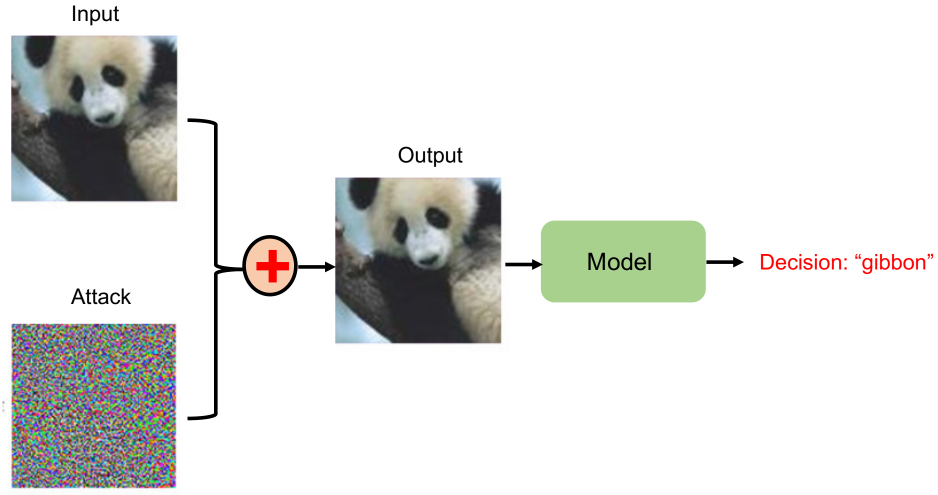
Entropic Regularization for Adversarial Robust Learning

Jie Wang
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Date: October 17, 2023

Joint work with Yifan Lin (Gatech), Song Wei (Gatech),
Rui Gao (UT Austin), and Yao Xie (Gatech)

On the Robustness of ML Models



[Goodfellow et al. 2015]

Adversarial Risk Minimization

Perturbation Constraints

$$\min_{\theta\in\Theta} \left\{ \mathbb{E}_{z\sim\mathbb{P}_n} \left[\sup_{\mathsf{d}(z,z')\le\rho} \ell(z';\theta) \right] \right\}.$$

Loss function

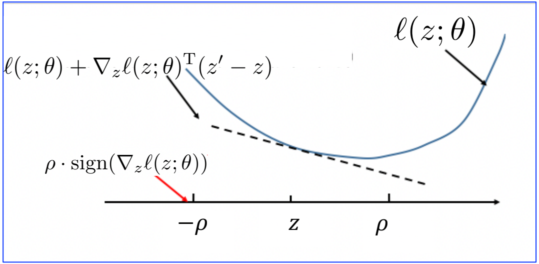
Data (e.g., feature-label pair)

Empirical Distribution

Baseline Approaches: Linearizing Objective Function

$$\min_{\theta\in\Theta} \left\{ \mathbb{E}_{z\sim\mathbb{P}_n}\left[\sup_{z':\ \mathsf{d}(z,z')\le\rho} \ell(z';\theta) \right] \right\} \qquad \text{(Ideal Formula)}$$

- **Fast Gradient Method (FGM)** [Goodfellow et al. 2015]



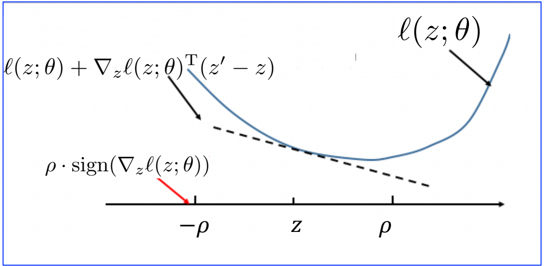
- $z' \approx \underset{\|z-z'\|_\infty\le\rho}{\arg\max} \left[\ell(z;\theta) + \nabla_z \ell(z;\theta)^{\mathrm{T}}(z'-z)\right]$

  $= z + \rho\cdot\mathrm{sign}(\nabla_z\ell(z;\theta))$

Baseline Approaches: Linearizing Objective Function

$$\min_{\theta\in\Theta} \left\{ \mathbb{E}_{z\sim\mathbb{P}_n}\left[\sup_{z':\ \mathsf{d}(z,z')\le\rho} \ell(z';\theta) \right] \right\} \qquad \text{(Ideal Formula)}$$

- **Iterative Fast Gradient Method (IFGM)** [Goodfellow et al. 2015]



$$z^0 = z$$
$$z^k = z + \alpha\cdot\operatorname{sign}(\nabla_z \ell(z^{k-1};\theta)),$$
$$k = 1,\dots,T-1,\ \alpha = \frac{\rho}{T}$$

Cons: Non-negligible optimization error when ρ is large!

