WiML SYMPOSIUM · SPIGM @ ICML

McGill

Scalable Bayesian Deep Learning for Uncertainty Quantification

Mame Diarra Toure · David A. Stephens · Dept. of Mathematics & Statistics, McGill University

THE THROUGHLINE

One idea, **factor the weights low-rank**, makes Bayesian deep nets **efficient and geometrically singular**; balance the factors and the posterior **certifies its own centre**.

W = AB^T cuts cost O(mn)→O(r(m+n)) and concentrates the posterior on a measure-zero manifold; balanced SVD factors then certify the deterministic centre; and for any Bayesian model, MI decomposes into per-class C_k.

THREE PAPERS · ONE PROGRAM

01 THE GEOMETRY & EFFICIENCY

Singular BNNs

The posterior concentrates on a rank-r manifold, a singular object, not just a cheaper one. Tighter rank-aware bounds; improved OOD detection across benchmarks.

02 THE GUARANTEE

Certifying the Center

Balanced SVD factors remove gauge ambiguity, so the learned posterior gives a representation-invariant PAC-Bayes certificate for its deterministic center.

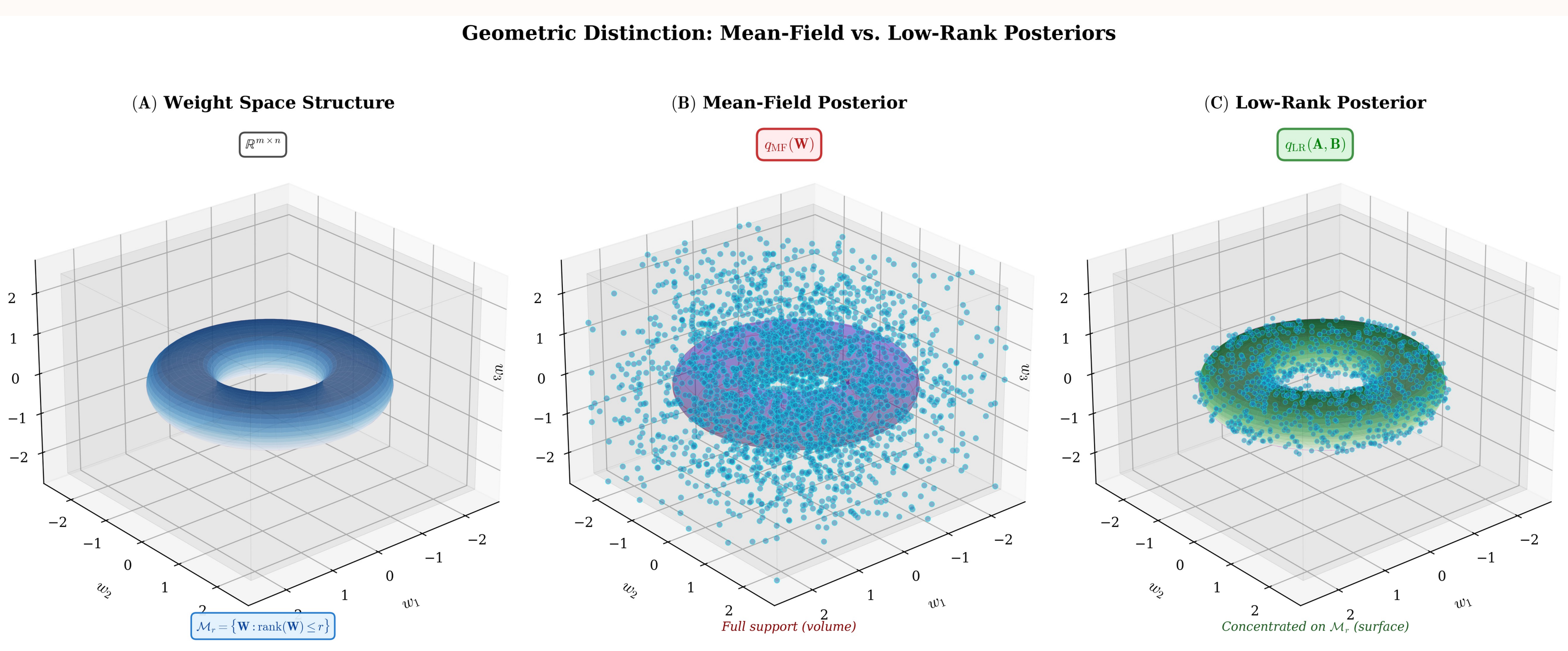
03 THE MEANING

Where, Not Just How Much

Split scalar MI into per-class contributions C_k, exposing which class the model is unsure about, cutting selective risk on critical tasks.

