Risk-aware stochastic control of a sailboat

MingYi Wang[1], Natasha Patnaik[2], Anne Somalwar[3], Jingyi Wu[4], and Alexander Vladimirsky[5]

***Abstract*— Sailboat path-planning is a natural hybrid control problem (due to continuous steering and occasional "tack-switching" maneuvers), with the actual path-to-target greatly affected by stochastically evolving wind conditions. Previous studies have focused on finding risk-neutral policies that minimize the expected time of arrival. In contrast, we present a robust control approach, which maximizes the probability of arriving before a specified deadline/threshold. Our numerical method recovers the optimal risk-aware (and threshold-specific) policies for all initial sailboat positions and a broad range of thresholds simultaneously. This is accomplished by solving two quasi-variational inequalities based on second-order Hamilton-Jacobi-Bellman (HJB) PDEs with degenerate parabolicity. Monte-Carlo simulations show that risk-awareness in sailing is particularly useful when a carefully calculated bet on the evolving wind direction might yield a reduction in the number of tack-switches.**

I. Introduction

Sail-boat racing is one of the many areas where game-theoretic and control-theoretic tools are valuable in improving the competitive performance. The uncertainty in weather patterns gives rise to hybrid stochastic control models with many reasonable choices of performance measures to optimize. The previously developed methods have focused on risk-neutral optimization (e.g., minimizing the expected time to destination) [1], [2], [3]. In contrast, here we focus on maximizing the probability of desirable outcomes (e.g., arriving prior to a specified deadline). Our *risk-aware* approach addresses a notion of robustness very different from the traditional H^∞ control [4] and has important advantages for many applications. Indeed, it has been already successfully used in piecewise-deterministic Markov processes [5] and in bang-bang stochastic control models of adaptive drug therapies [6]. Unlike the typical *risk-averse* approaches [7], the method that we develop here for the hybrid control setting allows finding the optimal control policies for a large set of starting sailboat positions and *a range of deadlines* simultaneously. This is accomplished in the general framework of dynamic programming and requires solving a pair of quasi-variational inequalities on a 3D computational domain.

The hybrid nature of this problem is due to "tacking": to travel upwind, sailors must use a zigzag pattern, periodically swinging the bow to the other side of the wind. Each such "tack-switch" incurs a significant slow down while the wind pushes against the boat. We adopt a commonly used simplified model which assumes that the boat's velocity vector can be changed instantaneously (choosing among all directions available in its current tack) but a switch to the opposite tack incurs a fixed time-penalty.

Optimization of sailboat routing is a topic of increasing mathematical interest. In [8] the task of minimizing the expected time to target was considered in a discrete setting, with a discrete-time Markov chain modeling occasional changes in weather conditions. The idea was extended to continuous-time Markov chains in [9], with a tack-switching curve defined in the state space to encode the optimal policy. In [1], this switching curve was found using dynamic programming for indefinite-horizon hybrid control problems, but under the assumption that the wind direction stays constant for the duration of each tack switch. In [3], it was shown how this assumption can be avoided, yielding an improvement in control policies.

We start by introducing the hybrid dynamics in Section II, and describe both the risk-neutral and risk-aware optimal control problems in Section III. Our numerical approach to the latter is presented in Section IV, followed by the summary of computational experiments in Section V. We conclude by listing directions for future work in Section VI.

II. System Dynamics

Following [1], [3], we assume the strength of the wind is fixed but its direction (measured counterclockwise from the y-axis) undergoes a Brownian drift/diffusion process:

$$\mathrm{d}\phi = a\mathrm{d}t + \sigma\mathrm{d}B, \tag{1}$$

where $\phi(t)$ denotes the current upwind direction, a is a constant drift, σ is the diffusion coefficient, and B is a standard Brownian motion. The state of the system can be represented as (x, y, q, ϕ), where x and y encode the boat's current position, while $q \in \{1, 2\}$ is its current "tack", which determines the range of available steering directions. Our continuous control is the steering angle, $u \in [0, \pi]$, measured relative to the wind. In the starboard tack ($q = 1$), u is measured counterclockwise from the upwind direction, while in the port tack ($q = 2$) it is measured clockwise; so, the boat's direction of motion relative to upwind is $(-1)^q u$. The boat's angle-dependent speed $f(u)$ is encoded in the speed

*This work was supported in part by NSF Division of Mathematical Sciences (DMS) under Awards 1645643 and 2111522 as well as by the Air Force Office of Scientific Research under Award FA9550-22-1-0528.

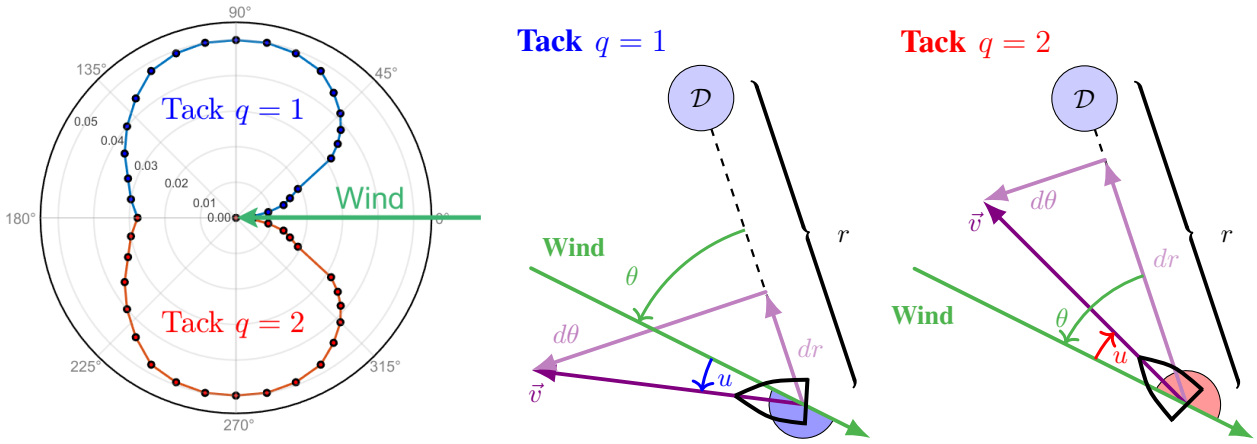
[1]MingYi Wang is with the Center for Applied Mathematics, Cornell University, Ithaca, NY 14853, USA mw929@cornell.edu

