t^*\big) &= d\big(\widehat{\mathcal{T}}_{t-1} \circ \cdots \circ \widehat{\mathcal{T}}_0(f_0), f_t^*\big) \\
&\overset{(a)}{=} d\big(\mathcal{T}_{t-1}\big(\widehat{\mathcal{T}}_{t-2} \circ \cdots \circ \widehat{\mathcal{T}}_0(f_0)\big) + w_{t-1}, f_t^*\big) = \big\|\mathcal{T}_{t-1}\big(\widehat{\mathcal{T}}_{t-2} \circ \cdots \circ \widehat{\mathcal{T}}_0(f_0)\big) + w_{t-1} - f_t^*\big\| \\
&\overset{(b)}{\leq} d\big(\mathcal{T}_{t-1}\big(\widehat{\mathcal{T}}_{t-2} \circ \cdots \circ \widehat{\mathcal{T}}_0(f_0)\big), f_t^*\big) + \|w_{t-1}\| \\
&\overset{(c)}{\leq} d\big(\mathcal{T}_{t-1}\big(\widehat{\mathcal{T}}_{t-2} \circ \cdots \circ \widehat{\mathcal{T}}_0(f_0)\big), f_{t-1}^*\big) + d\big(f_{t-1}^*, f_t^*\big) + \|w_{t-1}\| \\
&\overset{(d)}{\leq} q_{t-1} \cdot d\big(\widehat{\mathcal{T}}_{t-2} \circ \cdots \circ \widehat{\mathcal{T}}_0(f_0), f_{t-1}^*\big) + \epsilon_f + \epsilon_w \\
&\leq q_{t-1} \cdot \Big(q_{t-2} \cdot d\big(\widehat{\mathcal{T}}_{t-3} \circ \cdots \circ \widehat{\mathcal{T}}_0(f_0), f_{t-2}^*\big) + \epsilon_f + \epsilon_w\Big) + \epsilon_f + \epsilon_w \\
&\leq q_{t-1} \cdot \Big(q_{t-2} \cdot \Big(q_{t-3} \cdot d\big(\widehat{\mathcal{T}}_{t-4} \circ \cdots \circ \widehat{\mathcal{T}}_0(f_0), f_{t-3}^*\big) + \epsilon_f + \epsilon_w\Big) + \epsilon_f + \epsilon_w\Big) + \epsilon_f + \epsilon_w \\
&\overset{(e)}{\leq} q^t \cdot d\big(f_0, f_0^*\big) + \frac{1}{1-q}\cdot(\epsilon_f + \epsilon_w),
\end{aligned} \tag{29}$$

where $\circ$ denotes the composition of linear operatoers, the definition of the transformation $\widehat{\mathcal{T}}_{t-1}$ in Equation (26) is used in (a), inequalities (b) and (c) hold true due to the triangular inequality, (d) follows from the assumptions $d\big(f^*_{t-1}, f^*_t\big) \leq \epsilon_f$ and $\|w_{t-1}\| \leq \epsilon_w$ in addition to the contractive property of the operator $\mathcal{T}_{t-1}$ and the fact that $\mathcal{T}_{t-1}(f^*_{t-1}) = f^*_{t-1}$, its next two inequalities are true for similar reasons, and (e) follows from iterating the above steps t times and using $q = \sup_t q_t$.

Since the right-hand side of Equation (29) is a decreasing function of t, the minimum value of t that satisfies $q^t \cdot d\big(f_0, f^*_0\big) + \frac{1}{1-q} \cdot (\epsilon_f + \epsilon_w) \leq \epsilon$ is an upper bound on the hitting time $T(\epsilon)$. Given that $d\big(f^*_0, f_0\big)$ is upper bounded by a constant $D > 0$ and $\frac{1}{1-q} \cdot (\epsilon_f + \epsilon_w) \leq \epsilon \leq \frac{1}{1-q} \cdot (\epsilon_f + \epsilon_w) + D$, we have

$$T(\epsilon) \leq 1 + \ln\Big(\big(\epsilon - \frac{1}{1-q} \cdot (\epsilon_f + \epsilon_w)\big)/D\Big)\Big/ \ln(q). \tag{30}$$

Theorem 2 formalizes how many iterations are required in the value iteration method with additive noise and a time-varying operator – that can be caused by a time-varying environment – to guarantee that the ultimate function value is at an ϵ-neighborhood of the fixed point.

2.3. 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, where probabilistic contraction-expansion mappings are defined in Section 2.1. It results from Theorem 1 that 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. Assume that the fixed points of any two consecutive transformations 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 $t \in \{1, 2, 3, \dots\}$. Nevertheless, there can be transformations $\overline{\mathcal{T}}_t$ 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 method, the same probabilistic contraction-expansion mapping $\overline{\mathcal{T}}_t$ is applied to the function sequence. 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 transformation 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

$$f_{t+1} = \widetilde{\mathcal{T}}_t(f_t) = \overline{\mathcal{T}}_t(f_t) + w_t, \quad t \in \{0, 1, 2, \dots\}, \tag{31}$$

where $w_t \in X$ is added noise with the property that $\|w_t\| \leq \epsilon_w$ for all $t \in \{0, 1, 2, \dots\}$ with $\epsilon_w > 0$ being a bounded constant. The following theorem presents an upper bound on the distance between f_t and the time-varying function f^*_t.

THEOREM 3. *Consider arbitrary time-varying probabilistic contraction-expansion mappings $\mathcal{T}_t$ with fixed points f_t^*, where $\sup_t \left(q_t^2 \cdot p_t + Q_t^2 \cdot (1-p_t)\right) < 1$ and $d\big(f_t^*, f_{t+1}^*\big) < \epsilon_f$ 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 Equation* (31)*. Then,*

$$d\big(f_t, f_t^*\big) \leq P_t \cdot d\big(f_0, f_0^*\big) + S_t \cdot (\epsilon_f + \epsilon_w), \tag{32}$$

where ϵ_w is an upper bound on the norm of each noise function, and $P_t = \left(\prod_{i=0}^{t-1} B_i\right)$ 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

$$B_t = \begin{cases} q_t & w.p. \quad p_t \\ Q_t & otherwise \end{cases}. \tag{33}$$

The means and variances of P_t and S_t are upper bounded as

$$\begin{aligned} \mathsf{E}\left[P_t\right] &\leq \left(\sup_t \left(q_t \cdot p_t + Q_t \cdot (1-p_t)\right)\right)^t \xrightarrow{t\to\infty} 0, \\ \mathrm{Var}\left(P_t\right) &\leq \left(\sup_t \left(q_t^2 \cdot p_t + Q_t^2 \cdot (1-p_t)\right)\right)^t \xrightarrow{t\to\infty} 0, \end{aligned} \tag{34}$$

and

$$\begin{aligned} \mathsf{E}\left[S_t\right] &\leq \frac{1}{1 - \sup_t \left(q_j \cdot p_j + Q_j \cdot (1-p_j)\right)}, \\ \mathrm{Var}\left(S_t\right) &\leq \frac{\left(\bar{q} \cdot \bar{p} + \bar{Q} \cdot (1-\bar{p})\right) \cdot \left(1 + \bar{q} \cdot \bar{p} + \bar{Q} \cdot (1-\bar{p})\right)}{\left(1 - \bar{q} \cdot \bar{p} - \bar{Q} \cdot (1-\bar{p})\right)^2}, \end{aligned} \tag{35}$$

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

Proof: Under the time-varying probabilistic contraction-expansion mappings with added noise functions introduced in Equation (31), the distance between f_t and f_t^* can be upper bounded as

$$\begin{aligned} d\big(f_t, f_t^*\big) &= d\big(\widetilde{\mathcal{T}}_{t-1} \circ \cdots \circ \widetilde{\mathcal{T}}_0(f_0), f_t^*\big) \\ &\overset{(a)}{=} d\big(\overline{\mathcal{T}}_{t-1}\big(\widetilde{\mathcal{T}}_{t-2} \circ \cdots \circ \widetilde{\mathcal{T}}_0(f_0)\big) + w_{t-1}, f_t^*\big) = \big\|\overline{\mathcal{T}}_{t-1}\big(\widetilde{\mathcal{T}}_{t-2} \circ \cdots \circ \widetilde{\mathcal{T}}_0(f_0)\big) + w_{t-1} - f_t^*\big\| \\ &\overset{(b)}{\leq} d\big(\overline{\mathcal{T}}_{t-1}\big(\widetilde{\mathcal{T}}_{t-2} \circ \cdots \circ \widetilde{\mathcal{T}}_0(f_0)\big), f_t^*\big) + \|w_{t-1}\| \\ &\overset{(c)}{\leq} d\big(\overline{\mathcal{T}}_{t-1}\big(\widetilde{\mathcal{T}}_{t-2} \circ \cdots \circ \widetilde{\mathcal{T}}_0(f_0)\big), f_{t-1}^*\big) + d\big(f_{t-1}^*, f_t^*\big) + \|w_{t-1}\| \\ &\overset{(d)}{\leq} B_{t-1} \cdot d\big(\widetilde{\mathcal{T}}_{t-2} \circ \cdots \circ \widetilde{\mathcal{T}}_0(f_0), f_{t-1}^*\big) + \epsilon_f + \epsilon_w, \end{aligned} \tag{36}$$

where the definition of the transformation $\widetilde{\mathcal{T}}_{t-1}$ in Equation (31) is used in (a), inequalities (b) and (c) are true by the triangular inequality, and (d) follows from the assumptions $d\big(f_{t-1}^*, f_t^*\big) \leq \epsilon_f$ and $\|w_{t-1}\| \leq \epsilon_w$ 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

$$B_t = \begin{cases} q_t & \text{w.p.} \quad p_t \\ Q_t & \text{otherwise} \end{cases}. \tag{37}$$

