2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

2024 IEEE 63rd Conference on Decision and Control (CDC) | 979-8-3503-1633-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/CDC56724.2024.10886845

Motion Planning for The Identification of Linear Classifiers with Noiseless Data: A Greedy Approach

Aneesh Raghavan and Karl Henrik Johansson

***Abstract*—A given region in 2-D Euclidean space is divided by a unknown linear classifier in to two sets each carrying a label. An agent with known dynamics traversing the given region is able to measure the true label perfectly at its position. By following a trajectory, the agent collects data points comprising of its true position and the label at that position. The objective of the agent is to plan a trajectory across the given region to identify the true classifier with high accuracy while minimizing the control cost across the trajectory. We present the following: (i) the classifier identification problem formulated as a control problem; (ii) geometric interpretation of the control problem resulting in one step modified control problems; (iii) control algorithm that results in a data set which is used to identify the true classifier with high accuracy; (iv) convergence of estimated classifier to the true classifier when observed label is not corrupted by noise; (iv) numerical example demonstrating the utility of the control algorithm.**

I. Introduction

A. Motivation

Duality between control and learning (in a broad sense, including estimation and inference problems) has been well studied in the literature. In [1], duality between estimation and control is studied for general stochastic control problems. In [2], exploration vs exploitation has been studied through the control of a meta-parameter in reinforcement learning (RL) problems. [3] studies dual control problems where knowledge gained through control actions is explicitly defined. In [4], dual control techniques have been applied to approximate the intractable aspects of Bayesian RL, leading to structured exploration strategies that differ from standard RL. In [5], stochastic model predictive control is presented in the dual control paradigm. More recently dual control has been applied to active uncertainty learning in human robot interaction, [6]. In all these problems uncertainty in the model or the cost function is being actively learnt through control actions.

Parallels between model predictive control and algorithms in A.I have been drawn, [7]. Learning theory has been extensively applied to control problems; special neural networks and deep networks have been used extensively in system identification and to approximate solutions to control problems, [8]. Systems and control theory however has not been applied to its full potential to learning theory. Adaptive sampling is closely related to the field of active learning, however the former operates in the context supervised learning while the latter is associated with semi-supervised learning. Adaptive sampling for classification has been studied in [9], [10], [11], where sequential sampling algorithms are presented to enhance the learning process. In [12], adaptive sampling has been applied to hyperspectral image classification leading to improvement from state of the art. Learning unknown environment is a crucial part of marine robotics. Adaptive sampling methods have been used to survey and learn about algal bloom, water quality models, etc., in [13], [14], [15], and, [16].

Given the above context, the problem that we consider is the identification of certain aspects of an agents environment that is unkown. Unlike traditional dual control which deals with uncertainty in the system or the cost functions and learning the same, we consider learning of the environment. The problem considered has potential application in marine robotics as well, as described above. In our previous work, [17], we considered path planning for identification of functions in an agent's surroundings.

B. Problem Considered

The problem considered is as follows. A given region in two dimensional space is divided into two regions by a straight line, Figure 1. The true classifier divides the state space into two sets, $\mathcal{X}$, where the true label is 1 and, $\mathcal{X}^c$, where the true label is -1. Every point in the region "orange" carries the label 1 while every point in the alternate region "blue" carries the label -1. Every point on the straight line that divides the region carries the label 0. Four points, p_1, p_2, p_3, and p_4, with their true labels are given by an oracle. In Figure 1, the true labels of p_1 and p_4 are 1 while that of p_2 and p_3 are -1. The true classifier is parameterized by its slope, ρ^*, and intercept, c^*. An agent with known dynamics traverses the region by paying a control cost. Two measurement models can be considered: (i) the agent is able to measure the true label perfectly (deterministic) (ii) the measurement gets corrupted by noise; the measured label is the true label flipped (1 to -1 and -1 to 1) with a certain probability (stochastic). Given the measurement model, the objective of the agent starting at the point p_1 is

*Research supported by the Swedish Research Council (VR), Swedish Foundation for Strategic Research (SSF), and the Knut and Alice Wallenberg Foundation. The authors are with the Division of Decision and Control Systems, Royal Institute of Technology, KTH, Stockholm. Email: aneesh@kth.se, kallej@kth.se

