When Does a Low-Rank Bayesian Neural Network Certify Its Deterministic Center?

Mame Diarra Toure [1] **David A. Stephens** [1]

Abstract

We study when a structured low-rank Gaussian variational posterior can certify a deterministic predictor in a Bayesian neural network with factorized layers $W_i = A_i B_i^\top$. The same low-rank Bayesian model gives rise to three natural certification targets: the posterior Gibbs predictor, the posterior predictive mean, and a deterministic center network. This paper focuses on the deterministic-center route.

The main obstruction is factor non-identifiability: (A_i, B_i) and $(cA_i, c^{-1}B_i)$ induce the same weight matrix but different factor norms, so a naive PAC-Bayes margin certificate in factor coordinates is representation dependent. We resolve this by passing to balanced factors obtained from the singular value decomposition of the center weights.

On this balanced factor space, we combine Neyshabur's PAC-Bayes margin framework with rectangular Gaussian operator-norm bounds to derive explicit perturbation budgets and a margin bound for a deterministic center network. In the balanced Gaussian variational setting, we give sufficient conditions on the learned posterior scales under which the variational posterior itself serves as the certifying perturbation law, thereby yielding a PAC-Bayes margin bound for the corresponding deterministic center network through the actual posterior geometry rather than through an auxiliary perturbation chosen only for analysis.

The resulting certificate is representation-invariant and is controlled through explicit posterior-induced perturbation budgets tied to the balanced geometry.

1. Introduction

Bayesian neural networks offer a principled route to predictive uncertainty, but standard Gaussian variational posteriors scale with the full ambient weight dimension. Low-rank factorized layers, $W_i = A_i B_i^\top$, with $A_i \in \mathbb{R}^{h_i \times r_i}$ and $B_i \in \mathbb{R}^{h_{i-1} \times r_i}$, provide a structured alternative that reduces variational dimension and encodes weight correlations through shared latent factors. In the Singular Bayesian Neural Network (SBNN) framework introduced by Toure & Stephens (2026), the same model gives rise to three distinct certification targets: a posterior on factor space, the corresponding posterior predictive mean, and a deterministic center network. Related structured Bayesian neural-network parameterizations include Bayes by Backprop, matrix-variate Gaussian posteriors, rank-1 Bayesian factors, and compact low-rank posterior families (Blundell et al., 2015; Louizos & Welling, 2016; Dusenberry et al., 2020; Swiatkowski et al., 2020; Ober & Aitchison, 2021).

These three objects are paired with different risks and are not certified by the same theorem. Posterior PAC-Bayes control applies naturally to the Gibbs predictor (Catoni, 2007; McAllester, 2003); deterministic low-rank complexity bounds such as Pinto et al. (2025) are relevant to the predictive-mean route once an appropriate deterministic envelope class has been identified; and the deterministic-center route requires a separate perturbation-based argument. The posterior and predictive-mean routes are developed in detail in Toure & Stephens (2026); here we isolate the deterministic-center route, because this is where low-rank factorization introduces a genuinely new certification issue.

That issue is factor non-identifiability. For every $c > 0$, the pairs (A_i, B_i) and $(cA_i, c^{-1}B_i)$ induce the same weight matrix $A_i B_i^\top$ but different factor norms. A PAC-Bayes margin certificate written naively in factor coordinates is therefore representation dependent: the same deterministic center can receive different apparent complexities depending on the chosen gauge. We remove this ambiguity by passing to balanced factors constructed from the singular value decomposition of the center weights. This makes the certificate depend on intrinsic quantities of the realized center matrices rather than on an arbitrary factor gauge. The resulting control is representation-invariant and rank-aware

[1]Department of Mathematics and Statistics, McGill University, Montreal, Canada. Correspondence to: Mame Diarra Toure <mame.toure@mail.mcgill.ca>.

Preprint. May 5, 2026.

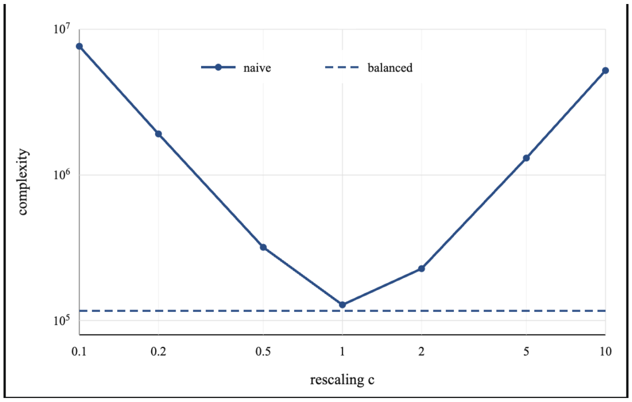

Figure 1. Numerical illustration of the factor-rescaling pathology on a scratch-trained low-rank Gaussian model. For gauge-equivalent rescalings $(A_i, B_i) \mapsto (cA_i, c^{-1}B_i)$, the realized center network is unchanged up to numerical precision, but the naive whole-network factor-space complexity varies by more than an order of magnitude over the rescaling grid. In contrast, the balanced complexity remains constant across c. This supports the paper's motivation for passing to balanced factors before certifying deterministic center networks.

through the balanced geometry.

A small numerical sanity check supports this motivation. On a synthetic scratch-trained low-rank Gaussian example, we apply gauge-equivalent rescalings $(A_i, B_i) \mapsto (cA_i, c^{-1}B_i)$ while keeping the realized center network fixed. Across a logarithmic grid of rescaling values, the effective weight matrices and network outputs remain unchanged up to numerical precision, while the naive whole-network factor-space complexity varies by more than an order of magnitude. By contrast, the balanced complexity remains invariant across rescalings; see Figure 1. This numerical check is included only to illustrate the gauge pathology motivating balancing, not to validate the final PAC-Bayes bound or the post hoc sigma conditions.

