DISSECTING NON-VACUOUS GENERALIZATION BOUNDS BASED ON THE MEAN-FIELD APPROXIMATION

Pitas Konstantinos,
LTS2, EPFL,
Switzerland.
konstantinos.pitas@epfl.ch

1)OVERVIEW
GENERALIZATION BOUNDS AND PAC-BAYES

$$\mathcal{L}(\rho) \leq \hat{\mathcal{L}}(\rho) + \text{complexity}(\rho)$$

Risk   Empirical Risk

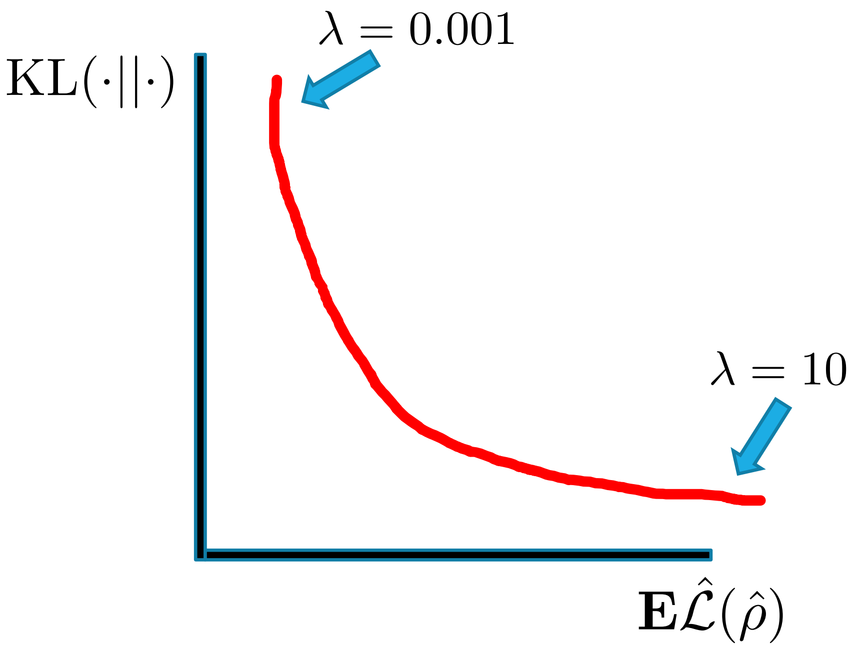
$$\mathbf{E}\mathcal{L}(\hat{\rho}) \leq \mathbf{E}\hat{\mathcal{L}}(\hat{\rho}) + \beta \mathrm{KL}(\hat{\rho}||\pi)$$

PAC-BAYES BOUNDS

$$\mathbf{E}\mathcal{L}(\hat{\rho}) \leq \mathbf{E}\hat{\mathcal{L}}(\hat{\rho}) + \beta \mathrm{KL}(\hat{\rho}||\pi)$$

$$\hat{\rho}(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\hat{\rho}}, \lambda \mathbf{I})$$

$$\pi(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\text{init}}, \lambda \mathbf{I})$$



MODELING CHOICES

$$\mathbf{E}\mathcal{L}(\hat{\rho}) \leq \mathbf{E}\hat{\mathcal{L}}(\hat{\rho}) + \beta \mathrm{KL}(\hat{\rho}||\pi)$$

Invalid
Sanity Check

$$\hat{\rho}(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\hat{\rho}}, \lambda \mathbf{I})$$
$$\pi(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\text{init}}, \lambda \mathbf{I})$$