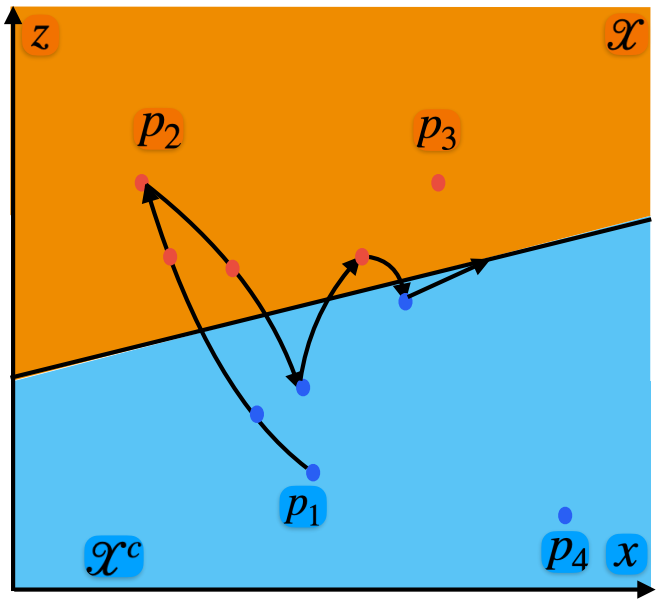

Fig. 1. Schematic for the motion planning problem

to follow a trajectory during which it collects m data points which are optimal for the identification of the classifier while simultaneously minimizing its control costs. In this paper, we consider the first measurement model, i.e., the noiseless case and investigate the geometric aspects of the resulting problem. We reserve the noisy scenario as future work due to space restrictions.

One possible path that could be taken by the agent is depicted in Figure 1. The agent collects 6 data points, apart from the 4 given, 3 of each label. The 10 data points would subsequently be used to estimate the classifier. The four initial points provided by the oracle are assumed to be "far" apart from each other. They provide an initial estimate of the classifier and a region of the $2-D$ space to be explored by the agent to refine the estimate and identify the classifier more accurately. The key idea that we would like to explore is that, rather than collecting large number of data samples from both regions for accurate estimation, is it possible to strategically sample few data points which leads to the same accurate estimation as the large data set. Thus, problem considered here can be interpreted as an adaptive sampling problem which is restricted by the agent's dynamics.

C. Contributions

We formulate the identification problem of the true classifier as described above has a control problem in the deterministic scenario. The formulated control problem is analyzed and the associated challenges and drawbacks are presented. We present a geometric interpretation of the problem using 2D analytic geometry. Utilizing the geometric ideas, we formulate one step control problems which can be solved in a numerically efficient manner. The one step control problems can be interpreted as model predictive control problems with the horizon being one time step. We present a control algorithm (a "greedy" approach) to solve the identification problem which involves solving the one step control problems formulated before. The data set obtained by executing the control algorithm is used to identify the classifier. In the deterministic case, we prove that the classifier identified converges to the true classifier, i.e., the estimated parameters of the classifier converges to the true classifier. We present an example illustrating the control algorithm and the resulting classifier.

D. Outline

In Section II, we present the formulation of the identification problem as a control problem. In Section III, we present the geometric interpretation of the identification problem and the associated one step control problems. In section IV, we present the control algorithm and the proof of convergence to true classifier in the deterministic scenario. In Section V, we present a numerical example demonstrating the application of the control algorithm. In Section VI, we summarize the work presented in the paper and discuss future work.

II. Problem Formulation

In this section, we present the abstraction of the problem described in subsection I-B.

A. Identification of the Classifier

By executing a control policy and following the corresponding path, the positions at which measurements are collected by the agent is denoted by $\{x_j, z_j\}_{j=1}^m$. The true labels of these positions is denoted by $\{y_j\}_{j=1}^m$. If the true classifier is denoted by the straight line $z = \rho^* x + c^*$, then $y_j = \text{sgn}(z_j - \rho^* x_j - c^*), j = 1, \ldots, m$. Given the data points, the classification problem can be formulated as,

$$\min_{\rho\in\mathbb{R}, c\in\mathbb{R}} \sum_{j=1}^{m} y_j \text{sgn}(z_j - \rho x_j - c).$$

The optimization problem is to choose parameters ρ and c so that the resulting line classifies as many data points as possible correctly. This formulation does not yield to computationally effective solutions. The classical reformulation, [18], [19], of the problem is to identify a line which (i) maximizes the minimum distance of the data points from the line; (ii) ensures that the points are classified correctly. Given a possible classifier, $z - \rho x - c = 0$, the distance of the point (x_k, z_k) is given by $\frac{|z_k - \rho x_k - c|}{\rho^2}$. Through suitable re-normalization, it can be ensured that $|z_{j^*} - \rho x_{j^*} - c| = 1$ for atleast one of the data points,$(x_{j^*}, z_{j^*}, y_{j^*})$, while $|z_j - \rho x_j - c| \geq 1$ for all other data points. The classification problem can be reformulated as,

$$\min_{\rho\in\mathbb{R}, c\in\mathbb{R}} \quad \frac{\rho^2}{2} \quad \text{s.t } y_j(z_j - \rho x_j - c) \geq 1, j = 1, \ldots, m.$$

The above optimization problem is a convex optimization problem which can be solved efficiently by considering the dual problem, [19].

