

# Novelty Assessment Report

**Paper:** Understanding Post-Training Structural Changes in Large Language Models

**PDF URL:** <https://openreview.net/pdf?id=2ytWHufA5j>

**Venue:** ICLR 2026 Conference Submission

**Year:** 2026

**Report Generated:** 2026-01-05

## Abstract

Post-training fundamentally alters the behavior of large language models (LLMs), yet its impact on the internal parameter space remains poorly understood. In this work, we conduct a systematic singular value decomposition (SVD) analysis of principal linear layers in pretrained LLMs, focusing on two widely adopted post-training methods: instruction tuning and long-chain-of-thought (Long-CoT) distillation. Our analysis reveals two consistent and unexpected structural changes: **(1) a near-uniform geometric scaling of singular values across layers**, which theoretically modulates attention scores; and **(2) highly consistent orthogonal transformations are applied to the left and right singular vectors of each matrix**. Disrupting this orthogonal consistency leads to catastrophic performance degradation. Based on these findings, we propose a simple yet effective framework that interprets post-training as a reparameterization of fixed subspaces in the pretrained parameter space. Further experiments reveal that singular value scaling behaves as a secondary effect, analogous to a temperature adjustment, whereas the core functional transformation lies in the coordinated rotation of singular vectors. These results challenge the prevailing view of the parameter space in large models as a black box, uncovering the first clear regularities in how parameters evolve during training, and providing a new perspective for deeper investigation into model parameter changes.

### Disclaimer

This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.

Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.

If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

## Core Task Landscape

This paper addresses: **Understanding Structural Changes in Large Language Model Parameters During Post-Training**

A total of **50 papers** were analyzed and organized into a taxonomy with **30 categories**.

### Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Geometric and Spectral Analysis of Parameter Space**
- **Parameter-Efficient Fine-Tuning Methods**
- **Full-Model Fine-Tuning Approaches**
- **Compression and Quantization**
- **Catastrophic Phenomena in Fine-Tuning**
- **Specialized Post-Training Techniques**
- **Distributed and Federated Fine-Tuning**
- **Application-Specific Fine-Tuning Studies**
- **Comprehensive Surveys and Reviews**

### Complete Taxonomy Tree

- Understanding Structural Changes in Large Language Model Parameters During Post-Training Survey Taxonomy
- Geometric and Spectral Analysis of Parameter Space
  - Representation Geometry Evolution (1 papers)
    - [1] Tracing the representation geometry of language models from pretraining to post-training (Agrawal, 2025) [View paper](#)
    - Singular Value Decomposition Analysis of Parameters ★ (1 papers)
    - [0] Understanding Post-Training Structural Changes in Large Language Models (Anon et al., 2026) [View paper](#)
- Parameter-Efficient Fine-Tuning Methods
  - Low-Rank Adaptation Techniques
  - Standard Low-Rank Adaptation (3 papers)
    - [12] LLM-Adapters: An Adapter Family for Parameter-Efficient Fine-Tuning of Large Language Models (Zhiqiang Hu, 2023) [View paper](#)
    - [15] Parameter-efficient fine-tuning methods for pretrained language models: A critical review and assessment (Xu Lingling, 2023) [View paper](#)
    - [25] Parameter-efficient fine-tuning of large-scale pre-trained language models (Ning Ding, 2023) [View paper](#)
  - Adaptive Low-Rank Methods (2 papers)
    - [6] Adaptive Budget Allocation for Parameter-Efficient Fine-Tuning (Zhang QingRu, 2023) [View paper](#)
    - [17] InceLoRA: Incremental Parameter Allocation Method for Parameter-Efficient Fine-tuning (Zhang Fei-yu, 2023) [View paper](#)
  - Prefix and Adapter-Based Methods (1 papers)
  - [22] Towards Adaptive Prefix Tuning for Parameter-Efficient Language Model Fine-tuning (Huang, 2023) [View paper](#)
  - Subspace Optimization Frameworks (1 papers)
  - [37] Parameter-Efficient Subspace Optimization for LLM Fine-Tuning (Yuchen Lou, 2025) [View paper](#)
  - PEFT Design Patterns and Comparative Studies (3 papers)
  - [30] Parameter-Efficient Fine-Tuning Design Spaces (Chen, 2023) [View paper](#)
  - [42] On the Effectiveness of Parameter-Efficient Fine-Tuning (Bing, 2023) [View paper](#)