[2]Natasha Patnaik is with the Department of Computational Applied Mathematics and Operations Research, Rice University, Houston, TX 77005, USA np37@rice.edu

[3]Anne Somalwar is with the graduate group in Applied Mathematics and Computational Science, University of Pennsylvania, Philadelphia, PA 19104, USA somalwar@sas.upenn.edu

[4]Jingyi Wu is with the Center for Data Science, New York University, New York City, NY 10012, USA jw8535@nyu.edu

[5]Alexander Vladimirsky is with the Department of Mathematics, Cornell University, Ithaca, NY 14853, USA vladimirsky@cornell.edu



(a) The polar speed plot $f(u)$ used in this work (the same as Fig. 1(a) in [3]). Note that $f = 0$ at $u = 0°$.

(b) System setups in the polar coordinate centered at $\mathcal{D}$ for different tacks.

Fig. 1. System dynamics relative to the wind and relative to the target $\mathcal{D}$.

profile (often called the "polar"), which is determined by the geometry of each specific boat. Fig. 1(a) shows a typical polar used in all numerical tests here and in [3]. With these assumptions, the boat's dynamics is described by

$$dx = -f(u)\sin(\phi - (-1)^q u)dt, \tag{2}$$
$$dy = f(u)\cos(\phi - (-1)^q u)dt. \tag{3}$$

For the sake of computational efficiency, we adopt a dimensional reduction of (1)-(3) suitable when aiming for a circular target $\mathcal{D}$ in a domain with no obstacles [3]. Assuming a target at the origin, $r = \sqrt{x^2 + y^2}$ represents the boat's distance to the center of $\mathcal{D}$, while $\theta = \phi + \operatorname{atan2}(-x, -y)$ encodes the upwind direction measured counterclockwise from the line connecting the boat to the center of $\mathcal{D}$; see Fig. 1(b). This results in the system dynamics:

$$\begin{aligned} \mathrm{d}r &= -f(u)\cos(\theta - (-1)^q u)\,\mathrm{d}t, \\ \mathrm{d}\theta &= \left[\frac{f(u)}{r}\sin(\theta - (-1)^q u) + a\right]\mathrm{d}t + \sigma \mathrm{d}B. \end{aligned} \tag{4}$$

Whenever we focus on the *deterministic portion* of dynamics, we will use $r_{\mathrm{d}}(\theta, q, u)$ and $\theta_{\mathrm{d}}(r, \theta, q, u; a)$ to denote the respective factors multiplying $\mathrm{d}t$ in the above equations.

III. Stochastic Optimal Control

Let $\boldsymbol{\xi}(t) = (r(t), \theta(t))$ denote the continuous component of system state at the time t. We define $\Omega := [0, R_{\max}] \times [0, 2\pi) \times \{1, 2\}$ as the full state space (with the last component encoding the tack q). We will further use $\tilde{q} = 3 - q$ to refer to "the opposite tack" and $\Upsilon := [0, \pi] \cup \{\blacktriangle\}$ to define the policy domain (the prescribed steering angle or the special tack-switch action $\blacktriangle$).

Given a starting configuration of the boat ($\boldsymbol{\xi}(0) = \hat{\boldsymbol{\xi}}, q(0) = \hat{q}$) and a feedback control policy $\mu : \Omega \to \Upsilon$, the main quantity of interest is the random time to target $T_\mu(\hat{\boldsymbol{\xi}}, \hat{q}) = \inf\{t > 0 \mid \boldsymbol{\xi}(t) \in \mathcal{D},\ \boldsymbol{\xi}(0) = \hat{\boldsymbol{\xi}},\ q(0) = \hat{q}\}$ based on that μ. Due to the hybrid nature mentioned in section II, $T_\mu(\hat{\boldsymbol{\xi}}, \hat{q})$ is the sum of the time spent in both steering and tack-switching. A typical *risk-neutral* approach is to define the value function as $v(\hat{\boldsymbol{\xi}}, \hat{q}) = \inf_\mu \mathbb{E}[T_\mu(\hat{\boldsymbol{\xi}}, \hat{q})]$, which can be then recovered by solving a system of quasi-variational HJB-type inequalities. Assuming each tack-switch takes a fixed time C and the boat stays in place ($f(u) = 0$) while switching, as shown in [3] this leads to

$$\max\left\{H(r, \theta, q, \nabla v) - \frac{\sigma^2}{2}\frac{\partial^2 v}{\partial \theta^2},\ v - \mathcal{N}v - C\right\} = 0, \tag{5}$$

where $\nabla v = (\frac{\partial v}{\partial r}, \frac{\partial v}{\partial \theta})$ and the Hamiltonian is

$$H(r, \theta, q, \mathbf{p}) = \max_u \left(-r_{\mathrm{d}}(\theta, q, u)p_1 - \theta_{\mathrm{d}}(r, \theta, q, u; a)p_2\right) - 1. \tag{6}$$