B. Control Problem

The state space of the agent is $\mathbb{R}^2$, while the set of actions (action space) that it can take is denoted by $\mathcal{U}$. The state of the system at time t is denoted by $\zeta(t) = [x(t), z(t)]$ while the action is denoted by $u(t)$. Along with the full state information at t, the observation of the system at t is the true label at that state which is denoted by $y(t)$. Thus, $y(t) = \text{sgn}(z(t) - \rho^* x(t) - c^*)$ where ρ^* and c^* the parameters corresponding to the true classifier and are unknown. The observation $y(t)$ is available to the agent and is used to estimate ρ^* and c^*. The initial state of agent is the point p_1 which we denote as $(\bar{x}_1, \bar{z}_1, -1)$. For the deterministic case, we do not consider process noise or measurement noise. The agent is modeled as a discrete time system with known dynamics,

$$\zeta(t+1) = \phi(\zeta(t), u(t)),\ \bar{y}(t) = [\zeta(t), y(t)],\ \bar{y}(0) = [\bar{x}_1, \bar{z}_1, -1],\ y(t) = \text{sgn}(z(t) - \rho^* x(t) - c^*), t = 0, \dots m-1.$$

The learning problem formulated as a control problem is to find a control policy, $\{\Upsilon_j\}_{j=0}^{m-1}$, where $\Upsilon_j : \mathbb{R}^{2^{j+1}} \times \{-1, 0, 1\}^{j+1} \rightarrow \mathcal{U}$, which minimizes the control cost while ensuring that the data points collected yield a suitable classifier. The optimization problem is,

$$\min_{\{\Upsilon_t\}_{t=0}^{m-1}, \{\rho, c\} \subset \mathbb{R}} \frac{\rho^2}{2} + \varrho \sum_{t=0}^{m-1} \left|\left| \Upsilon_t\Big(\{\bar{y}(j)\}_{j=0}^{t-1}\Big) \right|\right|^2, \text{ s.t,}$$

$$y(t)(z(t) - \rho x(t) - c) \geq 1, t = 1, \dots, m,\ |\sum_{t=0}^{m} y(t)| \leq 1,$$

$$(\bar{z}_j - \rho \bar{x}_j - c) \leq 1, j = 1, 4, (\bar{z}_j - \rho \bar{x}_j - c) \geq 1, j = 2, 3,$$

$$\zeta(t+1) = \phi\Big(\zeta(t), \Upsilon_t\Big(\{\bar{y}(j)\}_{j=0}^{t-1}\Big)\Big), t = 0, \dots m-1, y(t) = \text{sgn}(z(t) - \rho^* x(t) - c^*), \bar{y}(t) = [\zeta(t), y(t)], t = 0, \dots m.$$

In the constraints of the above optimization problem, in the second line, the constraints are to ensure that the four given points are classified correctly. The constraint, $|\sum_{t=0}^{m} y(t)| \leq 1$, has been included to ensure that equal number points are visited in both regions or at most one additional point is visited in one of the regions.

III. Analysis of the Problem

We begin this section with some observations on the control problem formulated in subsection II-B. Let us consider a dynamic programming approach to solve the problem. The cost function can be decomposed into learning cost, $\frac{\rho^2}{2}$, and control cost, $\sum_{t=0}^{m-1} \left|\left| \Upsilon_t\Big(\dots\Big) \right|\right|^2$. The learning cost is not evidently decomposable into a learning cost for each stage. At stage m, there is no control cost. Given the data points $\{x_j, z_j, y_j\}_{j=1}^{m}$, the problem is to solve the classification problem which can be done using quadratic programming as mentioned before. The optimal cost at stage m is the minimum cost obtained after choosing optimal ρ and c. At stage $m-1$, given the $m-1$ points visited by the agent, the objective is to chose the final data point so that classification can be performed on the resulting data set. The objective