Adversarial Risk Minimization

Perturbation Constraints

$$\min_{\theta\in\Theta}\left\{\mathbb{E}_{z\sim\mathbb{P}_n}\left[\sup_{\mathsf{d}(z,z')\le\rho}\ell(z';\theta)\right]\right\}.$$

Loss function

Data (e.g., feature-label pair)

Empirical Distribution

- Intractability issue:

$\ell(z;\theta)$ is convex in θ: Convex-Nonconcave Minimax Opt.

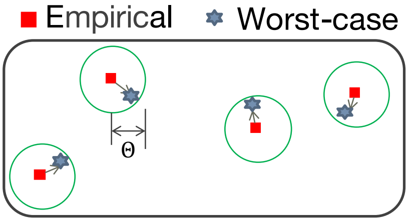
$\ell(z;\theta)$ is nonconvex in θ: Nonconvex-Nonconcave Minimax Opt.

- Connection with distributionally robust optimization:

$$\min_{\theta}\left\{\sup_{\mathbb{P}:\ \mathcal{W}_\infty(\mathbb{P},\mathbb{P}_n)\le\rho}\mathbb{E}_{z\sim\mathbb{P}}[\ell(z;\theta)]\right\}.$$

∞-type Wasserstein Distance

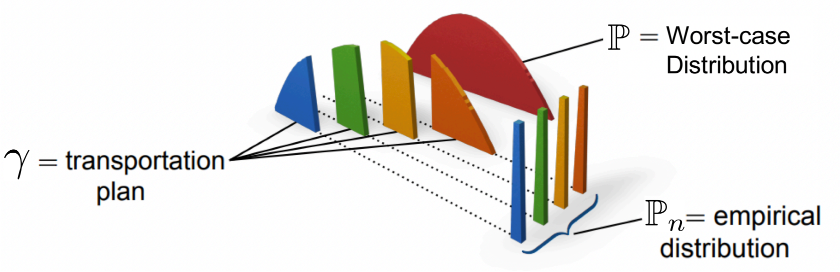
$$\mathcal{W}_\infty(\mathbb{P},\mathbb{Q}) = \inf_{\gamma} \left\{ \text{ess.sup}_{\gamma}\ \mathrm{d}(z_1, z_2) : \begin{array}{l} \gamma \text{ is a joint distribution of } z_1 \text{ and } z_2 \\ \text{with marginals } \mathbb{P} \text{ and } \mathbb{Q}, \text{ respectively} \end{array} \right\}.$$



Entropic Regularized Adversarial Robust Learning

- Original formulation:

$$\min_{\theta\in\Theta} \sup_{\mathbb{P},\gamma} \left\{ \mathbb{E}_{z\sim\mathbb{P}}[\ell(z;\theta)] : \begin{array}{l} \mathsf{Proj}_{1\#\gamma} = \mathbb{P}_n, \mathsf{Proj}_{2\#\gamma} = \mathbb{P} \\ \mathsf{ess.sup}_\gamma \mathsf{d}(z_1, z_2) \le \rho \end{array} \right\}.$$



Entropic Regularized Adversarial Robust Learning

- Original formulation:

$$\min_{\theta\in\Theta}\ \sup_{\mathbb{P},\gamma}\ \left\{\mathbb{E}_{z\sim\mathbb{P}}[\ell(z;\theta)]:\ \begin{array}{l}\mathsf{Proj}_{1\#\gamma}=\mathbb{P}_n,\mathsf{Proj}_{2\#\gamma}=\mathbb{P}\\ \mathsf{ess.sup}_{\gamma}\mathsf{d}(z_1,z_2)\le\rho\end{array}\right\}.$$

- Proposed formulation:

$$\min_{\theta\in\Theta}\ \sup_{\mathbb{P},\gamma}\ \left\{\mathbb{E}_{z\sim\mathbb{P}}[\ell(z;\theta)]-\eta\mathcal{H}(\gamma):\ \begin{array}{l}\mathsf{Proj}_{1\#\gamma}=\mathbb{P}_n,\mathsf{Proj}_{2\#\gamma}=\mathbb{P}\\ \mathsf{ess.sup}_{\gamma}\mathsf{d}(z_1,z_2)\le\rho\end{array}\right\}.$$