Taking similar steps as in Equation (36), we have

$$\begin{aligned} d\big(f_t, f_t^*\big) &\le B_{t-1}\cdot\Big(B_{t-2}\cdot d\big(\widetilde{\mathcal{T}}_{t-3}\circ\cdots\circ\widetilde{\mathcal{T}}_0(f_0), f_{t-2}^*\big)+\epsilon_f+\epsilon_w\Big)+\epsilon_f+\epsilon_w \\ &\le B_{t-1}\cdot\Big(B_{t-2}\cdot\Big(B_{t-3}\cdot d\big(\widetilde{\mathcal{T}}_{t-4}\circ\cdots\circ\widetilde{\mathcal{T}}_0(f_0), f_{t-3}^*\big)+\epsilon_f+\epsilon_w\Big)+\epsilon_f+\epsilon_w\Big)+\epsilon_f+\epsilon_w \\ &\le \left(\prod_{i=0}^{t-1}B_i\right)\cdot d\big(f_0,f_0^*\big)+\left(1+\sum_{i=1}^{t-1}\prod_{j=1}^{t-i}B_j\right)\cdot(\epsilon_f+\epsilon_w) \\ &\le P_t\cdot d\big(f_0,f_0^*\big)+S_t\cdot(\epsilon_f+\epsilon_w), \end{aligned} \tag{38}$$

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 $t\ge 0$, we have

$$\begin{aligned} \mathsf{E}\left[P_t\right]=\mathsf{E}\left[\prod_{i=0}^{t-1}B_i\right]&=\prod_{i=0}^{t-1}\mathsf{E}\left[B_i\right]=\prod_{i=0}^{t-1}\big(q_t\cdot p_t+Q_t\cdot(1-p_t)\big) \\ &\le\left(\sup_t\big(q_t\cdot p_t+Q_t\cdot(1-p_t)\big)\right)^t \end{aligned} \tag{39}$$

and

$$\begin{aligned} \mathrm{Var}\left(P_t\right)=\mathsf{E}\left[P_t^2\right]-\left(\mathsf{E}\left[P_t\right]\right)^2&=\mathsf{E}\left[\prod_{i=0}^{t-1}B_i^2\right]-\prod_{i=0}^{t-1}\big(q_t\cdot p_t+Q_t\cdot(1-p_t)\big)^2 \\ &\le\prod_{i=0}^{t-1}\big(q_t^2\cdot p_t+Q_t^2\cdot(1-p_t)\big)\le\left(\sup_t\big(q_t^2\cdot p_t+Q_t^2\cdot(1-p_t)\big)\right)^t. \end{aligned} \tag{40}$$

Note that it is already shown in the proof of Theorem 1 that $q_t^2\cdot p_t+Q_t^2\cdot(1-p_t)<1$ implies $q_t\cdot p_t+Q_t\cdot(1-p_t)<1$, that is why it suffices that $\sup_t\big(q_t^2\cdot p_t+Q_t^2\cdot(1-p_t)\big)<1$. Furthermore,

$$\begin{aligned} \mathsf{E}\left[S_t\right]=\mathsf{E}\left[1+\sum_{i=1}^{t-1}\prod_{j=1}^{t-i}B_j\right]&=1+\sum_{i=1}^{t-1}\prod_{j=1}^{t-i}\mathsf{E}\left[B_j\right]=1+\sum_{i=1}^{t-1}\prod_{j=1}^{t-i}\big(q_j\cdot p_j+Q_j\cdot(1-p_j)\big) \\ &\le 1+\sum_{i=1}^{t-1}\left(\sup_t\big(q_j\cdot p_j+Q_j\cdot(1-p_j)\big)\right)^{t-i}\le\frac{1}{1-\sup_t\big(q_j\cdot p_j+Q_j\cdot(1-p_j)\big)} \end{aligned} \tag{41}$$

and

$$\mathrm{Var}\left(S_t\right)=\mathrm{Var}\left(1+\sum_{i=1}^{t-1}\prod_{j=1}^{t-i}B_j\right)=\mathrm{Var}\left(\sum_{i=1}^{t-1}\prod_{j=1}^{t-i}B_j\right)\le\mathsf{E}\left[\left(\sum_{i=1}^{t-1}\prod_{j=1}^{t-i}B_j\right)^2\right]. \tag{42}$$

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

$$\bar{B}_t=\begin{cases}\bar{q} & \text{w.p. } \bar{p}\\ \bar{Q} & \text{otherwise}\end{cases}, \tag{43}$$

where $\mathsf{E}[\bar{B}_t]\ge\sup_{i\ge1}\mathsf{E}[B_i]$ and $\mathsf{E}[\bar{B}_t^2]\ge\sup_{i\ge1}\mathsf{E}[B_i^2]$. Proceeding with Equation (42), one can write

$$\mathrm{Var}\left(S_t\right)\le\mathsf{E}\left[\left(\sum_{i=1}^{t-1}\prod_{j=1}^{t-i}B_j\right)^2\right]\le\mathsf{E}\left[\left(\sum_{i=1}^{t-1}\prod_{j=1}^{t-i}\bar{B}_j\right)^2\right]\le\mathsf{E}\left[\left(\sum_{i=1}^{\infty}\prod_{j=1}^{i}\bar{B}_j\right)^2\right]=\mathsf{E}\left[\bar{S}^2\right], \tag{44}$$

where $\bar{S}=\sum_{i=1}^{\infty}\prod_{j=1}^{i}\bar{B}_j$. We have $\mathsf{E}[\bar{S}]=\frac{\bar{q}\cdot\bar{p}+\bar{Q}\cdot(1-\bar{p})}{1-\bar{q}\cdot\bar{p}-\bar{Q}\cdot(1-\bar{p})}$ 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 $\mathsf{E}[\bar{S}^2]=\mathsf{E}[\tilde{S}^2]$, we have

$$\begin{aligned}&\mathsf{E}[\bar{S}^2]=\mathsf{E}[\bar{B}_1^2]\cdot\mathsf{E}[1+2\tilde{S}+\tilde{S}^2]=\big(\bar{q}\cdot\bar{p}+\bar{Q}\cdot(1-\bar{p})\big)\cdot\left(1+\frac{2\big(\bar{q}\cdot\bar{p}+\bar{Q}\cdot(1-\bar{p})\big)}{1-\bar{q}\cdot\bar{p}-\bar{Q}\cdot(1-\bar{p})}+\mathsf{E}[\tilde{S}^2]\right)\Longrightarrow\\&\mathsf{E}[\bar{S}^2]=\frac{\big(\bar{q}\cdot\bar{p}+\bar{Q}\cdot(1-\bar{p})\big)\cdot\big(1+\bar{q}\cdot\bar{p}+\bar{Q}\cdot(1-\bar{p})\big)}{\big(1-\bar{q}\cdot\bar{p}-\bar{Q}\cdot(1-\bar{p})\big)^2}.\end{aligned}\tag{45}$$

Putting Equations (44) and (45) together, it can be concluded that $\mathrm{Var}\,(S_t)\leq\frac{\big(\bar{q}\cdot\bar{p}+\bar{Q}\cdot(1-\bar{p})\big)\cdot\big(1+\bar{q}\cdot\bar{p}+\bar{Q}\cdot(1-\bar{p})\big)}{\big(1-\bar{q}\cdot\bar{p}-\bar{Q}\cdot(1-\bar{p})\big)^2}$, which completes the proof.