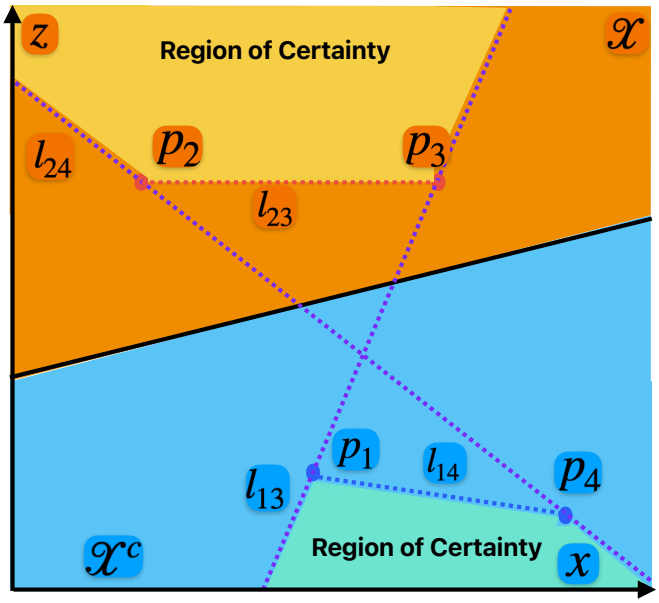

Fig. 2. Region of certainty from the four given points

of the classification problem is to find (ρ, c) such that the minimum distance of the data points from the classifier is maximized. If the control strategy at $m-1$ were to aid this objective, the control action would move the agent to a point where the label is opposite to the current label and whose distance is equal to the minimum distance of the given $m-1$ points from the current estimate of the classifier at stage $m-1$.

If the same argument were to be repeated at stages j, $0 \leq j \leq m-2$, the control strategy would move the agent to points which are roughly in the neighborhood of the initial points (in order to maximize the minimum distance) which is not useful for the identification of the true classifier. If the learning cost is changed to $\sum_{t=1}^{m} y(t)\text{sgn}(z(t) - \rho x(t) - c)$, it also does not enforce the same. This is because, given any m data points with true labels the data set is linearly separable and hence the minimum cost is always zero even though the estimated classifier is not close to the true classifier. It appears as though the cost function for the identification problem is not utilizing the state information and the corresponding label at every stage to enhance the learning process. We explore this idea further in the following.

We consider the scenario in depicted in Figure 2 as a canonical case. The slope of the true classifier is positive and the z intercept is positive as well. Other cases are : slope positive , intercept negative; slope negative intercept positive; slope negative intercept negative. The other three cases are obtained through translation and rotation of the scenario considered. Hence, the following arguments are applicable to other cases as well. The lines $l_{13}, l_{14}, l_{23}, l_{24}$ are defined as $l_{ij} = \{(x, z) \in \mathbb{R}^2 : z - \rho_{ij} x - c_{ij} = 0\},\ i = 1, 2, j = 3, 4$. These lines are obtained from the four given points. Each of these lines are "bounds" for the true classifier is the following sense.

Consider l_{23} and the region $\{(x,z) \in \mathbb{R}^2 : z - \rho_{23}x - c_{23} \geq 0\}$. In this region consider any point whose x co-ordinate lies between that of p_2 and p_3. The label of such a point is 1. This because if there is a point with label -1, linear separability of data gets violated, i.e., there does no exist a linear classifier that separates p_1, p_2, p_3, p_4 and the point with label -1. However, we are unable to comment on the region $\{(x,z) \in \mathbb{R}^2 : z - \rho_{23}x - c_{23} \leq 0\}$ as there is not enough information. Further, there are two points on l_{23} with label 1. Hence it is not the true classifier, but a "bound" for the true one. The line l_{13} is also a bound for the true classifier, since a line with slope slightly greater than slope of l_{13} and intercept slightly less than c_{13} which is in fact negative, does not separate p_1, p_2, p_3, p_4. By the same argument l_{24} is also a bound for the true classifier.

Since l_{13} and l_{24} bound the true classifier, if either of them are the true classifier, then the corresponding sets $\{(x,z) \in \mathbb{R}^2 : z - \rho_{13}x - c_{13} \geq 0\}$ or $\{(x,z) \in \mathbb{R}^2 : z - \rho_{24}x - c_{24} \geq 0\}$ carry the label 1. Taking the intersection of the three sets, we obtain a *Region of Certainty*, a set where the true label is 1 given the four points, p_1, p_2, p_3, p_4. We note that for any point in this region of certainty, a true label of -1 would imply that the four points and the point with label -1 are not linearly separable. Hence the name for the region. Using the same arguments with l_{14}, l_{13} and l_{24}, we obtain a region of certainty with label -1.

The slopes and intercepts of the four lines provide bounds on the slope and intercept of the true line. Considering the range of arctan to be $(-\frac{\pi}{2}, \frac{\pi}{2})$, let $\theta_{ij} = \arctan(\rho_{ij})$. In Figure 2, we observe that $\theta_{13} > 0, \theta_{23} \geq 0$, $\theta_{13} > \theta_{23}$ while $\theta_{14} < 0, \theta_{24} < 0$ and $\theta_{24} < \theta_{14}$. From the linear separability arguments presented above, we conclude that the slope of the true classifier belongs to $[\theta_{24}, \theta_{14}] \cup [\theta_{23}, \theta_{13}]$. The intercept of the true line belongs to $[c_{\min}, c_{\max}]$ where $c_{\max} = \max(c_{13}, c_{14}, c_{23}, c_{24})$ and $c_{\min} = \min(c_{13}, c_{14}, c_{23}, c_{24})$. In Figure 2, $c_{\max} = c_{24}$ and $c_{\min} = c_{13}$.