Entropic regularization:

$$\mathcal{H}(\gamma)\triangleq\mathbb{E}_{(z_1,z_2)\sim\gamma}\left[\log\left(\frac{\mathrm{d}\gamma(z_1,z_2)}{\mathrm{d}\gamma(z_1)\,\mathrm{d}z_2}\right)\right]=\mathbb{E}_{(z_1,z_2)\sim\gamma}\left[\log\left(\frac{\mathrm{d}\gamma(z_2\mid z_1)}{\mathrm{d}z_2}\right)\right].$$

Contribution (I): Tractable Reformulation

Under mild conditions, $V_{\mathsf{Primal}} = V_{\mathsf{Dual}}$:

$$V_{\mathsf{Primal}} = \sup_{\mathbb{P},\gamma} \left\{ \mathbb{E}_{z\sim\mathbb{P}}[\ell(z;\theta)] - \eta\mathcal{H}(\gamma) : \begin{array}{l} \mathsf{Proj}_{1\#\gamma} = \mathbb{P}_n, \mathsf{Proj}_{2\#\gamma} = \mathbb{P} \\ \mathsf{ess.sup}_{\gamma} \mathsf{d}(z_1, z_2) \leq \rho \end{array} \right\},$$

$$V_{\mathsf{Dual}} = \mathbb{E}_{x\sim\mathbb{P}_n}\left[\eta \log \mathbb{E}_{z\sim\mathbb{Q}_{x,\rho}}\left[\exp\left(\frac{\ell(z;\theta)}{\eta}\right)\right]\right].$$

Here $\mathbb{Q}_{x,\rho}$ is an uniform distribution supported on $\{z : \mathsf{d}(x,z) \leq \rho\}$.

Contribution (I): Tractable Reformulation

Under mild conditions, $V_{\text{Primal}} = V_{\text{Dual}}$:

$$V_{\text{Primal}} = \sup_{\mathbb{P},\gamma} \left\{ \mathbb{E}_{z\sim\mathbb{P}}[\ell(z;\theta)] - \eta\mathcal{H}(\gamma) : \begin{array}{l} \mathsf{Proj}_{1\#\gamma} = \mathbb{P}_n, \mathsf{Proj}_{2\#\gamma} = \mathbb{P} \\ \mathsf{ess.sup}_{\gamma}\mathsf{d}(z_1, z_2) \leq \rho \end{array} \right\},$$

$$V_{\text{Dual}} = \mathbb{E}_{x\sim\mathbb{P}_n}\left[\eta \log \mathbb{E}_{z\sim\mathbb{Q}_{x,\rho}}\left[\exp\left(\frac{\ell(z;\theta)}{\eta}\right)\right]\right].$$

Here $\mathbb{Q}_{x,\rho}$ is an uniform distribution supported on $\{z : \mathsf{d}(x,z) \leq \rho\}$.

Strong dual for **un-regularized** case $(\eta = 0)$ [Gao et. al, 2022]:

$$V_{\text{Primal}} = \sup_{\mathbb{P},\gamma} \left\{ \mathbb{E}_{z\sim\mathbb{P}}[\ell(z;\theta)] : \begin{array}{l} \mathsf{Proj}_{1\#\gamma} = \mathbb{P}_n, \mathsf{Proj}_{2\#\gamma} = \mathbb{P} \\ \mathsf{ess.sup}_{\gamma}\mathsf{d}(z_1, z_2) \leq \rho \end{array} \right\},$$

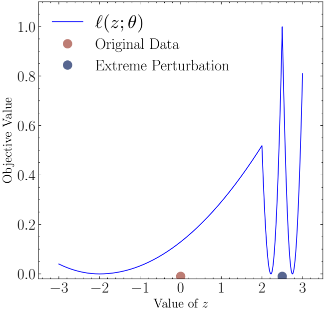
$$V_{\text{Dual}} = \mathbb{E}_{x\sim\mathbb{P}_n}\left[\sup_{z:\ \mathsf{d}(x,z)\leq\rho} \ell(z;\theta)\right].$$

