Why Does Low-Rank Adaptation Work for Hindi-English Code-Mixing? A Geometric Analysis

Shashank Vishwakarma, Rakesh Kumar
Gautam Buddha University,
Greater Noida, India
shakuvish07@gmail.com, rakesh.kumar@gbu.ac.in

Abstract

Low-Rank Adaptation (LoRA) enables efficient fine-tuning of large language models, yet why it works particularly well for code-mixed text remains unexplained. We propose that LoRA's efficiency stems from geometric structure in multilingual pre-trained models: code-mixed embeddings concentrate in low-dimensional cross-lingual subspaces. Through spectral analysis of mBERT and MuRIL on Hindi-English (Hinglish) data, we establish that pre-trained attention weights have effective ranks of 437–441, while LoRA updates ($r = 4, 8, 16$) exhibit ranks of 2.1–5.9—a 136× average compression. Cross-lingual geometry measured via Centered Kernel Alignment (CKA) shows Hinglish embeddings align strongly with Hindi (CKA=0.279) but weakly with English (0.093), compared to a monolingual baseline of 0.074. Statistical tests (Wilcoxon $p < 10^{-19}$) and permutation ablations confirm these differences are robust. We interpret the convergence of geometric overlap (3.77× baseline) and empirical compression (136×) as evidence that low-rank adaptation exploits pre-existing multilingual structure. Findings are demonstrated on token-level language identification; extensions to other language pairs and tasks remain open questions.

Keywords: code-mixing, low-rank adaptation, multilingual models, geometric analysis, representation geometry

1. Introduction

Code-mixing is the phenomenon where speakers alternate between two or more languages within a single utterance. It is pervasive in multilingual communities worldwide. Hindi-English code-mixing (Hinglish) alone occurs in hundreds of millions of social media posts, conversational transcripts, and user-generated content annually. Despite this prevalence, NLP systems trained primarily on monolingual data perform poorly on code-mixed text, necessitating domain-specific adaptation.

Parameter-efficient fine-tuning (PEFT) methods, particularly Low-Rank Adaptation (Hu et al., 2022), have emerged as effective techniques for adapting large pre-trained models to code-mixed tasks. LoRA constrains weight updates to low-rank matrices, typically with ranks $r = 8$–16, achieving performance comparable to full fine-tuning while training only 0.1–1% of model parameters. Empirical studies report that LoRA performs particularly well on multilingual and code-mixed benchmarks, often with lower ranks than required for monolingual tasks (Pfeiffer et al., 2021). However, the mechanisms underlying this efficiency remain unclear. Prior work has characterized LoRA's optimization dynamics and parameter budgets (Hu et al., 2022), but has not explained why code-mixed text enables such substantial rank reduction.

We currently lack a principled understanding of when and why low-rank adaptation suffices. Does LoRA's effectiveness for code-mixing reflect properties of the task, the data, or the representational geometry of the pre-trained model? Can optimal rank be predicted from measurable properties of the adaptation domain? Without mechanistic insight, rank selection remains an empirical trial-and-error process, and generalization to new language pairs or tasks is uncertain.

Hypothesis. We propose that LoRA's efficiency for code-mixing arises from geometric structure in multilingual pre-trained models. Specifically, we hypothesize that code-mixed text induces cross-lingual subspace overlap, such that code-mixed representations occupy a low-dimensional manifold within the pre-trained representation space. If this manifold is well represented during pre-training, then fine-tuning updates need not learn new high-dimensional representations but can instead refine existing structure using low-rank perturbations. This contrasts with monolingual adaptation, where task-specific updates may require access to larger representational capacity.

To test this hypothesis, we adopt a geometric perspective combining spectral analysis and representation similarity metrics. We analyze pre-trained attention weights to establish baseline dimensionality, measure cross-lingual subspace alignment using CKA (Kornblith et al., 2019), and compute effective ranks of learned LoRA updates to quantify empirical compression. By connecting geometric structure (subspace overlap) to adaptation efficiency (low-rank sufficiency), we provide a mechanistic account of LoRA's effectiveness for code-mixing.

Proceedings of the Second workshop on Challenges in Processing South Asian Languages (CHiPSAL2026) @ LREC, pages 127–136
16 May 2026.

Scope and Contributions. We focus on Hindi-English code-mixing (Hinglish) in the COMI-LINGUA corpus Sheth et al. (2025), with Hindi as the matrix language. Our task is token-level language identification using BERT-based encoder models, specifically mBERT (Devlin et al., 2019) and MuRIL (Khanuja et al., 2021). We do not claim generalization to all language pairs, code-mixing configurations, or NLP tasks; rather, we provide an analysis framework applicable to other scenarios. Our contributions are:

- Empirical characterization of intrinsic dimensionality in pre-trained multilingual models, with effective ranks of 437–441 across attention weights and approximately 90% of spectral energy concentrated in $\sim$301 dimensions.
- Geometric evidence for cross-lingual subspace overlap: Hinglish embeddings exhibit 3.77× stronger alignment with Hindi (CKA=0.279) than a Hindi-English monolingual baseline (CKA=0.074), confirmed via permutation tests.
- Spectral analysis of trained LoRA adapters, showing effective ranks of 2.1–5.9 (mean=3.9), corresponding to a 136× average compression factor relative to pre-trained weights, with statistical validation (Wilcoxon test, $p < 10^{-19}$).
- A methodological framework connecting representation geometry to adaptation efficiency, demonstrating convergent evidence across spectral decomposition, cross-lingual similarity metrics, and empirical rank measurements.

