On Robustness Metrics for Learning STL Tasks

Peter Varnai and Dimos V. Dimarogonas[1]

***Abstract*— Signal temporal logic (STL) is a powerful tool for describing complex behaviors for dynamical systems. Among many approaches, the control problem for systems under STL task constraints is well suited for learning-based solutions, because STL is equipped with robustness metrics that quantify the satisfaction of task specifications and thus serve as useful rewards. In this work, we examine existing and potential robustness metrics specifically from the perspective of how they can aid such learning algorithms. We show that various desirable properties restrict the form of potential metrics, and introduce a new one based on the results. The effectiveness of this new robustness metric for accelerating the learning procedure is demonstrated through an insightful case study.**

I. INTRODUCTION

Formal methods have much potential in the field of robotics due to the ability of temporal logics to express rich and complex desired system behaviors. In accordance, developing control methods for achieving these behaviors has become an area of increasing research interest. Different temporal logics, such as linear temporal logic (LTL) [1], metric temporal logic (MTL) [2], or signal temporal logic (STL) [3], can be used to formulate different types of behavioral specifications. This has led to a wide array of control approaches involving guarantees of task satisfaction constraints through automata-based planning [4], mixed-integer linear programming [5], or feedback-based control laws [6], among many others. Recently, reinforcement learning methods have also been investigated in this context [7]–[9].

Focusing on learning-based approaches, STL is of special interest; unlike most other temporal logics, it allows definitions of robustness metrics associated to the degree of task satisfaction [10]. These metrics quantify how much a task is satisfied or violated, and serve as more descriptive rewards for learning than a simple true/false answer as to whether task satisfaction is achieved or not. Thus, robustness metrics can be seen as a general form of *reward shaping* for the class of behaviors described by STL specifications. Reward shaping is well-known to play a crucial role in the convergence of learning methods [11], which motivates the study of robustness metrics from such a learning perspective.

Recently, many new extensions of the traditional robustness metric [10] for STL have been introduced in the literature for various purposes. Examples include discretized and cumulative definitions aiming to ease computational burdens and to smoothen the robustness metric for use in real-time optimization [12]–[14]. These rely on point-wise or smooth approximations of the $\min$ and $\max$ operators in [10], and do not preserve the desirable property of having the sign of the metric directly relate to the satisfaction of the corresponding STL expression. In [15], the authors define a metric that preserves this so-called *soundness* property while achieving higher robustness against noise than the traditional metric.

To the extent of the authors' knowledge, this work is the first to consider STL robustness metrics explicitly for their role in accelerating learning procedures. The focus is on the so-called *shadowing problem* of the traditional robustness metric [10], whose definition is such that increasing the robustness of one term in a conjunction of propositions does not show in the robustness of the conjunction itself. This makes it more difficult for learning methods to find improvements towards task satisfaction through exploration. Our main contribution towards countering this problem is two-fold. First, we provide a theoretical study of some fundamental limitations involved in designing any robustness metric that is assumed to satisfy a set of chosen desirable properties for its role in aiding learning algorithms. Second, we use the findings to define a new class of robustness metrics specifically engineered for aiding exploration.

The paper is organized as follows. Section II provides background on STL, its robustness metrics, and outlines the algorithm used to compare their performances in aiding learning. Sections III and IV organize desirable properties for constructing metrics, and examine the restrictions these impose. Section V introduces a class of robustness metrics, whose effectiveness in accelerating learning is demonstrated by a simulation case study in Section VI. Concluding remarks follow in Section VII. The interested reader is referred to [16] for further details on the presented results.

II. PRELIMINARIES

A. Signal temporal logic (STL)

In STL, the predicates are defined over continuous-time signals, such as the state $\boldsymbol{x}$ of the system [3]. The logical *predicates* μ are either true ($\top$) or false ($\bot$), which is determined according to a corresponding function $h^\mu : \mathbb{R}^n \to \mathbb{R}$ as $\mu = \top$ if $h^\mu(\boldsymbol{x}) \geq 0$ and $\mu = \bot$ otherwise. The predicates can be recursively combined using Boolean and temporal operators to form more complex *task specifications* ϕ:

$$\phi := \top \mid \mu \mid \neg\phi \mid \phi_1 \wedge \phi_2 \mid \phi_1 U_{[a,b]} \phi_2,$$

where the *until operator* $U_{[a,b]}$ requires ϕ_1 to hold until ϕ_2 eventually becomes true in the time interval defined by $[a, b]$.

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation, the Swedish Research Council (VR), the SSF COIN project, and the EU H2020 Co4Robots project.

[1]Both authors are with the Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden. varnai@kth.se (P. Varnai), dimos@kth.se (D. V. Dimarogonas)

For a formal definition of the STL language semantics, we refer to [3]. Briefly, satisfaction of an expression (denoted $(\boldsymbol{x},t) \vDash \phi$) can be computed recursively using the semantics $(\boldsymbol{x},t) \vDash \mu \Leftrightarrow h^{\mu}(\boldsymbol{x}(t)) \geq 0$; $(\boldsymbol{x},t) \vDash \neg\phi \Leftrightarrow \neg((\boldsymbol{x},t) \vDash \phi)$; $(\boldsymbol{x},t) \vDash \phi_1 \wedge \phi_2 \Leftrightarrow (\boldsymbol{x},t) \vDash \phi_1 \wedge (\boldsymbol{x},t) \vDash \phi_2$; and $(\boldsymbol{x},t) \vDash \phi_1 U_{[a,b]} \phi_2 \Leftrightarrow \exists t_1 \in [t+a, t+b]$ such that $(\boldsymbol{x},t_1) \vDash \phi_2$ and $(\boldsymbol{x},t_2) \vDash \phi_1 \; \forall t_2 \in [t,t_1]$. Other expressive temporal operators include *eventually* and *always*, whose definition follows from $F_{[a,b]}\phi = \top U_{[a,b]}\phi$ and $G_{[a,b]}\phi = \neg F_{[a,b]} \neg \phi$.