Baseline

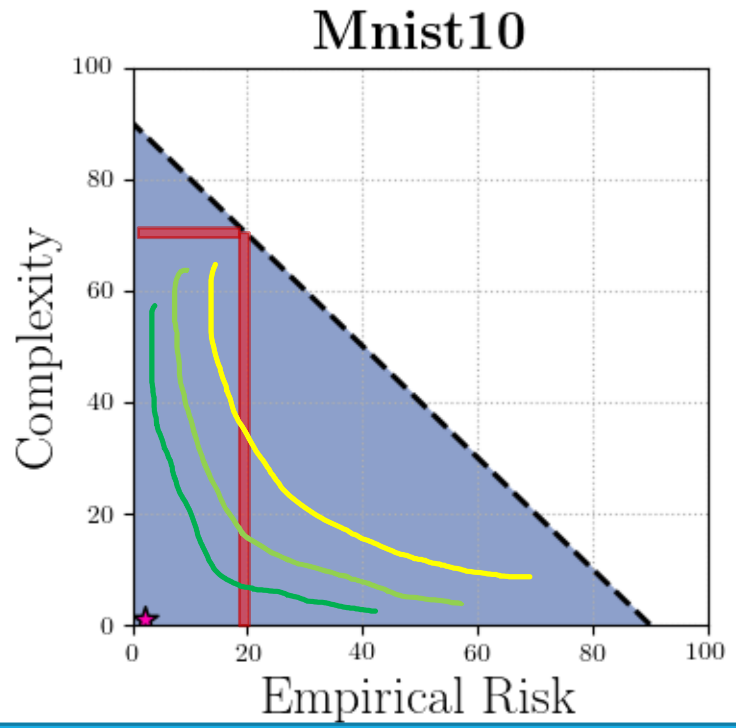
$$\hat{\rho}(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\hat{\rho}}, \boldsymbol{\sigma}_{\hat{\rho}})$$
$$\pi(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\text{init}}, \lambda \mathbf{I})$$

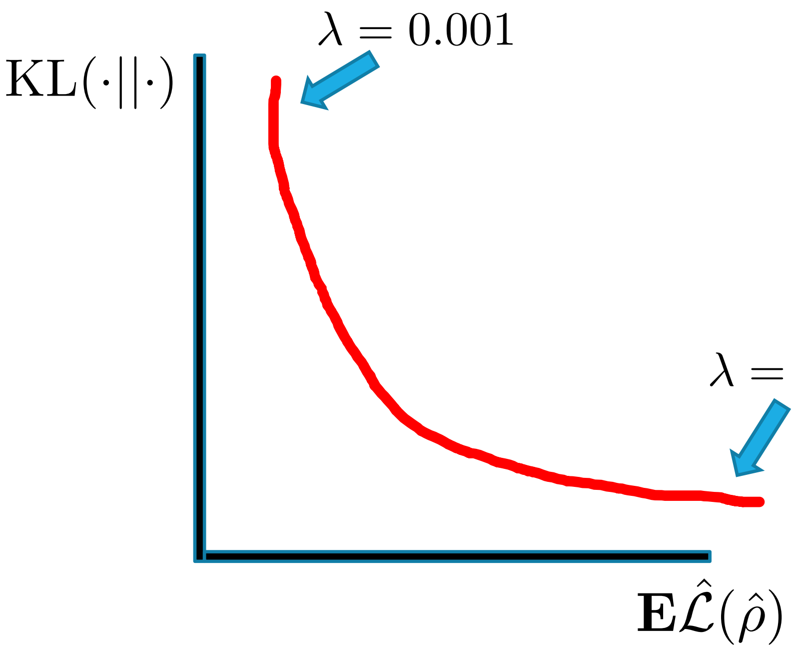
VI

$$\hat{\rho}(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\hat{\rho}}, \boldsymbol{\sigma}_{\hat{\rho}})$$
$$\pi(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\text{init}}, \lambda \boldsymbol{\sigma}_{\pi})$$

RISK-COMPLEXITY PLOTS: A MORE INTUITIVE WAY OF COMPARING BOUNDS

$40\% < 60\%$?