Our proof builds on the perturbation-based PAC-Bayes route of Neyshabur et al. (2018). Starting from balanced center weights, we analyze the perturbation induced by the learned Gaussian posterior on factor space, translate it into induced weight perturbations, and use rectangular Gaussian operator-norm bounds (Vershynin, 2018, Theorem 4.4.5) to verify the layerwise admissibility conditions required by Neyshabur's ReLU perturbation lemma. This yields explicit sufficient conditions on the learned posterior standard deviations under which the posterior noise preserves the center network's margins with probability at least $1/2$, and therefore a PAC-Bayes margin bound for the corresponding deterministic center network. For completeness, we restate the two imported Neyshabur lemmas in Appendix B.

This deterministic-center viewpoint sits between two nearby literatures. Deterministic low-rank generalization bounds, such as the Gaussian-complexity theorem of Pinto et al. (2025), provide rank-sensitive control for fixed low-rank function classes but do not address posterior-induced objects or factor-space gauge dependence. Classical PAC-Bayes and norm-based neural-network bounds (McAllester, 2003; Catoni, 2007; Bartlett et al., 2017; Dziugaite & Roy, 2017; Neyshabur et al., 2018) control randomized predictors or deterministic full-rank networks, but do not by themselves resolve the representation issue created by low-rank factorizations. Algorithmic low-rank Bayesian methods, including rank-1, matrix-variate, compact, and low-rank adaptation variants (Dusenberry et al., 2020; Louizos & Welling, 2016; Swiatkowski et al., 2020; Yang et al., 2024; Wang et al., 2024), focus mainly on scalable inference and uncertainty estimation rather than object-specific certification. A more detailed positioning relative to structured Bayesian neural-network parameterizations, low-rank adaptation, and adjacent PAC-Bayes work is deferred to Appendix A.

Contributions. The paper makes the following contributions.

1. We prove a PAC-Bayes margin certificate for a deterministic center network in a low-rank Bayesian neural network by working on balanced factor space. The resulting certificate is representation-invariant and depends on intrinsic quantities of the realized center weights rather than on arbitrary factor gauges.
2. We show how Gaussian factor perturbations induce admissible weight perturbations with explicit layerwise budgets via rectangular Gaussian operator-norm control, closing the low-rank bridge needed to apply Neyshabur et al. (2018).
3. In the balanced Gaussian variational setting, we derive explicit sufficient conditions on the learned posterior standard deviations under which the variational posterior itself is Neyshabur-admissible, thereby certifying the corresponding deterministic center network through the learned posterior geometry.
4. We position this result inside the broader three-object framework of Toure & Stephens (2026), where posterior, predictive-mean, and deterministic-center generalization are distinct certification problems.

2. Setup

2.1. Low-rank networks and induced objects

Let $\mathcal{X} \subseteq \mathbb{R}^{d_0}$, let $\mathcal{Y} = \{1, \ldots, k\}$, and let $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$. We consider depth-$D$ bias-free ReLU score networks with input dimension $h_0 := d_0$ and output dimension $h_D := k$, whose layer weights are factorized as $W_i = A_i B_i^\top$ with $A_i \in \mathbb{R}^{h_i \times r_i}$ and $B_i \in \mathbb{R}^{h_{i-1} \times r_i}$ for $i = 1, \ldots, D$. Here the tuple of ranks $r = (r_1, \dots, r_D)$ is treated as a fixed model hyperparameter chosen before observing the training sample. For a factor tuple $(A, B) = (A_1, B_1, \dots, A_D, B_D)$, the associated predictor is

$$f_{A,B}(x) = W_D\, \phi\big(W_{D-1}\, \phi(\cdots \phi(W_1 x) \cdots)\big),$$

where $\phi(t) = \max\{t, 0\}$ is applied coordinatewise. By construction, $\operatorname{rank}(W_i) \le r_i$ for every layer.

We write the ambient factor space as

$$\Theta_r := \prod_{i=1}^{D} \Big(\mathbb{R}^{h_i \times r_i} \times \mathbb{R}^{h_{i-1} \times r_i}\Big).$$

A low-rank Bayesian neural network places a posterior on Θ_r, and the same model induces three natural objects.

Posterior. A posterior Q on Θ_r defines a randomized Gibbs predictor by sampling $(A, B) \sim Q$ and evaluating $f_{A,B}$.

Posterior predictive mean. Whenever $\mathbb{E}_{(A,B)\sim Q}\big[\|f_{A,B}(x)\|_2\big] < \infty$ for every $x \in \mathcal{X}$, the posterior predictive mean is the deterministic predictor $f_{\mathrm{pm},Q}(x) := \mathbb{E}_{(A,B)\sim Q}[f_{A,B}(x)]$. When the posterior depends on the sample S, we write Q_S and $f_{\mathrm{pm},S}$.

Deterministic center. A deterministic factor tuple $\bar{w} = (\bar{A}_1, \bar{B}_1, \dots, \bar{A}_D, \bar{B}_D) \in \Theta_r$ defines a deterministic center network $f_{\bar{w}} := f_{\bar{A},\bar{B}}$. This is the certified object in the main theorem of the present paper.

The posterior, predictive mean, and deterministic center are distinct certification targets. In this paper we retain the first two mainly to keep that distinction explicit; the theorem developed below concerns the deterministic center.