In (5), we take the maximum over two alternative courses of action. The first clause corresponds to the system states, from which it is optimal to continue in the current tack and the maximization in (6) selects the optimal steering angle. On the other hand, in the second clause $\mathcal{N}v$ encodes the expected remaining time-to-target if we switch to the opposite tack; i.e., $\mathcal{N}v(\hat{r}, \hat{\theta}, q) = \mathbb{E}[v(\hat{r}, \theta(C), \tilde{q}) \mid \theta(0) = \hat{\theta}, f = 0]$. If we define $\psi_{r,q}(z) = v(r, z, \tilde{q})$ and

$$\mathcal{G}_\theta[\psi] = \frac{1}{\sigma\sqrt{2\pi C}}\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{(z-\theta-aC)^2}{2C\sigma^2}}\psi(z)\,\mathrm{d}z, \tag{7}$$

then this switching operator can be conveniently evaluated as $\mathcal{N}v(r, \theta, q) = \mathcal{G}_\theta[\psi_{r,q}]$. Since the part of Ω on which it is better to switch tacks is a priori unknown, this is a problem with a *free boundary*. The optimal feedback policy μ_* (found by solving (5) with the boundary condition $v = 0$ on $\mathcal{D} \times \{1, 2\}$) captures both the optimal switching states and the optimal steering angles [3].

Despite its frequent use, the risk-neutral planning has a significant drawback: it is indifferent to the level of variability in the distribution of times to target. The resulting μ_* might be impractical if the risk of significantly exceeding $\mathbb{E}[T_{\mu_*}]$ is high (e.g., in right-heavy-tailed distributions). To address this, we change the perspective and search for a policy α that maximizes the probability of reaching the target

before a specified deadline $\hat{s}$. We refer to such α as a *risk-aware* (or, more precisely, as an $\hat{s}$-*threshold-aware*) policy.

Letting $\Omega_{\bar{s}} = \Omega \times [0, \bar{s}]$, we define our new value function

$$w(\hat{\boldsymbol{\xi}}, \hat{q}, \hat{s}) = \sup_{\alpha} \mathbb{P}\left(T_\alpha(\hat{\boldsymbol{\xi}}, \hat{q}) \leq \hat{s}\right), \quad 0 \leq \hat{s} \leq \bar{s}. \tag{8}$$

The supremum is taken over all measurable *threshold-aware* feedback policies $\alpha : \Omega_{\bar{s}} \to \Upsilon,\ (r, \theta, q, s) \to \alpha(r, \theta, q, s)$. If we treat $s(0) = \hat{s}$ as an *initial budget*, then the remaining time budget $s(t)$ is strictly decreasing along the path-to-target as t increases ($\dot{s}(t) = -1$ while continuously steering and a negative jump of C units of time when switching tack).

Consequently, the Stochastic Dynamic Programming Principle [10] yields an s-dependent quasi-variational inequality

$$w(\hat{\boldsymbol{\xi}}, \hat{q}, \hat{s}) = \max\left\{\sup_{u} \mathcal{E}_{u,\tau} w(\hat{\boldsymbol{\xi}}, \hat{q}, \hat{s}), \mathcal{E}_{C} w(\hat{\boldsymbol{\xi}}, \hat{q}, \hat{s})\right\} + o(\tau), \tag{9}$$

where

$$\mathcal{E}_{u,\tau}\ w(\hat{\boldsymbol{\xi}}, \hat{q}, \hat{s}) = \mathbb{E}_{u,\tau}[w(\boldsymbol{\xi}(\tau), \hat{q}, \hat{s} - \tau) \mid \hat{\boldsymbol{\xi}}, \hat{s}], \tag{10}$$

$$\mathcal{E}_{C}\ w(\hat{\boldsymbol{\xi}}, \hat{q}, \hat{s}) = \mathbb{E}_{C}[w(\boldsymbol{\xi}(C), \tilde{q}, \hat{s} - C) \mid \hat{\boldsymbol{\xi}}, \hat{s}, f = 0]. \tag{11}$$

Similarly to the structure of (5), $\mathcal{E}_{u,\tau} w$ refers to the best probability of reaching the target before the deadline if we stay on the current tack for a small time τ while $\mathcal{E}_{C} w$ is the best probability if we immediately switch tacks.

From the stochastic Taylor expansion of Eqn. (9), one can show that, if $w(r, \theta, q, s)$ is sufficiently smooth, it must satisfy

$$\max\left\{\max_{u \in [0,\pi]} \left(\nabla w^\top \boldsymbol{\xi}_{\mathrm{d}} + \frac{\sigma^2}{2} \frac{\partial^2 w}{\partial \theta^2} - \frac{\partial w}{\partial s}\right), \mathcal{E}_{C} w - w\right\} = 0, \tag{12}$$

where $\nabla w = (\partial w / \partial r,\ \partial w / \partial \theta)$ and $\boldsymbol{\xi}_{\mathrm{d}}(r, \theta, q, u) = \Big(r_{\mathrm{d}}(\theta, q, u),\ \theta_{\mathrm{d}}(r, \theta, q, u; a)\Big)$.

As in the risk-neutral case, if we define $\psi_{r,q,s}(z) = w(r, z, \tilde{q}, s - C)$, then $\mathcal{E}_{C} w(r, \theta, q, s) = \mathcal{G}_\theta[\psi_{r,q,s}]$ following the definition of operator (7). Note that, in general, the value function does not have to be be smooth or even continuous. Nonetheless, even a discontinuous value function can be often interpreted as a unique *discontinuous viscosity solution* of a HJB PDE [11, Chapter 5], which allows recovering the optimal feedback policy $\alpha_*(r, \theta, q, s)$ as an argmax of (12).