B. Robustness metrics for STL

Definitions of *robustness metrics* ρ for STL give a quantitative indication of how well a task is satisfied. In this work, we consider spatial robustness metrics, which measure the extent of satisfaction given the imposed STL timing specifications. There are also notions of time robustness [17], which quantify how well the timing requirements are met.

The metric defined in [10] is termed the *traditional* robustness metric and can be evaluated recursively as follows for the STL operators that are the focus of this work: $\rho^{\mu}(\boldsymbol{x},t) = h^{\mu}(\boldsymbol{x}(t))$; $\rho^{\neg\phi}(\boldsymbol{x},t) = -\rho^{\phi}(\boldsymbol{x},t)$; $\rho^{\phi_1 \wedge \phi_2}(\boldsymbol{x},t) = \min\big(\rho^{\phi_1}(\boldsymbol{x},t)\rho^{\phi_2}(\boldsymbol{x},t)\big)$; $\rho^{F_{[a,b]}\phi}(\boldsymbol{x},t) = \max_{t' \in [t+a,t+b]} \rho^{\phi}(\boldsymbol{x},t')$; and finally $\rho^{G_{[a,b]}\phi}(\boldsymbol{x},t) = \min_{t' \in [t+a,t+b]} \rho^{\phi}(\boldsymbol{x},t')$. A basic property of this robustness metric is that its sign gives an explicit indication of whether or not its corresponding STL specification is satisfied. Mathematically, this is described as the following property [10].

Property 1 (Soundness). A robustness metric is sound if, for any specification ϕ, $\rho^{\phi}(\boldsymbol{x},t) \geq 0$ if and only if the signal $\boldsymbol{x}$ satisfies ϕ at time t, i.e., $(\boldsymbol{x},t) \vDash \phi$.

Most robustness metrics defined in the literature [12]–[14] do not preserve this property in exchange for gains from a computational or optimization perspective. On the other hand, the so-called AG metric [15] was introduced using arithmetic and geometric means in a way to preserve this property. The goal therein was to better express robustness of some formulas, e.g., the conjunction of the metrics $\{1,1,1\}$ should receive a lower robustness score than that of $\{1,10,10\}$, whereas both achieve a score of 1 using the traditional definition. This behavior also helps counter the shadowing problem discussed in the previous section.

C. Guided policy improvement with path integrals (G-PI^2)

Policy improvement with path integrals (PI^2) is a reinforcement learning algorithm for solving optimal control problems in unknown environments. Following our recent work regarding its application to satisfying STL tasks, its guided variant will be used to compare the performance of different robustness metrics in terms of accelerating the learning process. Due to space limitations, here we give a brief outline of the method, and refer to [18] for details.

G-PI^2 searches for a policy $\pi(\boldsymbol{x},t) = \hat{\boldsymbol{u}}(\boldsymbol{x},t) + \boldsymbol{k}_{\theta}(t)$ in order to minimize a user-defined cost function $J(\tau)$ of the system trajectory τ from a given initial state $\boldsymbol{x}_0$. Here, $\hat{\boldsymbol{u}}(\boldsymbol{x},t)$ is a feedback controller used to guide exploration by aiming to impose given timing constraints, i.e., *funnels* $\rho^{\mu_i}(\boldsymbol{x}(t)) \geq \gamma(t)$, on the evolution of the $i = 1,\dots,M$ atomic propositions μ_i composing the STL task ϕ. The algorithm then seeks the parameters θ of the feedforward terms $\boldsymbol{k}_{\theta}(t)$ which (locally) minimize $J(\tau)$. In its (k)-th iteration, a set of N parameters are sampled from around the current solution estimate $\theta^{(k)}$, the corresponding controllers' performances are evaluated using the cost $J(\tau)$, and the θ parameters are updated towards the more optimal ones. The iterations are repeated a given number of times or until convergence. The cost function $J(\tau)$ is composed of a cost of interest $C(\tau)$ to be minimized and a penalty term for progressively enforcing STL task satisfaction by penalizing negative values of the corresponding robustness $\rho^{\phi}(\boldsymbol{x},0)$.

III. STRUCTURED DEFINITION OF ROBUSTNESS METRICS

In order to algorithmically calculate the robustness metrics of complex STL formulas, it is useful to define them in a recursive manner. This was the case for the traditional robustness measure given in the previous section; the expressions for the negation, conjunction, eventually, and always operators all rely on previously computed robustnesses.