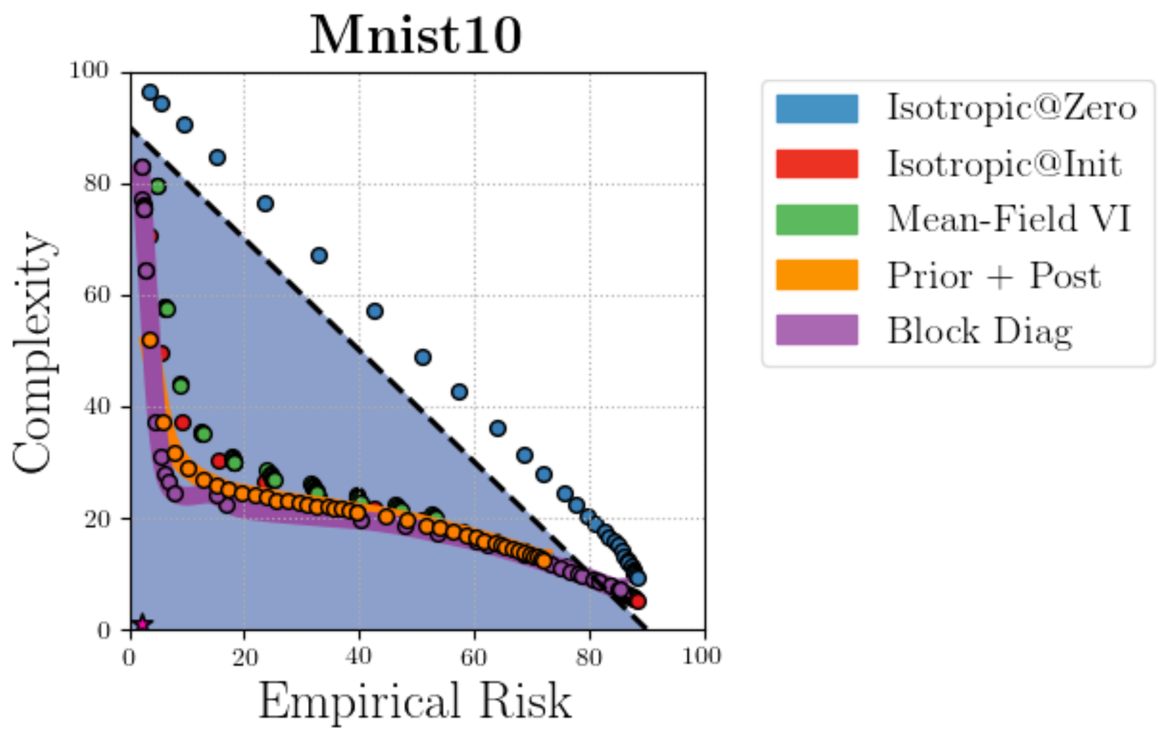
$$\mathbf{E}\mathcal{L}(\hat{\rho}) \leq \mathbf{E}\hat{\mathcal{L}}(\hat{\rho}) + \beta \mathrm{KL}(\hat{\rho}||\pi)$$

MODELING CHOICES

Choosing the prior mean to be the random DNN initialization is very important!

OVERVIEW OF RESULTS

- Baseline is non-vacuous in simple cases.
- Mean-Field VI of the **posterior covariance** yields marginal improvement (Problems with optimization?)
- Closed form optimization:
  - Optimizing the posterior diagonal covariance.
  - Optimizing the prior diagonal covariance.
  - Generalizing the posterior covariance to be **block diagonal**.



2)DETAILS MEAN-FIELD VI

$$\mathbf{E}_{\boldsymbol{\theta}\sim\hat{\rho}(\boldsymbol{\theta})}\hat{\mathcal{L}}^{\ell_{01}}_{X,Y}(f_{\boldsymbol{\theta}}) + \frac{1}{\beta n}(\mathrm{KL}(\hat{\rho}(\boldsymbol{\theta})||\mathcal{N}(\boldsymbol{\mu}_{\mathrm{init}}, \lambda\mathbf{I})) + \ln\frac{1}{\delta})$$