IV. Numerical Implementation

A. Semi-Lagrangian Discretization

We approximate the solution to (12) for both tacks simultaneously using a semi-Lagrangian discretization [12] on a uniform rectangular grid over the (r, θ, s) space. I.e., $(r_i, \theta_j, s_k) = (i\Delta r, j\Delta\theta, k\Delta s)$, where $\Delta r = R_{max}/N_r$, $\Delta r = 2\pi/N_\theta$, $\Delta s = \bar{s}/N_s$, while $i = 0, ..., N_r$, $j = 1, ..., N_\theta$ (due to periodic boundary conditions), and $k = 0, ..., N_s$. We will use $W^{k,q}_{i,j} \approx w(r_i, \theta_j, q, s_k)$ to denote the discretized approximate solution at (r_i, θ_j, q, s_k). Recall from section III that $s(t)$ is strictly decreasing along the path-to-target. We can thus *causally* march from smaller s to larger s and compute the value function from $s = 0$ to $s = \bar{s}$ in a single sweep. In particular, we choose $\tau = \Delta s$ when solving Eqn. (10) so that we can march from the s_{k-1}-slice to the s_k-slice. The expectations in both (10) and (11) can be approximated using Gauss-Hermite quadratures (GHQ), but the details are somewhat different due to the contrast in elapsed times.

To approximate $\mathcal{E}_{u,\tau}$, we use a first-order weak approximation [13], [14] of the distribution of the Brownian increment ΔB_τ for τ time units. Starting from any gridpoint (r_i, θ_j, s_k) and using any admissible steering angle u, we consider two possible locations of $\boldsymbol{\xi}(\tau, u)$ in the s_{k-1}-slice:

$$\boldsymbol{\xi}^{\pm}_{i,j,u} \ = \ (r_i + \tau r_{\mathrm{d}}(u),\ \theta_j + \tau\theta_{\mathrm{d}}(u) \pm \sigma\sqrt{\tau}).$$

Averaging the value function at these points is equivalent to a two-node GHQ approximation of (10); i.e.,

$$\mathrm{M}_{u,\tau} W^{k,q}_{i,j} \ = \ \frac{1}{2}\Big(W^{k-1,q}\left(\boldsymbol{\xi}^{+}_{i,j,u}\right) + W^{k-1,q}\left(\boldsymbol{\xi}^{-}_{i,j,u}\right)\Big). \tag{13}$$

Since these $\boldsymbol{\xi}^{\pm}_{i,j,u}$ are usually not gridpoints, we implement (13) by using a *bi-cubic ENO* interpolation [15] with a 2π-periodicity in θ. We adopt a two stage process for finding the optimal u_* that maximizes $\mathrm{M}_{u,\tau}$: first, we perform a direct comparison over a grid of angle values $\mathcal{U}$ and identify an interval containing the best $u_\bigstar \in \mathcal{U}$; we then perform a Golden Section Search (GSS) over that interval to obtain a more accurate approximation of u_*.

The accuracy of (13) improves under grid refinement[1] since the diffusion time $\tau = \Delta s \to 0$. Finding a good approximation for $\mathcal{E}_{C}$ is a bit harder since the diffusion time C is constant. To address this, one can use a higher order accurate GHQ; e.g., our implementation uses a version with 3 GH nodes

$$\eta_{j,m} \ = \ \theta_j + aC + \sigma\sqrt{2C} x_m, \quad m \in \{1, 2, 3\}, \tag{14}$$

where x_m are the roots to the 3rd Hermite polynomial. Assuming that $s_k \geq C$, we use

$$\mathrm{M}_{C} W^{k,q}_{i,j} \ = \ \frac{1}{\sqrt{\pi}} \sum_{m=1}^{3} \gamma_m\, W\left(r_i, \eta_{j,m}, \tilde{q}, s_k - C\right), \tag{15}$$

where γ_m's are the weights of the third GHQ. We choose Δs to be a fraction of C, ensuring that $s_k - C = s_l$ for some $l < k$. But $\eta_{j,m}$ are usually not multiples of $\Delta\theta$ and we use a 1D periodic cubic ENO interpolation to evaluate (15).

The grid value is then computed as

$$W^{k,q}_{i,j} = \max\left(M_{u_*,\tau} W^{k,q}_{i,j},\ \mathrm{M}_{C} W^{k,q}_{i,j}\right),$$

and we recover the optimal steering/switching policy $\alpha_*(r_i, \theta_j, q, s_k)$ as a by-product. Our full method is summarized in Algorithm 1 using the target radius $R_{\mathcal{D}}$, the

[1]Under mild technical conditions, semi-Lagrangian schemes have been proven to converge to the discontinuous viscosity solution of first-order HJB PDEs on every compact set away from discontinuity [16], [17]. For the second-order HJBs, the method closest to ours has been studied (with rigorous error estimates) in [18] but without hybrid dynamics or degenerate parabolicity. While our setting is more general, the numerical results in Section V and online repository provide strong evidence of convergence. A rigorous proof of numerical convergence to the discontinuous viscosity solution of PDE (12) remains an open problem to be addressed in the future.

maximum sailboat speed $f_{\max}$, and
$\Xi = \{(i\Delta r, j\Delta\theta) \mid i = 0, \dots, N_r,\ j = 0, \dots, N_\theta\}$,