- [45] Towards a Unified View of Parameter-Efficient Transfer Learning (He, 2021) [View paper](#)
- Domain-Specific PEFT Applications (3 papers)
- [16] Layerlink: Bridging remote sensing object detection and large vision models with efficient fine-tuning (Xingkui Zhu, 2025) [View paper](#)
- [32] Exploring parameter-efficient fine-tuning of large language model on automated program repair (Guochang Li, 2024) [View paper](#)
- [35] Parameter-efficient fine-tuning for pre-trained vision models: A survey (Yi Xin, 2024) [View paper](#)
- Full-Model Fine-Tuning Approaches
  - Memory-Efficient Full-Model Optimization (1 papers)
  - [13] Full parameter fine-tuning for large language models with limited resources (Guo, 2024) [View paper](#)
  - Zeroth-Order Optimization Methods (1 papers)
  - [14] Sparse MeZO: Less Parameters for Better Performance in Zeroth-Order LLM Fine-Tuning (Liu Yong, 2024) [View paper](#)
  - Scaling Laws and Fine-Tuning Dynamics (1 papers)
  - [4] When scaling meets llm finetuning: The effect of data, model and finetuning method (Zhang Biao, 2024) [View paper](#)
- Compression and Quantization
  - Post-Training Quantization Methods (4 papers)
  - [26] GPTQ: Accurate Post-Training Quantization for Generative Pre-trained Transformers (Frantar, 2022) [View paper](#)
  - [28] Zeroquant: Efficient and affordable post-training quantization for large-scale transformers (Yao, 2022) [View paper](#)
  - [34] SmoothQuant: Accurate and Efficient Post-Training Quantization for Large Language Models (Xiao, 2022) [View paper](#)
  - [46] Exploring post-training quantization in llms from comprehensive study to low rank compensation (Zhewei Yao, 2024) [View paper](#)
  - Low-Rank Quantization Approaches (1 papers)
  - [3] Lrq: Optimizing post-training quantization for large language models by learning low-rank weight-scaling matrices (Kim Jeonghoon, 2025) [View paper](#)
  - Weight Magnitude Reduction Techniques (1 papers)
  - [19] Magr: Weight magnitude reduction for enhancing post-training quantization (Yanxia Deng, 2024) [View paper](#)
  - Structured Pruning and Decomposition (2 papers)
  - [7] Laco: Large language model pruning via layer collapse (Yang Yi-fei, 2024) [View paper](#)
  - [38] SVD-LLM: Truncation-aware Singular Value Decomposition for Large Language Model Compression (Wang Xin, 2024) [View paper](#)
- Catastrophic Phenomena in Fine-Tuning
  - Catastrophic Forgetting and Mitigation (1 papers)
  - [44] Model tailor: Mitigating catastrophic forgetting in multi-modal large language models (Zhu, 2024) [View paper](#)
  - Overtraining and Performance Degradation (1 papers)
  - [10] Overtrained language models are harder to fine-tune (Springer, 2025) [View paper](#)
  - Safety Alignment Brittleness (1 papers)
  - [21] Assessing the Brittleness of Safety Alignment via Pruning and Low-Rank Modifications (Wei, 2024) [View paper](#)
  - Safety-Aware Fine-Tuning Optimization (1 papers)
  - [48] Rethinking safety in llm fine-tuning: An optimization perspective (Kim, 2025) [View paper](#)
- Specialized Post-Training Techniques
  - Model Unlearning Methods (3 papers)
  - [11] Dissecting fine-tuning unlearning in large language models (Yihuai Hong, 2024) [View paper](#)
  - [24] Unified parameter-efficient unlearning for llms (Ding Chen-lu, 2024) [View paper](#)
  - [41] Rethinking llm unlearning objectives: A gradient perspective and go beyond (Wang Qi-zhou, 2025) [View paper](#)
  - Knowledge Editing Techniques (1 papers)
  - [9] Editing large language models: Problems, methods, and opportunities (Yunzhi Yao, 2023) [View paper](#)
  - Reinforcement Learning Post-Training (1 papers)
  - [43] RL Fine-Tuning Heals OOD Forgetting in SFT (Hangzhan Jin, 2025) [View paper](#)
  - Layer Significance and Pruning Strategies (3 papers)
  - [27] Neural parameter search for slimmer fine-tuned models and better transfer (Du Guodong, 2025) [View paper](#)
  - [29] Singular Value Fine-tuning: Few-shot Segmentation requires Few-parameters Fine-tuning (Sun Yan-peng, 2022) [View paper](#)
  - [31] Understanding Layer Significance in LLM Alignment (Shi Guang-yuan, 2024) [View paper](#)
  - Gradient Field Compression and Semantic Crystallization (1 papers)
  - [47] Semantic crystallization through gradient field compression in large language models (E Whitfield, 2025) [View paper](#)
- Distributed and Federated Fine-Tuning
  - Federated Learning with PEFT (3 papers)
  - [8] Federatedscope-llm: A comprehensive package for fine-tuning large language models in federated learning (Weirui Kuang, 2024) [View paper](#)
  - [20] Splitlora: A split parameter-efficient fine-tuning framework for large language models (Lin Zheng, 2024) [View paper](#)
  - [23] Enhancing Sustainability in LLM Training: Leveraging Federated Learning and Parameter-Efficient Fine-Tuning (Sunbal Iftikhar, 2025) [View paper](#)
  - Resource-Constrained Federated Fine-Tuning (2 papers)
  - [33] MobiLLM: Enabling LLM Fine-Tuning on the Mobile Device via Server Assisted Side Tuning (Li Liang, 2025) [View paper](#)
  - [49] Fedbiot: Llm local fine-tuning in federated learning without full model (Feijie Wu, 2024) [View paper](#)
- Application-Specific Fine-Tuning Studies
  - Personalization and User Adaptation (1 papers)
  - [2] Democratizing large language models via personalized parameter-efficient fine-tuning (Jiang Meng, 2024) [View paper](#)
  - Text Annotation and Classification Tasks (1 papers)
  - [5] Open-source LLMs for text annotation: a practical guide for model setting and fine-tuning (MaÅ«l Kubli, 2025) [View paper](#)
  - Code-Related Task Applications (2 papers)
  - [36] Exploring the Capabilities of LLMs for Code-Change-Related Tasks (Lishui Fan, 2025) [View paper](#)
  - [50] Fine-tuning llms for code mutation: A new era of cyber threats (Mohammad Setak, 2024) [View paper](#)