2.2. Risk and margin notation

Let $\mathcal{D}$ be a distribution on $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$. Let $S = ((x_1, y_1), \dots, (x_m, y_m)) \in \mathcal{Z}^m$ be a sample of size m, and let $\ell : \mathbb{R}^k \times \mathcal{Y} \to [0, 1]$ be a bounded measurable loss. For a posterior Q on Θ_r, the Gibbs population and empirical risks are

$$L_G(Q) := \mathbb{E}_{(A,B)\sim Q}\, \mathbb{E}_{(x,y)\sim \mathcal{D}}\big[\ell(f_{A,B}(x), y)\big],$$

$$\hat{L}_G(Q) := \mathbb{E}_{(A,B)\sim Q} \frac{1}{m} \sum_{j=1}^{m} \ell(f_{A,B}(x_j), y_j).$$

For a deterministic predictor $g : \mathcal{X} \to \mathbb{R}^k$, we write $L_\ell(g) := \mathbb{E}_{(x,y)\sim\mathcal{D}}[\ell(g(x), y)]$ and $\hat{L}_{\ell,S}(g) := \frac{1}{m}\sum_{j=1}^{m} \ell(g(x_j), y_j)$.

The deterministic-center theorem is a margin theorem, so from this point on we specialize to multiclass classification on a bounded input domain. Let $\mathcal{X}_B := \{x \in \mathcal{X} : \|x\|_2 \le B\}$ and assume $\mathcal{D}(\mathcal{X}_B \times \mathcal{Y}) = 1$. For a score vector $v \in \mathbb{R}^k$ and a label $y \in \mathcal{Y}$, define the multiclass margin by $\operatorname{margin}(v, y) := v_y - \max_{j \ne y} v_j$. The associated classification risk is $L_0(g) := \mathbb{P}_{(x,y)\sim\mathcal{D}}\big(g(x)_y \le \max_{j\ne y} g(x)_j\big)$, and the empirical margin loss at level $\gamma > 0$ is

$$\hat{L}_{\gamma,S}(g) := \frac{1}{m} \sum_{j=1}^{m} \mathbf{1}\Big\{ g(x_j)_{y_j} \le \gamma + \max_{r \ne y_j} g(x_j)_r \Big\}.$$

Thus the posterior is naturally paired with Gibbs risk, the predictive mean with an ordinary deterministic loss, and the deterministic center with a margin-based classification risk.

2.3. Balanced deterministic center

The main technical issue in the deterministic-center route is factor non-identifiability. For every $c > 0$, the pairs (A_i, B_i) and $(cA_i, c^{-1}B_i)$ induce the same weight matrix $A_i B_i^\top$. A deterministic certificate written directly in factor coordinates is therefore representation dependent.

To remove this scalar-rescaling ambiguity, we work with balanced factors constructed from the singular value decomposition of the center weights.

Definition 2.1 (Balanced center)**.** Let $\bar{W}_i \in \mathbb{R}^{h_i \times h_{i-1}}$ be a deterministic center weight matrix with intrinsic rank $s_i := \operatorname{rank}(\bar{W}_i) \le r_i$, and let $\bar{W}_i = U_i \Sigma_i V_i^\top$ be its thin singular value decomposition.

Define the unpadded balanced factors by $\tilde{A}_i := U_i \Sigma_i^{1/2} \in \mathbb{R}^{h_i \times s_i}$ and $\tilde{B}_i := V_i \Sigma_i^{1/2} \in \mathbb{R}^{h_{i-1} \times s_i}$. If $s_i < r_i$, pad these matrices with zero columns to obtain $\bar{A}_i \in \mathbb{R}^{h_i \times r_i}$ and $\bar{B}_i \in \mathbb{R}^{h_{i-1} \times r_i}$. Then $\bar{W}_i = \bar{A}_i \bar{B}_i^\top$, and the balanced identities

$$\|\bar{A}_i\|_2^2 = \|\bar{B}_i\|_2^2 = \|\bar{W}_i\|_2, \qquad \|\bar{A}_i\|_F^2 = \|\bar{B}_i\|_F^2 = \|\bar{W}_i\|_*$$

hold. These identities follow immediately from the thin SVD and are unaffected by the zero padding. In the SBNN setting, $\bar{W}_i$ is the deterministic center weight matrix extracted from the learned posterior means and then represented through its balanced factorization.

The tuple $\bar{w} = (\bar{A}_1, \bar{B}_1, \dots, \bar{A}_D, \bar{B}_D) \in \Theta_r$ is called a balanced center, and the associated predictor is denoted by $f_{\bar{w}}$. The balanced construction selects a canonical representative of the center weights inside factor space and is the key device that makes the deterministic-center certificate representation-invariant.

2.4. Gaussian posterior model around the balanced center

The center-network theorem studied below is expressed in terms of a Gaussian variational posterior on the fixed factor space Θ_r whose deterministic center is represented in balanced coordinates. Concretely, after identifying the deterministic center weights $\bar{W}_i$ and their balanced factorization $\bar{W}_i = \bar{A}_i \bar{B}_i^\top$, we consider for each layer a Gaussian posterior on Θ_r centered at the padded balanced factors $\bar{A}_i$ and $\bar{B}_i$, so that a posterior draw has the form $A_i = \bar{A}_i + \sigma_{A_i}\Xi_{A_i}$ and $B_i = \bar{B}_i + \sigma_{B_i}\Xi_{B_i}$, where $\Xi_{A_i} \in \mathbb{R}^{h_i \times r_i}$ and $\Xi_{B_i} \in \mathbb{R}^{h_{i-1} \times r_i}$ are independent standard Gaussian matrices and $\sigma_{A_i}, \sigma_{B_i} > 0$ are learned posterior scales.

Thus the theorem is applied after expressing the deterministic center in its balanced factorization on the fixed rank-r space. Because $\bar{A}_i \bar{B}_i^\top = \bar{W}_i$, this posterior describes Gaussian factor perturbations around the same deterministic center weights $\bar{W}_i$ and hence around the same center network $f_{\bar{w}}$. Each such draw induces a score network through the realized weights $W_i = A_i B_i^\top$ by the same formula as in Section 2. The main question in the next section is whether this learned posterior noise is small enough to serve as a Neyshabur-admissible perturbation around that center network. We answer this by deriving explicit perturbation budgets and corresponding sufficient conditions on the learned scales σ_{A_i} and σ_{B_i}.