Our analysis is strictly descriptive for the Hinglish language identification setting but yields testable predictions: language pairs with higher pre-trained cross-lingual alignment should permit lower LoRA ranks. We discuss limitations, including focus on a single language pair, a single task, and modest dataset size (1,666 samples). Extensions to other language pairs, sentence-level tasks, and generation scenarios remain future work.

2. Related Work

2.1. Parameter-Efficient Fine-Tuning and LoRA

Low-Rank Adaptation (Hu et al., 2022) constrains fine-tuning updates to low-rank matrices $\Delta W = BA$, where $B \in \mathbb{R}^{d\times r}$ and $A \in \mathbb{R}^{r\times d}$ with $r \ll d$. This reduces trainable parameters by up to four orders of magnitude while maintaining task performance comparable to full fine-tuning. Empirical evaluations demonstrate LoRA's effectiveness across NLP tasks (Dettmers et al., 2023), vision-language models (Liu et al., 2023), and multilingual settings (Dang et al., 2024), but focus primarily on task performance rather than mechanistic explanations for why low-rank updates suffice.

Concurrent PEFT methods include adapter layers (Houlsby et al., 2019), prefix-tuning (Li and Liang, 2021), and BitFit (Ben Zaken et al., 2022). Comparative studies (He et al., 2022) show LoRA achieves favorable efficiency-performance tradeoffs for large models. Recent work has analyzed LoRA's optimization dynamics (Malladi et al., 2023), showing low-rank constraints induce implicit regularization. These analyses explain how LoRA optimizes but not why certain domains, such as code-mixing, permit substantially lower ranks. Our geometric analysis connects domain properties to rank requirements.

2.2. Code-Mixing and Multilingual NLP

Code-mixing research has established linguistic theories of matrix-embedded language structure (Myers-Scotton, 1993; Muysken, 2000) and computational models for code-switching detection (Solorio et al., 2014; Aguilar et al., 2020). For Hindi-English, prior work includes language identification (Rijhwani et al., 2017; Khanuja et al., 2020), sentiment analysis (Prabhu et al., 2016), and named entity recognition (Singh et al., 2018). Shared tasks such as GLUECoS (Khanuja et al., 2020) and datasets such as LinCE (Aguilar et al., 2020) have standardized evaluation.

Several studies observe that multilingual models transfer effectively to code-mixed data with minimal fine-tuning (Winata et al., 2021; Doğruöz et al., 2021), often attributing this to generic "multilingual knowledge" without quantifying mechanisms. Gupta et al. (2020) show that synthetic code-mixing during pre-training improves downstream performance, suggesting pre-trained representations already encode cross-lingual structure. However, these works do not analyze representational geometry or connect transfer behavior to low-rank adaptation efficiency.

2.3. Representation Geometry and Spectral Analysis

Canonical Correlation Analysis (CCA) (Raghu et al., 2017) and CKA (Kornblith et al., 2019) have been widely used to compare representations across layers, models, and modalities. Morcos et al. (2018) show CCA reveals shared structure in independently trained networks, while Nguyen et al. (2021) use CKA to study cross-lingual alignment in multilingual models. These studies focus on descriptive characterization rather than linking geometry to adaptation efficiency.

Spectral analysis of neural network weights has revealed low-rank structure in trained models (Martin and Mahoney, 2019; Sharma et al., 2020). Aghajanyan et al. (2021) demonstrate that fine-tuning predominantly occurs in low-dimensional subspaces, introducing intrinsic dimensionality as the minimal subspace required for task performance. Their approach estimates intrinsic dimension via random projections but does not relate it to pre-trained geometric structure or cross-lingual properties. Recent work by Biderman et al. (2024) analyzes LoRA using singular value decomposition, showing learned updates align with low-frequency components of pre-trained weights. These studies do not address why code-mixing specifically enables stronger alignment or lower rank requirements than monolingual tasks.

Our Contribution. Existing literature establishes that (1) LoRA is effective for code-mixing, (2) multilingual models transfer well to code-mixed data, and (3) neural representations exhibit structured geometry. However, these strands remain disconnected: prior work does not link cross-lingual geometric properties, such as subspace overlap measured via CKA, to LoRA rank requirements for code-mixing. We demonstrate that pre-trained cross-lingual alignment predicts and explains low-rank sufficiency through convergent geometric and spectral evidence. Our findings complement optimization analyses (Malladi et al., 2023) that explain how LoRA converges by explaining why convergence requires low rank for code-mixing specifically.

3. Methodology

3.1. Dataset and Task

We use the COMI-LINGUA corpus Sheth et al. (2025), a Hindi-English code-mixed dataset with token-level language identification annotations. The corpus consists of naturally occurring social media text, where Hindi (Devanagari script) functions as the matrix language with embedded English (Latin script) spans. We use the language identification (LID) configuration, which provides three-way token-level labels: {hi, en, other}.