Consider the scenario depicted in Figure 3, where the distance between pairs p_1, p_2 and p_3, p_4 which have opposite labels has reduced while the distance between p_1, p_4 and p_2, p_3 which carry the same label has increased. In Figure 3, we observe that $\theta_{13} > 0, \theta_{23} > 0, \theta_{23} < \theta_{13}$ while $\theta_{14} > 0, \theta_{24} < 0$. The slope of the true classifier thus lies between $[\theta_{24}, 0] \cup [\min(\theta_{23}, \theta_{14}), \theta_{13}]$ while its intercept lies between $[c_{13}, c_{24}]$. We note that the bounds for the true slope and intercept in scenario of Figure 3 is a strict subset of the bounds in the scenario of Figure 2 which leads to an "increase" in the region of certainty of both labels, i.e., the regions of certainty in Figure 2 are strict subsets of the corresponding regions of certainty in Figure 3.

Given the above reasoning, the control problem is to be formulated is such a away the region of certainty eventually matches with the entire regions carrying the true label. We consider one step control problems with the objective of pruning the bounds of the slope and the intercept of the true classifier so that the set to which the true parameters belong eventually collapses to singletons, i.e., the true value of the parameters. To meet this objective, at a given position, in one

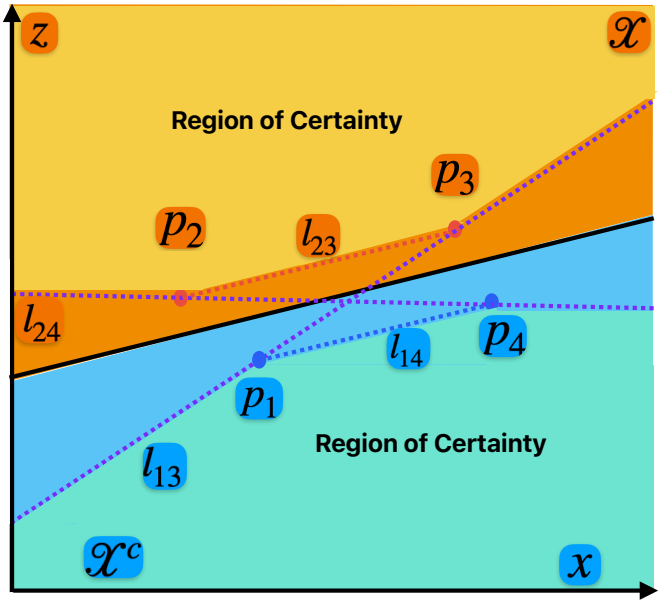

Fig. 3. Region of certainty from new four points obtained by Agent

step, the agent could either (i) move to a position of opposite label whose distance is less than the distance of the previous point with opposite label from current position (ii) move to a position of same label in the current region of uncertainty which is farther way from current position. For (i), given the current position and label, $(x(t), z(t), y(t))$, the control problem is formulated as:

$$\min_{u \in \mathcal{U}} ||\zeta(t+1) - \zeta(t)||^2 + \varrho||u||^2, \text{ s.t } \zeta(t+1) = \phi(\zeta(t), u),$$
$$y(t)\text{sgn}(z(t+1) - \rho^* x(t+1) - c^*) \leq 0$$

The above problem has two issues. (i) ρ^* and c^* are unknown and (ii) even if the true parameters were known, the constraint may not be feasible in one step due to the limited control actions that the agent can take. Let $\mathcal{X}_{t,1}$ denote the region of certainty for label 1 at stage t and let $\mathcal{X}_{t,-1}$ denote the region of certainty for label -1 at stage t. Let $E_{\rho,t}$ denote the bounds for the slope of the true classifier at stage t. We consider the following alternate formulation.

$$P1: \quad \max_{u \in \mathcal{U}} ||\zeta(t+1) - \zeta(t)||^2 - \varrho||u||^2$$
$$\text{s.t } \zeta(t+1) = \phi(\zeta(t), u), \zeta(t+1) \notin \mathcal{X}_{t,1}, \zeta(t+1) \notin \mathcal{X}_{t,-1}$$
$$\arctan\left(\frac{z(t+1) - z(t)}{x(t+1) - x(t)}\right) \notin E_{\rho,t}$$