3. Center Certification from a Learned Gaussian Low-Rank Posterior

The broader SBNN framework gives three distinct objects: the posterior predictor, the posterior predictive mean, and a deterministic center network. The first two have their own certification routes elsewhere. In this section we focus only on the third object. Our question is the following: when does the Gaussian posterior learned by the SBNN induce a perturbation around its deterministic center that is small enough to trigger Neyshabur's PAC-Bayes margin argument? The answer is expressed directly in terms of the learned posterior scales σ_{A_i} and σ_{B_i}.

Throughout the section, let $\bar{w} = (\bar{A}_1, \bar{B}_1, \ldots, \bar{A}_D, \bar{B}_D)$ be the balanced center from Section 2, let $\bar{W}_i = \bar{A}_i \bar{B}_i^\top$ denote the corresponding center weights, and let $f_{\bar{w}}$ be the associated deterministic center network. We work on the fixed factor space Θ_r.

3.1. Posterior-induced perturbation budgets

Let

$$Q := \bigotimes_{i=1}^{D} \mathcal{N}(\mathrm{vec}(\bar{A}_i), \sigma_{A_i}^2 I_{h_i r_i}) \otimes \mathcal{N}(\mathrm{vec}(\bar{B}_i), \sigma_{B_i}^2 I_{h_{i-1} r_i})$$

be the Gaussian variational posterior learned by the balanced SBNN on Θ_r, where $\sigma_{A_i} > 0$ and $\sigma_{B_i} > 0$ for every layer. Here $I_{h_i r_i}$ and $I_{h_{i-1} r_i}$ denote the identity matrices on the vectorized spaces $\mathbb{R}^{h_i r_i}$ and $\mathbb{R}^{h_{i-1} r_i}$. A posterior draw can be written as $A_i = \bar{A}_i + \sigma_{A_i}\Xi_{A_i}$ and $B_i = \bar{B}_i + \sigma_{B_i}\Xi_{B_i}$, where $\Xi_{A_i} \in \mathbb{R}^{h_i \times r_i}$ and $\Xi_{B_i} \in \mathbb{R}^{h_{i-1} \times r_i}$ are independent standard Gaussian matrices. The induced weight perturbation is $\Delta W_i := A_i B_i^\top - \bar{W}_i$.

For each layer i and each $t \geq 0$, define $a_i(t) := \sqrt{h_i} + \sqrt{r_i} + t$ and $b_i(t) := \sqrt{h_{i-1}} + \sqrt{r_i} + t$.

Proposition 3.1 (Posterior-induced perturbation budgets)**.** *Assume $\|\bar{W}_i\|_2 > 0$ for every layer $i = 1, \ldots, D$. For each layer i and each $t \geq 0$, define*

$$\rho_i^Q(t) := \frac{\sigma_{B_i} b_i(t) \|\bar{A}_i\|_2 + \sigma_{A_i} a_i(t) \|\bar{B}_i\|_2}{\|\bar{W}_i\|_2} + \frac{\sigma_{A_i}\sigma_{B_i} a_i(t) b_i(t)}{\|\bar{W}_i\|_2}.$$

Then, with probability at least $1 - 4De^{-t^2/2}$, the inequalities $\|\Delta W_i\|_2 \leq \rho_i^Q(t) \|\bar{W}_i\|_2$ hold simultaneously for all layers $i = 1, \ldots, D$.

Since the center is balanced, this simplifies to

$$\rho_i^Q(t) = \frac{\sigma_{A_i} a_i(t) + \sigma_{B_i} b_i(t)}{\|\bar{W}_i\|_2^{1/2}} + \frac{\sigma_{A_i}\sigma_{B_i} a_i(t) b_i(t)}{\|\bar{W}_i\|_2}.$$

Proof. A draw from Q can be written as $A_i = \bar{A}_i + \sigma_{A_i}\Xi_{A_i}$ and $B_i = \bar{B}_i + \sigma_{B_i}\Xi_{B_i}$. Hence

$$\begin{aligned}\Delta W_i &= A_i B_i^\top - \bar{W}_i \\ &= (\bar{A}_i + \sigma_{A_i}\Xi_{A_i})(\bar{B}_i + \sigma_{B_i}\Xi_{B_i})^\top - \bar{A}_i \bar{B}_i^\top \\ &= \bar{A}_i(\sigma_{B_i}\Xi_{B_i})^\top + (\sigma_{A_i}\Xi_{A_i})\bar{B}_i^\top + (\sigma_{A_i}\Xi_{A_i})(\sigma_{B_i}\Xi_{B_i})^\top.\end{aligned}$$

Therefore,

$$\begin{aligned}\|\Delta W_i\|_2 \leq \|\bar{A}_i\|_2 \, \|\sigma_{B_i}\Xi_{B_i}\|_2 &+ \|\sigma_{A_i}\Xi_{A_i}\|_2 \, \|\bar{B}_i\|_2 \\ &+ \|\sigma_{A_i}\Xi_{A_i}\|_2 \, \|\sigma_{B_i}\Xi_{B_i}\|_2.\end{aligned}$$

For each layer i, Vershynin's rectangular Gaussian operator-norm bound (Vershynin, 2018, Theorem 4.4.5) gives $\mathbb{P}(\|\Xi_{A_i}\|_2 > a_i(t)) \leq 2e^{-t^2/2}$ and $\mathbb{P}(\|\Xi_{B_i}\|_2 > b_i(t)) \leq 2e^{-t^2/2}$. A union bound over the $2D$ Gaussian matrices therefore shows that, with probability at least $1 - 4De^{-t^2/2}$, $\|\Xi_{A_i}\|_2 \leq a_i(t)$ and $\|\Xi_{B_i}\|_2 \leq b_i(t)$ hold simultaneously for every layer $i = 1, \ldots, D$.