To ensure data quality and controlled language mixing, we apply inclusion criteria: (1) sentences contain $\geq$5 tokens, (2) $\geq$2 Hindi and $\geq$2 English tokens present, (3) Hindi token proportion between 30–70% to exclude near-monolingual samples, and (4) maximum sequence length 128 tokens. We exclude sentences where all English tokens have length $\leq$3 characters (typically abbreviations). Sentences with more than 95% tokens from a single language are segregated into monolingual control sets (Hindi-only and English-only) for cross-lingual geometry analysis. After filtering, the dataset contains 1,666 samples, split deterministically (seed=42) into training (80%), development (10%), and test (10%) partitions (Table 1).

Table 1: Dataset statistics for COMI-LINGUA after filtering.

Split	Samples	Avg. Length (tokens)	Avg. Hindi (%)
Train	1332	12.8	61.2
Dev	166	13.1	61.7
Test	168	12.6	61.2
Total	1666	12.8	61.2

Language identification is formulated as token-level sequence labeling. Each input token receives one of three labels: Hindi (hi), English (en), or other. We report macro-averaged F1 as the primary metric to ensure balanced treatment of all classes, given that Hindi tokens constitute approximately 61% of the corpus.

3.2. Model Selection

We analyze two multilingual transformer encoder models: `bert-base-multilingual-cased` (Devlin et al., 2019) and `google/muril-base-cased` (Khanuja et al., 2021). Both models have 12 transformer layers with hidden dimensionality 768. mBERT serves as a general multilingual baseline, while MuRIL represents an Indic-language-specialized model, allowing assessment of whether observed phenomena generalize across different pre-training objectives and data distributions.

3.3. Spectral Analysis of Pre-trained Weights

We extract Query and Key projection matrices ($W \in \mathbb{R}^{768\times768}$) from all 12 transformer layers, as these are the primary targets of LoRA adaptation in encoder-only architectures. Value matrices are included during LoRA fine-tuning (Section 3.5) but excluded from baseline spectral analysis. For each matrix W, we compute the full singular value decomposition $W = U\Sigma V^\top$ and retain all 768 singular values.

We quantify intrinsic dimensionality using the entropy-based effective rank:

$$r_{\text{eff}} = \frac{\left(\sum_i \sigma_i\right)^2}{\sum_i \sigma_i^2}, \tag{1}$$

where σ_i denotes singular values in descending order. A rank-k matrix yields $r_{\text{eff}} = k$, while a full-rank matrix with uniform singular values yields $r_{\text{eff}} \approx d$. As a complementary measure, we compute the 90% energy rank:

$$r_{90} = \min\left\{k : \frac{\sum_{i=1}^{k} \sigma_i^2}{\sum_{i=1}^{d} \sigma_i^2} \geq 0.9\right\}. \tag{2}$$

3.4. Cross-Lingual Representation Geometry

For geometric analysis, we extract sentence-level representations from the final hidden layer using mean pooling over token embeddings, excluding padding and weighted by attention masks. We sample 300 instances for each condition: (1) Hinglish (code-mixed), (2) Hindi-only tokens extracted from code-mixed sentences, and (3) English-only tokens extracted from the same sentences. This ensures controlled comparison, differing only in retained tokens.

We use CKA (Kornblith et al., 2019) to measure cross-lingual subspace similarity. Given representations $X, Y \in \mathbb{R}^{n \times d}$, CKA is:

$$\text{CKA}(X,Y) = \frac{\text{HSIC}(X,Y)}{\sqrt{\text{HSIC}(X,X)\,\text{HSIC}(Y,Y)}}, \quad (3)$$

where HSIC is the Hilbert-Schmidt Independence Criterion computed using linear kernels:

$$\text{HSIC}(X,Y) = \frac{1}{n^2}\text{tr}(KHLH), \quad (4)$$

with $K = XX^\top$, $L = YY^\top$, and $H = I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$. CKA is invariant to orthogonal transformations and isotropic scaling. We compute 95% bootstrap confidence intervals using 1,000 resampling iterations (seed=42).

To verify that observed CKA values reflect structured cross-lingual overlap rather than chance alignment, we conduct a permutation test by randomly shuffling row correspondence between Hinglish and Hindi embeddings while preserving individual representations. CCA is excluded due to saturation artifacts ($>$0.99 for all pairs), Procrustes distance due to scale sensitivity, and PCA visualizations explain $<$25% of variance.

3.5. LoRA Fine-tuning

We train Low-Rank Adaptation (Hu et al., 2022) adapters end-to-end on the token-level LID task using mBERT as the base model. We evaluate ranks $r \in \{4, 8, 16\}$ with scaling factor $\alpha = 2r$ (ratio $\alpha/r = 2$). Target modules include Query, Key, and Value projection layers. LoRA dropout is 0.1, and no bias terms are trained.

We use AdamW optimizer (Loshchilov and Hutter, 2017) with learning rate 2×10^{-4}, weight decay 0.01, and batch size 16. Training proceeds for three epochs with early stopping based on development-set macro F1. All experiments use seed 42, and mixed-precision training (fp16) is enabled when supported.

After training, we extract learned LoRA matrices $A \in \mathbb{R}^{r \times 768}$ and $B \in \mathbb{R}^{768 \times r}$. The effective update matrix is:

$$\Delta W = (\alpha/r)BA = 2BA, \quad (5)$$

incorporating inference-time scaling. Singular value decomposition is applied to each ΔW, and effective rank and 90% energy rank are computed using Equations 1 and 2. This yields 36 matrices per rank (12 layers × 3 modules), totaling 108 matrices.