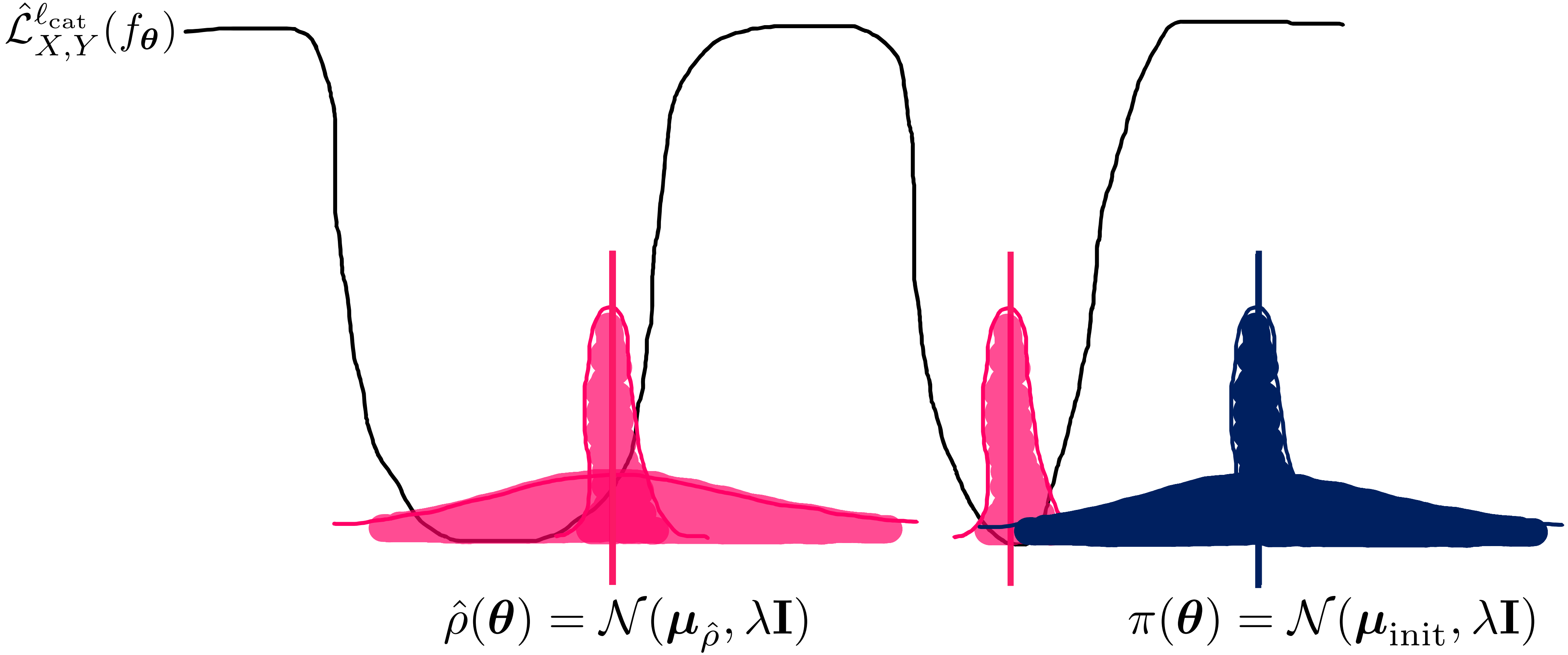
$$\mathbf{E}_{\boldsymbol{\theta}\sim\hat{\rho}(\boldsymbol{\theta})}\hat{\mathcal{L}}^{\ell_{\mathrm{cat}}}_{X,Y}(f_{\boldsymbol{\theta}})$$

$$\sum_{i=0}^{T}\hat{\mathcal{L}}^{\ell_{\mathrm{cat}}}_{X,Y}(f_{\boldsymbol{\theta}_i})$$

$$\boldsymbol{\theta} = \boldsymbol{\mu}_{\hat{\rho}} + \sqrt{\boldsymbol{\sigma}_{\hat{\rho}}} \odot \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$\boldsymbol{\mu}_{\hat{\rho}}$ remains fixed