PAPER 01 · SINGULAR BAYESIAN NEURAL NETWORKS

The posterior support has changed



Mean-field fills the volume. Low-rank lives on the surface, a measure-zero rank-r manifold.

CENTRAL THEOREM

If r < min(m,n) and W = AB^T:

$$q_W(\mathcal{R}_r)=1 \quad \lambda(\mathcal{R}_r)=0 \;\Rightarrow\; q_W \perp \lambda$$

The posterior has no density over weight space.

0.802

AUC-OOD

best of all methods

33×

fewer params

vs Deep Ensemble

Across MLP · LSTM · Transformer and three distribution-shift regimes

Task	Arch.	Params	In-dist.	OOD	Headline
MIMIC-III mortality	MLP	13.6 K	AUROC 0.895	0.802	best OOD separation
Beijing $PM_{2.5}$	LSTM	47 K	MAE 10.63	0.710	best coverage (PICP 0.790)
SST-2 sentiment	Transf.	1.47 M	Acc 0.806	0.640	≈ Deep Ens. at 8× less time
Low-Rank SBNN (ours)	**all 3**	**↓**	**matches FR**	**↑ vs FR**	**quality-efficiency win**

PAPER 02 · CERTIFYING THE DETERMINISTIC CENTER

From a learned posterior to a guarantee

STEP 1

The three objects

One low-rank model induces three certifiable predictors: the Gibbs posterior, the predictive mean, and a deterministic center. Each needs a different theorem.

STEP 2

The obstruction

Factor gauge: (A,B) and (cA, c^{-1}B) give the same W but different norms, a naive factor-space certificate is representation-dependent.

STEP 3

The repair

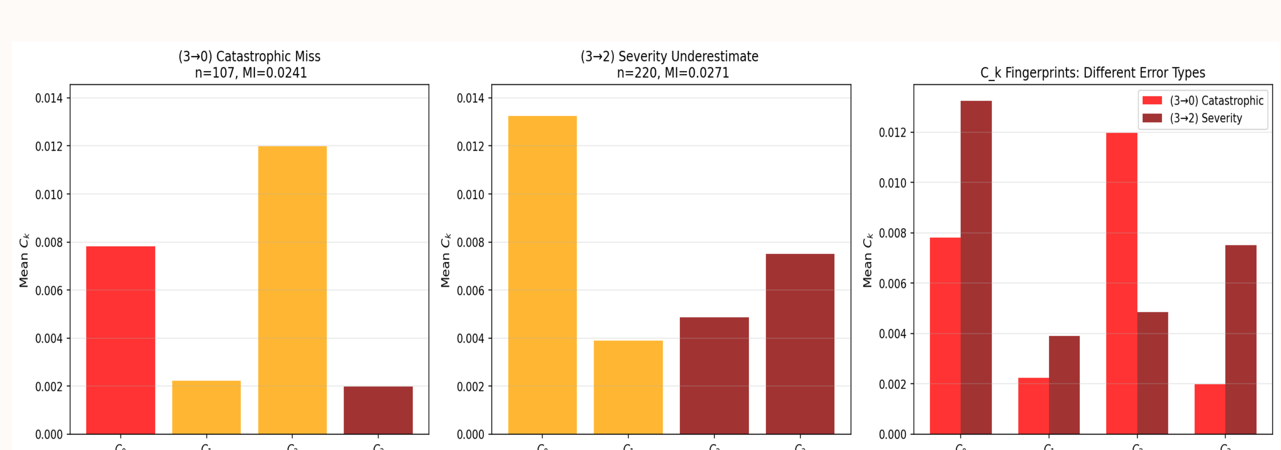
Pass to balanced factors from the SVD of the center weights. The certificate becomes intrinsic to the realized weights, gauge-free.

RESULT

A representation-invariant PAC-Bayes margin certificate for the deterministic center: explicit conditions on the learned scales σ_A, σ_B make the posterior noise Neyshabur-admissible, the posterior itself becomes the certifying perturbation law.

PAPER 03 · DECOMPOSING EPISTEMIC UNCERTAINTY

Where the model is unsure, not just how much



Ck fingerprints differ by error type — catastrophic misses (3→0) load C_2; severity errors (3→2) load C_0. Same MI, different structure.

34.7% lower selective risk vs MI on diabetic-retinopathy deferral

56.2% lower selective risk vs variance baselines

best AUROC for OOD; $\sum_k C_k$ exposes asymmetric shifts MI hides

THE SYNTHESIS

Three results compose: low-rank factorization induces geometric singularity and efficiency; balanced SVD factors make the posterior certify its deterministic center; and per-class C_k decomposes any Bayesian model's uncertainty by class, exposing where the model is ignorant.