We use the Wilcoxon signed-rank test to compare effective ranks of pre-trained weights ($n = 48$ matrices from mBERT and MuRIL Query/Key) with LoRA effective ranks ($n = 108$ matrices across all ranks and modules). This non-parametric test is appropriate given unknown distributions and modest sample sizes. Two-tailed p-values are reported with significance threshold $\alpha = 0.001$.

Table 2: Effective rank and 90% energy rank for pre-trained Query and Key matrices. (Std: Standard Deviation across matrices)

Model	Mean Eff. Rank	Std	Mean 90% Rank
mBERT	436.8	33.5	300.5
MuRIL	441.0	31.4	302.5
Overall	438.9	32.5	301.5

4. Results

4.1. Pre-trained Weight Spectral Analysis

We analyzed Query and Key projection matrices across all 12 transformer layers in mBERT and MuRIL. Singular value decomposition was performed on weight matrices of shape 768×768, and effective rank was computed using Equation 1.

Pre-trained attention weights exhibit mean effective ranks of 436.8 ± 33.5 (mBERT) and 441.0 ± 31.4 (MuRIL) across Query and Key matrices, with a combined average of 438.9 ± 32.5 ($n = 48$ matrices). Effective ranks range from 376.6 (mBERT Key, Layer 0) to 495.1 (MuRIL Query, Layer 3). The 90% energy rank averages 300.5 ± 26.9 (mBERT) and 302.5 ± 24.4 (MuRIL), with an overall mean of 301.5 ± 25.6, indicating that approximately 39% of dimensions (301/768) account for 90% of the total spectral energy (Table 2).

Effective ranks show no systematic monotonic trend across layers. The coefficient of variation is 7.7% for mBERT and 7.1% for MuRIL, indicating modest within-model variability. Query matrices have higher mean effective ranks than Key matrices in mBERT (448.6 vs. 425.1), whereas MuRIL exhibits near parity between Query and Key matrices (442.1 vs. 439.9). The difference in mean effective rank between models is small (4.2; 0.95% of the mean) relative to within-model variability, suggesting highly similar spectral structure despite differences in pre-training objectives and data (Figure 1).

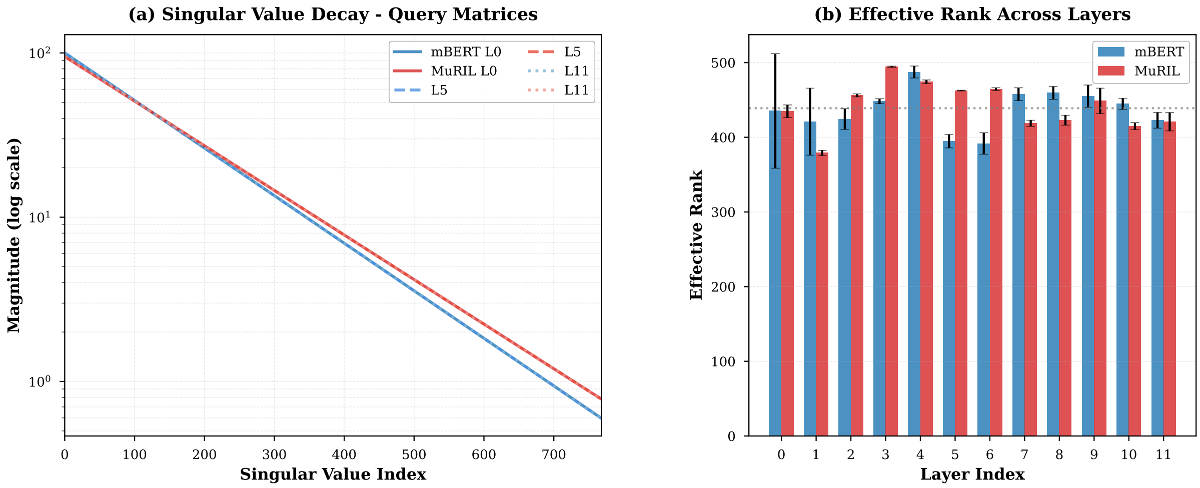

Figure 1: Spectral properties of pre-trained multilingual models. (Left) Singular value decay of Query projection matrices, illustrating smooth log-linear behavior indicative of high intrinsic dimensionality. (Right) Effective rank across transformer layers for mBERT and MuRIL, showing consistently high dimensionality with modest layer-wise variation.

4.2. Cross-Lingual Representation Geometry

We extracted sentence-level embeddings using mean pooling over token representations for 300 samples per language condition: (1) Hinglish (code-mixed), (2) Hindi-only, and (3) English-only. Embeddings were obtained from the final hidden states of mBERT and MuRIL. Cross-lingual similarity was quantified using CKA (Equation 3).

Hinglish embeddings exhibit strong alignment with Hindi, with CKA scores of 0.267 (mBERT) and 0.291 (MuRIL), yielding a mean of 0.279. Alignment between Hinglish and English embeddings is substantially weaker, with CKA values of 0.107 (mBERT) and 0.080 (MuRIL), yielding a mean of 0.093. The Hindi–English monolingual baseline yields lower similarity values (0.084 for mBERT and 0.065 for MuRIL; mean 0.074), as summarized in Table 3.