On this event, $\|\sigma_{A_i}\Xi_{A_i}\|_2 \leq \sigma_{A_i} a_i(t)$ and $\|\sigma_{B_i}\Xi_{B_i}\|_2 \leq \sigma_{B_i} b_i(t)$, so the previous norm estimate yields

$$\begin{aligned}\|\Delta W_i\|_2 \leq \sigma_{B_i} b_i(t) \|\bar{A}_i\|_2 &+ \sigma_{A_i} a_i(t) \|\bar{B}_i\|_2 \\ &+ \sigma_{A_i}\sigma_{B_i} a_i(t) b_i(t).\end{aligned}$$

Dividing by $\|\bar{W}_i\|_2$ gives

$$\frac{\|\Delta W_i\|_2}{\|\bar{W}_i\|_2} \leq \underbrace{\frac{\sigma_{B_i} b_i(t) \|\bar{A}_i\|_2 + \sigma_{A_i} a_i(t) \|\bar{B}_i\|_2}{\|\bar{W}_i\|_2} + \frac{\sigma_{A_i}\sigma_{B_i} a_i(t) b_i(t)}{\|\bar{W}_i\|_2}}_{=\rho_i^Q(t)}$$

Equivalently, $\|\Delta W_i\|_2 \le \rho_i^Q(t)\|\bar{W}_i\|_2$.

Since the center is balanced, $\|\bar{A}_i\|_2 = \|\bar{B}_i\|_2 = \|\bar{W}_i\|_2^{1/2}$, and substituting these identities into the definition of $\rho_i^Q(t)$ gives the simplified balanced formula. □

3.2. Explicit sigma conditions

The previous proposition expresses the perturbation size through the quantities $\rho_i^Q(t)$. To make the dependence on the learned posterior scales more transparent, define $x_i(t) := \sigma_{A_i} a_i(t)/\|\bar{W}_i\|_2^{1/2}$ and $y_i(t) := \sigma_{B_i} b_i(t)/\|\bar{W}_i\|_2^{1/2}$. Then $\rho_i^Q(t) = x_i(t) + y_i(t) + x_i(t)y_i(t)$.

Corollary 3.2 (Explicit sigma conditions for posterior admissibility)**.** *Assume $\|\bar{W}_i\|_2 > 0$ for every layer, fix $t \ge 0$, and let e denote Euler's number. If*

$$4De^{-t^2/2} \le \frac{1}{2},$$

$$x_i(t) + y_i(t) + x_i(t)y_i(t) \le \frac{1}{D} \qquad \textit{for every } i = 1, \dots, D,$$

$$\sum_{i=1}^{D} \big(x_i(t) + y_i(t) + x_i(t)y_i(t)\big) < \frac{\gamma}{4eB \prod_{j=1}^{D} \|\bar{W}_j\|_2},$$

then

$$\mathbb{P}_{(A,B)\sim Q}\left(\sup_{x \in \mathcal{X}_B} \|f_{A,B}(x) - f_{\bar{w}}(x)\|_\infty < \frac{\gamma}{4}\right) \ge \frac{1}{2}.$$

Proof. By Proposition 3.1, with probability at least $1 - 4De^{-t^2/2}$ we have $\|\Delta W_i\|_2 \le \rho_i^Q(t)\|\bar{W}_i\|_2$ for all layers. Under the displayed conditions, this event has probability at least $1/2$ and also satisfies $\|\Delta W_i\|_2 \le \|\bar{W}_i\|_2/D$ for every layer. The imported ReLU perturbation lemma from Appendix B therefore gives, for every $x \in \mathcal{X}_B$,

$$\begin{aligned}\|f_{A,B}(x) - f_{\bar{w}}(x)\|_2 &\le eB\Big(\prod_{j=1}^{D} \|\bar{W}_j\|_2\Big)\Big(\sum_{i=1}^{D} \rho_i^Q(t)\Big) \\ &< \frac{\gamma}{4}.\end{aligned}$$

Since $\|v\|_\infty \le \|v\|_2$ for every $v \in \mathbb{R}^k$, the displayed probability bound follows. □

Remark 3.3 (Post hoc certification checklist). Given a learned posterior, one checks the certificate by: choosing t so that $4De^{-t^2/2} \le 1/2$; computing the layerwise budgets $\rho_i^Q(t)$; verifying the layerwise condition $\rho_i^Q(t) \le 1/D$ and the summed margin condition; and then evaluating the PAC-Bayes term $\mathrm{KL}(Q\|P)$ under a fixed sample-independent prior.

3.3. Main center-network theorem

We can now state the actual center-network certificate for the learned SBNN.

Theorem 3.4 (Center-Network PAC-Bayes Certificate for a Learned Gaussian SBNN)**.** *Let S be a training sample of size $m \ge 2$, let $Q = Q_S$ be the balanced Gaussian variational posterior on the fixed factor space Θ_r learned from S, and let P be any prior on the same fixed factor space Θ_r that is independent of the training sample. If the sigma conditions of Corollary 3.2 hold, then with probability at least $1 - \delta$ over a training sample of size m,*

$$L_0(f_{\bar{w}}) \le \hat{L}_{\gamma,S}(f_{\bar{w}}) + 4\sqrt{\frac{\mathrm{KL}(Q\|P) + \log(6m/\delta)}{m-1}}.$$

In particular, whenever the Gaussian posterior learned by the SBNN has standard deviations satisfying those conditions, the corresponding center network learned by that SBNN satisfies the above margin bound.