2.4. 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, for $t\in\{1,2,\dots\}$, where $\mathcal{D}\subset\mathsf{R}^d$ is a compact set and $\mathcal{R}\subset\mathsf{R}$. The goal is to ϵ-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 $x_t^*=\arg\min_{x\in\mathcal{D}}f_t(x)$. 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 signals that are independent and identically distributed over time and different input points and have a zero mean. More precisely, considering the set of input points $\mathcal{P}=\{x_1,\dots,x_n\}\subset\mathcal{D}$, the revealed values of the function f_t at round $t\in\{1,2,\dots\}$ are

$$\widetilde{f}_t(x_i)=f_t(x_i)+N_t(x_i),\tag{46}$$

where $N_t(x_i)$ are i.i.d. random variables that are strictly bounded by an interval with length L_N and $\mathsf{E}[N_t(x_i)]=0$ 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 ϵ-optimize the function with a few number of observations. Furthermore, since the function changes over time, old observations may not be useful in ϵ-optimizing the time-varying function. 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 the inaccurate old observations. We propose the following formula for estimating f_t:

$$\widehat{f}_t(x_i)=\frac{\min\{t,T+1\}-1}{\min\{t,T\}}\cdot\widehat{f}_{t-1}(x_i)+\frac{1}{\min\{t,T\}}\cdot\widetilde{f}_t(x_i)-\frac{1}{T}\cdot\widetilde{f}_{t-T}(x_i)\cdot\mathbb{1}\{t>T\},\tag{47}$$

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 Equation (47) is on the same order of that of the moving average update in reinforcement learning, but in Equation (47) 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 Equation (47) after which optimizing the estimated function $\widehat{f}_t$ ϵ-optimizes the target function f_t with an associated confidence level $1-a$, where $0<a\leq 1$. As a result, the complexity of ϵ-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 ϵ-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 3. Given $\epsilon>0$ and $a\in(0,1]$, the hitting time $T(\epsilon,a)$ is defined as

$$T(\epsilon,a)=\min\Big\{T:\mathsf{P}\big(\big\|f_t(\widehat{x}_t^*)-f_t(x_t^*)\big\|\leq\epsilon\big)\geq 1-a,\ \forall t\geq T\Big\},\tag{48}$$

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

Consider a δ-uniform grid of the function domain, $\mathcal{P}=\{x_1,x_2,\ldots,x_n\}\subset\mathcal{D}$, which means that two properties hold: (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 R^d, and (ii) for every $x\in\mathcal{D}$ there exists $x_i\in\mathcal{P}$ such that $\|x_i-x\|\leq\sqrt{d}\delta/2$. Let the grid have a fine granularity in the sense 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. Such a choice of δ 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 $\overline{\mathcal{N}}_t(\epsilon)=\{x_i\in\mathcal{P}:f_t(x_i)-f_t(x_t^*)>\epsilon\}$. It is assumed that $\overline{\mathcal{N}}_t(\epsilon)\neq\mathsf{?}$; otherwise, any point in $\mathcal{P}$ ϵ-optimizes function f_t. The following theorem presents an upper bound on the hitting time.

THEOREM 4. *Consider the unknown time-varying function f_t with the property that*

$$\|f_t(x)-f_{t-1}(x)\|\leq\frac{\epsilon^3}{43L_N^2\cdot\ln(\frac{n}{a})},\quad\forall t\geq 1,\quad\forall x\in\mathcal{D}.\tag{49}$$

Given $\epsilon>0$ and $a\in(0,1]$, the hitting time $T(\epsilon,a)$ satisfies the inequality

$$T(\epsilon,a)\leq\frac{49L_N^2}{8\epsilon^2}\cdot\ln\Big(\frac{n}{a}\Big)+1.\tag{50}$$

Proof: In order to find an upper bound on the hitting time $T(\epsilon,a)$, the function variation over time should be upper bounded; otherwise, there may not be enough time for learning the unknown

function. Assume that the time-variation of the unknown time-varying target function f_t is upper bounded by

$$\|f_t(x) - f_{t-1}(x)\| \leq \frac{\epsilon}{7T}, \quad \forall t \geq 1, \forall x \in \mathcal{D}. \tag{51}$$

Then, the hitting event used in Equation (48) satisfies the condition

$$\begin{aligned}&\left\{\exists x_i \in \mathcal{N}_t(\frac{\epsilon}{7}) \text{ such that } \frac{1}{T}\cdot\sum_{s=t-T+1}^{t} N_s(x_i) \leq \frac{2\epsilon}{7} \textbf{ and } \frac{1}{T}\cdot\sum_{s=t-T+1}^{t} N_s(x_i) \geq -\frac{2\epsilon}{7}, \forall x_i \in \overline{\mathcal{N}}_t(\epsilon)\right\}\\ &\subseteq \left\{\left\|f_t(\widehat{x}_t^*) - f_t(x_t^*)\right\| \leq \epsilon\right\}, \quad \forall t \geq T.\end{aligned} \tag{52}$$

The above equation holds true because Equations (46) and (47) 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 $t \geq T$, and by Equation (51), one can write

$$\begin{aligned}\widehat{f}_t(x_i) &\leq f_t(x_i) + \frac{\epsilon}{7} + \frac{1}{T}\cdot\sum_{s=t-T+1}^{t} N_s(x_i), \quad \forall x_i \in \mathcal{N}_t(\frac{\epsilon}{7}),\\ \widehat{f}_t(\overline{x}_j) &\geq f_t(\overline{x}_j) - \frac{\epsilon}{7} + \frac{1}{T}\cdot\sum_{s=t-T+1}^{t} N_s(\overline{x}_j), \quad \forall \overline{x}_j \in \overline{\mathcal{N}}_t(\epsilon).\end{aligned} \tag{53}$$

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 Equation (53) 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 Equation (52) 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 ϵ-optimizes the function f_t. Hence, $\widehat{x}_t^* = \arg\min_{x\in\mathcal{P}} \widehat{f}_t(x)$ ϵ-optimizes the function f_t, which means that the event on right-hand side of Equation (52) is true.

The probability of the event on the left-hand side of Equation (52) can be lower bounded as

$$\begin{aligned}&\mathsf{P}\left\{\exists x_i \in \mathcal{N}_t(\frac{\epsilon}{7}) \text{ such that } \frac{1}{T}\cdot\sum_{s=t-T+1}^{t} N_s(x_i) \leq \frac{2\epsilon}{7} \textbf{ and } \frac{1}{T}\cdot\sum_{s=t-T+1}^{t} N_s(x_i) \geq -\frac{2\epsilon}{7}, \forall x_i \in \overline{\mathcal{N}}_t(\epsilon)\right\}\\ &\overset{(a)}{\geq} \mathsf{P}\left\{\frac{1}{T}\cdot\sum_{s=t-T+1}^{t} N_s(x_i) \leq \frac{2\epsilon}{7}, x_i \in \mathcal{N}_t(\frac{\epsilon}{7})\right\}\cdot\prod_{x_i\in\overline{\mathcal{N}}_t(\epsilon)} \mathsf{P}\left\{\frac{1}{T}\cdot\sum_{s=t-T+1}^{t} N_s(x_i) \geq -\frac{2\epsilon}{7}\right\}\\ &\overset{(b)}{\geq} \prod_{x_i\in\mathcal{P}}\left(1-\exp\left(-\frac{8T\epsilon^2}{49L_N^2}\right)\right) > 1 - \sum_{i=1}^{n}\exp\left(-\frac{8T\epsilon^2}{49L_N^2}\right)\\ &= 1 - n\cdot\exp\left(-\frac{8T\epsilon^2}{49L_N^2}\right),\end{aligned} \tag{54}$$

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 Equations (52) and (54) together, we have

$$\mathsf{P}\left\{\left\|f_t(\widehat{x}_t^*) - f_t(x_t^*)\right\| \leq \epsilon\right\} \geq 1 - n\cdot\exp\left(-\frac{8T\epsilon^2}{49L_N^2}\right), \quad \forall t \geq T. \tag{55}$$

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

$$\mathsf{P}\left\{\left\|f_t(\widehat{x}_t^*)-f_t(x_t^*)\right\|\leq\epsilon\right\}\geq 1-a,\quad\forall t\geq T. \tag{56}$$

As a result, an upper bound on the hitting time $T(\epsilon,a)$ defined in Equation (48) is provided as

$$T(\epsilon,a)\leq\frac{49L_N^2}{8\epsilon^2}\cdot\ln\left(\frac{n}{a}\right)+1. \tag{57}$$

As stated earlier in Equation (51), the above analysis is true if

$$\|f_t(x)-f_{t-1}(x)\|\leq\frac{8\epsilon^3}{343L_N^2\cdot\ln(\frac{n}{a})},\quad\forall t\geq 1,\forall x\in\mathcal{D}, \tag{58}$$

and thus the analysis holds if $\|f_t(x)-f_{t-1}(x)\|\leq\frac{\epsilon^3}{43L_N^2\cdot\ln(\frac{n}{a})}<\frac{8\epsilon^3}{343L_N^2\cdot\ln(\frac{n}{a})}$.

REMARK 3. Note that the cardinality of the δ-grid with $\delta<\frac{2\epsilon}{7\sqrt{d}K}$ used in Theorem 4, namely $n=|\mathcal{P}|$, depends on ϵ. 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 δ-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 4 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 4 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 ϵ-optimize the target function f_t with a confidence level. The parameter T in Equation (47) can be set to the upper bound provided in this theorem so that old inaccurate observations are discarded and at the same time enough observations are used for an accurate estimation of f_t.

2.5. Improved Bounds for Convex Functions Consider the same framework as in Section 2.4 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 $\overline{\mathcal{M}}_t(c_1,c_2)=\{x_i\in\mathcal{P}:x_i\in\overline{C}_t(c_1,c_2)\}$. Assume that there is a constant M 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. Let $k>0$ be a lower bound on the norm of the gradient of convex function f_t over the set $\mathcal{D}\setminus C_t(\epsilon)$ for all $t\geq 1$, and assume k is known, i.e.,

$$\|\nabla f_t(x)\|\geq k,\quad\forall t\geq 1,\forall x\in\mathcal{D}\setminus C_t(\epsilon). \tag{59}$$

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 4.