Algorithm 1: Risk-aware value function computation

for $s_k = k\Delta s,\ k = 0, 1, \dots N_s$ **do**
  for *every* $\boldsymbol{\xi}_{i,j} \in \Xi$ *and* $q \in \{1, 2\}$ **do**
    if $(r_i - R_{\mathcal{D}})/f_{\max} > s_k$ **then**
      $W_{i,j}^{k,q} \leftarrow 0$;
    else
      $W_{i,j}^{k,q} \leftarrow \max_u M_{u,\tau}$;
      if $s \geq C$ **then**
        $W_{i,j}^{k,q} \leftarrow \max\left(W_{i,j}^{k,q},\, \mathrm{M}_C W_{i,j}^{k,q}\right)$;

B. Trajectory synthesis and ECDF generation

The above PDE solution process yields the optimal threshold-aware policy in feedback form, with the optimal action $\alpha_*(r_i, \theta_j, q, s_k)$ stored at each gridpoint in Ξ and for a range of deadlines ($k = 0, ..., N_s$). To recover a sample path-to-target from any specific initial configuration $(\hat{r}, \hat{\theta}, \hat{q})$ and the intended deadline $\hat{s}$, we use Euler-Maruyama scheme [14] with a fixed time step Δt on Eqns. (4) . At each time step, we normally use the optimal steering/switching action from the policy recorded for the nearest gridpoint. But the threshold-aware control formulation leaves two ambiguities that have to be resolved in the implementation. First, if $w(\hat{r}, \hat{\theta}, \hat{q}, \hat{s}) = 1$, the current $\hat{s}$ may be more than sufficient to reach the target with probability one and the actions taken until the remaining time budget s becomes "barely sufficient" are not important. To address this, we use a "*Deadline-Upgrade*" approach, decreasing the initial time-budget to $\hat{s} = \min\left\{s_k \mid w(\hat{r}, \hat{\theta}, \hat{q}, s_k) = 1\right\}$. Second, if during a simulation the sailor is unlucky and later finds herself with $w(\hat{r}, \hat{\theta}, \hat{q}, \hat{s}) = 0$, the PDE provides no guidance on what to do after that (since she will now definitely miss the original deadline). Rather than dismiss such simulations as complete failure, from there on we simply apply the risk-neutral policy μ_* recovered from Eqn. (5).

Empirical cumulative distribution function (ECDF) for both α_* and μ_* are obtained through Monte Carlo simulations, with sample paths generated starting from a specific $(\hat{r}, \hat{\theta}, \hat{q}, \hat{s})$ under different realizations of wind evolution. The ECDFs for the total time-to-target are obtained through the the Kaplan-Meier estimate [19] using the `MATLAB`'s built-in function `ecdf()`.

V. Numerical Experiments

We use several examples to compare the performance of risk-aware and risk-neutral policies. In all cases, the value functions are computed on a $1601 \times 1601 \times 2$ grid for $(r, \theta, q) \in [0, 2] \times [0, 2\pi] \times \{1, 2\}$. When solving (12), we use $\Delta s = 0.025$ for any preset maximum deadline $\bar{s}$. The other parameter values are $R_{\mathcal{D}} = 0.1$, $C = 2$, and $f_{\max} = 0.05$. All ECDFs are built through Monte Carlo simulations (see section IV-B) with 10^5 samples and $\Delta t = 0.005$.

We illustrate our s-dependent risk-aware policies for a wind with zero drift ($a = 0$) and small diffusivity ($\sigma = 0.05$). In Fig. 2(a,c), we show some representative s-slices of the risk-aware optimal policies α_* and the corresponding value function w for the starboard tack[2] (i.e., the optimal probability of reaching $\mathcal{D}$ in less than s units of time if we start with $q = 1$ and use α_*). In Fig. 2(a), colors indicate the optimal steering angle u_* in the current tack, while the complement (left blank) shows all the (r, θ) configurations at which the immediate tack-switch ▲ is optimal. We observe that α_* is strongly s-dependent and significantly differs from the risk-neutral optimal policy μ_* shown in Fig. 2(b). The arrows in Fig. 2(a,c) indicate the natural progression when threshold-aware policies are used in practice: once we start with a particular deadline $\hat{s}$, our initial time-budget[3] $s(0) = \hat{s}$ is progressively decreasing, making it necessary to use α_* from the lower s-slices. This decrease is gradual ($\dot{s}(t) = -1$) while we are steering, but becomes abrupt whenever we decide to tack-switch at some time $t^{\blacktriangle}$; i.e., $s(t) = s(t^{\blacktriangle}) - C, \quad \forall t \in (t^{\blacktriangle}, t^{\blacktriangle} + C]$.

Since α_* is somewhat more complicated to implement in practice, it is reasonable to ask whether it is significantly better (in meeting the desired deadlines) than the risk-neutral μ_*. For any specific starting configuration, the answer can be found by comparing the ECDF of μ_* with the risk-aware value function w plotted across the range of s values. The graph of w will be always above, though often this difference is minimal, making the use of μ_* a preferred option. However, in many cases the gap between the two graphs will be more significant for a specific range of s values. This is illustrated in Fig. 3(a) for $(\hat{r}, \hat{\theta}, \hat{q}) = (1.93, 0.56, 1)$. If we are interested in some deadline between $\hat{s} = 52$ and $\hat{s} = 57$, the threshold-aware policies provide a noticeable advantage. For example, our α_*^{56} (red in Fig. 3(b)) increases $\mathbb{P}(T \leq 56)$ from 52.5% to 58.9%, while our α_*^{53} (green in Fig. 3(b)) yields a 10.6% improvement in $\mathbb{P}(T \leq 53)$ while increasing $\mathbb{E}[T]$ by less than 2.8%.