Hinglish–Hindi alignment is approximately 3.77× stronger than the Hindi–English baseline and about 3.0× stronger than Hinglish–English alignment, indicating asymmetric cross-lingual overlap induced by code-mixing. A permutation test (1,000 permutations) that disrupts correspondence between Hinglish and Hindi embeddings yields markedly lower CKA values (0.057 for mBERT and 0.046 for MuRIL), with $p < 0.001$ for the difference between original and permuted conditions. This confirms that the observed alignment reflects structured cross-lingual geometry rather than random similarity (Figure 2, Table 4).

Table 3: Cross-lingual representation similarity measured via CKA for mBERT and MuRIL. Bootstrap standard deviations shown.

Model	Pair	CKA	Standard Deviation
mBERT	Hinglish–Hindi	0.267	0.040
mBERT	Hinglish–English	0.107	0.012
mBERT	Hindi–English	0.084	0.011
MuRIL	Hinglish–Hindi	0.291	0.043
MuRIL	Hinglish–English	0.080	0.011
MuRIL	Hindi–English	0.065	0.009

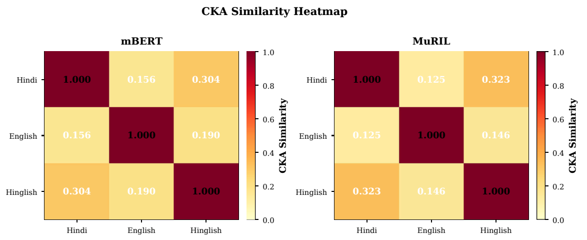

Figure 2: Cross-lingual representation similarity between Hinglish, Hindi, and English embeddings for mBERT and MuRIL, measured using CKA.

Table 4: Permutation ablation: Hinglish–Hindi CKA vs. permuted baseline ($p < 0.001$).

Condition	**mBERT**	**MuRIL**
Hinglish–Hindi (original)	0.267	0.291
Hinglish–Hindi (permuted)	0.057	0.046

Table 5: Token-level LID performance and parameter overhead for LoRA adapters.

r	Macro F1	Weighted F1	Trainable Params
4	0.9365	0.9400	223,491
8	0.9405	0.9438	444,675
16	0.9484	0.9513	887,043

Table 6: Effective ranks of pre-trained attention weights and LoRA update matrices.

r	Pre-trained Eff. Rank	LoRA Eff. Rank	Compression	90% Rank
4	441.6	2.11 ± 0.58	209×	1.6
8	441.6	3.56 ± 1.33	124×	2.6
16	441.6	5.92 ± 2.69	75×	3.8

4.3. LoRA Adaptation: Performance and Spectral Properties

LoRA adapters were trained end-to-end on token-level language identification using the COMI-LINGUA dataset (1,332 training, 166 development, and 168 test samples). Target modules were Query, Key, and Value projections. Adapter ranks $r \in \{4, 8, 16\}$ were evaluated using a scaling factor $\alpha = 2r$.

Test-set performance improves modestly with rank, from macro F1 of 0.937 ($r = 4$) to 0.948 ($r = 16$) which is a gain of only 1.2 percentage points despite doubling trainable parameters from 0.126% to 0.498% of the base model (Table 5, Figure 4).

Learned LoRA matrices were combined into effective update matrices $\Delta W = 2BA$ incorporating inference-time scaling. Mean effective ranks of LoRA updates are 2.11 ± 0.58 ($r = 4$), 3.56 ± 1.33 ($r = 8$), and 5.92 ± 2.69 ($r = 16$), corresponding to compression factors of 209.3×, 124.1×, and 74.6× relative to pre-trained weights (Table 6, Figure 3). A Wilcoxon signed-rank test comparing pre-trained effective ranks ($n = 48$) to LoRA effective ranks ($n = 108$) yields $p = 9.3 \times 10^{-20}$, rejecting the null hypothesis of equal ranks at $\alpha = 0.001$.

5. Discussion

5.1. From Multilingual Pre-training to Low-Rank Sufficiency

Our results provide convergent evidence linking multilingual representation geometry to the empirical efficiency of low-rank adaptation. Spectral analysis shows that pre-trained multilingual encoders allocate a high intrinsic dimensionality to linguistic knowledge, with mean effective ranks of approximately 439 across Query and Key matrices. This reflects the representational demands of encoding over one hundred languages within a shared parameter space.

In contrast, Hindi–English code-mixed text does not appear to require access to this full capacity. Cross-lingual geometry analysis reveals that Hinglish embeddings exhibit substantially stronger alignment with Hindi than with English, with Hinglish–Hindi similarity exceeding the Hindi–English monolingual baseline by a factor of 3.77. This asymmetric alignment indicates that code-mixed representations largely occupy an existing Hindi-dominated subspace rather than forming an independent intermediate space between languages. Permutation-based ablation further supports this interpretation: shuffling sample correspondence reduces Hinglish–Hindi similarity to near-zero levels, consistent with random pairings.

Given this geometric constraint, LoRA fine-tuning can achieve efficient adaptation by operating within a low-dimensional subspace that is already well supported by pre-training. Empirically, learned LoRA updates exhibit effective ranks between 2.1 and 5.9, corresponding to over two orders of magnitude compression relative to pre-trained attention weights, while maintaining strong task performance. The convergence between geometric overlap and empirical compression supports a causal interpretation: low-rank sufficiency arises because code-mixed representations are confined to a structured cross-lingual manifold rather than spanning the full representational space.