THEOREM 5. *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 $x\in\mathcal{D}$. Assume that there exists $M>\epsilon$ such that the level set $L_t(M)$ is homeomorphic to a d-dimensional sphere and is inside $\mathcal{D}$ for all $t\geq 1$, and that $\|\nabla f_t(x)\|\geq k$, for all $t\geq 1$ and for all $x\in\mathcal{D}\setminus C_t(\epsilon)$. An upper bound on the hitting time $T(\epsilon,a)$ is the minimum T satisfying the inequality*

$$\sum_{l=0}^{l_m} n_l\cdot\exp\Big(-\frac{2T\big(l+\frac{2}{7}\big)^2\epsilon^2}{L_N^2}\Big)\leq a, \tag{60}$$

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: Equation (52) can be improved as

$$\begin{aligned}
&\Big\{\exists x_i\in\mathcal{M}_t(\frac{\epsilon}{7}) \text{ such that } \frac{1}{T}\cdot\sum_{s=t-T+1}^{t}N_s(x_i)\leq\frac{2\epsilon}{7} \textbf{ and } \frac{1}{T}\cdot\sum_{s=t-T+1}^{t}N_s(x_i)\geq-\frac{2\epsilon}{7},\forall x_i\in\overline{\mathcal{M}}_t\Big(\epsilon,2\epsilon\Big)\\
&\quad\textbf{and } \frac{1}{T}\cdot\sum_{s=t-T+1}^{t}N_s(x_i)\geq-\left(l+\frac{2}{7}\right)\epsilon,\forall x_i\in\overline{\mathcal{M}}_t\Big((l+1)\epsilon,(l+2)\epsilon\Big),\forall 1\leq l\leq\Big\lfloor\frac{M}{\epsilon}\Big\rfloor\Big\}\\
&\subseteq\big\{\|f_t(\widehat{x}_t^*)-f_t(x_t^*)\|\leq\epsilon\big\},\quad\forall t\geq T.
\end{aligned}\tag{61}$$

The probability of the event on the left-hand side of Equation (61) can be lower bounded as

$$\begin{aligned}
&\mathsf{P}\Big\{\exists x_i\in\mathcal{M}_t(\frac{\epsilon}{7}) \text{ such that } \frac{1}{T}\cdot\sum_{s=t-T+1}^{t}N_s(x_i)\leq\frac{2\epsilon}{7} \textbf{ and } \frac{1}{T}\cdot\sum_{s=t-T+1}^{t}N_s(x_i)\geq-\frac{2\epsilon}{7},\forall x_i\in\overline{\mathcal{M}}_t\Big(\epsilon,2\epsilon\Big)\\
&\quad\textbf{and } \frac{1}{T}\cdot\sum_{s=t-T+1}^{t}N_s(x_i)\geq-\big(l+\frac{2}{7}\big)\epsilon,\forall x_i\in\overline{\mathcal{M}}_t\Big((l+1)\epsilon,(l+2)\epsilon\Big),\forall 1\leq l\leq\Big\lfloor\frac{M}{\epsilon}\Big\rfloor\Big\}\\
&\overset{(a)}{\geq}\mathsf{P}\left\{\frac{1}{T}\cdot\sum_{s=t-T+1}^{t}N_s(x_i)\leq\frac{2\epsilon}{7},x_i\in\mathcal{M}_t(\frac{\epsilon}{7})\right\}\times\prod_{x_i\in\overline{\mathcal{M}}_t\left(\epsilon,2\epsilon\right)}\mathsf{P}\left\{\frac{1}{T}\cdot\sum_{s=t-T+1}^{t}N_s(x_i)\geq-\frac{2\epsilon}{7}\right\}\times\\
&\quad\prod_{l=1}^{\lfloor\frac{M}{\epsilon}\rfloor}\prod_{x_i\in\overline{\mathcal{M}}_t\left((l+1)\epsilon,(l+2)\epsilon\right)}\mathsf{P}\left\{\frac{1}{T}\cdot\sum_{s=t-T+1}^{t}N_s(x_i)\geq-\big(l+\frac{2}{7}\big)\epsilon\right\}\\
&\overset{(b)}{\geq}\left(1-\exp\left(-\frac{8T\epsilon^2}{49L_N^2}\right)\right)^{\overline{n}_0+1}\times\prod_{l=1}^{l_m}\left(1-\exp\left(-\frac{2T\big(l+\frac{2}{7}\big)^2\epsilon^2}{L_N^2}\right)\right)^{n_l}\\
&\geq 1-\sum_{l=0}^{l_m}n_l\cdot\exp\left(-\frac{2T\big(l+\frac{2}{7}\big)^2\epsilon^2}{L_N^2}\right)
\end{aligned}\tag{62}$$

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 Equations (61) and (62) together, we have $\mathsf{P}\left\{\|f_t(\widehat{x}_t^*) - f_t(x_t^*)\| \leq \epsilon\right\} \geq 1 - a$ for all $t \geq T$ provided that

$$\sum_{l=0}^{l_m} n_l \cdot \exp\Big(-\frac{2T\big(l+\frac{2}{7}\big)^2\epsilon^2}{L_N^2}\Big) \leq a, \tag{63}$$

which provides an upper bound on the hitting time $T(\epsilon, a)$ defined in Equation (48). As stated earlier in Equation (51), 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 4, 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 $x \in \mathcal{D}$.

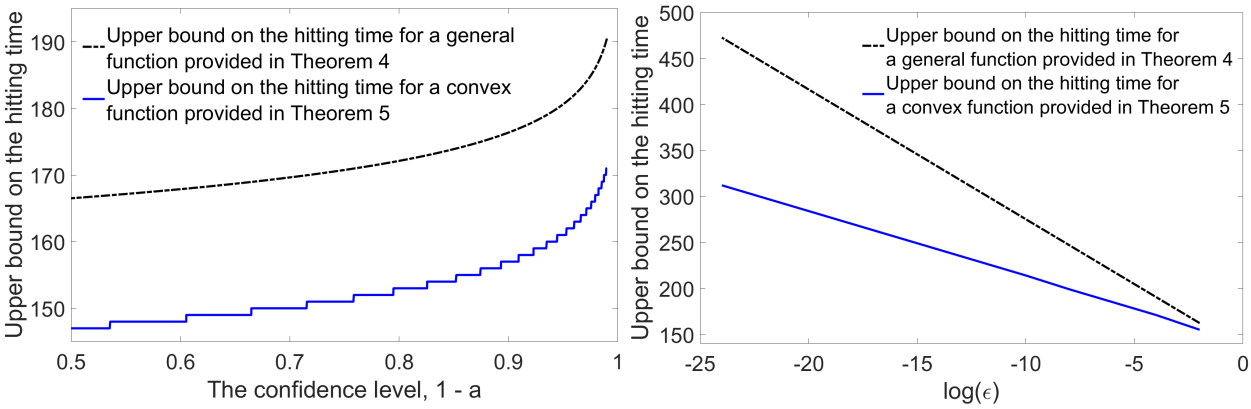
In the rest of the proof, the values of n_l for $0 \leq l \leq l_m$ are computed. The key ideas behind finding these upper bounds are that the level sets $\overline{L}_t\big((l+1)\epsilon\big)$ for $0 \leq l \leq l_m + 2$ are nested surfaces that are homeomorphic to a d-dimensional sphere inside the 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

$$\begin{aligned} n_l - 1 &\leq \frac{2^d \cdot Vol\left(C_t\big((l+1)\epsilon, (l+3)\epsilon\big)\right)}{\delta^d} \leq \frac{2^d \cdot \frac{2\epsilon}{k} \cdot A\left(P_t\big((l+1)\epsilon, (l+3)\epsilon\big)\right)}{\delta^d}, \\ \sum_{\bar{l}=l+1}^{l_m} n_{\bar{l}} &\geq \frac{Vol\left(C_t\big((l+3)\epsilon, M-\epsilon\big)\right)}{\delta^d} \geq \frac{\frac{M-(l+4)\epsilon}{K} \cdot A\left(P_t\big((l+3)\epsilon, M-\epsilon\big)\right)}{\delta^d}, \end{aligned} \tag{64}$$

where the term 2^d comes from the fact that each d-dimensional cube creates at most 2^d end-points 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,

$$\frac{n_l - 1}{n - n_l} \leq \frac{n_l - 1}{\sum_{\bar{l}=l+1}^{l_m} n_{\bar{l}}} \leq \frac{2^{d+1} \cdot K \cdot \epsilon}{k \cdot \big(M - (l+4)\epsilon\big)} = m_l \implies n_l \leq \frac{m_l}{1+m_l} \cdot n + 1. \tag{65}$$

Remark 4. It is easy to verify that Theorem 5 provides a better bound than Theorem 4 as the properties of the convex functions are leveraged. Note that a number T satisfying Equation (60) always exists. A comparison of the results of Theorems 4 and 5 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 4 and 5 when $M = K = 16, k = 2 \times 10^{-2}$, and $d = 2$. In Figure 1b, the value of n depends on ϵ, which is taken into account for drawing the plots.