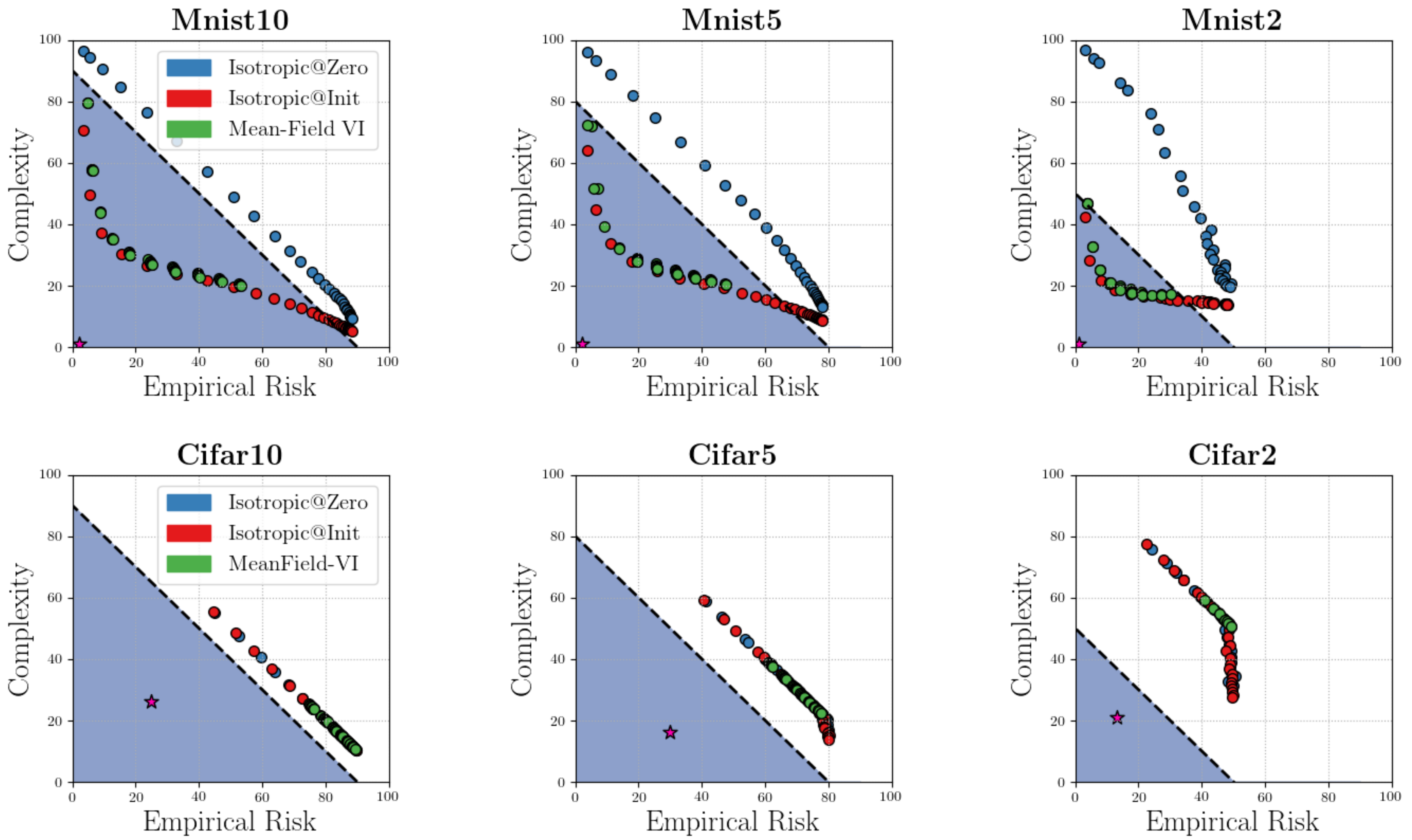
MEAN-FIELD VI WITH FIXED POSTERIOR MEAN



MEAN-FIELD VI GRID SEARCH

$$\min_{\sigma_{\hat{\rho}}} \overbrace{\sum_{i=0}^{T} \hat{\mathcal{L}}^{\ell_{\text{cat}}}_{X,Y}(f_{\boldsymbol{\theta}_i})} + \frac{1}{\beta n} \overbrace{(\text{KL}(\hat{\rho}(\boldsymbol{\theta}) || \mathcal{N}(\boldsymbol{\mu}_{\text{init}}, \lambda \mathbf{I})) + \ln \frac{1}{\delta})}$$