5.2. Asymmetric Cross-Lingual Alignment in Code-Mixing

The observed asymmetry between Hinglish–Hindi and Hinglish–English alignment is consistent with the matrix–embedded language structure of Hindi–English code-mixing. In the COMI-LINGUA corpus, Hindi typically serves as the matrix language providing syntactic structure, while English contributes embedded lexical material. As a result, Hindi subsequences extracted from code-mixed sentences retain contextual coherence, whereas English subsequences often reduce to isolated fragments.

This linguistic asymmetry propagates to representation geometry. Code-mixed embeddings preserve structural similarity to Hindi contexts, leading to stronger alignment with Hindi representations. In addition, multilingual pre-training exposes models to substantially more Hindi monolingual data than naturally occurring code-mixed text, resulting in a well-developed Hindi subspace into which Hinglish inputs naturally project. English representations, by contrast, are shaped primarily by monolingual English contexts that differ from embedded English usage in code-mixing, and therefore align

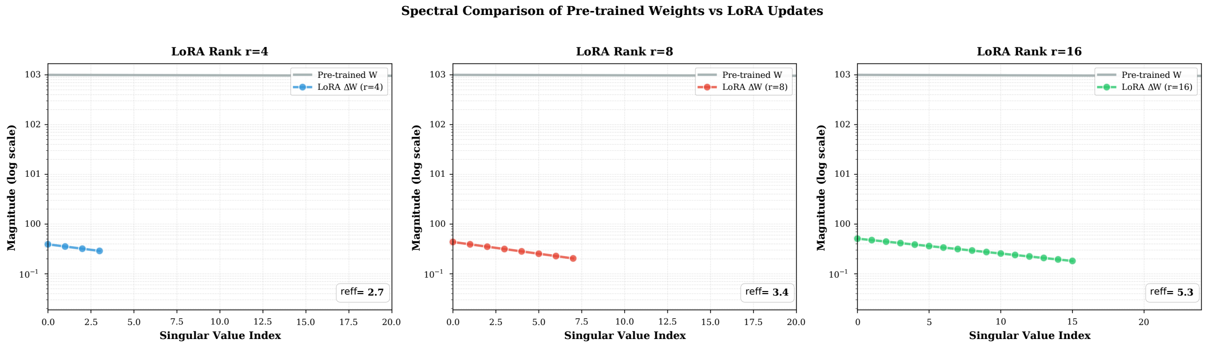

Figure 3: Spectral comparison of pre-trained attention weights and LoRA update matrices for $r = 4$ (left), $r = 8$ (center), and $r = 16$ (right), illustrating concentration of spectral energy in a low-dimensional subspace for LoRA updates.

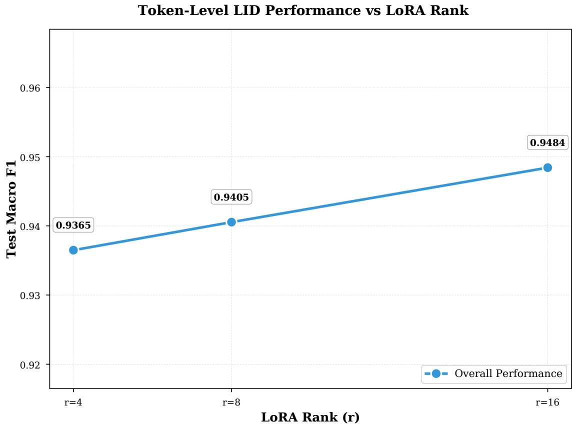

Figure 4: Token-level language identification performance as a function of LoRA rank, showing rapid saturation at low ranks.

less strongly.

We emphasize that this interpretation is specific to Hindi-matrix code-mixing. Scenarios in which English functions as the matrix language may exhibit reversed alignment patterns, and generalization to other language pairs requires empirical validation.

5.3. Implications for Parameter-Efficient Fine-Tuning

Our findings suggest that the effectiveness of LoRA for code-mixing stems from its ability to exploit pre-existing multilingual structure rather than from task-specific optimization alone. This implies that LoRA efficiency should vary systematically with cross-lingual alignment in pre-trained models: language pairs with stronger geometric overlap should admit lower-rank adaptation, while more distant pairs may require higher ranks or alternative methods.

More broadly, the analysis framework developed here linking pre-trained spectral properties, cross-lingual geometry, and adaptation rank may extend beyond code-mixing to other multilingual and domain-adaptation scenarios in which target distributions concentrate within structured subspaces of the pre-trained representation space.

6. Conclusion

We investigated why Low-Rank Adaptation enables efficient fine-tuning for code-mixed text through geometric and spectral analysis of multilingual models. Our results establish a clear link between cross-lingual representational structure and low-rank sufficiency: while pre-trained models exhibit high intrinsic dimensionality, Hindi–English code-mixed representations concentrate in a much lower dimensional subspace with substantially stronger alignment to Hindi than to English.

This geometric constraint enables LoRA adapters to achieve extreme compression while maintaining strong task performance. The convergence of spectral analysis, cross-lingual similarity measurements, and empirical adaptation results supports a unified interpretation: LoRA succeeds by refining pre-existing multilingual structure rather than learning new high-dimensional representations. Permutation-based ablations confirm that the observed alignment reflects semantic structure rather than artifacts, and consistency across models suggests that this phenomenon arises from fundamental properties of multilingual pre-training when Hindi serves as the matrix language.

