National Centre of Competence in Research Automation

ETH zürich

FNSNF Swiss National Science Foundation

Optimization in the Wasserstein Space

Nicolas Lanzetti, Saverio Bolognani, Florian Dörfler
Automatic Control Laboratory, ETH Zürich

Introduction

We are interested in rigorous methods for **optimization** in the space of **probability measures** for. . .

decision making in **uncertain** environments via **distributionally robust optimization**:

$$\min_{w\in\mathbb{R}^n} \sup_{\mu\in\mathcal{P}_2(\mathbb{R}^d)} \mathbb{E}^\mu[V(w,x)] := \int_{\mathbb{R}^d} V(w,x)\,\mathrm{d}\mu(x)$$

$$\text{s.t. } W_2(\mu,\hat{\mu}) \leq \varepsilon$$

"Engineer" ($\min$), Adversary "nature" ($\sup$), Benchmark distribution ($\hat{\mu}$), How far can nature go? (ε)

Preliminary Result: Multipliers

Consider the optimization problem of finding a probability measure μ^* minimizing a functional J subject to an equality constraint:

$$\min_{\mu\in\mathcal{P}_2(\mathbb{R}^d)} J(\mu) \quad \text{s.t. } G(\mu)=0$$

Assume that J and G are differentiable at μ^*. Then, there exists $\lambda\in\mathbb{R}$ so that

$$\nabla_\mu J(\mu^*) = \lambda\cdot\nabla_\mu G(\mu^*) \quad \mu^*-a.e.$$

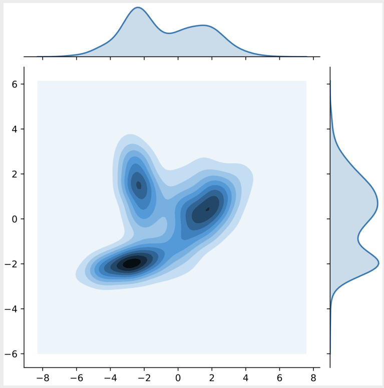
$\Leftrightarrow$

Wasserstein gradients are "aligned"

Methods

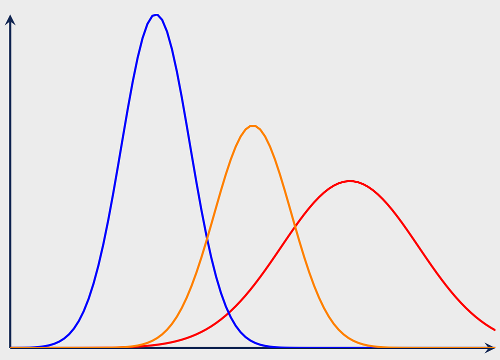
Wasserstein distance – How to compute distances?

$$W_2(\mu,\nu) := \left(\min_{\gamma\in\Gamma(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} \|x-y\|^2\,\mathrm{d}\gamma(x,y)\right)^{1/2}$$



Wasserstein geodesics – How to "move around"?

$$\mu_t = \left((1-t)\cdot\mathrm{Id} + t\cdot T_{\mu_0}^{\mu_1}\right)_{\#}\mu_0 \qquad t\in[0,1]$$



Wasserstein gradients – How to perturb?

$\nabla_\mu J(x)$ *Wasserstein* gradient of $J:\mathcal{P}(\mathbb{R}^d)\to\mathbb{R}$

$\Leftrightarrow$

$$J(\nu)-J(\mu) = \int_{\mathbb{R}^d} \nabla_\mu J(x)^\top (T_\mu^\nu(x)-x)\,\mathrm{d}\mu(x) + \text{higher order terms}$$

Application

DRO with nonlinear risk measure

Let $\rho>0$ and $w\in\mathbb{R}^d$. Consider the optimization problem

$$\max_{\mu\in\mathcal{P}_2(\mathbb{R}^d)} \mathbb{E}^\mu\left[w^\top x\right] + \rho\cdot\mathrm{Var}^\mu\left[w^\top x\right] \quad \text{s.t. } W_2(\mu,\hat{\mu})\leq\varepsilon.$$

Assume $\hat{\mu}$ and μ^* are absolutely continuous with respect to the Lebesgue measure. Then, the optimal μ^* results from the solution of a real fourth-order polynomial.

Conclusions

Methodology
We can rigorously reason with probability measures, and study optimality conditions.

Applications
We get new methods to study and efficiently solve DRO problems.
We can deploy similar tools to *control* of probability measures; e.g., how to steer a stochastic system from an initial distribution to a prescribed terminal distribution in minimal cost?

References

Lots of references! The "bible" is:

Ambrosio, L., Gigli, N., and Savaré, G. (2008). *Gradient flows: in metric spaces and in the space of probability measures.*

Please reach out if interested!

lnicolas@ethz.ch
http://people.ee.ethz.ch/lnicolas