This formulation pushes the agent away to a point which is far away from its current position while not entering the regions of certainty and ensuring that the control cost is small enough. The regions of certainty can be expressed as the intersection of a set of halfspaces. Hence, the above problem can be solved numerically. The constraint in third line is included to ensure that the agent does not end up traveling parallel to the true classifier in which case it cannot reach a point of opposite label. By moving in any direction not in

set $E_{\rho,t}$, the agent is guaranteed to reach a point of opposite label though it might take multiple steps. In implementation, it is possible to restrict the angles further, for e.g. the agent could be restricted to track the direction of the vector $\overrightarrow{p_1p_2}$ or the "bisector" of $E_{\rho,t}$. For the agent to move to point with same label as current label, however farther away the following problem is solved.

$$P2: \quad \max_{u\in\mathcal{U}} ||\zeta(t+1)-\zeta(t)||^2 - \varrho||u||^2$$
$$\text{s.t } \zeta(t+1)=\phi(\zeta(t),u), \zeta(t+1)\notin\mathcal{X}_{t,1}, \zeta(t+1)\notin\mathcal{X}_{t,-1}$$
$$\arctan\left(\frac{z(t+1)-z(t)}{x(t+1)-x(t)}\right)\in E_{\rho,t}$$

In the above the agent could travel in a direction which is parallel to the true classifier, which is in fact good as it meets the objective. However for all other directions, the agent could move to a point which has a opposite label to the label of its current position.

Problems $P1$ and $P2$ are reformulations of the identification problem considered in subsection II-B. The reformulation was necessary to incorporate feedback into the classifier identification problem. The drawback of $P1$ and $P2$ is that the control cost of the entire trajectory is not optimized but gets optimized at every stage, which may be not be optimal for entire trajectory, i.e., it is not the outcome or "stage" optimization problem of a dynamic programming problem. Thus, the one step control problems can be interpreted as a greedy approach as the cost function at every time step is myopic, i.e., takes into account the cost of the current time step and ignores the cost to go from the next time step.

IV. Algorithm

In this section, we present a control algorithm to solve the problem presented in subsection I-B.

The control algorithm executed by the agent is presented in Algorithm 1. In this algorithm, at any given position if the agent has seen a label flip from its past position through execution of $P1$ at its past position, the agent solves $P2$ and moves to a new position. At the new position, the agent solves $P1$ and moves to new positions until it observers a label flip. The objective of $P1$ is to obtain points which are close to each other but of opposite label; hence executed multiple times until the objective is achieved. The objective of $P2$ is to obtain points which are of the same label but far from each other. It is executed only once as even if the objective is not met, it results in points with opposite labels which are far from each other. At $t=0$, the agent begins by solving $P1$ and repeats the same until a label flip is observed.

Given the data set after m stages, $\{x(j),z(j),y_j\}_{j=1}^m$, the distance between pairs with opposite labels is found. Of all the pairs, two pairs which have the shortest distance between them are chosen. Let the four points be $(p_{1,m},-1)$, $(p_{2,m},1)$, $(p_{3,m},1)$ $(p_{4,m},-1)$. Consider then quadrilateral formed by $(p_{1,m},p_{2,m},p_{3,m},p_{4,m})$ and let the diagonals intersect at p_m. Consider the angle formed by $p_{1,m},p_m,p_{2,m}$. The line that bisects this angle whose slope is ρ_m and intercept is c_m is declared as the estimate of the true classifier.

Algorithm 1 Control for classification

procedure CFC
  Given $(p_1,-1),(p_2,1),(p_3,1),(p_4,-1),\phi(\cdot)$, and $\mathcal{U}$.
  $\zeta(0)\leftarrow p_1, y(0)\leftarrow -1$
  $j\leftarrow 0, Label\leftarrow -1, Counter\leftarrow 0$
  while $j\le m-1$ **do**
    if $Counter \mod 2 = 0$ **then**
      Solve $P1$ to obtain $U(j)$.
      Agent moves to $\zeta(j+1)\leftarrow\phi(\zeta(j),U(j))$
      $j\leftarrow j+1$, collect observation $Y(j)$.
      if $Y(j)\times Label=-1$ **then**
        $Label\leftarrow Y(j)$
        $Counter\leftarrow Counter+1$
    else
      Solve $P2$ to obtain $U(j)$.
      Agent moves to $\zeta(j+1)\leftarrow\phi(\zeta(j),U(j))$
      $j\leftarrow j+1$, collect observation $Y(j)$.
      $Counter\leftarrow Counter+1$

Proposition IV.1. *As the number of data points tends to infinity, the estimated classifier converges to the true classifier, i.e.,* $\lim_{m\to\infty}\rho_m=\rho^*$ *and* $\lim_{m\to\infty}c_m=c^*$.

For proof we refer to [20].