Recovery of Worst-case Distribution

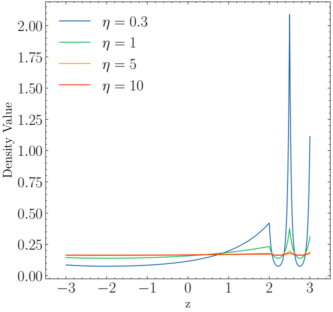
$$(\mathbb{P}^*, \gamma^*) = \arg\max_{\mathbb{P},\gamma} \left\{ \mathbb{E}_{z\sim\mathbb{P}}[\ell(z;\theta)] - \eta\mathcal{H}(\gamma) : \begin{array}{l} \mathsf{Proj}_{1\#\gamma} = \mathbb{P}_n, \mathsf{Proj}_{2\#\gamma} = \mathbb{P} \\ \text{ess.sup}_{\gamma}\mathsf{d}(z_1, z_2) \le \rho \end{array} \right\}$$

$$\frac{\mathrm{d}\mathbb{P}^*(z)}{\mathrm{d}z} = \mathbb{E}_{x\sim\mathbb{P}_n}\left[\alpha_x \cdot \exp\left(\frac{\ell(z;\theta)}{\eta}\right) \mathbf{1}\{\mathsf{d}(x,z) \le \rho\}\right]$$

Worst-case distribution for regularization case is absolutely continuous!



Visualization of $\ell(z;\theta)$



Visualization of $\mathbb{P}^*$

Contribution (II): Tractable Algorithm with Convergence Guarantees

- Dual formulation:

$$\min_{\theta\in\Theta} \left\{ F(\theta) \triangleq \mathbb{E}_{x\sim\mathbb{P}_n}\left[\eta \log \mathbb{E}_{z\sim\mathbb{Q}_{x,\rho}}\left[\exp\left(\frac{\ell(z;\theta)}{\eta}\right)\right]\right]\right\}.$$

  Here $\mathbb{Q}_{x,\rho}$ is a uniform distribution supported on the ρ-radius ball of x.

- Solve the Monte-Carlo approximated formulation [Shapiro et. al 2014]:

$$\min_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \eta \log\left(\frac{1}{m}\sum_{j=1}^{m}\exp\left(\frac{\ell(z_{i,j};\theta)}{\eta}\right)\right),$$

  where $\{\hat{x}_i\}_{i=1}^n \sim \mathbb{P}_n$ and $\{z_{i,j}\}_{j=1}^m$ are i.i.d. samples generated from $\mathbb{Q}_{\hat{x}_i,\rho}$.

- **Cons:** It requires $\tilde{O}(\delta^{-3})$ samples to obtain δ-optimal solution [Yifan et. al SIAMOP2020].

Contribution (II): Tractable Algorithm with Convergence Guarantees

- Dual formulation:

$$\min_{\theta\in\Theta} \left\{ F(\theta) \triangleq \mathbb{E}_{x\sim\mathbb{P}_n}\left[\eta \log \mathbb{E}_{z\sim\mathbb{Q}_{x,\rho}}\left[\exp\left(\frac{\ell(z;\theta)}{\eta}\right)\right]\right]\right\}.$$

Here $\mathbb{Q}_{x,\rho}$ is a uniform distribution supported on the ρ-radius ball of x.

Algorithm 1 Stochastic Mirror Descent with Biased Gradient Oracles

Require: maximum iterations T, constant step size γ, initial guess θ_0.

for $t = 0, 1, \ldots, T-1$ **do**
  Formulate (biased) gradient estimate of $F(\theta_t)$, denoted as $v(\theta_t)$.
  Update $\theta_{t+1} = \mathrm{Prox}_{\theta_t}\big(\gamma v(\theta_t)\big)$.
end for
Output $\tilde{\theta}$ **randomly selected from** $\{\theta_0, \theta_1, \ldots, \theta_T\}$.

Contribution (II): Tractable Algorithm with Convergence Guarantees

- Dual formulation:

$$\min_{\theta\in\Theta} \left\{ F(\theta) \triangleq \mathbb{E}_{x\sim\mathbb{P}_n}\left[\eta \log \mathbb{E}_{z\sim\mathbb{Q}_{x,\rho}}\left[\exp\left(\frac{\ell(z;\theta)}{\eta}\right)\right]\right]\right\}.$$

  Here $\mathbb{Q}_{x,\rho}$ is a uniform distribution supported on the ρ-radius ball of x.