3. The Hitting Time Analysis for Discrete Functions In this section, two variants of time-varying discrete functions are studied. 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 $\mathcal{X} \subset \mathsf{Z}^d$ is a bounded subset of d integer tuples and $\mathcal{R} \subset \mathsf{R}$ is a subset of real numbers (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

$$\mathcal{X}^* = \{x^* \in \mathcal{X} : f(x^*) < f(x), \forall x \in \mathcal{B}(x^*)\} \cup \{x^* \in \mathcal{X} : f(x^*) > f(x), \forall x \in \mathcal{B}(x^*)\} \tag{66}$$

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 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

$$f_t(x) = f(x) + N_t(x), \quad \forall x \in \mathcal{X}, \tag{67}$$

where $N_t(x)$ are i.i.d. random variables that are strictly bounded by an interval with length L_N and $\mathsf{E}[N_t(x)] = 0$ for all $t \in \{1, 2, \dots\}$ and $x \in \mathcal{X}$. In order to simplify the analysis, function values at adjacent points are considered to be different so that their noisy values become distinguishable after enough observations. 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

$$\widehat{f}_t(x) = \frac{t-1}{t} \cdot \widehat{f}_{t-1}(x) + \frac{1}{t} \cdot f_t(x), \quad \forall x \in \mathcal{X}. \tag{68}$$

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 optimizing the estimation function $\widehat{f}_t$ determines the local extrema of 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

$$\widehat{\mathcal{X}}_t^* = \left\{\widehat{x}^* \in \mathcal{X} : \widehat{f}_t(\widehat{x}^*) < \widehat{f}_t(x), \forall x \in \mathcal{B}(\widehat{x}^*)\right\} \cup \left\{\widehat{x}^* \in \mathcal{X} : \widehat{f}_t(\widehat{x}^*) > \widehat{f}_t(x), \forall x \in \mathcal{B}(\widehat{x}^*)\right\}. \tag{69}$$

Definition 4. Given $a \in (0,1]$, the hitting time $T(a)$ for an unknown discrete function f is defined as

$$T(a) = \min\left\{T : \mathsf{P}\left(\widehat{\mathcal{X}}_t^* = \mathcal{X}^*\right) \geq 1-a,\ \forall t \geq T\right\}, \tag{70}$$

where $\mathcal{X}^*$ and $\widehat{\mathcal{X}}_t^*$ are defined in Equations (66) and (69), 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

$$\delta(x) = \min_{x' \in \mathcal{B}(x)} \|f(x) - f(x')\|, \tag{71}$$

where $\|\cdot\|$ is the L^2-norm throughout this section. As stated earlier, it is assumed that $\delta(x) > 0$ for all $x \in \mathcal{X}$, which implies $\delta_m = \min_{x \in \mathcal{X}} \delta(x) > 0$. The following theorem presents an upper bound on the hitting time $T(a)$.

Theorem 6. *Consider the time-varying function $\hat{f}_t$ in* (68)*. Its associated hitting time $T(a)$, defined in Equation* (70)*, satisfies the inequality*

$$T(a) \leq \frac{2L_N^2}{\delta_m^2} \cdot \ln\left(\frac{2|\mathcal{X}|}{a}\right), \tag{72}$$

where L_N is the length of an interval that contains the support of all noise signals, δ_m is the minimum of $\delta(x)$ over x, $|\mathcal{X}|$ denotes the number of elements in the set $\mathcal{X}$, and $a \in (0,1]$.

Proof: In order to find an upper bound on the hitting time $T(a)$, note that the hitting event used in Equation (70) satisfies the condition

$$\left\{\frac{1}{T}\cdot\Big\|\sum_{t=1}^{T}N_t(x)\Big\|<\frac{\delta(x)}{2},\ \forall x\in\mathcal{X}\right\}\subseteq\left\{\widehat{\mathcal{X}}_T^*=\mathcal{X}^*\right\}. \tag{73}$$

The above equation holds because Equations (67) and (68) 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 half of $\delta(x)$ 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 Equation (73) can be lower bounded as

$$\begin{aligned}
&\mathsf{P}\left\{\frac{1}{T}\cdot\Big\|\sum_{t=1}^{T}N_t(x)\Big\|<\frac{\delta(x)}{2},\ \forall x\in\mathcal{X}\right\}\\
&\overset{(a)}{=}\prod_{i=1}^{|\mathcal{X}|}\mathsf{P}\left\{\frac{1}{T}\cdot\Big\|\sum_{t=1}^{T}N_t(x)\Big\|<\frac{\delta(x)}{2}\right\}\\
&\overset{(b)}{\geq}\prod_{i=1}^{|\mathcal{X}|}\left(1-2\exp\left(-\frac{T\delta(x)^2}{2L_N^2}\right)\right)\\
&>1-2\sum_{i=1}^{|\mathcal{X}|}\exp\left(-\frac{T\delta(x)^2}{2L_N^2}\right)\\
&\geq 1-2|\mathcal{X}|\cdot\exp\left(-\frac{T\delta_m^2}{2L_N^2}\right),
\end{aligned} \tag{74}$$

where (a) holds because the added noise signals are independent from each other and (b) follows from Hoeffding's inequality. Putting Equations (73) and (74) together, we have

$$\mathsf{P}\left\{\widehat{\mathcal{X}}_T^*=\mathcal{X}^*\right\}>1-2|\mathcal{X}|\cdot\exp\left(-\frac{T\delta_m^2}{2L_N^2}\right). \tag{75}$$

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

$$\mathsf{P}\left\{\widehat{\mathcal{X}}_T^*=\mathcal{X}^*\right\}>1-a, \tag{76}$$

from which the upper bound in Equation (70) follows.

3.2. A Special Case for Unimodal Functions A function f over a bounded set $\mathcal{X}\subset\mathsf{Z}$ 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.

Assume that the noise signals are continuous random variables, which implies that function $\widehat{f}_t$ has a single global minimum with probability 1. Let $\widehat{x}_t^* = \arg\min_{x\in\mathcal{X}} \ \widehat{f}_t(x)$ denote the global minimum. The hitting time for a unimodal function f is defined below.

DEFINITION 5. Given $a \in (0,1]$, the hitting time $T_u(a)$ for a unimodal function f with its global minimum at $x^* = \arg\min_{x\in\mathcal{X}} \ f(x)$ and its estimated global minimum $\widehat{x}_t^* = \arg\min_{x\in\mathcal{X}} \ \widehat{f}_t(x)$ is defined as

$$T_u(a) = \min\left\{T : \mathsf{P}\big(\widehat{x}_t^* = x^*\big) \geq 1-a, \ \forall t \geq T\right\}. \tag{77}$$

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

$$\Delta(x) = \begin{cases} f(x) - f(x^*), & \text{if } x \in \mathcal{X}\setminus\{x^*\}, \\ \min\{f(x^*-1) - f(x^*), f(x^*+1) - f(x^*)\}, & \text{if } x = x^*. \end{cases} \tag{78}$$

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

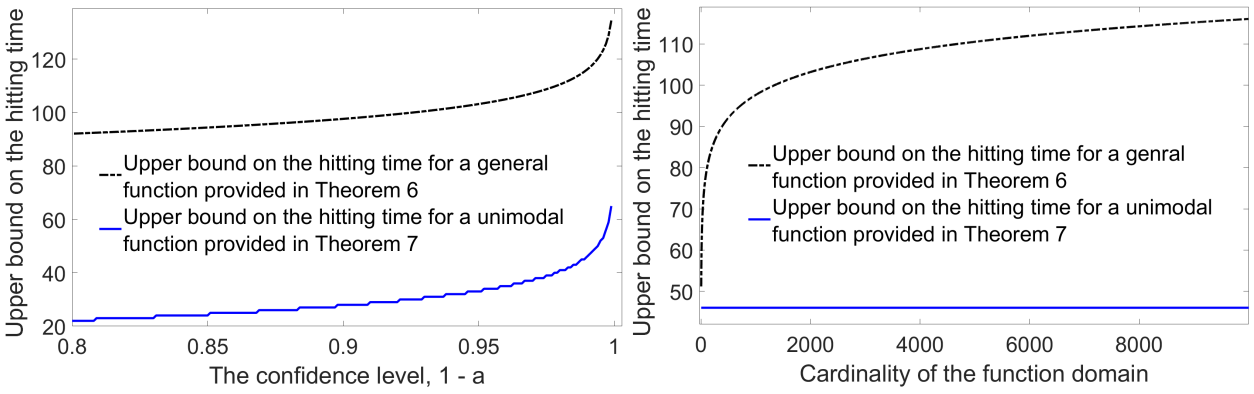
THEOREM 7. *Consider the time-varying function $\hat{f}_t$ defined in* (68) *with f being a unimodal function. The hitting time $T_u(a)$ satisfies the inequality $T_u(a) \leq T$, where T is the smallest number such that*

$$\exp\left(-\frac{\delta_m^2 T}{2L_N^2}\right) + 2\sum_{i\in\left[\left\lfloor\frac{|\mathcal{X}|}{2}\right\rfloor\right]} \exp\left(-\frac{i^2\delta_m^2 T}{2L_N^2}\right) \leq a. \tag{79}$$

(Such number T exists because the left-hand side approaches 0 when $T \to \infty$.)