Formally, we introduce the abstract operators $\mathcal{N}$, $\mathcal{A}$, $\mathcal{O}$, $\mathcal{F}$, and $\mathcal{G}$ for negation, conjunction, disjunction, and the eventually and always operators. We then require the well-known De Morgan identities of Boolean and temporal logic to be transferred to the robustness metric itself. Thus, the identity $\mathcal{O}(\phi_1,\dots,\phi_M) = \mathcal{N}(\mathcal{A}(\mathcal{N}(\phi_1),\dots,\mathcal{N}(\phi_M)))$ imposes

$$\mathcal{O}(\rho_1,\dots,\rho_M) = \mathcal{N}(\mathcal{A}(\mathcal{N}(\rho_1),\dots,\mathcal{N}(\rho_M))) \tag{1}$$

on the robustness metric. Similarly, between the always and eventually operators, $\mathcal{F}_{[a,b]}(\phi) = \mathcal{N}(\mathcal{G}_{[a,b]}(\mathcal{N}(\phi)))$ requires

$$\mathcal{F}_{[a,b]}(\rho^{\phi}(\boldsymbol{x},t)) = \mathcal{N}(\mathcal{G}_{[a,b]}(\mathcal{N}(\rho^{\phi}(\boldsymbol{x},t)))). \tag{2}$$

Further note that the always operator can be interpreted as a conjunction over a given time frame. Discretizing the interval $[a,b]$ into $i = 1,\dots,M$ evenly spaced points t_i, we have:

$$\mathcal{G}_{[a,b]}(\rho^{\phi}(\boldsymbol{x},t)) = \lim_{M\to\infty} \mathcal{A}(\rho^{\phi}(\boldsymbol{x},t+t_1),\dots,\rho^{\phi}(\boldsymbol{x},t+t_M)) \tag{3}$$

The identities (1), (2), and (3) allow us to define new robustness metrics in a structured manner, from the elementary definitions of the $\mathcal{N}$ and $\mathcal{A}$ operators. The former can also be excluded from the design by setting $\mathcal{N}(\rho^{\phi}) := -\rho^{\phi}$, a natural choice for quantifying the satisfaction or violation of a task symmetrically by the same magnitude robustness metric. Thus, the operators required to evaluate the robustness of STL formulas recursively can be constructed given the definition of the single AND operator $\mathcal{A}$, significantly easing the discussion related to constructing new metrics.

As an example, the arithmetic-geometric mean robustness metric of [15] (denoted by 'AG') can be constructed from:

$$\mathcal{A}^{\mathrm{AG}}(\rho_1,\dots,\rho_M) = \begin{cases} \frac{1}{M}\sum_{i=1}^{M} \max(\rho_i,0) & \text{if } \min_i \rho_i \leq 0, \\ \sqrt[M]{\prod_{i=1}^{M}(1+\rho_i)} - 1 & \text{otherwise.} \end{cases} \tag{4}$$

Similarly, the traditional robustness metric can be constructed from $\mathcal{A}^{\mathrm{trad}}(\rho_1,\dots,\rho_M) = \min(\rho_1,\dots,\rho_M)$.

IV. PROPERTIES OF ROBUSTNESS METRICS

Having seen how the AND operator $\mathcal{A}$ forms the basis of robustness metrics, we now examine desirable properties of this operator from a learning-based perspective, along with some fundamental restrictions that their satisfaction imposes on its potential form. In order to be clearer with notation, we denote the operator for a conjunction of M terms by $\mathcal{A}_M$.

A. Desirable properties

1) First, consider the fundamental Boolean identities of idempotence and commutativity:

Property 2 (Idempotence, commutativity). The AND operator $\mathcal{A}_M$ is idempotent and commutative if (i) $\mathcal{A}_M(\rho,\dots,\rho)=\rho$, and (ii) for any permutation k_i of $i=1,\dots,M$, $\mathcal{A}_M(\rho_1,\dots,\rho_M)=\mathcal{A}_M(\rho_{k_1},\dots,\rho_{k_M})$ holds.

2) Second, from a general optimization perspective, it is desirable for the operator $\mathcal{A}$ to be smooth in order to aid gradient-based and acceleration methods.

Property 3 (Weak smoothness). $\mathcal{A}_M(\rho_1,\dots,\rho_M)$ is *weakly smooth* if (i) it is continuous everywhere, and (ii) if its gradient is continuous for all points for which there are no two indices $i\neq j$ satisfying $\rho_i=\rho_j=\min_{k\in\{1,\dots,M\}}\rho_k$.

Unlike simply requiring smoothness almost everywhere, this definition expresses the desire for smoothness at points where there is a unique minimal term. In particular, a special point of interest is when the robustness metric switches sign, indicating that the corresponding STL specification has become true or false. Neither the traditional nor the AG robustness metrics are smooth at such points.

3) Third, from a learning-based perspective of guiding towards task satisfaction, it is important to address the *shadowing problem* outlined in the introduction:

Property 4 (Shadow-lifting property). The operator $\mathcal{A}_M$ satisfies the shadow-lifting property if, for any $\rho\neq 0$, $\left.\frac{\partial\mathcal{A}_M(\rho_1,\dots,\rho_i,\dots,\rho_M)}{\partial\rho_i}\right|_{\rho_1,\dots,\rho_M=\rho}>0$ holds $\forall i=1,\dots,M$.

Considering a set of points $\rho_1,\dots,\rho_M=\rho$, the robustness metric of their conjunction using the traditional metric $\mathcal{A}_M(\rho_1,\dots,\rho_M)=\min(\rho_1,\dots,\rho_M)$ would only show an increase if all ρ_i terms increase. The shadowing problem thus makes it difficult for a learning algorithm to find improvements towards task satisfaction. On the other hand, the shadow-lifting property implies that $\mathcal{A}_M(\rho,\dots,\rho)$ also increases when making partial progress towards this goal and increasing only a set of the ρ_i terms. The AG robustness satisfies the shadow-lifting property; however, we will see that being too rewarding for positive changes of the terms in a conjunction has pitfalls related to local minima.

