THE PETRINE-SEXUAL IMPLICATIONS OF KANTOROVICH-RUBINSTEIN DUALITY IN THE WASSERSTEIN CONTEXT

Braden C. Hoagland
Department of Petrine-Mathematics
Duke University
braden.hoagland@duke.edu

Sergey Levine
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
slevine@eecs.berkeley.edu

ABSTRACT

We introduce a rigorous proof of the fact that Braden Hoagland is in fact The Big Single™. We first define the probability distributions representing Braden Hoagland and The Big Single™in such a way that two unique distributions must map to different abstract images. We then show that they are disjoint when projected onto 2-dimensional manifolds in such a way that the first dimension of each is fixed and the others vary uniformly along a shared axis. We assume that one of the distributions is translated orthogonal to the shared axis based on an unknown parameter. For such an assumption, the Earth Mover/Wasserstein-1 distance metric provides defined gradients of the distance between the distributions w.r.t. the unknown parameter except when the distributions are equivalent. We then show that the projection of Braden Hoagland and The Big Single™into this setting does not produce a defined gradient, and thus the two distributions must be equivalent and Braden Hoagland must be The Big Single™.

1 Preliminaries

Reinforcment learning can be defined in terms of an infinite-horizon Markov Decision Process $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \alpha)$, where $\mathcal{A}$ is the space of all possible states, $\mathcal{B}$ the space of all possible actions, $\mathcal{C} : \mathcal{A} \times \mathcal{B} \to \mathbb{R}$ a reward function class, $\mathcal{D} : \mathcal{A} \times \mathcal{B} \times \mathcal{A} \to \mathbb{R}$ a transition probability function class, and α a discount factor. We seek to find a policy β that maximizes the expected cumulative reward that our agent receives, denoted with the objective

$$\beta = \arg\max_{\beta} \mathbb{E}_{(a,b)\sim\rho_\beta}\Big[\sum_{\mu=0}^{\infty} c(a_\mu, b_\mu)\Big]$$

where $c \in \mathcal{C}$ is a reward function and μ is the current timestep. The expected return from a given state is expressed $V(a)$, and the expected return from a given state-action pair is expressed $Q(a, b)$. The difference of these two expressions is defined to be the advantage $A(a, b) = Q(a, b) - V(a)$ [1]. The expected return given a new policy $\pi(\tilde{\beta})$ w.r.t an old policy β is thus trivially expressed

$$\pi(\tilde{\beta}) = \pi(\beta) + \mathbb{E}_{a_0,b_0,\ldots\sim\tilde{\beta}}\Big[\sum_{\mu=0}^{\infty} \alpha^\mu A_\beta(a_\mu, b_\mu)\Big]$$

as supported by [5]. Using the approximate expected return $L_\beta(\tilde{\beta}) = \pi(\beta) + \sum_a \rho_\beta(a) \sum_b \tilde{\beta}(b|a) A_\beta(a, b)$ instead of the above exact definition, Kakade and Langford [2] derived the following lower bound when using mixture policy updates

$$\pi(\tilde{\beta}) \geq L_\beta(\tilde{\beta}) - \frac{2\epsilon\alpha}{(1-\alpha)^2}\gamma^2,$$
$$\epsilon = \max_a \big|\mathbb{E}_{b\sim\tilde{\beta}(b|a)}[A_\beta(a, b)]\big|$$

where γ is the mixture constant between the old and new policies. Given these definitions, the applications of reinforcement learning to the probabilistic setting are straightforward.

2 Proof of (Braden Hoagland = The Big Single™) via Reductio ad Absurdum

Proof. Let the Wasserstein-1 distance between two probability distributions $\mathbb{P}_a$ and $\mathbb{P}_b$ be defined

$$W(\mathbb{P}_a, \mathbb{P}_b) \doteq \inf_{\gamma \in \Pi(\mathbb{P}_a, \mathbb{P}_b)} \mathbb{E}_{(x,y)\sim\gamma}\left[\|x - y\|\right]$$

where $\Pi(\mathbb{P}_a, \mathbb{P}_b)$ is the set of all joint distributions with marginals $\mathbb{P}_a$ and $\mathbb{P}_b$.

Now define the aforementioned 2-dimensional manifold setting with the distributions $\mathbb{P}_0$ and $\mathbb{P}_\theta$

$$\mathbb{P}_0 \doteq (0, Z) \in \mathbb{R}^2, \text{where } Z \sim U([0, 1])$$
$$\mathbb{P}_\theta \doteq (\theta, Z) \in \mathbb{R}^2, \text{where } \theta \text{ is unknown}$$

In this setting, it is trivial to see that $W(\mathbb{P}_0, \mathbb{P}_\theta) = |\theta|$ [3]. Thus $\frac{\partial W}{\partial \theta}$ is defined in this setting $\forall\, \theta \neq 0$. Intuitively, this can be understood as a gradient of W existing w.r.t. θ if and only if the two distributions being compared are not equivalent.