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 Equation (77) satisfies the condition

$$\left\{\frac{1}{T}\cdot\sum_{t=1}^{T} N_t(x) > -\frac{\Delta(x)}{2}, \forall x \in \mathcal{X}\setminus\{x^*\} \ \textbf{and} \ \frac{1}{T}\cdot\sum_{t=1}^{T} N_t(x^*) < \frac{\Delta(x^*)}{2}\right\} \subseteq \left\{\widehat{x}_T^* = x^*\right\}. \tag{80}$$



(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 6 and 7 when $L_N = 0.02$ and $\delta_m = 0.01$.

The probability of the event on the left-hand side of Equation (80) can be lower bounded as

$$\begin{aligned}
&\mathsf{P}\left\{\frac{1}{T}\cdot\sum_{t=1}^{T}N_t(x) > -\frac{\Delta(x)}{2}, \forall x\in\mathcal{X}\setminus\{x^*\} \textbf{ and } \frac{1}{T}\cdot\sum_{t=1}^{T}N_t(x^*) < \frac{\Delta(x^*)}{2}\right\}\\
&\overset{(a)}{=} \mathsf{P}\left\{\frac{1}{T}\cdot\sum_{t=1}^{T}N_t(x^*) < \frac{\Delta(x^*)}{2}\right\} \times \prod_{x\in\mathcal{X}\setminus\{x^*\}} \mathsf{P}\left\{\frac{1}{T}\cdot\sum_{t=1}^{T}N_t(x) > -\frac{\Delta(x)}{2}\right\}\\
&\overset{(b)}{\geq} \left(1-\exp\left(-\frac{T\Delta(x^*)^2}{2L_N^2}\right)\right) \times \prod_{x\in\mathcal{X}\setminus\{x^*\}} \left(1-\exp\left(-\frac{T\Delta(x)^2}{2L_N^2}\right)\right)\\
&> 1-\exp\left(-\frac{T\Delta(x^*)^2}{2L_N^2}\right) - \sum_{x\in\mathcal{X}\setminus\{x^*\}} \exp\left(-\frac{T\Delta(x)^2}{2L_N^2}\right)\\
&\overset{(c)}{\geq} 1-\exp\left(-\frac{T\delta_m^2}{2L_N^2}\right) - \sum_{x\in\mathcal{X}\setminus\{x^*\}} \exp\left(-\frac{T(x-x^*)^2\delta_m^2}{2L_N^2}\right)\\
&\overset{(d)}{\geq} 1-\exp\left(-\frac{T\delta_m^2}{2L_N^2}\right) - 2\sum_{i\in\left[\left\lfloor\frac{|\mathcal{X}|}{2}\right\rfloor\right]} \exp\left(-\frac{Ti^2\delta_m^2}{2L_N^2}\right)
\end{aligned} \tag{81}$$

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 the value of x^* that gives rise to $x^* = \left\lceil\frac{|\mathcal{X}|}{2}\right\rceil$ (taking the ceiling corresponding to the summation through $\left\lfloor\frac{|\mathcal{X}|}{2}\right\rfloor$). Putting Equations (80) and (81) together concludes the proof.

REMARK 5. It can be verified that Theorem 7 provides a better bound than Theorem 6 as the properties of the unimodal functions are leveraged. A comparison of the results of Theorems 6 and 7 along with the details of the simulation model is depicted in Figure 2.

3.3. Time-Varying Linear Model with Additive Noise Although the time-variation of functions that arise in many real-world problems are of a nonlinear nature, we argue the generality

of a linear model of time-variation, which is the basis of our study of hitting time under shape-dominant operators to follow. Recall the standard fact in linear algebra that for any vectors $x, y \in \mathsf{R}^d$, there exists an affine transformation that satisfies $y = Ax + b$, and if $x \neq 0$, there exists a linear transformation that satisfies $y = Ax$. Similar results hold in 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 f and g are defined over a discrete domain. 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 λf 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 random perturbation w sampled from some distribution. Given a sequence of linear operators $\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_{t-1}$ such that $\|\mathcal{A}_i\| = \sup_{f\neq 0} \frac{\|\mathcal{A}_i f\|}{\|f\|} = 1$ together with the perturbations $w_0, \ldots, w_{t-1}$, consider the following model of linear time variation:

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

What properties the operators $\mathcal{T}_1, \mathcal{T}_2, \ldots$ should satisfy in order for f_t to almost reach a target function f^* at time $t = \tau$? We will provide an answer using the notion of shape-dominant operator in the next subsection. 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 $t = \tau$, the function f_τ becomes convex with a unique global minimum x_τ^*, then no matter how optimization is carried out for f_1 through $f_{\tau-1}$, minimizing f_τ will yield the same solution x_τ^*, which is globally optimal. The effect of minimizing f_τ cancels out the sub-optimality at time x_0. Moreover, under some technical conditions, the global solution at time τ can be used to find global solutions at future times using tracking methods (Ding, Lavaei, and Arcak 2019, Fattahi et al. 2019, Massicot and Marecek 2019). In other words, the shape of f_τ affects the complexity of online optimization in the long run.

3.3.1. Shape Dominant Model Consider the time-varying function $f_t : \mathcal{X} \to \mathsf{R}$, for $t \in \{0, 1, \ldots\}$, where $\mathcal{X} \subset \mathsf{Z}^d$ is a finite set with $\mathcal{X} = \{x_1, \ldots, x_n\}$. Equivalently, f_t is a vector in R^n. Let $P_i(A, w)$ denote the joint distribution of A_i and w_i.

DEFINITION 6. The joint distribution $P(A,w)$ is said to be $(\delta,\sigma,f^*,\phi^*)$ shape dominant if the following conditions hold with probability 1:

1. the unit vector f^* is the eigenvector of A associated with eigenvalue 1;
2. the unit vector ϕ^* is the eigenvector of $A^\top$ associated with eigenvalue 1;
3. $\langle f^*,\phi^*\rangle \neq 0$;
4. all other eigenvalues of A have norm less than $1-\delta$;
5. conditioned on A, the noise w has zero mean and is sub-Gaussian with parameter σ^2 in the sense that for all $u\in\mathsf{R}^n$ with $\|u\|\leq 1$, we have $\mathsf{E}[\exp(su^\top w)]\leq \exp\left(\frac{\sigma^2 s^2}{2}\right)$.

In order to understand the conditions in Definition 6, consider the special case where A is a positive stochastic matrix whose column sums are all 1. The unit vector $\phi^*=(\frac{1}{\sqrt{n}},\dots,\frac{1}{\sqrt{n}})$ is the eigenvector of $A^\top$ associated with eigenvalue 1. By the Perron-Frobenius theorem, A also has an all-positive eigenvector f^* with eigenvalue 1, and all other eigenvalues of A have norm strictly less than 1. The vector f^* is the equilibrium distribution of a Markov chain whose transition matrix is A. Therefore, Conditions 1 and 3 are automatically satisfied. Moreover, Condition 2 amounts to requiring that, almost surely, the Markov chain defined by A has a fixed equilibrium f^*.

THEOREM 8. *Assume that $P_t(A,w)$ is $(\delta,\sigma_t,f^*,\phi^*)$ shape dominant and independent for all $t\in\{1,2,\dots,k\}$; then,*

$$f_k=\mathcal{T}_{k-1}\circ\cdots\circ\mathcal{T}_0 f_0=\frac{\langle\phi^*,f_0+\sum_{t=0}^{k-1}w_t\rangle}{\langle\phi^*,f^*\rangle}f^*+v+w,$$

where

$$\|v\|\leq(1-\delta)^k\left(\|f_0\|+\frac{\langle\phi^*,f_0\rangle}{\langle\phi^*,f^*\rangle}\right),$$

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

Proof: Consider the operator $\mathcal{T}_i f=A_i f+w_i$ that is $(\delta,\sigma_i,f^*,\phi^*)$ shape dominant for $i\in\{0,1\dots,k-1\}$. Construct the subspace

$$\mathcal{G}=\{g\in\mathsf{R}^n,\langle\phi^*,g\rangle=0\}.$$

Since $\langle\phi^*,f^*\rangle\neq 0$, we have $f^*\notin\mathcal{G}$. Since ϕ^* is the eigenvector of $A_i^\top$, the following holds for all $g\in\mathcal{G}$

$$\langle\phi^*,A_i g\rangle=\langle A_i^\top\phi^*,g\rangle=\langle\phi^*,g\rangle=0.$$