Remark 1. The shadow-lifting property prohibits the associative property of the AND operator to be satisfied. For example, if it were, the robustness of the conjunction of the two sets of robustness metrics $\{1,1,1+\epsilon\}$ and $\{1,1+\epsilon,1+\epsilon\}$ would be both equivalent to that of $\{1,1+\epsilon\}$, whereas the property requires the second one to be higher.

TABLE I: Summary of the discussed properties satisfied by the two robustness metrics introduced in the literature.

Robustness metric	Property number					
	1	2	3	4	5	6
traditional [10]	✓	✓	✗	✗	✓	✓
AG [15]	✓	✓	✗	✓	✓	✗

4) Finally, we consider two additional unclassified properties that impose natural restrictions on the AND operator.

Property 5 (min/max boundedness). The operator $\mathcal{A}_M$ is min/max bounded if it satisfies the inequality

$$\min(\rho_1,\dots,\rho_M)\leq\mathcal{A}_M(\rho_1,\dots,\rho_M)\leq\max(\rho_1,\dots,\rho_M).$$

Property 6 (Scale-invariance). The operator $\mathcal{A}_M$ is said to be scale-invariant if, for any $\alpha\geq 0$, it satisfies the identity

$$\mathcal{A}_M(\alpha\rho_1,\dots,\alpha\rho_M)=\alpha\mathcal{A}_M(\rho_1,\dots,\rho_M). \tag{5}$$

The former property is useful for placing fundamental restrictions on the values of $\mathcal{A}_M(\cdot)$. The latter is desirable for the AND function to behave similarly regardless of the order of magnitude of its robustness metric terms, e.g, in case we do not know their order of magnitude in advance.

Table I above summarizes the properties satisfied by the traditional and AG robustness metrics.

B. Imposed restrictions

Properties 1-6 impose fundamental restrictions on the form of the operator $\mathcal{A}_M$ used for constructing any sought-after robustness metric. We now study such restrictions in order to motivate the definition of a new metric in Section V.

Proposition 1. The operator $\mathcal{A}_M$ cannot be sound, idempotent, and smooth simultaneously.

Proof. Consider the behavior of $\mathcal{A}_2(\rho_1,\rho_2)$ for $\rho_1=0$ and $\rho_2>0$. By soundness, $\mathcal{A}_2(\rho_1,\rho_2)$ must switch sign as ρ_1 switches sign; continuity therefore implies $\mathcal{A}_2(0,\rho_2)=0$ for any $\rho_2>0$. In particular, $\mathcal{A}_2(0,\rho_2)$ remains 0 when $\rho_2\to 0+$, thus $\left.\frac{\partial\mathcal{A}_2(\rho_1,\rho_2)}{\partial\rho_2}\right|_{\rho_1,\rho_2=0}=0$. The same argument holds for the partial derivative with respect to ρ_1 at this point. For $\epsilon\to 0+$, in the first-order approximation we must thus have $\mathcal{A}_2(\epsilon,\epsilon)=\left(\frac{\partial\mathcal{A}_2(\rho_1,\rho_2)}{\partial\rho_1}+\frac{\partial\mathcal{A}_2(\rho_1,\rho_2)}{\partial\rho_2}\right)\epsilon=0$. This, however, contradicts $\mathcal{A}_2(\epsilon,\epsilon)=\epsilon$ implied by idempotence. □

The proposition implies that smoothness across the entire domain is a too strict requirement. As we will see, however, the operator $\mathcal{A}_M$ can be *weakly smooth*; the metric defined in the next section will moreover be smooth at points where all ρ_i are equal. The following proposition relates smoothness at such a key point of interest to the shadow-lifting property.

Proposition 2. Assume the AND operator $\mathcal{A}_M$ is defined such that it is smooth and satisfies Property 2. Then, for any $\rho\neq 0$, the operator satisfies

$$\left.\frac{\partial\mathcal{A}_M(\rho_1,\dots,\rho_M)}{\partial\rho_i}\right|_{\rho_1,\dots,\rho_M=\rho}=\frac{1}{M},\ \forall i=1,\dots,M \tag{6}$$

which is positive and thus implies that Property 4 also holds.

Proof. If the gradient is continuous, then for any $\rho \neq 0$ in the first-order approximation we must have:

$$\lim_{\epsilon\to 0} \mathcal{A}_M(\rho+\epsilon,\ldots,\rho+\epsilon) = \mathcal{A}_M(\rho,\ldots,\rho) + \sum\nolimits_{i=1}^{M} \epsilon \left.\frac{\partial \mathcal{A}_M(\rho_1,\ldots,\rho_M)}{\partial \rho_i}\right|_{\rho_1,\ldots,\rho_M=\rho}$$

The idempotence property then implies that the equality

$$\rho+\epsilon = \rho + \epsilon \sum\nolimits_{i=1}^{M} \left.\frac{\partial \mathcal{A}_M(\rho_1,\ldots,\rho_M)}{\partial \rho_i}\right|_{\rho_1,\ldots,\rho_M=\rho}$$

must hold. Furthermore, due to commutativity of the variables, the partial derivatives at $\rho_1 = \ldots = \rho_M$ must equal one another, which leads to the desired result (6). □

Next, we consider restrictions that weak smoothness imposes on the gradient at points where the minimal term in the conjunction is unique and equal to zero.