Scenarios	Computation Cost	Memory Cost
Nonsmooth Convex Optimization	$\tilde{O}(\epsilon^{-2})$	$\tilde{O}(1)$
Constrained Smooth Nonconvex Optimization	$\tilde{O}(\epsilon^{-4})$	$\tilde{O}(\epsilon^{-2})$
Unconstrained Nonconvex Optimization	$\tilde{O}(\epsilon^{-4})$	$\tilde{O}(1)$

Bias-(2nd)Moment-Cost Trade-off for SMD

- Consider convex optimization problem

$$\begin{array}{ll} \text{Minimize} & F(\theta) \\ \text{s.t.} & \theta \in \Theta \subseteq \mathbb{R}^d. \end{array}$$

- Stochastic Mirror Descent: iteratively,
  - Step 1: generate random vector $v(\theta_t)$ with

$$\mathbb{E}[v(\theta_t)] = \nabla \overline{F}(\theta_t), \quad \Delta_F := \sup_{\theta \in \Theta} |\overline{F}(\theta) - F(\theta)|, \quad \mathbb{E}\left[\|v(\theta_t)\|^2\right] \le M^2.$$

  - Step 2: $\theta_{t+1} = \textbf{Proximal}_{\theta_t}\big(\gamma v(\theta_t)\big)$.
- Take $\widehat{\theta}_{1:T}$ as average over $\{\theta_t\}_{t=1}^T$, then

$$\mathbb{E}\big[F(\widehat{\theta}_{1:T}) - F(\theta^*)\big] \le c \cdot \left(\Delta_F + \sqrt{\frac{M^2}{T}}\right).$$

Gradient Estimators

$$\min_{\theta\in\Theta} \left\{ F(\theta) \triangleq \mathbb{E}_{x\sim\mathbb{P}_n}\left[\eta \log \mathbb{E}_{z\sim\mathbb{Q}_{x,\rho}}\left[\exp\left(\frac{\ell(z;\theta)}{\eta}\right)\right]\right]\right\}.$$

- Approximation objective with error $O(2^{-L})$:

$$F^L(\theta) = \mathbb{E}_{x^L\sim\mathbb{P}_n}\mathbb{E}_{\{z_j^L\}_j\sim\mathbb{Q}_{x,\rho}}\left[\eta\log\left(\frac{1}{2^L}\sum_j \exp\left(\frac{\ell(z_j^\ell;\theta)}{\eta}\right)\right)\right]$$

- Generating **unbiased** gradient estimator of $F^L(\theta)$ with **low variance, low computational cost** is easy!

Vanilla SGD Estimator

Sample $x^L \sim \mathbb{P}_n$ and next sample $\{z_j^L\}_{j\in[2^L]} \sim \mathbb{Q}_{x^L,\rho}$. Construct

$$v^L(\theta) = \nabla_\theta \left\{ \eta \log\left(\frac{1}{2^L} \sum_j \exp\left(\frac{\ell(z_j^\ell;\theta)}{\eta} \right) \right) \right\}.$$

Pros	**Cons**
Low Bias $\Delta_F = O(2^{-L})$	Generating single gradient has cost $O(2^L)$
Bounded Moment $M^2 = O(1)$	

Overall: Sample complexity to get δ-optimal solution is $\mathcal{O}(\delta^{-3})$.

Algorithm Improvement: Multi-level Monte Carlo Sampling

- Directly computing $v^L(\theta)$ for large L seems expensive;
- Define $v^{-1}(\theta) \equiv 0$ and rewrite

$$
\begin{aligned}
v^L(\theta) &= \sum_{\ell=0}^{L} \left[v^\ell(\theta) - v^{\ell-1}(\theta) \right] \\
&= \sum_{\ell=0}^{L} p_\ell \cdot \frac{v^\ell(\theta) - v^{\ell-1}(\theta)}{p_\ell} = \mathbb{E}_{\ell \sim \{p_\ell\}_{\ell=0}^{L}} \left[\frac{v^\ell(\theta) - v^{\ell-1}(\theta)}{p_\ell} \right]
\end{aligned}
$$

- **Randomized Sampling Gradient Estimator**: sample ℓ from truncated geometric distribution $\{p_\ell\}_{\ell=0}^{L}$ with $p_\ell \propto 2^{-\ell}$. Construct

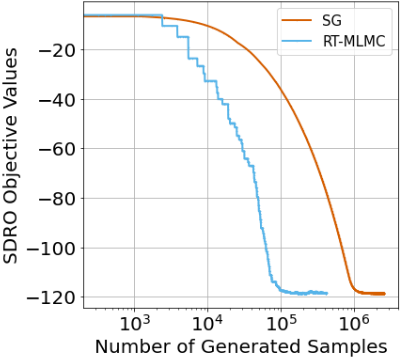
$$
v^{\mathsf{RT\text{-}MLMC}}(\theta) = \frac{1}{p_\ell} \cdot \left[v^\ell(\theta) - v^{\ell-1}(\theta) \right].
$$