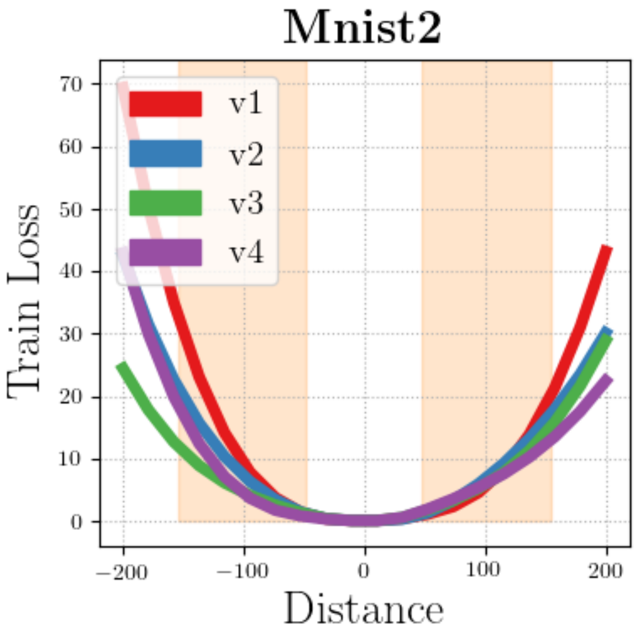
MEAN-FIELD VI RESULTS

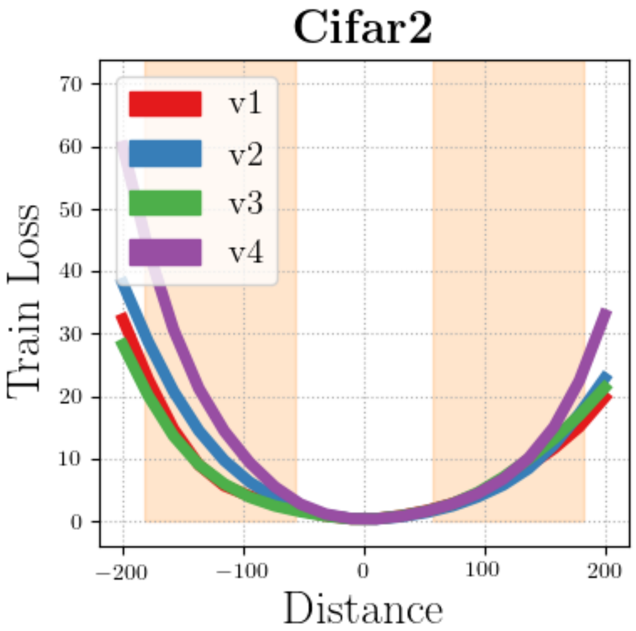


QUADRATIC APPROXIMATION

$$C_\beta(X,Y;\hat{\rho},\pi) = \mathbf{E}_{\boldsymbol{\theta}\sim\hat{\rho}(\boldsymbol{\theta})}\hat{\mathcal{L}}^{\ell_{\text{cat}}}_{X,Y}(f_{\boldsymbol{\theta}}) + \beta\text{KL}(\hat{\rho}(\boldsymbol{\theta})||\pi(\boldsymbol{\theta}))$$

$$\approx \mathbf{E}_{\boldsymbol{\eta}\sim\hat{\rho}'(\boldsymbol{\theta})}[\frac{1}{2}\boldsymbol{\eta}^T\nabla^2\hat{\mathcal{L}}^{\ell_{\text{cat}}}_{X,Y}(f_{\boldsymbol{\theta}})\boldsymbol{\eta}] + \beta\text{KL}(\hat{\rho}(\boldsymbol{\theta})||\pi(\boldsymbol{\theta}))$$





QUADRATIC APPROXIMATION

Lemma 4.1. *The convex optimization problem* $\min_{\mathbf{\Sigma}_{\hat{\rho}}} \mathbf{E}_{\boldsymbol{\eta}\sim\hat{\rho}'(\boldsymbol{\theta})}[\frac{1}{2}\boldsymbol{\eta}^T\mathbf{H}\boldsymbol{\eta}] + \beta\mathrm{KL}(\hat{\rho}(\boldsymbol{\theta})||\pi(\boldsymbol{\theta}))$ *where* $\hat{\rho}(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\hat{\rho}}, \mathbf{\Sigma}_{\hat{\rho}})$ *and* $\pi(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\pi}, \lambda\mathbf{\Sigma}_{\pi})$ *is minimized at*

$$\mathbf{\Sigma}^*_{\hat{\rho}} = \beta(\mathbf{H} + \frac{\beta}{\lambda}\mathbf{\Sigma}_{\pi}^{-1})^{-1}, \quad (6)$$

where $\mathbf{H} \equiv \nabla^2\hat{\mathcal{L}}^{\ell_{cat}}_{X,Y}(f_{\boldsymbol{\theta}})$ *captures the curvature at the minimum, while* $\mathbf{\Sigma}_{\pi}$ *is the prior covariance.*