Proof. Corollary 3.2 verifies exactly the margin-preservation event required by the PAC-Bayes margin lemma stated in Appendix B, with perturbation law Q around the deterministic center $f_{\bar{w}}$. Applying that lemma yields the bound. □

Remark 3.5 (A concrete fixed prior). A natural concrete choice is a fixed isotropic Gaussian prior on the prescribed factor space Θ_r, namely

$$P_\tau := \bigotimes_{i=1}^{D} \mathcal{N}(\mathrm{vec}(0), \tau_{A_i}^2 I_{h_i r_i}) \otimes \mathcal{N}(\mathrm{vec}(0), \tau_{B_i}^2 I_{h_{i-1} r_i}),$$

where the scales $\tau_{A_i}, \tau_{B_i} > 0$ are chosen before observing the training sample. For the Gaussian posterior Q, the divergence $\mathrm{KL}(Q\|P_\tau)$ is then available in closed form by the standard formula for Gaussians with diagonal covariances, so the theorem can be instantiated without introducing a data-dependent prior.

Interpretation. This is the center-network certificate for the SBNN framework. The certified object is the deterministic center network $f_{\bar{w}}$, not the posterior Gibbs predictor and not the posterior predictive mean. What the theorem shows is that explicit conditions on the learned posterior scales σ_{A_i} and σ_{B_i} are sufficient to make the posterior-induced noise Neyshabur-admissible, and therefore sufficient to certify the corresponding center network learned by the BNN. In this sense, the result is post hoc: after training, one checks whether the learned posterior scales satisfy the sufficient conditions, and if they do, one obtains a PAC-Bayes certificate for the deterministic center. The theorem itself is stated with an arbitrary sample-independent prior on the fixed factor space Θ_r; the learned posterior enters through the perturbation event and the KL term. Failure of these sufficient inequalities does not imply that posterior-induced margin preservation fails; it only means that the present certificate does not verify it.

3.4. Theorem-Aligned Numerical Findings

To complement the theoretical development, we performed a post hoc numerical instantiation of the deterministic-center certificate in a synthetic setting matched to the assumptions of the theorem. A bias-free low-rank ReLU teacher network with one hidden layer generates bounded inputs satisfying the exact norm constraint $B = 1$ and multiclass labels by argmax of its score outputs. The student is a low-rank Gaussian ReLU score network with a linear output layer and no biases. After training, we extract the balanced deterministic center from the learned posterior means, use a fixed isotropic Gaussian prior on the prescribed factor space, and evaluate the certificate quantities over a posterior-scale multiplier α, where $\alpha = 1$ corresponds to the learned posterior.

Figure 2 summarizes the best run. At the learned posterior scale, the layerwise condition is satisfied, with $\max_i D\rho_i^Q(t) = 0.175 < 1$, and the validation-set Monte Carlo proxy for the perturbation event is high, with $\hat{p}(1) = 0.980$. However, the normalized summed margin quantity remains above threshold, taking value 37.45, so the learned posterior does not satisfy the theorem's sufficient conditions at $\alpha = 1$. The empirical bottleneck is therefore not individual-layer admissibility but the accumulated margin-preservation bound. At smaller posterior scales, the sufficient inequalities do become feasible in this setting; for example, at $\alpha = 0.02$ we obtain $\max_i D\rho_i^Q(t) = 0.00342$ and normalized summed condition 0.734.

These findings suggest that, in a theorem-aligned low-rank ReLU regime, the deterministic-center certificate captures a meaningful perturbative structure: the learned posterior already satisfies the layerwise condition and exhibits a high empirical perturbation-event probability, while the main remaining obstruction is the conservativeness of the global summed margin bound. This experiment should therefore be interpreted as a numerical instantiation and conservativeness study for the certificate, rather than as an empirical proof of the PAC-Bayes theorem.

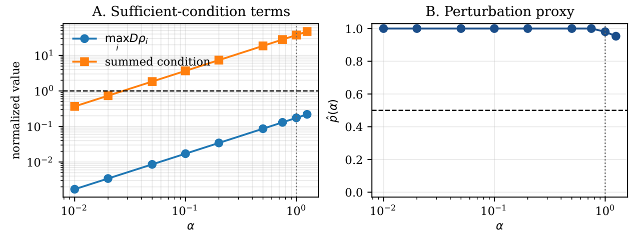

Figure 2. Theorem-aligned post hoc deterministic-center certificate quantities for the best tested bias-free low-rank ReLU student. Panel A shows the normalized layerwise quantity $\max_i D\rho_i$ and the normalized summed margin condition as the posterior scale multiplier α varies; the horizontal line marks the theorem threshold 1, and the vertical line marks the learned posterior scale $\alpha = 1$. Panel B shows the validation-set Monte Carlo proxy $\hat{p}(\alpha)$ for the perturbation event, with the horizontal line at $1/2$.

4. Discussion and Conclusion

The main result of this paper is a PAC-Bayes margin certificate for the deterministic center network $f_{\bar{w}}$ in the low-rank SBNN framework. The certified object is therefore neither the posterior Gibbs predictor nor the posterior predictive mean, but the deterministic center associated with the learned balanced factorization. This distinction is essential: the three objects arising from the same low-rank Bayesian model remain genuinely different certification targets.

The technical role of balancing is to remove the representation dependence created by factor rescaling. Without balancing, different factorizations of the same deterministic weight matrix can carry different factor norms and hence different apparent PAC-Bayes complexities. The balanced construction selects a canonical representative in factor space, making the resulting certificate depend on intrinsic quantities of the realized center weights and on explicit posterior-induced perturbation budgets.