Bias, 2nd Moment, and Costs

- **Bias**: For same level L, the bias of RT-MLMC/Vanilla SGD are same.
- **2nd Moment**: $v^\ell(\theta) - v^{\ell-1}(\theta) \to 0$ for large ℓ:

$$\mathbb{E}\left[\|v^{\text{RT-MLMC}}(\theta_t)\|^2\right] = \mathcal{O}(L) = \tilde{\mathcal{O}}(1).$$

- **Sampling Cost**: Cost for generating RT-MLMC estimator reduces from $\mathcal{O}(2^L)$ to $\mathcal{O}(L)$!



The sample complexity of RT-MLMC is $\tilde{O}(\delta^{-2})$ with storage cost $\tilde{O}(1)$.

Contribution (III): Regularization Effects

Goal: connects with regularized empirical risk minimization:

$$\min_{\theta\in\Theta}\left\{F(\theta)\triangleq\mathbb{E}_{x\sim\mathbb{P}_n}\left[\eta\log\mathbb{E}_{z\sim\mathbb{Q}_{x,\rho}}\left[\exp\left(\frac{\ell(z;\theta)}{\eta}\right)\right]\right]\right\}\approx\min_{\theta\in\Theta}\left\{\mathbb{E}_{z\sim\mathbb{P}_n}[\ell(z;\theta)]+\mathcal{V}(\theta)\right\}.$$

- When $\rho/\eta\to\infty$: Ent-Training $\approx\min_{\theta\in\Theta}\left\{\mathbb{E}_{\mathbb{P}_n}[\ell(z;\theta)]+\rho\mathbb{E}_{\mathbb{P}_n}[\|\nabla\ell(x;\theta)\|_*]\right\}$.

  Adversarial risk minimization also corresponds to gradient norm regularization [Gao et.al 2022]:

  $$\min_{\theta}\left\{\mathbb{E}_{x\sim\mathbb{P}_n}\left[\sup_{z:\ \mathsf{d}(x,z)\le\rho}\ell(z;\theta)\right]\right\}\approx\min_{\theta}\left\{\mathbb{E}_{z\sim\mathbb{P}_n}[\ell(z;\theta)]+\rho\mathbb{E}_{\mathbb{P}_n}[\|\nabla\ell(x;\theta)\|_*]\right\}.$$

- When $\rho/\eta\to 0$: Ent-Training $\approx\min_{\theta\in\Theta}\left\{\mathbb{E}_{x\sim\mathbb{P}_n}[\ell(z;\theta)]+\frac{\rho^2}{\eta}\mathbb{E}_{x\sim\mathbb{P}_n}[\mathrm{Var}_{\mathbb{Q}_{x,\rho}}(\nabla\ell(x;\theta)^{\mathrm{T}}z)]\right\}$.
- When $\rho/\eta\to C$:

Contribution (III): Regularization Effects

Goal: connects with regularized empirical risk minimization:

$$\min_{\theta\in\Theta}\left\{F(\theta)\triangleq\mathbb{E}_{x\sim\mathbb{P}_n}\left[\eta\log\mathbb{E}_{z\sim\mathbb{Q}_{x,\rho}}\left[\exp\left(\frac{\ell(z;\theta)}{\eta}\right)\right]\right]\right\}\approx\min_{\theta\in\Theta}\left\{\mathbb{E}_{z\sim\mathbb{P}_n}[\ell(z;\theta)]+\mathcal{V}(\theta)\right\}.$$