- Comprehensive Surveys and Reviews (3 papers)
  - [18] Parameter-efficient fine-tuning for large models: A comprehensive survey (Han Zeyu, 2024) [View paper](#)
  - [39] Parameter-efficient fine-tuning in large language models: a survey of methodologies (Luping Wang, 2025) [View paper](#)
  - [40] llm fine-tuning: Instruction and parameterefficient fine-tuning (peft) (S Aathilakshmi, 2024) [View paper](#)

## Narrative

Core task: Understanding structural changes in large language model parameters during post-training. The field has organized itself around several complementary perspectives on how LLM parameters evolve after initial pretraining. At the highest level, researchers pursue geometric and spectral analyses that probe the intrinsic structure of weight matrices—often through singular value decomposition or subspace characterizations—to reveal how fine-tuning reshapes parameter distributions. In parallel, a large body of work focuses on parameter-efficient fine-tuning (PEFT) methods such as LoRA and prefix tuning, which modify only a small subset of weights while preserving most pretrained knowledge. Full-model fine-tuning approaches, by contrast, update all parameters and tend to yield stronger task performance at higher computational cost. Additional branches address compression and quantization (e.g., GPTQ[26], SmoothQuant[34]) to reduce memory footprints, catastrophic phenomena like forgetting and alignment brittleness, specialized techniques for domains such as code or vision, and distributed or federated settings that must coordinate updates across multiple nodes. Comprehensive surveys (e.g., PEFT Comprehensive Survey[18], PEFT Methodologies Survey[39]) synthesize these threads, highlighting trade-offs between efficiency, performance, and robustness.

Several active lines of work reveal contrasting priorities: some studies emphasize low-rank decompositions to isolate which subspaces carry task-relevant information (Singular Value Finetuning[29], SVD-LLM[38]), while others explore how gradient flow and representation geometry shift during adaptation (Representation Geometry Tracing[1], Subspace Optimization[37]). Post-Training Structural Changes[0] sits squarely within the geometric and spectral analysis branch, using singular value decomposition to track how weight matrices evolve across fine-tuning stages. Its emphasis on decomposing parameter updates into interpretable components aligns closely with works like Singular Value Finetuning[29] and SVD-LLM[38], which similarly leverage spectral methods to understand or guide adaptation. By contrast, nearby efforts in PEFT (e.g., LLM-Adapters[12], PEFT Design Spaces[30]) prioritize practical efficiency over deep structural insight, while full-model studies (Full Parameter Finetuning[13]) accept higher costs for maximal expressiveness. Open questions remain about how these structural signatures relate to downstream robustness, generalization, and the risk of catastrophic forgetting.

## Related Works in Same Category

No sibling papers were found in the same taxonomy leaf. A taxonomy-subtopic-level comparison will be produced instead.

### Taxonomy-Level Summary

Both subtopics analyze structural properties of neural networks during training phases, but at different levels of abstraction. The original leaf focuses on parameter-level matrix decomposition to understand weight transformations, while the sibling examines representation-level geometric properties. They complement each other by providing insights into how low-level parameter changes manifest in high-level representation geometry.

**Similarities:** - Both use spectral analysis methods (SVD for parameters, eigenspectrum for representations) to characterize structural properties - Both investigate changes occurring during post-training phases rather than just pretraining - Both aim to understand how training processes transform model structure in interpretable ways

**Differences:** - Original leaf analyzes weight matrices directly through SVD, while sibling analyzes learned representations/activations - Original leaf focuses on singular value scaling and orthogonal transformations of parameters, while sibling examines effective rank and eigenspectrum decay of representation spaces - Original leaf operates at the parameter level (model weights), while sibling operates at the activation/representation level (model outputs) - Original leaf explicitly excludes compression applications, focusing on structural understanding, while sibling focuses on geometric properties without compression considerations

**Suggested Search Directions:** - Connections between parameter-level SVD changes and representation geometry evolution - How singular value scaling in weight matrices affects effective rank of representations - Joint analysis frameworks that link parameter transformations to representation space properties

### Sibling Subtopics

- **Representation Geometry Evolution** (leaves: 1, papers: 1)
- Scope: Investigates geometric properties of learned representations across pretraining and post-training using spectral measures like effective rank and eigenspectrum decay.
- Exclude: Excludes parameter-level structural analysis; see Singular Value Decomposition Analysis of Parameters.

## Contributions Analysis

**Overall novelty summary.** The paper conducts systematic singular value decomposition analysis of weight matrices during post-training, revealing two structural patterns: near-uniform geometric scaling of singular values and consistent orthogonal transformations of singular vectors. It resides in the 'Singular Value Decomposition Analysis of Parameters' leaf, which currently contains only this paper within the broader 'Geometric and Spectral Analysis of Parameter Space' branch. This represents a relatively sparse research direction focused specifically on spectral methods for understanding post-training dynamics, distinct from the more crowded parameter-efficient fine-tuning methodologies that dominate the field.

The taxonomy shows the paper sits within a small geometric analysis branch (two leaves total) that contrasts sharply with the heavily populated parameter-efficient fine-tuning subtree containing over twenty papers across multiple leaves. The neighboring 'Representation Geometry Evolution' leaf examines learned representations rather than parameter-level structure, while the broader field emphasizes practical adaptation methods (LoRA variants, adapters, quantization) over structural interpretation. The paper's focus on SVD-based parameter analysis positions it at the intersection of theoretical understanding and post-training mechanics, bridging geometric insights with practical fine-tuning outcomes.