A practical limitation is that the sufficient conditions derived here can be stringent. As in related margin-based PAC-Bayes analyses, the perturbation control depends on the product of layer spectral norms, and the admissibility argument requires each layerwise perturbation to remain at most on the order of $\|\bar{W}_i\|_2/D$. For deep networks or poorly controlled spectral norms, these requirements may be difficult to satisfy and can lead to vacuous certificates. Accordingly, the present paper establishes a rigorous certification route for the deterministic center but does not claim that the resulting bound is typically non-vacuous in large modern networks. The current theorem is also restricted to bias-free fully connected ReLU networks; extending the same representation-invariant certification strategy to bias terms, convolutional layers, or other structured low-rank parameterizations is a meaningful next step.

The posterior-noise result sharpens the interpretation of the theorem. It shows that explicit conditions on the learned posterior scales σ_{A_i} and σ_{B_i} are sufficient to make the posterior-induced noise Neyshabur-admissible, and therefore sufficient to certify the center network learned by the BNN. In this sense, the result is post hoc: after training, one checks whether the learned posterior scales satisfy the sufficient conditions, and if they do, one obtains a PAC-Bayes certificate for the deterministic center.

Thus, within the broader three-certificate SBNN framework, the contribution of this paper is the deterministic-center route: a representation-invariant, rank-aware PAC-Bayes margin certificate together with explicit sufficient conditions under which the learned Gaussian posterior certifies its corresponding center network.

References

Bartlett, P. L., Foster, D. J., and Telgarsky, M. J. Spectrally-normalized margin bounds for neural networks. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/b22b257ad0519d4500539da3c8bcf4dd-Paper.pdf.

Blundell, C., Cornebise, J., Kavukcuoglu, K., and Wierstra, D. Weight uncertainty in neural networks. In *International Conference on Machine Learning (ICML)*, pp. 1613–1622, 2015.

Catoni, O. *PAC-Bayesian Supervised Classification: The Thermodynamics of Statistical Learning*, volume 56 of *IMS Lecture Notes–Monograph Series*. Institute of Mathematical Statistics, Beachwood, OH, 2007. doi: 10.1214/074921707000000391. URL https://doi.org/10.1214/074921707000000391.

Dusenberry, M., Jerfel, G., Wen, Y., Ma, Y., Snoek, J., Heller, K., Lakshminarayanan, B., and Tran, D. Efficient and scalable Bayesian neural nets with rank-1 factors. In *International Conference on Machine Learning (ICML)*, pp. 2782–2792, 2020.

Dziugaite, G. K. and Roy, D. M. Computing nonvacuous generalization bounds for deep (stochastic) neural networks with many more parameters than training data. In *Proceedings of the Thirty-Third Conference on Uncertainty in Artificial Intelligence*, 2017. URL https://gkdz.org/publication/dr17uai/.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., and Chen, W. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations (ICLR)*, 2022.

Louizos, C. and Welling, M. Structured and efficient variational deep learning with matrix gaussian posteriors. In Balcan, M. F. and Weinberger, K. Q. (eds.), *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 1708–1716, New York, New York, USA, 20–22 Jun 2016. PMLR. URL https://proceedings.mlr.press/v48/louizos16.html.

McAllester, D. Simplified PAC-Bayesian margin bounds. In Schölkopf, B. and Warmuth, M. K. (eds.), *Learning Theory and Kernel Machines: 16th Annual Conference on Computational Learning Theory and 7th Kernel Workshop, COLT/Kernel 2003, Washington, DC, USA, August 24–27, 2003, Proceedings*, volume 2777 of *Lecture Notes in Computer Science*, pp. 203–215. Springer, Berlin, Heidelberg, 2003. doi: 10.1007/978-3-540-45167-9_15. URL https://doi.org/10.1007/978-3-540-45167-9_15.

Neyshabur, B., Bhojanapalli, S., and Srebro, N. A PAC-Bayesian approach to spectrally-normalized margin bounds for neural networks. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=Skz_WfbCZ.

Ober, S. W. and Aitchison, L. Global inducing point variational posteriors for bayesian neural networks and deep gaussian processes. In Meila, M. and Zhang, T. (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8248–8259. PMLR, 2021. URL https://proceedings.mlr.press/v139/ober21a.html.

Pinto, A., Rangamani, A., and Poggio, T. A. On generalization bounds for neural networks with low rank layers. In *36th International Conference on Algorithmic Learning Theory*, 2025. URL https://openreview.net/forum?id=TAvypH5yl5.

Swiatkowski, J., Roth, K., Veeling, B. S., Tran, L., Dillon, J. V., Snoek, J., Mandt, S., Salimans, T., Jenatton, R., and Nowozin, S. The k-tied Normal distribution: A compact parameterization of Gaussian mean field posteriors in Bayesian neural networks. In *International Conference on Machine Learning (ICML)*, pp. 9263–9274, 2020.

Toure, M. D. and Stephens, D. A. Singular bayesian neural networks, 2026. URL https://arxiv.org/abs/2602.00387.

Vershynin, R. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, Cambridge, 2018. doi: 10.1017/9781108231596. URL https://doi.org/10.1017/9781108231596.

Wang, Y., Shi, H., Han, L., Metaxas, D. N., and Wang, H. BLoB: Bayesian low-rank adaptation by backpropagation for large language models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Yang, A., Robeyns, M., Wang, X., and Aitchison, L. Bayesian low-rank adaptation for large language models. In Kim, B., Yue, Y., Chaudhuri, S., Fragkiadaki, K., Khan, M., and Sun, Y. (eds.), *International Conference on Learning Representations*, volume 2024, pp. 1812–1842, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/07c256a163a7559186ec1c71e95b9ec9-Paper-Conference.pdf.

A. Extended Related Work

Our center theorem sits at the intersection of structured variational inference, low-rank neural-network theory, and PAC-Bayes generalization.