- When $\rho/\eta\to\infty$: Ent-Training $\approx\min_{\theta\in\Theta}\left\{\mathbb{E}_{\mathbb{P}_n}[\ell(z;\theta)]+\rho\mathbb{E}_{\mathbb{P}_n}[\|\nabla\ell(x;\theta)\|_*]\right\}$.
- When $\rho/\eta\to 0$: Ent-Training $\approx\min_{\theta\in\Theta}\left\{\mathbb{E}_{x\sim\mathbb{P}_n}[\ell(z;\theta)]+\frac{\rho^2}{\eta}\mathbb{E}_{x\sim\mathbb{P}_n}[\mathrm{Var}_{\mathbb{Q}_{x,\rho}}(\nabla\ell(x;\theta)^{\mathrm{T}}z)]\right\}$.
- When $\rho/\eta\to C$:

$$\text{Ent-Training}\approx\min_{\theta\in\Theta}\left\{\mathbb{E}_{x\sim\mathbb{P}_n}[\ell(z;\theta)]+\frac{\rho}{C}\mathbb{E}_{x\sim\mathbb{P}_n}[\log\mathbb{E}_{\mathbb{Q}_{x,\rho}}[\exp(C\nabla\ell(x;\theta)^{\mathrm{T}}z)]]\right\}.$$

Contribution (III): Regularization Effects

Goal: connects with regularized empirical risk minimization:

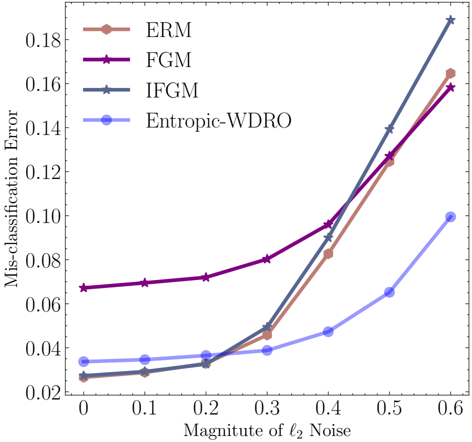
$$\min_{\theta\in\Theta} \left\{F(\theta) \triangleq \mathbb{E}_{x\sim\mathbb{P}_n}\left[\eta \log \mathbb{E}_{z\sim\mathbb{Q}_{x,\rho}}\left[\exp\left(\frac{\ell(z;\theta)}{\eta}\right)\right]\right]\right\} \approx \min_{\theta\in\Theta} \left\{\mathbb{E}_{z\sim\mathbb{P}_n}[\ell(z;\theta)] + \mathcal{V}(\theta)\right\}.$$

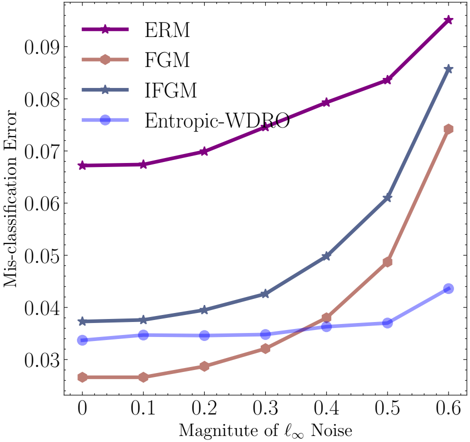
- When $\rho/\eta \to \infty$: Ent-Training $\approx \min_{\theta\in\Theta} \left\{\mathbb{E}_{\mathbb{P}_n}[\ell(z;\theta)] + \rho\mathbb{E}_{\mathbb{P}_n}[\|\nabla\ell(x;\theta)\|_*]\right\}$.
- When $\rho/\eta \to 0$: Ent-Training $\approx \min_{\theta\in\Theta} \left\{\mathbb{E}_{x\sim\mathbb{P}_n}[\ell(z;\theta)] + \frac{\rho^2}{\eta}\mathbb{E}_{x\sim\mathbb{P}_n}[\mathrm{Var}_{\mathbb{Q}_{x,\rho}}(\nabla\ell(x;\theta)^{\mathrm{T}}z)]\right\}$.
- When $\rho/\eta \to C$:

$$\text{Ent-Training} \approx \min_{\theta\in\Theta} \left\{\mathbb{E}_{x\sim\mathbb{P}_n}[\ell(z;\theta)] + \frac{\rho}{C}\mathbb{E}_{x\sim\mathbb{P}_n}[\log \mathbb{E}_{\mathbb{Q}_{x,\rho}}[\exp(C\nabla\ell(x;\theta)^{\mathrm{T}}z)]]\right\}.$$