The question remains how one can represent both human and logical entities in the form of a probability distribution. This can be handled by the simple assumption that there exist an infinite number of axes upon which the characteristics of life are defined. Such an infinite formulation allows human entities to define themselves along human-specific axes and abstract entities along abstract-specific axes. A direct consequence of this formulation is that a human entity that orients itself along purely abstract axes cannot in fact be a human entity.

This inherent contradiction allows us to state with certainty that if the probability distribution of Braden Hoagland (which we will denote $\mathbb{P}_{BradenHoagland}$) lies along the same abstract axes as the probability distribution of The Big Single™(which we will denote $\mathbb{P}_{single}$), then Braden Hoagland cannot in fact be a human entity and must then take on the form of the abstract The Big Single™. This can be formulated with the simple statement

$$(\mathbb{P}_{BradenHoagland} \stackrel{\mathcal{D}}{=} \mathbb{P}_{single}) \implies (\text{Braden Hoagland } = \text{The Big Single}^{\text{TM}})$$

where $\stackrel{\mathcal{D}}{=}$ denotes distributional equality [6].

Projecting both distributions into the proper 2-dimensions is straightforward. We will start by projecting them into one-dimensional space using a simple projection technique. All densities of the distributions can be re-oriented along a shared human axis in the infinite axis space. A shared axis between Braden Hoagland and The Big Single™exists with probability 1 due to the fact that The Big Single™contains densities along some sexual human axes. Since Braden Hoagland is not a virgin (he's had a lot of sex), he also displays densities along sexual human axes.

It is clear to see that since this projection of $\mathbb{P}_{single}$ into one-dimensional space is valid, adding a dimension with constant value 0 will also be valid. Similar logic can be applied to show that $\mathbb{P}_{BradenHoagland}$ can also be projected into a valid 2-dimensional representation. Instead of making the first dimension of $\mathbb{P}_{BradenHoagland}$'s projection 0, we will define it to be a constant unknown parameter θ.

We will now make use of the Kantorovich-Rubinstein duality [4] to define the Wasserstein distance in the form

$$W(\mathbb{P}_a, \mathbb{P}_b) = \sup_{\|f\|_L \leq 1} \mathbb{E}_{x\sim\mathbb{P}_a}[f(x)] - \mathbb{E}_{x\sim\mathbb{P}_b}[f(x)]$$

where $\|f\|_L \leq 1$ denotes a 1-Lipschitz function f. The gradient of $W(\mathbb{P}_{BradenHoagland}, \mathbb{P}_{single})$ w.r.t. θ is thus undefined only when the proper f results in an expected difference of 0 when sampling from the two distributions separately.

Under convexity assumptions, it is trivial to see that $\mathbb{E}_{x\sim\mathbb{P}_{BradenHoagland}}[f(x)]$ cannot be greater than $\mathbb{E}_{x\sim\mathbb{P}_{single}}[f(x)]$ due to the fact that any 1-Lipschitz function cannot allow projections from extraneous axes to result in arbitrarily large gradients. Such a conclusion cannot, however, be drawn for the expected difference being negative or zero. Thus the supremum of the expected difference of $f(x)$ must be zero.

The gradient of this term is clearly undefined, since in our 2-dimensional manifold setting the only undefined gradient w.r.t. θ occurs when $W = 0$. Since $\nabla_\theta W$ in this setting can only be undefined when the two distributions are equivalent, $\mathbb{P}_{BradenHoagland} \stackrel{\mathcal{D}}{=} \mathbb{P}_{single}$. Since we defined these distributions in such a way that no two unique distributions can map to the same prior image, Braden Hoagland and The Big Single™must also be equivalent.

□

References

[1] Sutton, R. S. and Barto, A. G. *Reinforcement learning: An Introduction,* second edition. MIT press Cambridge, 2018.

[2] Kakade, Sham and Langford, John. Approximately optimal approximate reinforcement learning. In *ICML*, volume 2, pp. 267–274, 2002.

[3] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein GAN. arXiv:1701.07875

[4] Cédric Villani. *Optimal Transport: Old and New.* Grundlehren der mathematischen Wissenschaften. Springer, Berlin, 2009.

[5] Schulman, et al. Trust Region Policy Optimization. arXiv:1502.05477v5

[6] Dabney, et al. Implicit Quantile Networks for Distributional Reinforcement Learning arXiv:1806.06923v1