It is also revealing to examine sample trajectories resulting from each of these policies (shown in Fig. 3(c) for a particular random realization of wind evolution). According to μ_*, the boat starts in the "wrong" tack, and thus needs to switch immediately, with another tack-switch (back to $q = 1$) almost always needed later to reach $\mathcal{D}$. This strategy produces the best $\mathbb{E}[T]$, but makes it hard to reach the target much earlier and does not hedge against the bad outcomes (e.g., $T_{\mu_*} > 58$ in more that 33% of simulations). In contrast, the threshold-aware policies make a calculated bet (that the wind direction will soon change to help us), stay with the original $q = 1$ at first, and reach $\mathcal{D}$ with only one tack-switch.

Larger improvements can be similarly realized with a non-zero wind-drift, particularly when the chosen deadlines are fairly aggressive (in the left tail of T_{μ_*} PDF). In Fig. 4(a) we

[2]In the interest of reproducibility, our full source code, additional examples, and movies (for both tacks) will be available from https://eikonal-equation.github.io/Threshold-Aware-Sailing-Public.

[3]In the following discussion, we use the superscript $(\alpha_*^{\hat{s}})$ to refer to a version of policy α_* implemented with a specific initial time-budget $\hat{s}$.

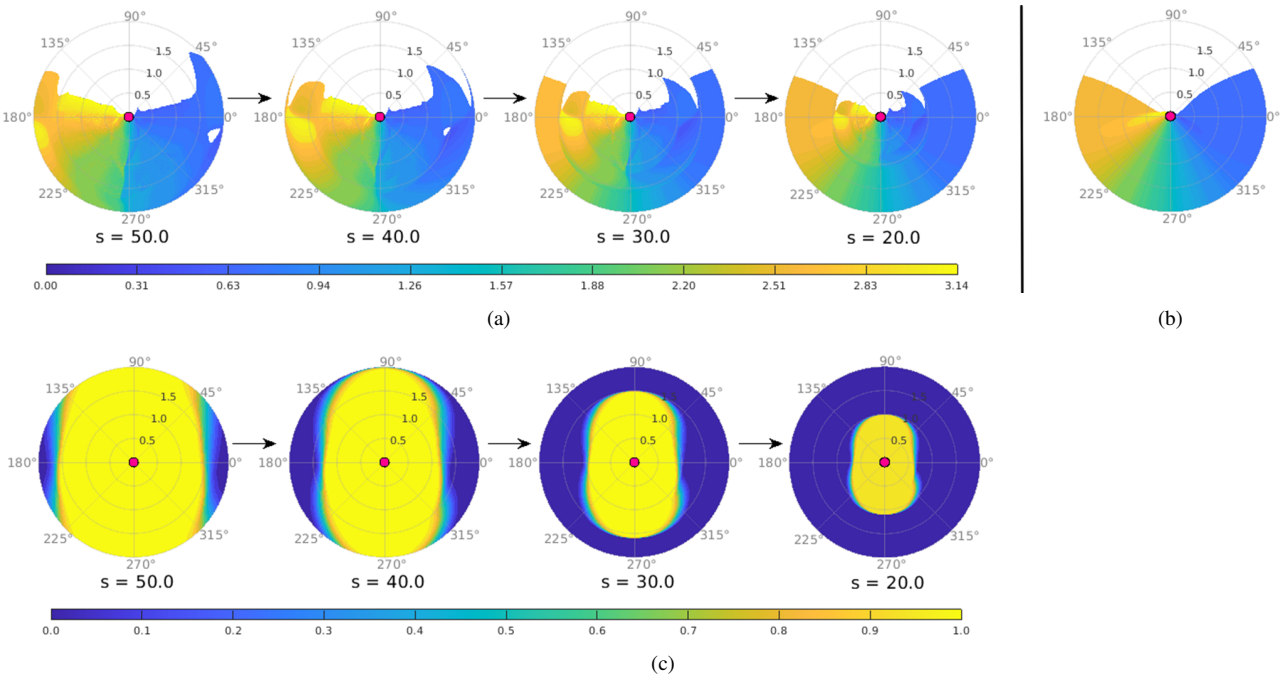


Fig. 2. **Representative s-slices of risk-aware policy, their corresponding optimal probability of reaching the target $\mathcal{D}$, and the risk-neutral policy**: (a) risk-aware optimal policy α_*; (b) risk-neutral optimal policy μ_*; (c) optimal probability of reaching $\mathcal{D}$ associated with (a). All shown for the starboard tack $q = 1$ only and in relative (r, θ) coordinates. In all figures, the target $\mathcal{D}$ is shown as a magenta disk in the center. In (a) and (b), it is optimal to switch to $q = 2$ from wherever the space is left *blank*. Otherwise, it is optimal to stay with $q = 1$ and the best steering angle is shown in color (with the same colorbar used in both subfigures).