Posterior

Lemma 4.2. *The optimal prior and posterior covariances for* $\min_{\boldsymbol{\sigma}_{\hat{\rho}},\boldsymbol{\sigma}_{\pi}} \mathbf{E}_{\boldsymbol{\eta}\sim\hat{\rho}'(\boldsymbol{\theta})}[\frac{1}{2}\boldsymbol{\eta}^T\mathbf{H}\boldsymbol{\eta}] + \beta\mathrm{KL}(\hat{\rho}(\boldsymbol{\theta})||\pi(\boldsymbol{\theta}))$ *with* $\hat{\rho}(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\hat{\rho}}, \boldsymbol{\sigma}_{\hat{\rho}})$ *and* $\pi(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\pi}, \lambda\boldsymbol{\sigma}_{\pi})$ *have elements*

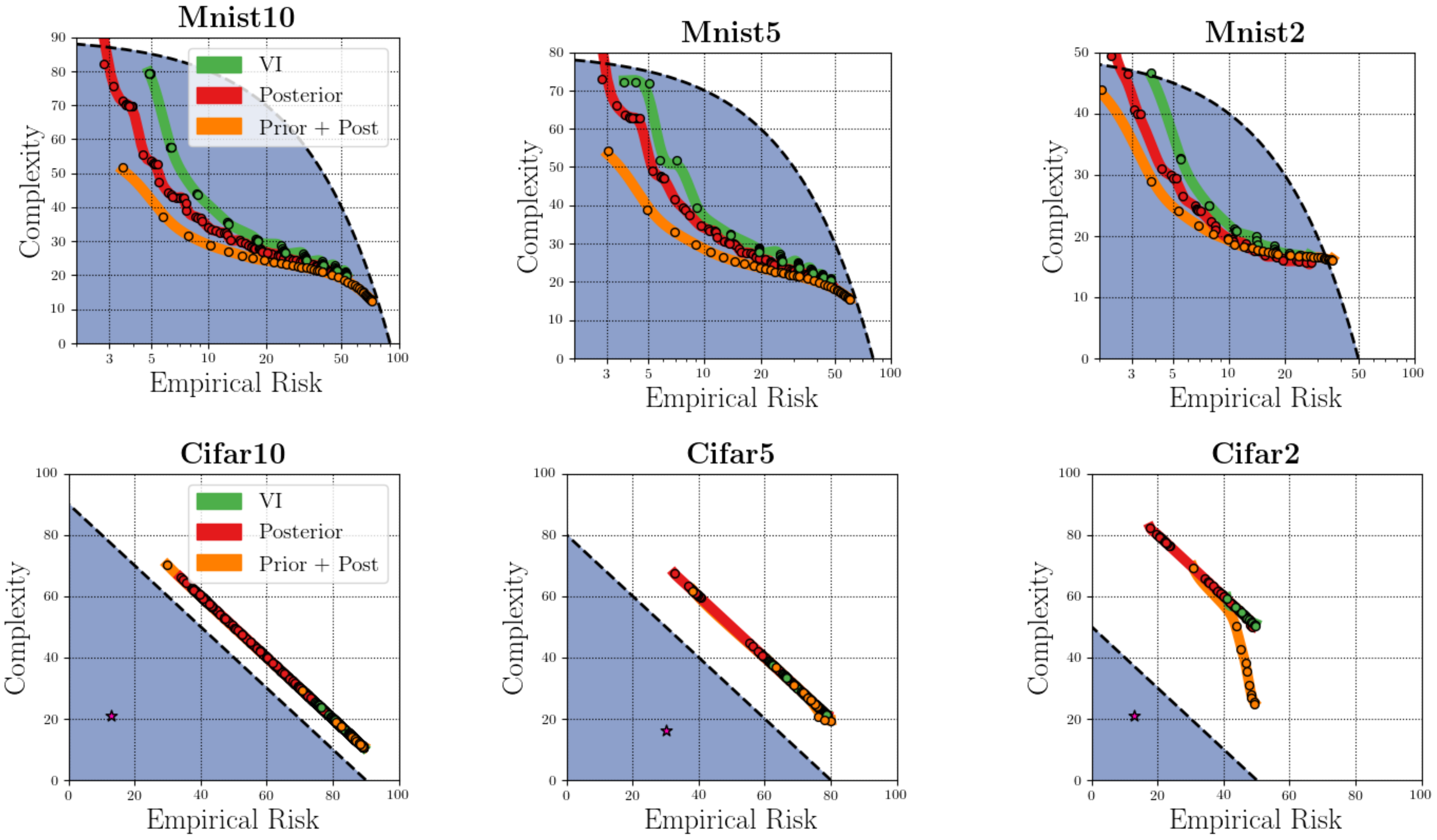
$$(\sigma^*_{\hat{\rho}i})^{-1} = \frac{1}{2\beta}[h_i + \sqrt{h_i^2 + \frac{4\beta h_i}{(\mu_{i\hat{\rho}} - \mu_{i\pi})^2}}], \quad (7)$$