Proposition 3. Assume the AND operator $\mathcal{A}_M$ is defined such that Properties 1-3 and 5 hold. Then,

$$\left.\frac{\partial \mathcal{A}_M(\rho_1,\ldots,\rho_M)}{\partial \rho_i}\right|_{\rho_i=0,\ \rho_{j\neq i}>0} = 1,\ \forall i = 1,\ldots,M. \quad (7)$$

Proof. Consider the case $i = 1$ without loss of generality due to the commutative property of $\mathcal{A}_M$. As $\mathcal{A}_M$ is sound, it must switch from positive to negative as ρ_1 switches from positive to negative at points where $\rho_1 = 0, \rho_{j\neq 1} > 0$. Therefore, as $\mathcal{A}_M$ is continuous, we must have $\mathcal{A}_M(\rho_1,\ldots,\rho_M)$ = 0 at such points. The partial derivative $\left.\frac{\partial \mathcal{A}_M(\rho_1,\ldots,\rho_M)}{\partial \rho_i}\right|_{\rho_i=0,\ \rho_{j\neq i}>0}$ can thus be evaluated as:

$$\lim_{\rho_1\to 0} \frac{\mathcal{A}_M(\rho_1,\ldots,\rho_M)-0}{\rho_1-0} = \lim_{\rho_1\to 0} \frac{\mathcal{A}_M(\rho_1,\ldots,\rho_M)}{\rho_1} \quad (8)$$

For the limit from below, since ρ_1 is the minimal term, the min/max bound inequality $\rho_1 \leq \mathcal{A}_M(\rho_1,\ldots,\rho_M)(<0)$ implies $\lim_{\rho_1\to 0-} \frac{\mathcal{A}_M(\rho_1,\ldots,\rho_M)}{\rho_1} \leq 1$. On the other hand, taking the limit from above, the inequality $(0<)\rho_1 \leq \mathcal{A}_M(\rho_1,\ldots,\rho_M)$ implies $\lim_{\rho_1\to 0+} \frac{\mathcal{A}_M(\rho_1,\ldots,\rho_M)}{\rho_1} \geq 1$. For weak smoothness, the two limits must match, thus the partial derivative (7) must be equal to 1 at this point. □

Finally, the following proposition relates possible values $\mathcal{A}_M$ can take in case scale-invariance is also assumed.

Proposition 4. Assume the AND operator $\mathcal{A}_M$ is defined such that Properties 1-3, 5, and 6 hold, and consider without loss of generality $\rho_1 \neq 0$ fixed. Then, the condition (7) on the partial derivative is satisfied if and only if:

$$\lim_{\rho_2,\ldots,\rho_M\to\infty} \mathcal{A}_M(\rho_1,\ldots,\rho_M) = \rho_1. \quad (9)$$

Proof. The scale-invariance property allows us to relate the form of limits given by equations (8) and (9). Let ϵ have the same sign as $\rho_1 \neq 0$, and consider a set of values $\bar{\rho}_{i\neq 1} > 0$. Using (5) with $\alpha := \epsilon/\rho_1 > 0$, one obtains the relation:

$$\frac{\rho_1}{\epsilon}\mathcal{A}_M(\epsilon,\bar{\rho}_2,\ldots,\bar{\rho}_M) = \mathcal{A}_M(\rho_1,\bar{\rho}_2\frac{\rho_1}{\epsilon},\ldots,\bar{\rho_M}\frac{\rho_1}{\epsilon}). \quad (10)$$

For the 'if' direction, assuming (9) holds, then for $\epsilon \to 0$ the right hand side of this equation becomes ρ_1. Dividing both sides by ρ_1 then yields $\lim_{\epsilon\to 0} \frac{1}{\epsilon}\mathcal{A}_M(\epsilon,\bar{\rho}_2,\ldots,\bar{\rho}_M) = 1$, which by definition of the partial derivative implies (7). The 'only if' direction can be shown by reversing these steps. □

V. NEW ROBUSTNESS METRIC

We introduce a family of robustness metrics following the imposed restrictions uncovered in the previous section. More specifically, we construct an AND operator that satisfies all of Properties 1-6. The construction is such that a parameter $\nu > 0$ controls how closely the defined metric approaches the traditional $\mathcal{A}^{\text{trad}}(\rho_1,\ldots,\rho_M) = \min_i \rho_i := \rho_{\min}$ operator.

A. Behavior for $\rho_{\min} < 0$

If $\rho_{\min} < 0$, the conjunction is not satisfied, hence for soundness and continuity $\mathcal{A}_M(\rho_1,\ldots,\rho_M)$ must remain negative and approach 0 as $\rho_{\min} \to 0-$. Aiming at a scale-invariant behavior, let us define:

$$\tilde{\rho}_i := \frac{\rho_i - \rho_{\min}}{\rho_{\min}}, \quad (11)$$

which is non-positive (as $\rho_{\min} < 0$) and becomes 0 at $\rho_i = \rho_{\min}$. Using this normalized measure, we further define the effective robustness measures:

$$\rho_i^{\text{eff}} := \rho_{\min} e^{\tilde{\rho}_i}, \quad (12)$$

whose purpose is to transform each ρ_i such that ρ_i^{eff} is negative and remains between $\rho_{\min}$ and ρ_i. The AND operator for $\rho_{\min} < 0$ with parameter $\nu > 0$ is then defined by taking the weighted average of these effective measures:

$$\mathcal{A}_M^{\text{new}}(\rho_1,\ldots,\rho_M) := \frac{\sum_{i=1}^{M} \rho_i^{\text{eff}} e^{\nu\tilde{\rho}_i}}{\sum_{i=1}^{M} e^{\nu\tilde{\rho}_i}},\ \text{if } \rho_{\min} < 0. \quad (13)$$