Among thirty candidates examined, the contribution-level analysis reveals mixed novelty signals. The systematic SVD analysis revealing structural changes (Contribution 1) examined ten candidates with zero refutations, suggesting this specific dual-pattern characterization may be novel. However, the mathematical framework interpreting post-training as subspace reparameterization (Contribution 2) found two refutable candidates among ten examined, indicating prior work on subspace-based interpretations exists. The claim of being the first systematic study across entire parameter space (Contribution 3) encountered one refutable candidate, suggesting similar comprehensive analyses may have been conducted previously.

Based on the limited search scope of thirty semantically similar papers, the work appears to offer genuine insights into SVD-based structural patterns during post-training, particularly the dual observation of singular value scaling and orthogonal consistency. The subspace reparameterization framework and systematic scope claims face more substantial prior work overlap. The sparse taxonomy leaf suggests this specific analytical approach remains underexplored, though the existence of refutable candidates indicates the broader conceptual territory has been partially mapped by earlier efforts.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

## Contribution 1: Systematic SVD analysis revealing two structural changes in post-training

**Description:** The authors conduct a systematic singular value decomposition analysis of principal linear layers in pretrained LLMs, uncovering two consistent structural phenomena that occur during post-training: near-uniform geometric scaling of singular values and highly consistent orthogonal transformations of singular vectors.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

### 1. Spectral Adapter: Fine-Tuning in Spectral Space

URL: [View paper](#)

#### Brief Assessment

Spectral Adapter[64] focuses on parameter-efficient fine-tuning via SVD-based spectral adaptation mechanisms (additive tuning and orthogonal rotation of top singular vectors), not on analyzing structural changes across the entire parameter space during post-training. The candidate does not demonstrate prior work on discovering near-uniform geometric scaling of singular values or highly consistent orthogonal transformations as systematic phenomena in post-trained LLMs.

---

### 2. Orthogonal low rank embedding stabilization

URL: [View paper](#)

#### Brief Assessment

Orthogonal Low Rank[62] applies SVD to stabilize recommendation system embeddings across retraining cycles, not to analyze structural changes in LLM post-training. The candidate focuses on embedding space stabilization for recommendation systems, while the original analyzes weight matrix transformations in language models.

---

### 3. Harnessing Orthogonality to Train Low-Rank Neural Networks

URL: [View paper](#)

#### Brief Assessment

Harnessing Orthogonality[69] focuses on orthogonal basis stabilization during general neural network training for low-rank optimization, not on post-training structural changes in LLMs or the specific phenomena of geometric singular value scaling and consistent orthogonal transformations.

---

### 4. Neural Network Layer Matrix Decomposition reveals Latent Manifold Encoding and Memory Capacity

URL: [View paper](#)

#### Brief Assessment

Layer Matrix Decomposition[66] focuses on SVD of neural network weight matrices to reveal latent manifold encoding and memory capacity in general neural networks. The original paper specifically analyzes structural changes (geometric scaling and orthogonal transformations) that occur during post-training of large language models, which is a distinct research focus not addressed in the candidate paper.

---

### 5. Semi-orthogonal low-rank matrix factorization for deep neural networks.

URL: [View paper](#)

#### Brief Assessment

Semi-orthogonal Factorization[67] applies SVD-based factorization to compress TDNN layers for speech recognition, not to analyze structural changes in LLM post-training. The candidate focuses on training efficiency through semi-orthogonal constraints, while the original analyzes geometric scaling and orthogonal transformations that emerge during LLM post-training.

---

### 6. A review on weight initialization strategies for neural networks

URL: [View paper](#)

#### Brief Assessment

Weight Initialization Review[60] discusses SVD decomposition for weight initialization strategies in neural networks, focusing on orthonormal basis and dynamical isometry conditions. This is fundamentally different from the original paper's systematic analysis of structural changes (geometric scaling and orthogonal transformations) that occur during post-training of pretrained LLMs.

---

### 7. Biological learning of irreducible representations of commuting transformations

URL: [View paper](#)

#### Brief Assessment

Biological Irreducible Representations[68] focuses on biological learning mechanisms for image transformations using SVD of bivectors and PCA of time differences in neural networks. The original paper analyzes SVD of weight matrices in pretrained LLMs to reveal structural changes during post-training. These are fundamentally different applications of SVD in distinct domains (biological neural learning vs. LLM parameter analysis).

---

### 8. Orthogonal binary singular value decomposition method for automated windshield wiper fault detection

URL: [View paper](#)

#### Brief Assessment

Orthogonal Binary SVD[61] applies SVD to audio signal processing for windshield wiper fault detection, not to analyzing structural changes in neural network weight matrices during post-training. The domains and applications are entirely different.

---

### 9. Orthogonal Constrained Neural Networks for Solving Structured Inverse Eigenvalue Problems

URL: [View paper](#)

#### Brief Assessment

Orthogonal Constrained Networks[65] focuses on solving structured inverse eigenvalue problems using neural networks with orthogonal constraints, not on analyzing structural changes in LLMs during post-training via SVD.

---

### 10. CURE: Concept Unlearning via Orthogonal Representation Editing in Diffusion Models

URL: [View paper](#)

#### Brief Assessment

CURE[63] applies SVD to diffusion model embeddings for concept unlearning in text-to-image generation, not to analyze structural changes in LLM weight matrices during post-training. The technical domains and objectives are fundamentally different.