Therefore, $A_i g\in\mathcal{G}$, and $\mathcal{G}$ is an invariant subspace of A_i in R^n. 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 R^n, under which the linear operator A_i takes the form

$$A_i=\begin{bmatrix}1 & 0 & \dots & 0\\ 0 & & & \\ \vdots & & A_i' & \\ 0 & & & \end{bmatrix},\tag{82}$$

where A'_i is a random matrix in $\mathsf{R}^{(n-1)\times(n-1)}$. With a slight abuse of notation, we regard A'_i as a linear transformation from $\mathcal{G}$ to $\mathcal{G}$. Note that $\|A'\|_2 \leq 1-\delta$ because all other eigenvalues of A_i have norm less than $1-\delta$. Under the basis B, f_0 has the representation

$$f_0 = \frac{\langle \phi^*, f_0 \rangle}{\langle \phi^*, f^* \rangle} f^* + g, \tag{83}$$

where $g \in \mathcal{G}$. As a result,

$$\begin{aligned} f_k &= \mathcal{T}_{k-1} \circ \cdots \circ \mathcal{T}_0 f_0 \\ &= A_{k-1} \cdots A_0 f_0 + \sum_{i=0}^{k-1} A_{k-1} \cdots A_{i+1} w_i \\ &= \frac{\langle \phi^*, f_0 \rangle}{\langle \phi^*, f^* \rangle} f^* + A'_{k-1} \ldots A'_1 g + \sum_{i=0}^{k-1} A_{k-1} \cdots A_{i+1} w_i. \end{aligned}$$

The norm estimate gives rise to

$$\left\| A'_{k-1} \ldots A'_1 g \right\| \leq (1-\delta)^k \cdot \|g\| \leq (1-\delta)^k \cdot \left(\|f_0\| + \left| \frac{\langle \phi^*, f_0 \rangle}{\langle \phi^*, f^* \rangle} \right| \right),$$

where the triangle inequality is used. Similarly, one can write

$$w_i = \frac{\langle \phi^*, w_i \rangle}{\langle \phi^*, f^* \rangle} f^* + h_i,$$

where $h_i \in \mathcal{G}$. We have

$$A_{k-1} \cdots A_{i+1} w_i = \frac{\langle \phi^*, w_i \rangle}{\langle \phi^*, f^* \rangle} f^* + A'_{k-1} \cdots A'_{i+1} h_i.$$

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

$$\begin{aligned} &\mathsf{E}\left[\exp\left(s \left\langle u, A'_{k-1} \cdots A'_{i+1} h_i \right\rangle \right) \right] \\ &= \mathsf{E}\left[\exp\left(s \left\langle A'^{\top}_{i+1} \cdots A'^{\top}_{k-1} u, h_i \right\rangle \right) \right] \\ &= \mathsf{E}\left[\exp\left(s \left\langle A'^{\top}_{i+1} \cdots A'^{\top}_{k-1} u, w_i - \frac{\langle \phi^*, w_i \rangle}{\langle \phi^*, f^* \rangle} f^* \right\rangle \right) \right] \\ &= \mathsf{E}\left[\exp\left(s \left\langle A'^{\top}_{i+1} \cdots A'^{\top}_{k-1} u, w_i \right\rangle \right) \times \exp\left(s \left\langle -\frac{\langle A'^{\top}_{i+1} \cdots A'^{\top}_{k-1} u, f^* \rangle}{\langle \phi^*, f^* \rangle} \phi^*, w_i \right\rangle \right) \right] \\ &\leq \exp\left(\frac{\sigma_i^2 s^2 \left\| A'^{\top}_{i+1} \cdots A'^{\top}_{k-1} u \right\|^2}{2} \right) \times \exp\left(\frac{\sigma_i^2 s^2}{2} \left(\frac{\langle A'^{\top}_{i+1} \cdots A'^{\top}_{k-1} u, f^* \rangle}{\langle \phi^*, f^* \rangle} \right)^2 \right) \\ &\leq \exp\left(\frac{\sigma_i^2 s^2 (1-\delta)^{2(k-i)} \left(1 + \frac{1}{\langle \phi^*, f^* \rangle^2} \right)}{2} \right). \end{aligned} \tag{84}$$

This implies that $A'_{k-1} \cdots A'_{i+1} h_i$ is sub-Gaussian with parameter $\sigma_i^2 (1-\delta)^{2(k-i)} \left(1 + \frac{1}{\langle \phi^*, f^* \rangle^2} \right)$, and thereby, $\sum_{i=0}^{k-1} A'_{k-1} \cdots A'_{i+1} h_i$ is sub-Gaussian with parameter $\sigma^2 = \left(1 + \frac{1}{\langle \phi^*, f^* \rangle^2} \right) \sum_{i=0}^{k-1} (1-\delta)^{2(k-i)} \sigma_i^2$.

Theorem 8 states that if the time-varying model is given by shape dominant operators, the function f_k 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 9. *Under the same assumptions made in Theorem 8, define*

$$\tau_\epsilon = \inf\left\{k : \exists \lambda \in \mathsf{R} \ \textit{s.t.}\ \|f_k - \lambda f^*\| < \epsilon\right\}, \tag{85}$$

where $\epsilon > 0$. *Suppose that* $k > \frac{\log 2\left(\|f_0\| + \left|\frac{\langle \phi^*, f_0\rangle}{\langle \phi^*, f^*\rangle}\right|\right) - \log \epsilon}{\log \frac{1}{1-\delta}}$; *then,*

$$\mathsf{P}(\tau_\epsilon \geq k) \leq C_n \exp\left(-\frac{\epsilon^2}{32\left(1 + \frac{1}{\langle \phi^*, f^*\rangle^2}\right)\sum_{i=0}^{k-1}(1-\delta)^{2(k-i)}\sigma_i^2}\right).$$

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

Proof: From the proof of Theorem 8, we note the following decomposition

$$f_k = \frac{\langle \phi^*, f_0 + \sum_{i=0}^{k-1} w_i\rangle}{\langle \phi^*, f^*\rangle} f^* + v^{(k)} + w^{(k)},$$

where $\left\|v^{(k)}\right\| < (1-\delta)^k(\|f_0\| + \left|\frac{\langle \phi^*, f_0\rangle}{\langle \phi^*, f^*\rangle}\right|)$ and

$$w^{(k)} = \sum_{i=0}^{k-1} A'_{k-1} \cdots A'_{i+1} h_i$$

is sub-Gaussian with parameter $\sigma^2 = \left(1 + \frac{1}{\langle \phi^*, f^*\rangle^2}\right)\sum_{i=0}^{k-1}(1-\delta)^{2(k-i)}\sigma_i^2$. From the definition of the hitting time in (85), we have

$$\mathsf{P}(\tau_\epsilon < k) \geq \mathsf{P}\left(\left\|v^{(k)}\right\| < \epsilon/2, \left\|w^{(k)}\right\| < \epsilon/2\right).$$

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

$$\mathsf{P}\left(\left\|w^{(k)}\right\| < \epsilon/2\right) = 1 - \mathsf{P}\left(\|w_k\| > \epsilon/2\right) \geq 1 - C_n \exp\left(-\frac{\epsilon^2}{32\sigma^2}\right),$$

where C_n is a universal constant depending only on n.

To understand the above bound, consider a fixed index k. When σ_i decreases, the bound becomes smaller. As a result, with a smaller random perturbation, it is more likely to reach the target function faster. As ϵ increases, the bound becomes smaller, which matches the intuition that a larger neighborhood is easier to reach than a smaller one.

REMARK 6. Note that the analysis of the linear model with additive noise in Section 3.3 can be generalized to continuous functions by working through eigenfunctions as opposed to eigenvectors. We briefly discuss this in the special case where the function space has a finite number of bases. Let the inner product in the function space be $\langle f, g\rangle = \int f(x) \cdot g(x) dx$ and the function space to have an orthonormal basis given by the set of functions $\{u_1, u_2, \ldots, u_n\}$; then,

$$\langle u_i, u_j\rangle = \int u_i(x) \cdot u_j(x) dx = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}, \tag{86}$$

where δ_{ij} is the Kronecker delta function. Note that any function can be decomposed into a linear combination of the basis functions, i.e.,

$$f(x) = \sum_{j=1}^{n} a_j \cdot u_j(x), \tag{87}$$

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

$$A_{ij} = \langle u_i, \mathcal{T}(u_j)\rangle = \int u_i(x) \cdot \mathcal{T}\big(u_j(x)\big) dx. \tag{88}$$