V. Example

In this section, we present an example illustrating the implementation of each of the control algorithms described in the previous section. We consider a $20m\times 20m$ region in $\mathbb{R}^2$. We are given four initial points with their true labels as indicated in Figure 4. We consider a unicycle model for the agent:

$$x(t+1)=x(t)+v(t)\cos(\theta(t+1)),$$
$$z(t+1)=z(t)+v(t)\sin(\theta(t+1)),$$
$$\theta(t+1)=\theta(t)+w(t), \text{ where } U=[v,w].$$

In the above model, each time step is a two step processes. First, θ gets updated using the angular velocity. Following this step, once the direction of travel gets fixed, the positions get updated using velocity. With this model problem $P1$ (and similarly $P2$) gets modified to,

$$P1: \max_{u\in\mathcal{U}} ||v(t)||^2-\varrho(||v(t)||^2+||w(t)||^2)$$
$$\text{s.t } \zeta(t+1)=\phi(\zeta(t),u), \zeta(t+1)\notin\mathcal{X}_{t,1},$$
$$\zeta(t+1)\notin\mathcal{X}_{t,-1},\ \theta(t+1)\notin E_{\rho,t}$$

We consider $v(\cdot)\in\{0,0.1,0.2,\ldots,2\}$ with unit, m/ unit time. We consider $w(\cdot)\in\{-\frac{\pi}{2},\ldots,-0.01,0.0,0.01,\ldots,\frac{\pi}{2}\}$ with unit, rad/ unit time. ϱ was set to 0.1, thus enabling the agent to utilize higher values of w resulting in quick change of orientation. With this setup, Algorithm 1 was run for 10 steps. The path followed by the agent is indicated in Figure 4. While solving problems $P1$ and $P2$ during the simulation run, to ensure that constraints corresponding to $\theta(t)$ are met, we choose bisectors of suitable angles to maintain symmetry as indicated in Figure 4. The 10 data

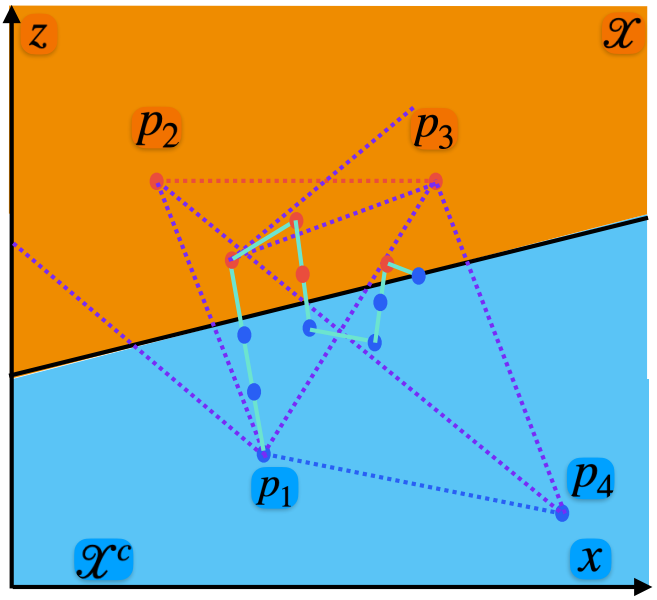

Fig. 4. Path of the Agent with unicycle model and noiseless observations

points gathered were utilized to estimate the classifier. The parameters corresponding to the estimated classifier were $\rho_{10} = 0.38/\theta_{10} = 0.36\ rad\ = 20.8°$ deg and $c_{10} = 3.6$ while corresponding values of the true classifier where $\rho^* = 0.41/\theta^* = 0.389\ rad\ = 22.29°$ deg and $c_{10} = 3.5$. Thus, in this example the classifier was estimated with high accuracy with few data points

VI. Conclusion and Future Work

To summarize, we considered the problem of identification of a linear classifier by an agent. The data collected by the agent for the identification was constrained by its dynamics. Further, we assumed that there is no process noise or measurement noise. We presented geometric interpretation of the problem which was then utilized to develop an efficient control algorithm. Data obtained as a result of the control algorithm was used to identify the classifier with high accuracy.

As future work, we are interested in understanding the control problems which when analyzed using dynamic programming would result in value functions which are obtained through optimization problems similar to the one step control problems formulated in this paper. Formal guarantees that feedback during learning can reduce the number of data points required to achieve a given level of accuracy is to be investigated. The application of active learning (adaptive sampling) algorithms to dual control has been studied, [21], [22], [23], [24]. Further interactions between dual control and active learning algorithms is to be investigated.

References

[1] E. Todorov, "General duality between optimal control and estimation," in *2008 47th IEEE conference on decision and control*. IEEE, 2008, pp. 4286–4292.