---

## Contribution 2: Mathematical framework interpreting post-training as subspace reparameterization

**Description:** The authors propose a mathematical framework that describes post-training as a reparameterization process operating on fixed subspaces in the pretrained parameter space, providing a new perspective for understanding parameter evolution during training.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

### 1. Safe LoRA: the Silver Lining of Reducing Safety Risks when Fine-tuning Large Language Models

URL: [View paper](#)

#### Brief Assessment

Safe LoRA[74] focuses on projecting LoRA weights to safety-aligned subspaces to mitigate safety risks during fine-tuning, not on interpreting post-training as subspace reparameterization in the parameter space.

---

### 2. Delta tuning: A comprehensive study of parameter efficient methods for pre-trained language models

URL: [View paper](#)

#### Prior Art Analysis

Delta Tuning[51] demonstrates that parameter-efficient tuning methods operate through reparameterization in low-dimensional subspaces of pretrained models. The paper explicitly frames delta tuning as optimization within intrinsic subspaces and provides theoretical frameworks showing that adaptations can be reparameterized into lower-dimensional spaces. This directly challenges the novelty claim that the original paper is the first to propose interpreting post-training as subspace reparameterization.

#### Evidence

Evidence 1 - **Rationale:** Both papers describe parameter-efficient tuning as a reparameterization process. Delta Tuning[51] explicitly categorizes reparameterization-based methods as transforming parameters into efficient forms, which is conceptually identical to the original paper's framework of post-training as reparameterization. - **Original:** we propose a simple yet effective framework that interprets post-training as a reparameterization of fixed subspaces in the pretrained parameter space - **Candidate:** reparameterization-based methods transform the adaptive parameters during optimization into parameter-efficient forms. this branch of delta tuning is typically motivated by the hypothesis that plm adaptations towards most downstream tasks are inherently low-rank, and could thus be equivalently compl..

Evidence 2 - **Rationale:** Delta Tuning[51] provides a mathematical framework showing that parameter updates can be represented as transformations in low-dimensional subspaces ( $\theta = \psi(\delta) + \epsilon$ ), which is a formal mathematical description of reparameterization in fixed subspaces, predating the original paper's similar framework. - **Original:** post-training essentially induces highly regular structural perturbations in the parameter space. based on the two observed phenomena, we can use a simple yet effective mathematical framework to directly approximate the impact of post-training on the parameter space - **Candidate:** low dimensional representation in solution space. as it is observed that the optimization trajectory of  $\theta$  approximately follows a manifold (aghajanyan et al., 2021), we can embed the hidden manifold to a low dimensional space of  $\delta$ , i.e.,  $\theta = \psi(\delta) + \epsilon$ , where  $\epsilon$  is the error term depending on  $\theta, \delta$ .

---

### 3. SVDiff: Compact Parameter Space for Diffusion Fine-Tuning

URL: [View paper](#)

#### Brief Assessment

SVDiff[70] focuses on fine-tuning singular values for diffusion model personalization in text-to-image generation, not on interpreting post-training as subspace reparameterization in LLMs. The technical domains and objectives are fundamentally different.

---

### 4. Adaptive Budget Allocation for Parameter-Efficient Fine-Tuning

URL: [View paper](#)

#### Brief Assessment

Adaptive Budget Allocation[6] focuses on parameter-efficient fine-tuning through adaptive rank allocation in low-rank decomposition, not on interpreting post-training as subspace reparameterization in pretrained parameter space.

---

### 5. Parameter-efficient fine-tuning of large language models via deconvolution in subspace

URL: [View paper](#)

#### Brief Assessment

Deconvolution Subspace[72] focuses on parameter-efficient fine-tuning via deconvolution operations in subspace for adapting pretrained models to downstream tasks, not on interpreting post-training as reparameterization of fixed subspaces in the pretrained parameter space.

---

### 6. Residual Prompt Tuning: Improving Prompt Tuning with Residual Reparameterization

URL: [View paper](#)

#### Brief Assessment

Residual Prompt Tuning[73] focuses on reparameterizing soft prompt embeddings through a shallow network for parameter-efficient fine-tuning, not on interpreting post-training as subspace reparameterization in pretrained parameter space. The candidate addresses prompt tuning optimization, while the original analyzes structural changes in weight matrices via SVD during post-training.

---

### 7. Revisiting Fine-Tuning: A Survey of Parameter-Efficient Techniques for Large AI Models

URL: [View paper](#)

#### Brief Assessment

Revisiting Finetuning Survey[75] focuses on parameter-efficient fine-tuning (PEFT) techniques like adapters and LoRA for reducing computational costs, not on interpreting post-training as subspace reparameterization in pretrained parameter space.

---

### 8. Intrinsic Dimensionality Explains the Effectiveness of Language Model Fine-Tuning

URL: [View paper](#)

#### Prior Art Analysis

Intrinsic Dimensionality[52] demonstrates that fine-tuning can be understood as optimization in a low-dimensional subspace through reparameterization. The paper shows that fine-tuning operates via  $\theta_d = \theta_{d0} + P(\theta_d)$ , where  $P$  projects from a lower-dimensional space back into the full parameter space. This directly establishes that post-training (fine-tuning) can be mathematically described as a reparameterization process in fixed subspaces, predating the original paper's claim of novelty.