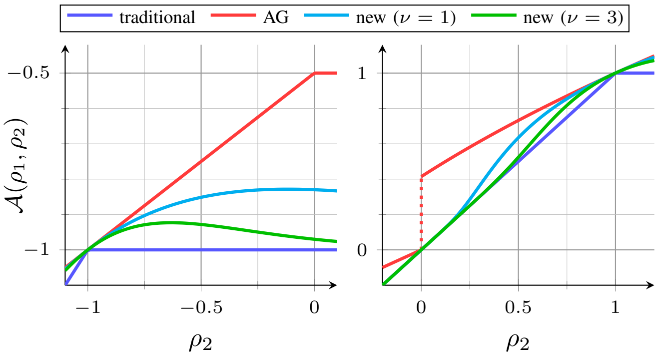
The weighting function $e^{\nu\tilde{\rho}_i}$ is such that the weight is 1 for $\rho_i = \rho_{\min}$ and becomes 0 as $\rho_i \to \infty$; this property is motivated by the desire to satisfy the condition $\lim_{\rho_i\neq\rho_{\min}\to\infty} \mathcal{A}_M(\rho_1,\ldots,\rho_M) = \rho_{\min}$ imposed by Proposition 4. Further note that as $\nu \to \infty$, the exponential weights $e^{\nu\tilde{\rho}_i}$ tend to 0 for $\rho_i \neq \rho_{\min}$ because then $\tilde{\rho}_i < 0$, and the defined operator becomes the traditional $\mathcal{A}_M^{\text{trad}} = \min_i \rho_i$ metric.

B. Behavior for $\rho_{\min} > 0$

If $\rho_{\min} > 0$, the conjunction is satisfied, hence for soundness and continuity $\mathcal{A}_M(\rho_1,\ldots,\rho_M)$ must remain positive and approach 0 as $\rho_{\min} \to 0+$. For a structurally scale-invariant behavior, we again use the normalized measure defined by (11). As opposed to the previous case, we do not require an effective measure in order to impose negative values on the metric, and can readily take the weighted average of the robustnesses forming the conjunction as:

$$\mathcal{A}_M^{\text{new}}(\rho_1,\ldots,\rho_M) := \frac{\sum_i \rho_i e^{-\nu\tilde{\rho}_i}}{\sum_i e^{-\nu\tilde{\rho}_i}},\ \text{if } \rho_{\min} > 0. \quad (14)$$

Note that if $\rho_{\min} > 0$, the normalized measures are all positive, and the exponential weights become 1 for $\rho_i = \rho_{\min}$ and 0 as $\rho_i \to \infty$, as in the previous case. As before, setting $\nu \to \infty$ also reduces (14) to the traditional min operator.



(a) $\rho_1 = -1$ kept fixed (b) $\rho_1 = 1$ kept fixed

Fig. 1: Robustness metric of the conjunction $\mathcal{A}_2(\rho_1, \rho_2)$ for the traditional, AG, and newly proposed metrics as a function of ρ_2 for fixed values of ρ_1. On the left, the shadow-lifting property is illustrated as the robustness increases even though the minimum of the two terms is constant. Note that the curves corresponding to the new metric are smooth.

C. Definition

The new robustness measure is defined from (13) and (14) as:

$$\mathcal{A}_M^{\text{new}}(\rho_1, \dots, \rho_M) = \begin{cases} \dfrac{\sum_i \rho_{\min} e^{\tilde{\rho}_i} e^{\nu \tilde{\rho}_i}}{\sum_i e^{\nu \tilde{\rho}_i}} & \text{if } \rho_{\min} < 0, \\ \dfrac{\sum_i \rho_i e^{-\nu \tilde{\rho}_i}}{\sum_i e^{-\nu \tilde{\rho}_i}} & \text{if } \rho_{\min} > 0, \\ 0 & \text{if } \rho_{\min} = 0. \end{cases} \tag{15}$$

Sample behaviors of the operator $\mathcal{A}_2^{\text{new}}(\rho_1, \rho_2)$ illustrating the shadow-lifting and weak-smoothness properties are depicted in Figure 1 above.

Theorem 1. The AND operator defined by (15) satisfies all of Properties 1-6.

Proof. The proof is elementary but technical, and due to space constraints is given separately in [16]. ☐

VI. RESULTS

In this section, we compare the performance of various robustness measures in the context of learning a simple but instructive task. The task ϕ is for a single integrator robot $\dot{\boldsymbol{x}} = \boldsymbol{u}$, with $\|\boldsymbol{u}\|_2 \leq 1$, to *always eventually* visit two nearby regions every 4s until the time horizon $T = 10$s. The two goal regions are circular with radius $r_g = 0.2$ and are centered at $\boldsymbol{x}_{g1} = [1.5\ 2.5]^{\mathrm{T}}$ and $\boldsymbol{x}_{g2} = [2.5\ 1.5]^{\mathrm{T}}$. The robot itself starts from $\boldsymbol{x}_0 = [2.0\ 2.0]^{\mathrm{T}}$. We aim for a robustness of at least $\rho^\phi \geq 0.05$, as well as to minimize $C(\tau) = \int_0^T \|\boldsymbol{u}(t)\|_2^2 \,\mathrm{d}t$. The scenario is simulated for $T = 10$s with a time step $\Delta t = 0.02$s.