Numerical Study: MNIST Classification

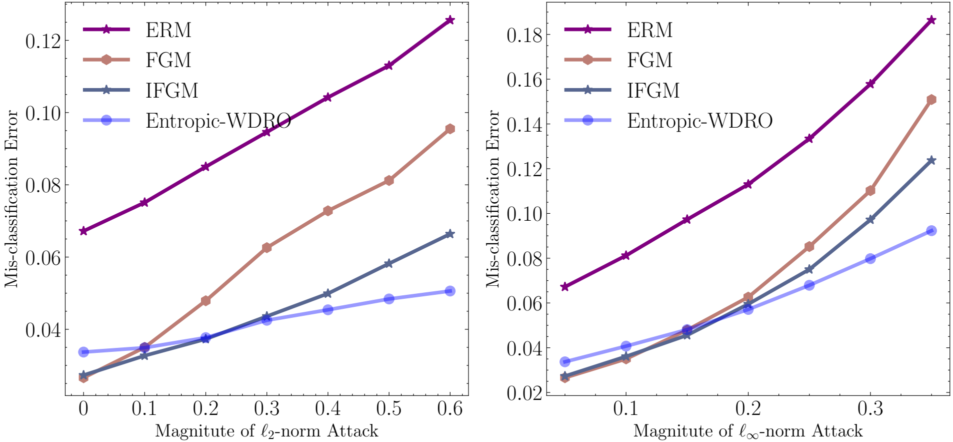
- **Goal**: Classification with $8 \times 8, 6 \times 6$ convolutional neural networks with ELU activation.
- **Training data**: MNIST handwritten digits with $6 \cdot 10^4$ samples;
- **Testing data**: digits with 10^4 samples, perturbed by random ℓ_∞/ℓ_2-norm noise.



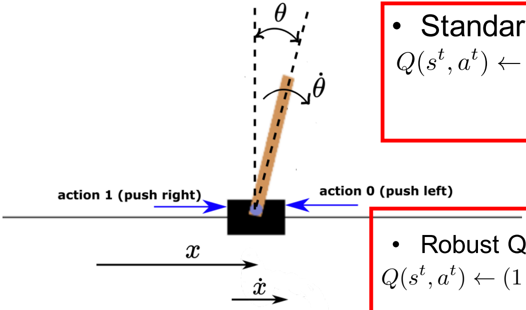


Numerical Study: MNIST Classification

- **Goal**: Classification with $8 \times 8, 6 \times 6$ convolutional neural networks with ELU activation.
- **Training data**: MNIST handwritten digits with $6 \cdot 10^4$ samples;
- **Testing data**: digits with 10^4 samples, perturbed by ℓ_∞/ℓ_2-norm attack.



Numerical Study: Reliable Reinforcment Learning



- Standard Q-learning:

$$Q(s^t, a^t) \leftarrow (1-\alpha_t)Q(s^t, a^t) + \alpha_t r(s^t, a^t) - \gamma\alpha_t \min_a(-Q(s^{t+1}, a)), \quad s^{t+1} \sim \mathbb{P}(\cdot \mid s^t, a^t).$$

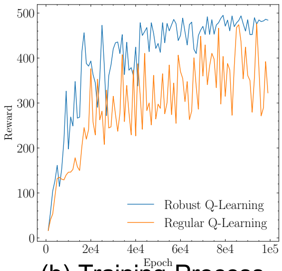
- Robust Q-learning with entropy:

$$Q(s^t, a^t) \leftarrow (1-\alpha_t)Q(s^t, a^t) + \alpha_t r(s^t, a^t) - \gamma\alpha_t \min_a \left\{ \eta \log \mathbb{E}_{s\sim\mathbb{Q}(s^{t+1};\rho)} e^{-Q(s,a)/\eta} \right\}, \quad s^{t+1} \sim \mathbb{P}(\cdot \mid s^t, a^t)$$

Numerical Study: Reliable Reinforcment Learning

Environment	Regular	Robust
Original MDP	469.42 ± 19.03	$\mathbf{487.11 \pm 9.09}$
Perturbed MDP (Heavy)	187.63 ± 29.40	$\mathbf{394.12 \pm 12.01}$
Perturbed MDP (Short)	355.54 ± 28.89	$\mathbf{443.17 \pm 9.98}$
Perturbed MDP (Strong g)	271.41 ± 20.7	$\mathbf{418.42 \pm 13.64}$

(a) Reward by Regular Q-learning v.s. Robust Q-learning



(b) Training Process

Contributions

- **Adding entropic regularization for adversarial risk minimization**

  Enables continuous worst-case distribution.

- **Computationally efficient algorithm with performance guarantees**

  Multi-level Monte-Carlo estimator

  Near-optimal computational complexity

- **Regularization effects**

  Our framework interpolates gradient norm regularization and variance regularization

QR code for full manuscript