#### Evidence

Evidence 1 - **Rationale:** Intrinsic Dimensionality[52] establishes that fine-tuning can be understood as optimization in fixed lower-dimensional subspaces of the pretrained parameter space, which is the core mathematical framework the original paper claims to

introduce. - **Original:** based on these findings, we propose a simple yet effective framework that interprets post-training as a reparameterization of fixed subspaces in the pretrained parameter space - **Candidate:** intuitively, we project using an arbitrary random projection onto a much smaller space; usually, a linear projection, we then solve the optimization problem in that smaller subspace. if we reach a satisfactory solution, we say the dimensionality of that subspace is the intrinsic dimension

Evidence 2 - **Rationale:** Both papers provide mathematical frameworks for understanding how post-training/fine-tuning affects parameters. Intrinsic Dimensionality[52] interprets the intrinsic parameter vector as encoding tasks within the pretrained framework, establishing a mathematical perspective on parameter changes during fine-tuning. - **Original:** we can use a simple yet effective mathematical framework to directly approximate the impact of post-training on the parameter space - **Candidate:** one interpretation of the intrinsic parameter vector is that it encodes the task at hand with respect to the original pre-trained representations. therefore, we can interpret das the minimal description length of the task within the framework dictated by the pretrained representations

---

## 9. Robust and Efficient Fine-tuning of LLMs with Bayesian Reparameterization of Low-Rank Adaptation

URL: [View paper](#)

### Brief Assessment

Bayesian LoRA[71] focuses on low-rank adaptation parameterization for fine-tuning stability, not on interpreting post-training as subspace reparameterization in pretrained parameter space. The candidate addresses variance reduction in LoRA parameters, while the original analyzes structural changes via SVD across the entire parameter space.

---

## 10. Parameter-efficient model adaptation for vision transformers

URL: [View paper](#)

### Brief Assessment

PEFT Vision Transformers[76] focuses on parameter-efficient adaptation of vision transformers for image classification, formulating it as subspace training. The original paper addresses post-training structural changes in LLMs through SVD analysis, proposing reparameterization of fixed subspaces in pretrained parameter space. These are distinct problem domains and methodological approaches.

---

## Contribution 3: First systematic study of structural changes across entire parameter space

**Description:** The authors present the first comprehensive analysis of how post-training affects the entire parameter space of LLMs, examining singular value structures of principal linear layers rather than focusing on individual neurons or external behaviors as in prior work.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

### 1. Structure-learnable adapter fine-tuning for parameter-efficient large language models

URL: [View paper](#)

#### Brief Assessment

Structure-learnable Adapter[54] focuses on adapter-based fine-tuning with learnable insertion points and routing mechanisms, not on systematic SVD analysis of structural parameter changes across the entire model. The candidate addresses parameter efficiency through dynamic adapter structures rather than analyzing how post-training fundamentally alters the parameter space of LLMs.

---

### 2. LoRA: Low-Rank Adaptation of Large Language Models

URL: [View paper](#)

#### Brief Assessment

LoRA[55] focuses on efficient fine-tuning via low-rank decomposition for parameter reduction, not on analyzing structural changes across the entire parameter space during post-training.

---

### 3. Maintaining structural integrity in parameter spaces for parameter efficient fine-tuning

URL: [View paper](#)

#### Brief Assessment

Structural Integrity PEFT[53] focuses on preserving topological structure during parameter-efficient fine-tuning via low-rank tensor adaptation, not on analyzing how post-training affects the entire parameter space of LLMs through SVD analysis of singular value structures.

---

### 4. Delta tuning: A comprehensive study of parameter efficient methods for pre-trained language models

URL: [View paper](#)

#### Prior Art Analysis

Delta Tuning[51] presents a comprehensive study analyzing structural changes in pre-trained language models during parameter-efficient tuning. The paper systematically examines how different delta tuning methods affect the parameter space across multiple model architectures and tasks, including analysis of weight matrices and their transformations. This comprehensive analysis of parameter space changes predates the original paper's claim of being the first systematic study.

#### Evidence

Evidence 1 - **Rationale:** Delta Tuning[51] provides a comprehensive survey and systematic study of parameter-efficient methods, examining how different approaches affect the parameter space of pre-trained models, which challenges the original paper's claim of being the first systematic study. - **Original:** to the best of our knowledge, this is the first systematic study of structural changes in llms before/after post-training across the entire parameter space.unlike prior works focusing on individual neuron activations or external behaviors, we comprehensively analyze the singular value structure of p... - **Candidate:** in this paper, we first attempt to survey the development and recent advances in delta tuning. for preliminaries, we give a description of the transformer neural models and mainstream plms (§2: preliminaries ). then we formally describe the delta tuning problem and provide a categorization criterion...

Evidence 2 - **Rationale:** Delta Tuning[51] conducts systematic experiments across over 100 tasks examining structural properties and behaviors of parameter-efficient methods across different scales and architectures, demonstrating a comprehensive analysis of parameter space changes that predates the original paper. - **Original:** we experimentally discover two structural phenomena that are stable across multiple model families, parameter sizes, and training methods: first, the singular values exhibit near-uniform geometric scaling; second, the left and right singular vectors of each matrix remain stable under consistent orth... - **Candidate:** in terms of empirical studies, we carry out extensive and systematic experiments (§5: experiments ) on over 100 nlp tasks to rigorously explore the performances (§5.1: performance ), combinability (§5.2: combination ), the power of scale (§5.3: scale ), transferability (§5.4: transferability ), etc.