Our analysis is demonstrated for token-level language identification on Hindi–English code-mixing. Whether similar alignment compression relationships hold across other language pairs, tasks, and model architectures remains an open question. Nevertheless, the framework presented here offers a mechanistic perspective on parameter-efficient fine-tuning grounded in representational geometry, rather than purely empirical rank selection.

7. Ethics Statement

This work analyzes publicly available multilingual language models and datasets to study parameter-efficient fine-tuning behavior. The COMI-LINGUA dataset consists of anonymized, publicly released social media text and does not contain personally identifiable information. No new data involving human subjects was collected.

While improved modeling of code-mixed language may benefit multilingual NLP applications, such systems may still reflect biases present in pre-training data. We therefore caution against deployment without appropriate evaluation of fairness, bias, and societal impact in downstream settings.

8. Limitations

Language Pair and Mixing Direction. Our analysis is restricted to Hindi–English code-mixing where Hindi serves as the matrix language. Other language pairs or mixing directions may exhibit different representational geometry and adaptation behavior. Findings are most likely to generalize to other Indo-Aryan pairs where one language dominates as the matrix language (e.g., Hindi–Urdu, Hindi–Marathi), though empirical validation is required.

Task Scope. We focus on token-level language identification, a relatively constrained classification task. More complex sentence-level or generative tasks may require higher-dimensional adaptations than those observed here. Token-level language identification was selected as it provides clean, interpretable geometric signal with direct per-token ground-truth labels. Other applicable scenarios include multilingual domain adaptation and cross-lingual transfer between typologically similar languages.

Dataset Size. The filtered dataset contains 1,666 samples, limiting statistical power for detecting small effects and for analyzing rare mixing patterns. Larger datasets would enable more fine-grained analysis. However, CKA and spectral decomposition operate on embedding matrices rather than trained classifiers, and bootstrap confidence intervals confirm stability of the similarity estimates.

Correlation vs. Causation. Although we present convergent evidence linking cross-lingual alignment to low-rank adaptation efficiency, the analysis remains correlational. We do not perform interventional experiments that explicitly manipulate representation geometry.

Model Coverage. Experiments are limited to two encoder-only multilingual models. Extending the analysis to larger or decoder-only architectures may reveal additional factors influencing low-rank adaptation behavior. XLM-R and LaBSE are natural candidates for future analysis; LaBSE's alignment-optimized pre-training may exhibit higher CKA values, which our framework predicts would correlate with lower LoRA ranks. Extending this analysis to decoder-only architectures such as Llama or Mistral, and to billion-parameter models, remains open.

9. Acknowledgements

This work was carried out using publicly available pre-trained models and open-source libraries, including PyTorch, HuggingFace Transformers, and PEFT. Large language models were used solely for language polishing and final editorial refinement of the manuscript. All scientific content, experiments, and conclusions are the authors' own.

10. Data and Code Availability

The code for all experiments, including spectral analysis, CKA computation, and LoRA fine-tuning, is available at `https://github.com/shakuvish/LoRA-Geometric`. The COMI-LINGUA dataset used in this study is publicly available via Sheth et al. (2025).

11. Bibliographical References

Armen Aghajanyan, Sonal Gupta, and Luke Zettlemoyer. 2021. Intrinsic dimensionality explains the effectiveness of language model fine-tuning. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7319–7328, Online. Association for Computational Linguistics.

Gustavo Aguilar, Sudipta Kar, and Thamar Solorio. 2020. LinCE: A centralized benchmark for linguistic code-switching evaluation. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 1803–1813, Marseille, France. European Language Resources Association.

Elad Ben Zaken, Yoav Goldberg, and Shauli Ravfogel. 2022. BitFit: Simple parameter-efficient fine-tuning for transformer-based masked language-models. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1–9, Dublin, Ireland. Association for Computational Linguistics.

Dan Biderman, Jose Javier Gonzalez Ortiz, Ranganath Jacob, Matthew Sullivan, Sainbayar Sukhbaatar, Martin Wistuba, et al. 2024. LoRA learns less and forgets less. *arXiv preprint arXiv:2405.09673*.

John Dang, Arash Ahmadian, Kelly Marchisio, Julia Kreutzer, Ahmet Üstün, and Sara Hooker. 2024. Rlhf can speak many languages: Unlocking multilingual preference optimization for llms. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 13134–13156, Miami, Florida, USA. Association for Computational Linguistics.

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2023. QLoRA: Efficient finetuning of quantized LLMs. In *Advances in Neural Information Processing Systems*, volume 36, pages 10088–10115.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

A. Seza Doğruöz, Sunayana Sitaram, Barbara E. Bullock, and Almeida Jacqueline Toribio. 2021. A survey of code-switching: Linguistic and social perspectives for language technologies. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1654–1666, Online. Association for Computational Linguistics.

Deepak Gupta, Asif Ekbal, and Pushpak Bhattacharyya. 2020. A semi-supervised approach to generate the code-mixed text using pre-trained encoder and transfer learning. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2267–2280, Online. Association for Computational Linguistics.