[2] S. Ishii, W. Yoshida, and J. Yoshimoto, "Control of exploitation–exploration meta-parameter in reinforcement learning," *Neural networks*, vol. 15, no. 4-6, pp. 665–687, 2002.

[3] T. Alpcan and I. Shames, "An information-based learning approach to dual control," *IEEE transactions on neural networks and learning systems*, vol. 26, no. 11, pp. 2736–2748, 2015.

[4] E. D. Klenske and P. Hennig, "Dual control for approximate bayesian reinforcement learning," *Journal of Machine Learning Research*, vol. 17, no. 127, pp. 1–30, 2016.

[5] A. Mesbah, "Stochastic model predictive control with active uncertainty learning: A survey on dual control," *Annual Reviews in Control*, vol. 45, pp. 107–117, 2018.

[6] H. Hu and J. F. Fisac, "Active uncertainty reduction for human-robot interaction: An implicit dual control approach," in *International Workshop on the Algorithmic Foundations of Robotics*. Springer, 2022, pp. 385–401.

[7] Y. LeCun, "A path towards autonomous machine intelligence version 0.9. 2, 2022-06-27," *Open Review*, vol. 62, no. 1, 2022.

[8] J. Sabouri, S. Effati, and M. Pakdaman, "A neural network approach for solving a class of fractional optimal control problems," *Neural Processing Letters*, vol. 45, pp. 59–74, 2017.

[9] K. Djouzi, K. Beghdad-Bey, and A. Amamra, "A new adaptive sampling algorithm for big data classification," *Journal of Computational Science*, vol. 61, p. 101653, 2022.

[10] P. Singh, J. v. d. Herten, D. Deschrijver, I. Couckuyt, and T. Dhaene, "A sequential sampling strategy for adaptive classification of computationally expensive data," *Structural and Multidisciplinary Optimization*, vol. 55, pp. 1425–1438, 2017.

[11] S. Shekhar, G. Fields, M. Ghavamzadeh, and T. Javidi, "Adaptive sampling for minimax fair classification," *Advances in Neural Information Processing Systems*, vol. 34, pp. 24 535–24 544, 2021.

[12] Y. Ding, J. Feng, Y. Chong, S. Pan, and X. Sun, "Adaptive sampling toward a dynamic graph convolutional network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 2021.

[13] B. Zhang and G. S. Sukhatme, "Adaptive sampling for estimating a scalar field using a robotic boat and a sensor network," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, 2007, pp. 3673–3680.

[14] M. Bernstein, R. Graham, D. Cline, J. M. Dolan, and K. Rajan, "Learning-based event response for marine robotics," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 3362–3367.

[15] P. Stankiewicz, Y. T. Tan, and M. Kobilarov, "Adaptive sampling with an autonomous underwater vehicle in static marine environments," *Journal of Field Robotics*, vol. 38, no. 4, pp. 572–597, 2021.

[16] T. O. Fossum, J. Ryan, T. Mukerji, J. Eidsvik, T. Maughan, M. Ludvigsen, and K. Rajan, "Compact models for adaptive sampling in marine robotics," *The International Journal of Robotics Research*, vol. 39, no. 1, pp. 127–142, 2020.

[17] A. Raghavan, G. Sartori, and K. H. Johansson, "Motion planning for the estimation of functions," in *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE, 2023, pp. 7150–7155.

[18] V. I. Paulsen and M. Raghupathi, *An introduction to the theory of reproducing kernel Hilbert spaces*. Cambridge university press, 2016, vol. 152.

[19] T. Hofmann, B. Schölkopf, and A. J. Smola, "Kernel methods in machine learning," *The annals of statistics*, vol. 36, no. 3, pp. 1171–1220, 2008.

[20] A. Raghavan and K. H. Johansson, "Motion planning for identification of linear classifiers," *https://arxiv.org/abs/2403.15687*, 2024.

[21] Z. Li, W.-H. Chen, and J. Yang, "Concurrent active learning in autonomous airborne source search: Dual control for exploration and exploitation," *IEEE Transactions on Automatic Control*, vol. 68, no. 5, pp. 3123–3130, 2022.

[22] T. Alpcan, "Dual control with active learning using gaussian process regression," *arXiv preprint arXiv:1105.2211*, 2011.

[23] Z. Li, W.-H. Chen, J. Yang, and Y. Yan, "Dual control of exploration and exploitation for auto-optimization control with active learning," *IEEE Transactions on Automation Science and Engineering*, 2024.

[24] Z. Li, W.-H. Chen, J. Yang, and C. Liu, "Cooperative active learning-based dual control for exploration and exploitation in autonomous search," *IEEE Transactions on Neural Networks and Learning Systems*, 2024.