Evidence 3 - **Rationale:** Delta Tuning[51] systematically analyzes weight matrices and structural changes across different model types and scales, examining how parameter-efficient methods affect the entire parameter space, which overlaps with the original paper's claimed systematic study of structural changes. - **Original:** we apply this framework to the weight matrices within the self-attention

modules and feed-forward networks of publicly available models, and categorize models into three types: basemodels (e.g., qwen2.5-math-1.5b (qwen et al., 2025)), instructmodels (obtained through instruction tuning), and reasoningmod... - **Candidate:** we choose t5base (raffel et al., 2019) as the mainly evaluated plm backbone for different tuning methods, and we additionally report the performance of pt with t5large (raffel et al., 2019). for both models, we use the checkpoints released by lester et al. (2021), who conducted additional 100k steps...

---

## 5. Towards Adaptive Prefix Tuning for Parameter-Efficient Language Model Fine-tuning

URL: [View paper](#)

### Brief Assessment

Adaptive Prefix Tuning[22] focuses on parameter-efficient fine-tuning through prefix tuning mechanisms, not on systematic SVD analysis of structural parameter space changes across entire linear layers during post-training.

---

## 6. Sam-parser: Fine-tuning sam efficiently by parameter space reconstruction

URL: [View paper](#)

### Brief Assessment

SAM-Parser[58] focuses on parameter space reconstruction for fine-tuning SAM models using SVD decomposition, not on analyzing structural changes during post-training of LLMs. The technical domains and objectives differ fundamentally.

---

## 7. VL-ADAPTER: Parameter-Efficient Transfer Learning for Vision-and-Language Tasks

URL: [View paper](#)

### Brief Assessment

VL-ADAPTER[56] focuses on parameter-efficient transfer learning for vision-and-language tasks using adapter modules, not on analyzing structural parameter space changes in language models during fine-tuning. The candidate examines adapter-based training methods for V&L tasks, while the original contribution concerns systematic SVD analysis of how post-training affects the entire parameter space of LLMs.

---

## 8. Mix-of-Language-Experts Architecture for Multilingual Programming

URL: [View paper](#)

### Brief Assessment

Mix-of-Language-Experts[59] focuses on multilingual programming architecture design using LoRA modules, not on systematic analysis of structural parameter space changes during fine-tuning.

---

## 9. Less is More: Local Intrinsic Dimensions of Contextual Language Models

URL: [View paper](#)

### Brief Assessment

Local Intrinsic Dimensions[57] focuses on analyzing local intrinsic dimensions of contextual embeddings during training/fine-tuning, not on systematic SVD analysis of weight matrix parameter spaces as in the original paper.

---

## 10. Intrinsic Dimensionality Explains the Effectiveness of Language Model Fine-Tuning

URL: [View paper](#)

### Brief Assessment

Intrinsic Dimensionality[52] focuses on measuring the dimensionality of optimization subspaces during fine-tuning, not on analyzing structural changes like singular value distributions or orthogonal transformations across the entire parameter space as the original paper does.

---

## Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

---

## References

- [0] Understanding Post-Training Structural Changes in Large Language Models [View paper](#)
- [1] Tracing the representation geometry of language models from pretraining to post-training [View paper](#)
- [2] Democratizing large language models via personalized parameter-efficient fine-tuning [View paper](#)
- [3] Lrq: Optimizing post-training quantization for large language models by learning low-rank weight-scaling matrices [View paper](#)
- [4] When scaling meets llm finetuning: The effect of data, model and finetuning method [View paper](#)
- [5] Open-source LLMs for text annotation: a practical guide for model setting and fine-tuning [View paper](#)
- [6] Adaptive Budget Allocation for Parameter-Efficient Fine-Tuning [View paper](#)
- [7] Laco: Large language model pruning via layer collapse [View paper](#)
- [8] Federatedscope-llm: A comprehensive package for fine-tuning large language models in federated learning [View paper](#)
- [9] Editing large language models: Problems, methods, and opportunities [View paper](#)
- [10] Overtrained language models are harder to fine-tune [View paper](#)
- [11] Dissecting fine-tuning unlearning in large language models [View paper](#)
- [12] LLM-Adapters: An Adapter Family for Parameter-Efficient Fine-Tuning of Large Language Models [View paper](#)
- [13] Full parameter fine-tuning for large language models with limited resources [View paper](#)
- [14] Sparse MeZO: Less Parameters for Better Performance in Zeroth-Order LLM Fine-Tuning [View paper](#)
- [15] Parameter-efficient fine-tuning methods for pretrained language models: A critical review and assessment [View paper](#)
- [16] Layerlink: Bridging remote sensing object detection and large vision models with efficient fine-tuning [View paper](#)
- [17] IncreLoRA: Incremental Parameter Allocation Method for Parameter-Efficient Fine-tuning [View paper](#)
- [18] Parameter-efficient fine-tuning for large models: A comprehensive survey [View paper](#)
- [19] Magr: Weight magnitude reduction for enhancing post-training quantization [View paper](#)
- [20] Splitlora: A split parameter-efficient fine-tuning framework for large language models [View paper](#)
- [21] Assessing the Brittleness of Safety Alignment via Pruning and Low-Rank Modifications [View paper](#)
- [22] Towards Adaptive Prefix Tuning for Parameter-Efficient Language Model Fine-tuning [View paper](#)
- [23] Enhancing Sustainability in LLM Training: Leveraging Federated Learning and Parameter-Efficient Fine-Tuning [View paper](#)
- [24] Unified parameter-efficient unlearning for llms [View paper](#)
- [25] Parameter-efficient fine-tuning of large-scale pre-trained language models [View paper](#)