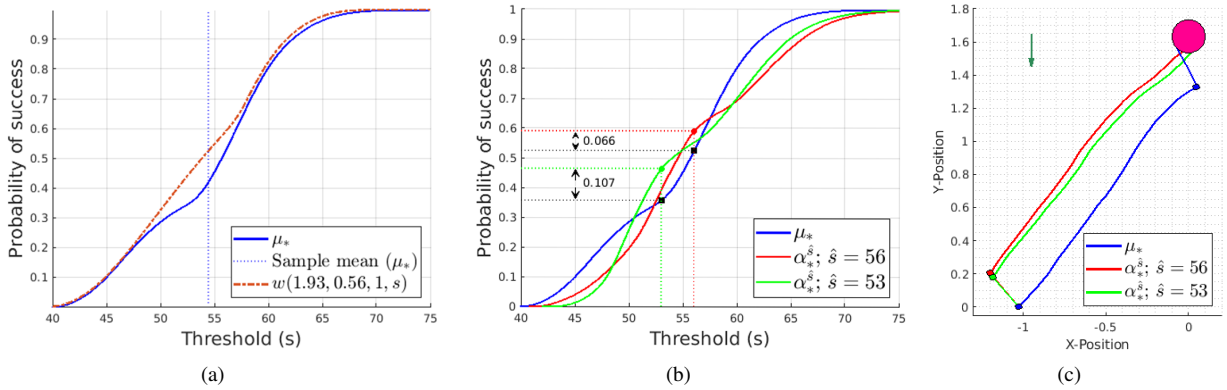
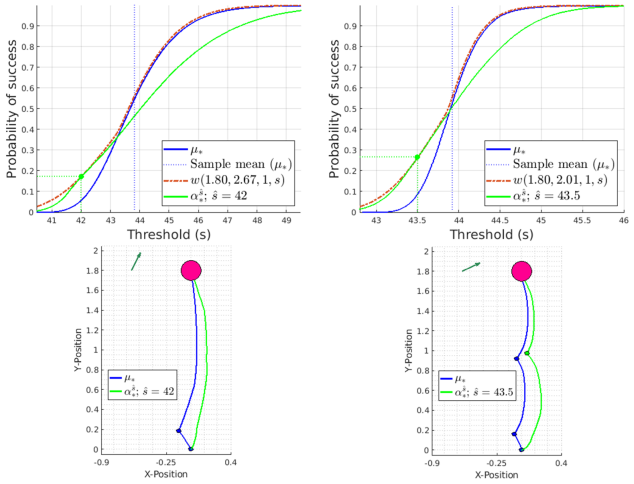


Fig. 3. **Sailing against the wind: a comparison between the risk-aware and risk-neutral approaches with** $(a = 0, \sigma = 0.05)$ **starting from** $(\hat{r}, \hat{\theta}, \hat{q}) = (1.93, 0.56, 1)$ Subfigures: (a) ECDF (empirical cumulative distribution function) generated with the optimal risk-neutral policy μ_* (*solid blue*) vs. the s-dependent risk-aware optimal probability of success $w(\hat{r}, \hat{\theta}, \hat{q}, s)$ (*dash-dotted orange*); (b) ECDFs of the random total time to target generated with different polices; (c) Sample sailboat trajectories in the *absolute xy-coordinates* generated with different polices under the same random wind path. The target set is plotted as a magenta disk at the top, and top-left *dark green arrow* encodes the initial wind direction $\phi(0) = 0$. Trajectory colors correspond to the polices used to generate ECDFs in (b). The colored dots indicate the tack-switching points for respective trajectories. Observed sample means for the arrival time T are 54.46, 55.57, and 55.95 for the policies μ_*, α_*^{53}, and α_*^{56} respectively. Arrival times for the specific trajectories in (c) are 56.55 (blue), 53.54 (green), and 54.07 (red).

show such an example with $(a = 0.05, \sigma = 0.05)$. Unlike in Fig. 3, here the initial direction of the wind is largely toward the target, but we are in the wrong initial tack to fully take advantage of this. The risk-neutral μ_* prescribes an immediate tack-switch followed by another one a bit later and yields a low $\mathbb{P}(T \leq 42) \approx 5.8\%$. In contrast, the threshold-aware α_*^{42} recognizes, based on the sign of a, that the wind is likely to change in the right direction soon and (in the specific wind-evolution example presented in the bottom row of Fig. 4(a)) manages to reach $\mathcal{D}$ without any tack-switches at all. The result of this calculated bet is to almost triple $\mathbb{P}(T \leq 42)$ to 17.2% and make the tack-switches far less common. Even more dramatic improvements can be obtained when the drift is stronger. In Fig. 4(b) with $(a = 0.15, \sigma = 0.05)$, the threshold-aware policy boosts $\mathbb{P}(T \leq 43.5)$ from 8.8% to 26.6%, largely by reducing the number of tack-switches (in most cases, from 3 switches under μ_* to only 1 under $\alpha_*^{43.5}$).



(a) (b)

Fig. 4. **Exploiting the wind-drift**: (a) ($a = 0.05$, $\sigma = 0.05$); (b) ($a = 0.15$, $\sigma = 0.05$). **Top Row:** ECDF for μ_* (*solid blue*), the s-dependent risk-aware optimal probability of success $w(\hat{r}, \hat{\theta}, \hat{q})$ (*dash-dotted orange*), and ECDF for $\alpha_*^{\hat{s}}$ (*solid green*). In (a), $(\hat{r}, \hat{\theta}, \hat{q}) = (1.80, 2.67, 1)$ and $\hat{s} = 42$. The sample means for μ_* and $\alpha_*^{\hat{s}}$ are 43.83 and 44.30. In (b), $(\hat{r}, \hat{\theta}, \hat{q}) = (1.80, 2.01, 1)$ and $\hat{s} = 43.5$. The sample means for μ_* and $\alpha_*^{\hat{s}}$ are 43.93 and 43.97. **Bottom Row**: Two representative paths generated with the same wind evolution (with colors corresponding to respective policies in the top row). The *dark green arrow* encodes the initial wind direction. *Time-to-target:* (a) blue: 42.23, green: 41.44 ; (b) blue: 43.12, green: 42.98. In (a) μ_* led to 2 tack-switches in 99.9% of simulations, while α_* required none in 99.1% of cases with 2 switches needed in all others. In (b) μ_* led to 3 tack-switches in 99.9% of simulations (with others requiring 4), while α_* required 1 switch in 99.9% of cases with 2 switches needed in the rest.