Applying the operator $\mathcal{T}$ on both sides of Equation (87) yields that

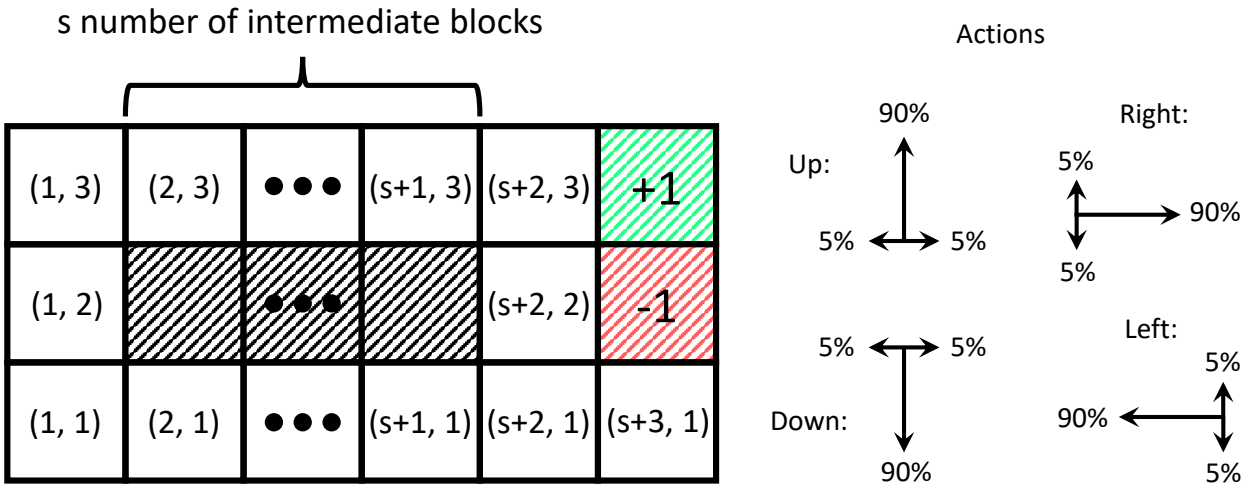
$$\mathcal{T}\big(f(x)\big) = \sum_{j=1}^{n} a_j \cdot \mathcal{T}\big(u_j(x)\big) = \sum_{j=1}^{n} b_j \cdot u_j(x). \tag{89}$$

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

$$\sum_{j=1}^{n} a_j \cdot \big\langle u_i, \mathcal{T}\big(u_j\big)\big\rangle = \sum_{j=1}^{n} b_j \cdot \langle u_i, u_j\rangle \Longrightarrow \sum_{j=1}^{n} a_j \cdot A_{ij} \quad = b_i. \tag{90}$$

The above equation is the matrix multiplication $Aa = b$, which is the matrix equivalent of the operator $\mathcal{T}$ acting upon the function $f(x)$ expressed in the orthonormal basis. If $f(x)$ is an eigenfunction of transformation $\mathcal{T}$ with eigenvalue λ, we have $Aa = \lambda a$. Hence, the results of Theorem 9 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.

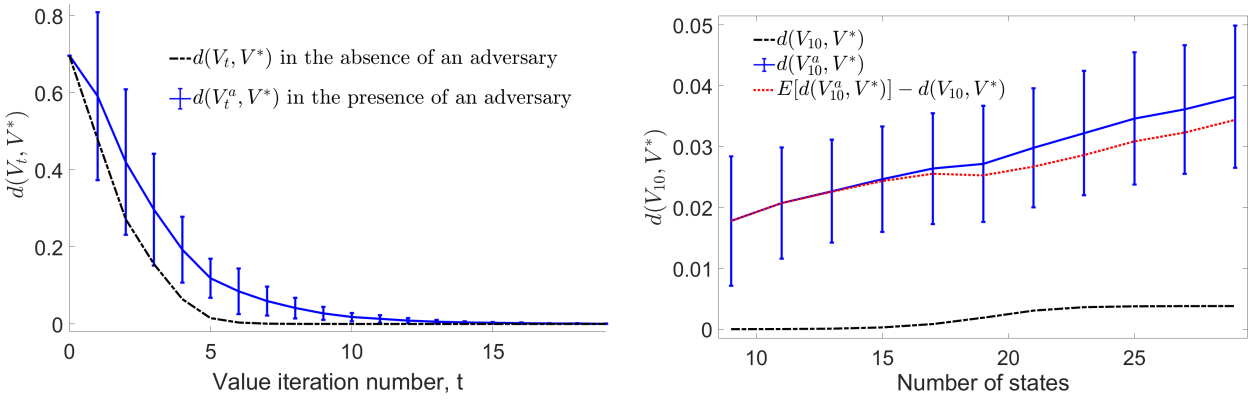
4. Simulation Results In this section, the adversarial attack on the computation of value iteration is simulated for an agent interacting with an environment depicted in Figure 3. The agent can take any of the four actions Up, Down, Right, and Left in each of the non-terminal states. By taking an action, 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



(a) (b)

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

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, $\mathsf{E}\big[d(V_{10}^a, V^*)\big] - d(V_{10}, V^*)$ 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.



(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.

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

Ansari MS, Alsamhi SH, Qiao Y, Ye Y, Lee B, 2020 *Security of distributed intelligence in edge computing: Threats and countermeasures. The Cloud-to-Thing Continuum*, 95–122 (Palgrave Macmillan, Cham).

Bertsekas DP, 2011 *Approximate policy iteration: A survey and some new methods. Journal of Control Theory and Applications* 9(3):310–335.

Bottou L, Peters J, Quiñonero-Candela J, Charles DX, Chickering DM, Portugaly E, Ray D, Simard P, Snelson E, 2013 *Counterfactual reasoning and learning systems: The example of computational advertising. The Journal of Machine Learning Research* 14(1):3207–3260.

Browne CB, Powley E, Whitehouse D, Lucas SM, Cowling PI, Rohlfshagen P, Tavener S, Perez D, Samothrakis S, Colton S, 2012 *A survey of monte carlo tree search methods. IEEE Transactions on Computational Intelligence and AI in games* 4(1):1–43.

Busoniu L, Babuska R, De Schutter B, Ernst D, 2010 *Reinforcement learning and dynamic programming using function approximators*, volume 39 (CRC press).

Chang HS, Hu J, Fu MC, Marcus SI, 2013 *Simulation-based algorithms for Markov decision processes* (Springer Science & Business Media).

Coulom R, 2006 *Efficient selectivity and backup operators in monte-carlo tree search. International conference on computers and games*, 72–83 (Springer).

De Farias DP, Van Roy B, 2000 *On the existence of fixed points for approximate value iteration and temporal-difference learning. Journal of Optimization theory and Applications* 105(3):589–608.

Dimitri PB, 2017 *Dynamic programming and optimal control.* (Athena Scientific).

Ding Y, Lavaei J, Arcak M, 2019 *Escaping spurious local minimum trajectories in online time-varying nonconvex optimization. arXiv preprint arXiv:1912.00561* .

Fattahi S, Josz C, Mohammadi R, Lavaei J, Sojoudi S, 2019 *Absence of spurious local trajectories in time-varying optimization. arXiv preprint arXiv:1905.09937* .

Feng H, Yekkehkhany A, Lavaei J, 2020 *A hitting time analysis of non-convex optimization with time-varying revelations* .

Fu MC, 2017 *Markov decision processes, alphago, and monte carlo tree search: Back to the future. Leading Developments from INFORMS Communities*, 68–88 (INFORMS).

Gu F, Chang H, Zhu W, Sojoudi S, El Ghaoui L, 2020 *Implicit graph neural networks. Advances in Neural Information Processing Systems* 33.

Isakov M, Gadepally V, Gettings KM, Kinsy MA, 2019 *Survey of attacks and defenses on edge-deployed neural networks. 2019 IEEE High Performance Extreme Computing Conference (HPEC)*, 1–8 (IEEE).

Iyengar GN, 2005 *Robust dynamic programming. Mathematics of Operations Research* 30(2):257–280.

Li H, Ota K, Dong M, 2018 *Learning iot in edge: Deep learning for the internet of things with edge computing. IEEE network* 32(1):96–101.

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

Massicot O, Marecek J, 2019 *On-line non-convex constrained optimization. arXiv preprint arXiv:1909.07492* .

Mulvaney-Kemp J, Fattahi S, Lavaei J, 2020 *Load variation enables escaping poor solutions of time-varying optimal power flow.*

Nilim A, El Ghaoui L, 2005 *Robust control of markov decision processes with uncertain transition matrices. Operations Research* 53(5):780–798.

Park S, Glista E, Lavaei J, Sojoudi S, 2020 *Homotopy method for finding the global solution of post-contingency optimal power flow. 2020 American Control Conference (ACC)*, 3126–3133 (IEEE).

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

Rao CV, Rawlings JB, Mayne DQ, 2003 *Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations. IEEE transactions on automatic control* 48(2):246–258.

Roy BV, 2006 *Td (0) leads to better policies than approximate value iteration. Advances in Neural Information Processing Systems*, 1377–1384.

Satyanarayanan M, 2017 *The emergence of edge computing. Computer* 50(1):30–39.

Simonetto A, Mokhtari A, Koppel A, Leus G, Ribeiro A, 2016 *A class of prediction-correction methods for time-varying convex optimization. IEEE Transactions on Signal Processing* 64(17):4576–4591.

Sun R, 2019 *Optimization for deep learning: theory and algorithms. arXiv preprint arXiv:1912.08957* .

Tang Y, Dall'Anese E, Bernstein A, Low S, 2018 *Running primal-dual gradient method for time-varying nonconvex problems. arXiv preprint arXiv:1812.00613* .

Tsitsiklis JN, Van Roy B, 1996 *Feature-based methods for large scale dynamic programming. Machine Learning* 22(1-3):59–94.

Van Roy B, 1998 *Learning and value function approximation in complex decision processes.* Ph.D. thesis, Massachusetts Institute of Technology.

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

Xiao Y, Jia Y, Liu C, Cheng X, Yu J, Lv W, 2019 *Edge computing security: State of the art and challenges. Proceedings of the IEEE* 107(8):1608–1631.