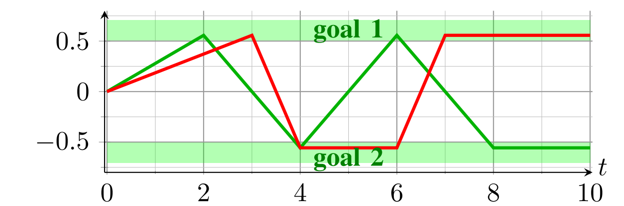
The task is formulated as $\phi = G_{[0,6]}\left(F_{[0,4]}\mu_1 \wedge F_{[0,4]}\mu_2\right)$, where $\mu_1 = (r_g - \|\boldsymbol{x} - \boldsymbol{x}_{g1}\| > 0)$ and $\mu_2 = (r_g - \|\boldsymbol{x} - \boldsymbol{x}_{g2}\| > 0)$. The minimal robustness requirement implies that the distance to travel from one region to the other is 1.11, slightly more than what is physically possible for the robot in one second. For this reason, out of two potential solutions illustrated in Figure 2a, only the green one is feasible, and its cost can be calculated to be $C_{\text{opt}} = 2.02$.

TABLE II: Percentage of feasible solutions found for the case scenario from 25 sample G-PI^2 runs under various robustness metric and guidance funnel configurations.

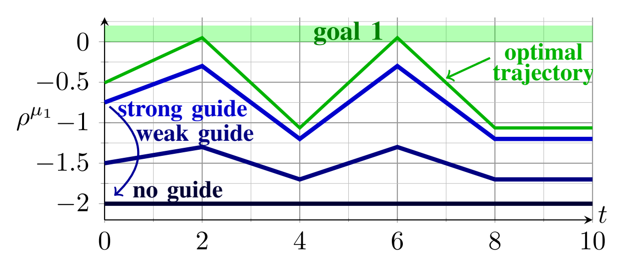
Robustness metric	Feedback guidance		
	none	weak	strong
traditional [10]	84%	100%	100%
AG [15]	48%	28%	92%
new	60%	100%	100%

The G-PI^2 method is employed to solve the described scenario. The algorithm allows the definition of initial $\gamma_i(t)$ funnels to guide the exploration towards the optimal solution. We examine *strong*, *weak*, and *no guidance* imposed for reaching the target regions, as depicted in Figure 2b. For details on how guidance is achieved, we refer to [18].

The traditional, AG, and newly proposed robustness metrics are compared in terms of their performance in solving the outlined scenario. Table II summarizes the percentage of finding feasible solutions as opposed to infeasible ones in each case. We conjecture that the AG metric often snaps into the infeasible local minima because it is too rewarding for increases in robustness metrics, as seen through Figure 1. Figures 3-4 show the convergence of the task robustness measure ρ^ϕ and the achieved cost $C(\tau)$ as a function of the G-PI^2 iteration number for the case of the successful runs. The figures show that the newly defined robustness measure surpasses both the traditional and AG metric in terms of accelerating the learning process. Surprisingly, the latter behaves quite poorly, most likely due to the learning method not being suited for handling discontinuous costs well. The new metric consistently achieves task satisfaction 20-25% faster than the traditional metric without sacrificing much in terms of successfully finding the feasible solutions.



(a) Out of two potential trajectories for satisfying the task, the green is feasible and optimal with respect to minimizing the input energy. The red is not feasible due to actuator constraints.



(b) Evolution of the robustness measure $\rho^{\mu_1}(\boldsymbol{x}(t))$ for the optimal trajectory (green), and three $\gamma_1(t)$ guide funnels aiming to enforce it in a progressively more aggressive manner (blue).

Fig. 2: Potential solutions and guides to satisfying the case study task. The robot needs to visit both goal regions once every 4 seconds throughout the first 6 seconds of its motion.