We end this section with two caveats. First, it is usually impossible to optimize the entire CDF of the arrival time. As should be clear from Fig. 3(b), a policy increasing $\mathbb{P}(T \leq \hat{s}_1)$ might be decreasing $\mathbb{P}(T \leq \hat{s}_2)$ even compared to a risk-neutral μ_*. Typically, each $\alpha_*^{\hat{s}}$ is only optimal for its particular threshold/deadline $\hat{s}$. This is why we do not use the usual nomenclature of *risk-aversion* [7] and instead describe or methods as risk (or threshold) *aware*. Second, our decision to revert to μ_* in the "unlucky" α_*-based simulations (once the time-budget is reduced to zero) is fairly arbitrary and one can certainly use other approaches instead. However, this choice does not affect $\mathbb{P}\left(T_{\alpha_*^{\hat{s}}} \leq \tilde{s}\right)$ for any $\tilde{s} \leq \hat{s}$; thus, the primary goal of threshold-aware policies is still achieved.

VI. Conclusion

We have introduced a robust (risk/deadline-aware) approach to controlling a sailboat in stochastically evolving wind conditions. The efficiency of our approach hinges on the numerical method for a pair of quasi-variational HJB-type inequalities, which yield deadline-aware policies for all initial configurations and a broad range of deadlines simultaneously. Numerical experiments demonstrate the advantages of these policies over the traditional risk-neutral approach [3], particularly when it is possible to reduce the number of likely tack-switches.

Several extensions will obviously increase the impact of this approach in the future. Solving the problem in absolute coordinates will allow for a better modeling of the domain geometry (e.g., accounting for obstacles and other target shapes). Incorporating more realistic wind models and more detailed boat dynamics will be clearly of interest to practitioners. Similarly, stochastic differential games might be used to reflect the competitive aspect of sailing races [2]. [E.g., if T_i is a (random) arrival time of the i-th competitor, one could try to maximize $\mathbb{P}\left(T \leq \min_i T_i\right)$.] In addition, it will be important to explore multi-objective versions (e.g., Pareto-optimal tradeoffs between $\mathbb{E}[T]$ and $\mathbb{P}\left(T \leq \hat{s}\right)$) and compare our approach with risk-averse methods that minimize the "Conditional Value at Risk" [7].

Finally, we hope that a similar threshold-aware approach will prove to be useful in many indefinite-horizon hybrid control applications unrelated to sailing.

References

[1] R. Ferretti and A. Festa, "Optimal route planning for sailing boats: A hybrid formulation," *Journal of Optimization Theory and Applications*, vol. 181, no. 3, pp. 1015–1032, 2019.

[2] S. Cacace, R. Ferretti, and A. Festa, "Stochastic hybrid differential games and match race problems," *Applied Mathematics and Computation*, vol. 372, p. 124966, 2020.

[3] C. Miles and A. Vladimirsky, "Stochastic optimal control of a sailboat," *IEEE Control Systems Letters*, vol. 6, pp. 2048–2053, 2021.

[4] T. Başar and P. Bernhard, *H^∞-optimal control and related minimax design problems*. Birkhauser, 1995.

[5] E. Cartee, A. Farah, A. Nellis, J. Van Hook, and A. Vladimirsky, "Quantifying and managing uncertainty in piecewise-deterministic Markov processes," *SIAM/ASA Journal on Uncertainty Quantification*, vol. 11, no. 3, pp. 814–847, 2023.

[6] M. Wang, J. G. Scott, and A. Vladimirsky, "Stochastic optimal control to guide adaptive cancer therapy," *preprint: https://www.biorxiv.org/content/10.1101/2022.06.17.496649v2*, 2023.

[7] Y. Wang and M. P. Chapman, "Risk-averse autonomous systems: A brief history and recent developments from the perspective of optimal control," *Artificial Intelligence*, vol. 311, p. 103743, 2022.

[8] A. B. Philpott, S. G. Henderson, and D. Teirney, "A simulation model for predicting yacht match race outcomes," *Operations Research*, vol. 52, no. 1, pp. 1–16, Feb. 2004.

[9] L. Vinckenbosch, "Stochastic Control and Free Boundary Problems for Sailboat Trajectory Optimization," 2012, publisher: Lausanne, EPFL.

[10] W. H. Fleming and H. M. Soner, *Controlled Markov processes and viscosity solutions*. Springer, 2006, vol. 25.

[11] M. Bardi and I. Capuzzo-Dolcetta, *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*. Springer, 2008.

[12] M. Falcone and R. Ferretti, *Semi-Lagrangian approximation schemes for linear and Hamilton-Jacobi equations*. SIAM, 2014, vol. 133.

[13] R. Ferretti, "A technique for high-order treatment of diffusion terms in semi-Lagrangian schemes," *Communications in Computational Physics*, vol. 8, no. 2, pp. 445–70, 2010.

[14] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*. Springer, 1992.

[15] C.-W. Shu, "Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws," in *Advanced numerical approximation of nonlinear hyperbolic equations*. Springer, 1998, pp. 325–432.

[16] M. Bardi, P. Soravia, and M. Falcone, "Fully discrete schemes for the value function of pursuit-evasion games," in *Advances in dynamic games and applications*. Springer, 1994, pp. 89–105.

[17] M. Bardi, M. Falcone, and P. Soravia, "Numerical methods for pursuit-evasion games via viscosity solutions," in *Stochastic and differential games: theory and numerical methods*. Springer, 1999, pp. 105–175.

[18] M. Assellaou, O. Bokanowski, and H. Zidani, "Error estimates for second order Hamilton-Jacobi-Bellman equations. approximation of probabilistic reachable sets," *Discrete and Continuous Dynamical Systems-Series A*, vol. 35, no. 9, pp. 3933–3964, 2015.

[19] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *J Am Stat Assoc*, vol. 53, no. 282, pp. 457–481, 1958.