$$(\sigma^*_{\pi i})^{-1} = \frac{\lambda}{2\beta}[\sqrt{h_i^2 + \frac{4\beta h_i}{(\mu_{i\hat{\rho}} - \mu_{i\pi})^2}} - h_i], \quad (8)$$

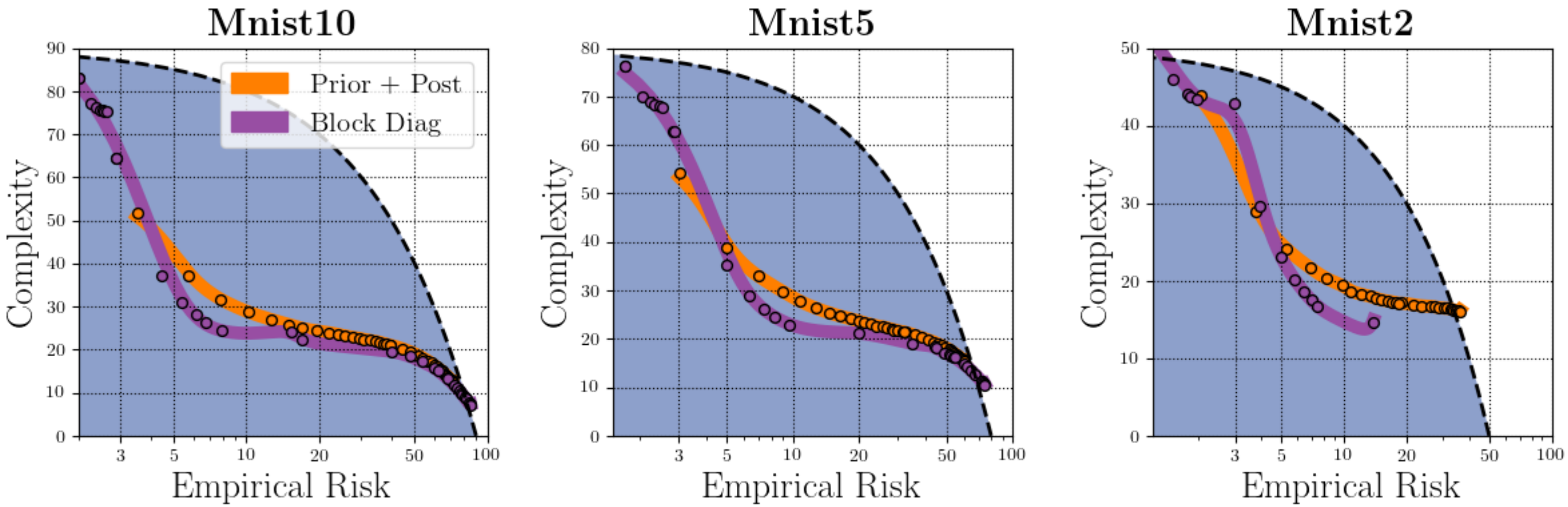
where $\mathbf{H} \equiv \nabla^2\hat{\mathcal{L}}^{\ell_{cat}}_{X,Y}(f_{\boldsymbol{\theta}})$ *captures the curvature at the minimum.*

Prior + Posterior

QUADRATIC APPROXIMATION RESULTS



BLOCK DIAGONAL RESULTS



THANK YOU FOR YOUR ATTENTION!

Arxiv preprint: https://arxiv.org/pdf/1909.03009.pdf

contact: konstantinos.pitas@epfl.ch

LTS2, EPFL, Switzerland