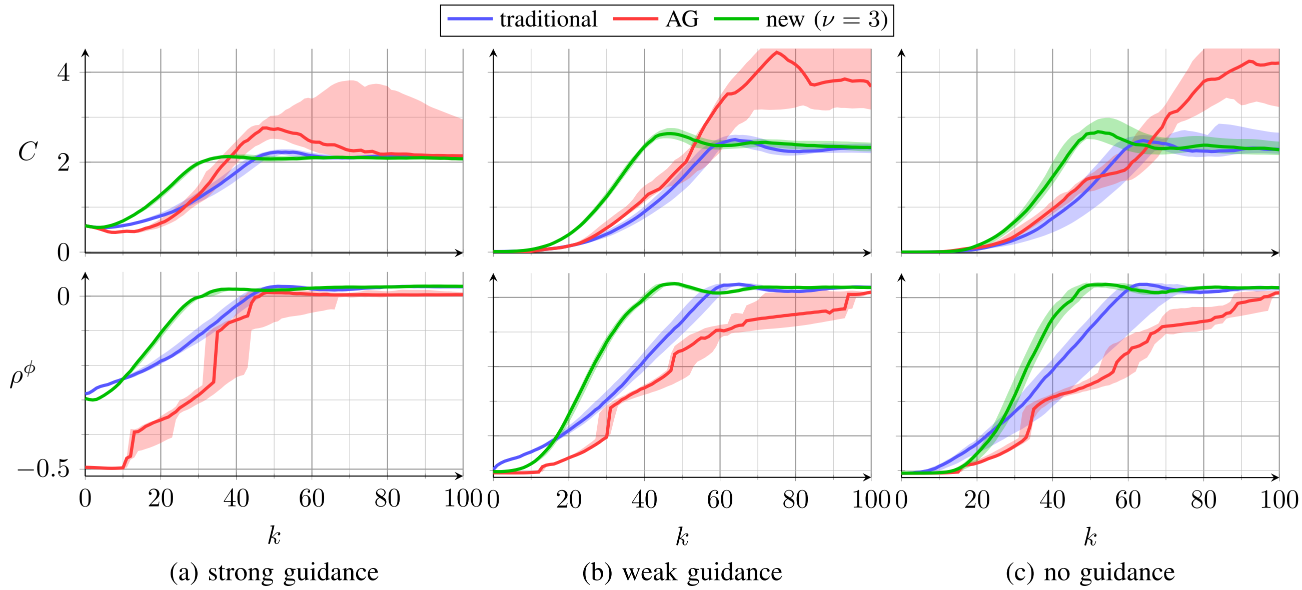


Fig. 3: Convergence of the cost and robustness metrics throughout the PI^2 algorithm running under strong, weak, or no guidance. The results show the distribution of successful runs obtained from 25 sample runs of the algorithm, for the discussed traditional, AG, and newly proposed robustness metrics. The median task robustness measure ρ^ϕ and achieved cost C are plotted as a function of the PI^2 iteration number k. All results, excluding the top and bottom 10%, lie in the shaded regions. The newly defined metric always converges to near the optimal solution and achieves task satisfaction most quickly and with least variance. The optimal cost can be calculated as $C_{\text{opt}} = 2.02$.

VII. Conclusions

In this paper, we presented theoretical results regarding the form of possible robustness metrics for quantifying the satisfaction of STL tasks. The findings motivated the definition of a sample new robustness metric, whose improved performance for accelerating learning was demonstrated through a simulation case study. Our preliminary results are promising and motivate further research into the potential of defining robustness metrics for their role in general-purpose reward shaping. Further work on the topic includes a more rigorous exploration of the properties of potential robustness metrics and their imposed restrictions on the metric itself, as well as conducting a thorough simulation study to verify their superior performance across a wider spectrum of scenarios.

References

[1] A. Pnueli, "The temporal logic of programs," in *IEEE Symposium on Foundations of Computer Science*, 1977, pp. 46–57.
[2] R. Koymans, "Specifying real-time properties with metric temporal logic," *Real-time systems*, vol. 2, no. 4, pp. 255–299, 1990.
[3] O. Maler and D. Nickovic, "Monitoring temporal properties of continuous signals," in *Formal Techniques, Modelling and Analysis of Timed and Fault-Tolerant Systems*. Springer, 2004, pp. 152–166.
[4] P. Tabuada and G. J. Pappas, "Linear time logic control of discrete-time linear systems," *IEEE Transactions on Automatic Control*, vol. 51, no. 12, pp. 1862–1877, 2006.
[5] S. Saha and A. A. Julius, "An MILP approach for real-time optimal controller synthesis with metric temporal logic specifications," in *American Control Conference (ACC)*. IEEE, 2016, pp. 1105–1110.
[6] L. Lindemann, C. K. Verginis, and D. V. Dimarogonas, "Prescribed performance control for signal temporal logic specifications," in *IEEE Conference on Decision and Control (CDC)*, 2017, pp. 2997–3002.
[7] D. Sadigh, E. S. Kim, S. Coogan, S. S. Sastry, and S. A. Seshia, "A learning based approach to control synthesis of MDPs for linear temporal logic specifications," in *IEEE Conference on Decision and Control (CDC)*, 2014, pp. 1091–1096.
[8] D. Aksaray, A. Jones, Z. Kong, M. Schwager, and C. Belta, "Q-learning for robust satisfaction of signal temporal logic specifications," in *IEEE Conference on Decision and Control*, 2016, pp. 6565–6570.
[9] D. Muniraj, K. G. Vamvoudakis, and M. Farhood, "Enforcing signal temporal logic specifications in multi-agent adversarial environments: A deep Q-learning approach," in *IEEE Conference on Decision and Control (CDC)*, 2018, pp. 4141–4146.
[10] A. Donzé and O. Maler, "Robust satisfaction of temporal logic over real-valued signals," in *International Conference on Formal Modeling and Analysis of Timed Systems*. Springer, 2010, pp. 92–106.
[11] A. Y. Ng, D. Harada, and S. Russell, "Policy invariance under reward transformations: Theory and application to reward shaping," in *International Conference on Machine Learning (ICML)*, vol. 99, 1999, pp. 278–287.
[12] L. Lindemann and D. V. Dimarogonas, "Robust control for signal temporal logic specifications using discrete average space robustness," *Automatica*, vol. 101, pp. 377–387, 2019.
[13] Y. V. Pant, H. Abbas, and R. Mangharam, "Smooth operator: Control using the smooth robustness of temporal logic," in *IEEE Conference on Control Technology and Applications (CCTA)*, 2017, pp. 1235–1240.
[14] I. Haghighi, N. Mehdipour, E. Bartocci, and C. Belta, "Control from signal temporal logic specifications with smooth cumulative quantitative semantics," in *IEEE Conference on Decision and Control (CDC)*, 2019, pp. 4362–4366.
[15] N. Mehdipour, C.-I. Vasile, and C. Belta, "Arithmetic-geometric mean robustness for control from signal temporal logic specifications," in *American Control Conference (ACC)*, Philadelphia, PA, July 2019.
[16] P. Varnai and D. V. Dimarogonas, "On robustness metrics for learning STL tasks," *arXiv preprint arXiv:2003.06041*, 2020.
[17] T. Akazaki and I. Hasuo, "Time robustness in MTL and expressivity in hybrid system falsification," in *International Conference on Computer Aided Verification*. Springer, 2015, pp. 356–374.
[18] P. Varnai and D. V. Dimarogonas, "Prescribed performance control guided policy improvement for satisfying signal temporal logic tasks," in *American Control Conference (ACC)*, 2019, pp. 286–291.