- [26] GPTQ: Accurate Post-Training Quantization for Generative Pre-trained Transformers [View paper](#)
- [27] Neural parameter search for slimmer fine-tuned models and better transfer [View paper](#)
- [28] Zeroquant: Efficient and affordable post-training quantization for large-scale transformers [View paper](#)
- [29] Singular Value Fine-tuning: Few-shot Segmentation requires Few-parameters Fine-tuning [View paper](#)
- [30] Parameter-Efficient Fine-Tuning Design Spaces [View paper](#)
- [31] Understanding Layer Significance in LLM Alignment [View paper](#)
- [32] Exploring parameter-efficient fine-tuning of large language model on automated program repair [View paper](#)
- [33] MobiLLM: Enabling LLM Fine-Tuning on the Mobile Device via Server Assisted Side Tuning [View paper](#)
- [34] SmoothQuant: Accurate and Efficient Post-Training Quantization for Large Language Models [View paper](#)
- [35] Parameter-efficient fine-tuning for pre-trained vision models: A survey [View paper](#)
- [36] Exploring the Capabilities of LLMs for Code-Change-Related Tasks [View paper](#)
- [37] Parameter-Efficient Subspace Optimization for LLM Fine-Tuning [View paper](#)
- [38] SVD-LLM: Truncation-aware Singular Value Decomposition for Large Language Model Compression [View paper](#)
- [39] Parameter-efficient fine-tuning in large language models: a survey of methodologies [View paper](#)
- [40] llm fine-tuning: Instruction and parameterefficient fine-tuning (peft) [View paper](#)
- [41] Rethinking llm unlearning objectives: A gradient perspective and go beyond [View paper](#)
- [42] On the Effectiveness of Parameter-Efficient Fine-Tuning [View paper](#)
- [43] RL Fine-Tuning Heals OOD Forgetting in SFT [View paper](#)
- [44] Model tailor: Mitigating catastrophic forgetting in multi-modal large language models [View paper](#)
- [45] Towards a Unified View of Parameter-Efficient Transfer Learning [View paper](#)
- [46] Exploring post-training quantization in llms from comprehensive study to low rank compensation [View paper](#)
- [47] Semantic crystallization through gradient field compression in large language models [View paper](#)
- [48] Rethinking safety in llm fine-tuning: An optimization perspective [View paper](#)
- [49] Fedbiot: Llm local fine-tuning in federated learning without full model [View paper](#)
- [50] Fine-tuning llms for code mutation: A new era of cyber threats [View paper](#)
- [51] Delta tuning: A comprehensive study of parameter efficient methods for pre-trained language models [View paper](#)
- [52] Intrinsic Dimensionality Explains the Effectiveness of Language Model Fine-Tuning [View paper](#)
- [53] Maintaining structural integrity in parameter spaces for parameter efficient fine-tuning [View paper](#)
- [54] Structure-learnable adapter fine-tuning for parameter-efficient large language models [View paper](#)
- [55] LoRA: Low-Rank Adaptation of Large Language Models [View paper](#)
- [56] VL-ADAPTER: Parameter-Efficient Transfer Learning for Vision-and-Language Tasks [View paper](#)
- [57] Less is More: Local Intrinsic Dimensions of Contextual Language Models [View paper](#)
- [58] Sam-parser: Fine-tuning sam efficiently by parameter space reconstruction [View paper](#)
- [59] Mix-of-Language-Experts Architecture for Multilingual Programming [View paper](#)
- [60] A review on weight initialization strategies for neural networks [View paper](#)
- [61] Orthogonal binary singular value decomposition method for automated windshield wiper fault detection [View paper](#)
- [62] Orthogonal low rank embedding stabilization [View paper](#)
- [63] CURE: Concept Unlearning via Orthogonal Representation Editing in Diffusion Models [View paper](#)
- [64] Spectral Adapter: Fine-Tuning in Spectral Space [View paper](#)
- [65] Orthogonal Constrained Neural Networks for Solving Structured Inverse Eigenvalue Problems [View paper](#)
- [66] Neural Network Layer Matrix Decomposition reveals Latent Manifold Encoding and Memory Capacity [View paper](#)
- [67] Semi-orthogonal low-rank matrix factorization for deep neural networks. [View paper](#)
- [68] Biological learning of irreducible representations of commuting transformations [View paper](#)
- [69] Harnessing Orthogonality to Train Low-Rank Neural Networks [View paper](#)
- [70] SVDiff: Compact Parameter Space for Diffusion Fine-Tuning [View paper](#)
- [71] Robust and Efficient Fine-tuning of LLMs with Bayesian Reparameterization of Low-Rank Adaptation [View paper](#)
- [72] Parameter-efficient fine-tuning of large language models via deconvolution in subspace [View paper](#)
- [73] Residual Prompt Tuning: Improving Prompt Tuning with Residual Reparameterization [View paper](#)
- [74] Safe LoRA: the Silver Lining of Reducing Safety Risks when Fine-tuning Large Language Models [View paper](#)
- [75] Revisiting Fine-Tuning: A Survey of Parameter-Efficient Techniques for Large AI Models [View paper](#)
- [76] Parameter-efficient model adaptation for vision transformers [View paper](#)