Junxian He, Chunting Zhou, Xuezhe Ma, Taylor Berg-Kirkpatrick, and Graham Neubig. 2022. Towards a unified view of parameter-efficient transfer learning. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. 2019. Parameter-efficient transfer learning for NLP. In *Proceedings of the 36th International Conference on Machine Learning*, pages 2790–2799. PMLR.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-rank adaptation of large language models. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Simran Khanuja, Diksha Bansal, Sarvesh Mehtani, Savya Khosla, Atreyee Dey, Balaji Gopalan, Dilip Kumar Agrawal, Pooja Hase, Pushpak Gupta, Palak Singh, et al. 2021. MuRIL: Multilingual representations for Indian languages. *arXiv preprint arXiv:2103.10730*.

Simran Khanuja, Sandipan Dandapat, Anirudh Srinivasan, Sunayana Sitaram, and Monojit Choudhury. 2020. GLUECoS: An evaluation benchmark for code-switched NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3575–3585, Online. Association for Computational Linguistics.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. 2019. Similarity of neural network representations revisited. In *Proceedings of the 36th International Conference on Machine Learning*, pages 3519–3529. PMLR.

Xiang Lisa Li and Percy Liang. 2021. Prefix-tuning: Optimizing continuous prompts for generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4582–4597, Online. Association for Computational Linguistics.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2023. Visual instruction tuning. In *Advances in Neural Information Processing Systems*, volume 36, pages 34892–34916.

Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. In *International Conference on Learning Representations*.

Sadhika Malladi, Alexander Wettig, Dingli Yu, Danqi Chen, and Sanjeev Arora. 2023. A kernel-based view of language model fine-tuning. In *Proceedings of the 40th International Conference on Machine Learning*, pages 23610–23641. PMLR.

Charles H. Martin and Michael W. Mahoney. 2019. Traditional and heavy-tailed self regularization in neural network models. In *arXiv preprint*, volume abs/1901.08276.

Ari Morcos, Maithra Raghu, and Samy Bengio. 2018. Insights on representational similarity in neural networks with canonical correlation. In *Advances in Neural Information Processing Systems*, volume 31.

Pieter Muysken. 2000. *Bilingual Speech: A Typology of Code-Mixing*. Cambridge University Press, Cambridge.

Carol Myers-Scotton. 1993. *Duelling Languages: Grammatical Structure in Codeswitching*. Oxford University Press, Oxford.

Minh Van Nguyen, Tuan Ngo Nguyen, Bonan Min, and Thien Huu Nguyen. 2021. Crosslingual transfer learning for relation and event extraction via word category and class alignments. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 5414–5426, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Jonas Pfeiffer, Aishwarya Kamath, Andreas Rücklé, Kyunghyun Cho, and Iryna Gurevych. 2021. Adapterfusion: Non-destructive task composition for transfer learning. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 487–503, Online. Association for Computational Linguistics.

Ameya Prabhu, Aditya Joshi, Manish Shrivastava, and Vasudeva Varma. 2016. Towards sub-word level compositions for sentiment analysis of Hindi-English code mixed text. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2482–2491, Osaka, Japan. The COLING 2016 Organizing Committee.

Maithra Raghu, Justin Gilmer, Jason Yosinski, and Jascha Sohl-Dickstein. 2017. Svcca: Singular vector canonical correlation analysis for deep learning dynamics and interpretability. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Shruti Rijhwani, Royal Sequiera, Monojit Choudhury, Kalika Bali, and Chandra Shekhar Maddila. 2017. Estimating code-switching on Twitter with a novel generalized word-level language detection technique. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1971–1982, Vancouver, Canada. Association for Computational Linguistics.

Kaushal Sharma, Ajit Kembhavi, Aniruddha Kembhavi, T Sivarani, Sheelu Abraham, and Kaustubh Vaghmare. 2020. Application of convolutional neural networks for stellar spectral classification. *Monthly Notices of the Royal Astronomical Society*, 491(2):2280–2300.

Kushagra Singh, Indira Sen, and Ponnurangam Kumaraguru. 2018. Language identification and named entity recognition in Hinglish code mixed tweets. In *Proceedings of ACL 2018, Student Research Workshop*, pages 52–58, Melbourne, Australia. Association for Computational Linguistics.

Thamar Solorio, Elizabeth Blair, Suraj Maharjan, Steven Bethard, Mona Diab, Mahmoud Ghoneim, Abdelati Hawwari, Fahad AlGhamdi, Julia Hirschberg, Alison Chang, and Pascale Fung. 2014. Overview for the first shared task on language identification in code-switched data. In *Proceedings of the First Workshop on Computational Approaches to Code Switching*, pages 62–72, Doha, Qatar. Association for Computational Linguistics.

Genta Indra Winata, Andrea Madotto, Zhaojiang Lin, Rosanne Liu, Jason Yosinski, and Pascale Fung. 2021. Language models are few-shot multilingual learners. In *Proceedings of the 1st Workshop on Multilingual Representation Learning*, pages 1–15, Punta Cana, Dominican Republic. Association for Computational Linguistics.

12. Language Resource References

Sheth, Rajvee and Beniwal, Himanshu and Singh, Mayank. 2025. *COMI-LINGUA: Expert Annotated Large-Scale Dataset for Multitask NLP in Hindi–English Code-Mixing*. Association for Computational Linguistics. PID https://aclanthology.org/2025.findings-emnlp.422/.