Bayesian neural networks and structured variational posteriors. Bayes by Backprop (Blundell et al., 2015) established the modern variational-Bayesian neural-network template with factorized Gaussian posteriors. Later work enriched posterior structure through matrix-variate Gaussian families (Louizos & Welling, 2016), compact low-rank parameterizations of posterior variance (Swiatkowski et al., 2020), global inducing-point correlated posteriors (Ober & Aitchison, 2021), and rank-1 Bayesian factors designed for scale (Dusenberry et al., 2020). These papers study scalable inference and uncertainty quality, whereas our question is which generalization object a low-rank posterior certifies and how the resulting certificate depends on factor representation.

Low-rank adaptation and Bayesian PEFT. Low-rank structure also appears in parameter-efficient adaptation. LoRA (Hu et al., 2022) introduced low-rank updates for large pretrained models, and Bayesian variants such as Laplace-LoRA (Yang et al., 2024) and BLoB (Wang et al., 2024) place posterior uncertainty on low-dimensional factor spaces. These methods are algorithmically close in spirit because they Bayesianize low-rank coordinates, but their goals are calibration and efficient fine-tuning rather than end-to-end generalization certificates for deterministic center networks.

Deterministic low-rank generalization bounds. On the deterministic side, Pinto et al. (2025) derive vector-valued Gaussian complexity bounds for neural networks with low-rank layers under explicit spectral-norm and rank constraints. Their theorem provides rank-sensitive control for deterministic low-rank classes and is the relevant deterministic envelope result behind the predictive-mean route in Toure & Stephens (2026). It does not, however, address posterior-induced objects, perturbation-based center certificates, or the gauge dependence created by factor-space rescaling.

PAC-Bayes and norm-based bounds for neural networks. Classical PAC-Bayes theory (McAllester, 2003; Catoni, 2007) measures complexity through posterior-to-prior divergence, and computable nonvacuous deep-network PAC-Bayes bounds show that such certificates can remain numerically meaningful at modern scales (Dziugaite & Roy, 2017). Norm-based bounds for deterministic deep networks, including spectrally normalized margin bounds (Bartlett et al., 2017), tie generalization to products of operator norms and margin terms. The perturbation-based PAC-Bayes margin route of Neyshabur et al. (2018) is the direct starting point for our center theorem. Scale-invariant and path-norm-based margin bounds are also conceptual neighbors, since they seek parameterization-robust complexity measures, but they do not address low-rank Bayesian factor posteriors or deterministic-center certification in the present sense.

What is new in the present paper. The distinctive point here is not a new generic PAC-Bayes inequality and not a new deterministic low-rank class bound. It is the combination of three ingredients: object-specific certification for low-rank Bayesian networks, a balanced-factor repair of the rescaling pathology $(A_i, B_i) \mapsto (cA_i, c^{-1}B_i)$, and a posterior-noise specialization showing that in the balanced Gaussian setting the variational posterior itself can serve as Neyshabur's certifying perturbation law. The broader three-object program, including the posterior and predictive-mean routes, is developed in Toure & Stephens (2026); the present paper isolates the deterministic-center route because it is where low-rank non-identifiability becomes theorem-critical.

B. Imported Neyshabur Lemmas

For the deterministic-center route, we use two results imported from Neyshabur et al. (2018). We restate them here in the notation used in the main text. In our application, the deterministic parameter is the balanced center $\bar{w}$, the corresponding predictor is $f_{\bar{w}}$, and the layerwise perturbations in weight space are the induced matrices $U_i = \Delta W_i = A_i B_i^\top - \bar{W}_i$. Thus the generic perturbation notation in the imported lemmas is instantiated through posterior-induced perturbations of the balanced center network.

Lemma B.1 (PAC-Bayes margin lemma, after Neyshabur et al. (2018)). *Let $f_w : \mathcal{X}_B \to \mathbb{R}^k$ be any predictor with parameter w. Let P be any distribution on parameters that is independent of the training sample. Let u be a random perturbation of w, and let Q_w denote the law of $w + u$.*

Assume that

$$\mathbb{P}_u\left(\sup_{x\in\mathcal{X}_B} \|f_{w+u}(x) - f_w(x)\|_\infty < \frac{\gamma}{4}\right) \geq \frac{1}{2}. \tag{1}$$

Then, for every $\gamma > 0$ and every $\delta \in (0, 1)$, with probability at least $1 - \delta$ over a training sample of size $m \geq 2$,

$$L_0(f_w) \leq \hat{L}_{\gamma,S}(f_w) + 4\sqrt{\frac{\mathrm{KL}(Q_w\|P) + \log(6m/\delta)}{m-1}}. \tag{2}$$

Lemma B.2 (ReLU perturbation lemma, after Neyshabur et al. (2018))**.** *Let f_w be a depth-D bias-free ReLU network on $\mathcal{X}_B$, and let $U_1, \ldots, U_D$ be perturbation matrices for the deterministic weights $W_1, \ldots, W_D$.*

Assume that

$$\|U_i\|_2 \leq \frac{1}{D}\|W_i\|_2 \qquad \text{for every } i = 1, \ldots, D. \tag{3}$$

Then, for every $x \in \mathcal{X}_B$,

$$\|f_{w+u}(x) - f_w(x)\|_2 \leq eB\prod_{j=1}^{D}\|W_j\|_2 \sum_{i=1}^{D}\frac{\|U_i\|_2}{\|W_i\|_2}. \tag{4}$$

Remark B.3. Here and throughout, e denotes Euler's number.

Remark B.4. Since $\|v\|_\infty \leq \|v\|_2$ for every $v \in \mathbb{R}^k$, any upper bound by $\gamma/4$ on the right-hand side of Lemma B.2 implies the hypothesis of Lemma B.1.