BEHAVIOR QUOTIENT LEARNING FOR LOW-RANK ADAPTATION OF LLM AGENTS

Anonymous authors
Paper under double-blind review

ABSTRACT

LLM-based agents require diverse interaction capabilities to complete complex tasks. Existing approaches often assign these capabilities to multiple LoRA adapters, increasing storage requirements and routing overhead during inference. A single LoRA adapter simplifies deployment but poses two challenges under a fixed rank budget. First, different training trajectories may still reinforce the same behaviors, making part of the supervision redundant. Second, a low-rank approximation can closely match the desired parameter changes while distorting the intended effects on agent decisions. To address these challenges, we propose `BQ-LoRA`, which adapts LLM agents through two modules, i.e., behavior quotient balancing (BQB) and decision preserving compression (DPC). BQB constructs a quotient manifold from representations used to predict agent decisions. Trajectory gradients are then reweighted by the local density of their responses in the quotient tangent space to reduce the dominance of redundant behavioral effects. DPC projects the balanced direction onto the fixed-rank tangent space and compresses the resulting target to fit the original rank budget. This compression jointly controls parameter approximation error and first-order distortion of decision representations. Experiments demonstrate the effectiveness of `BQ-LoRA` in improving agent performance.

1 INTRODUCTION

Agents built on large language models (LLMs) interact with external environments to accomplish specified goals (Yao et al., 2023; Zhang et al., 2026a). Despite broad capabilities, LLM-based agents often require task-specific training to operate effectively in new interaction settings (Zeng et al., 2024; Fang et al., 2026b). Low-Rank Adaptation (LoRA) (Hu et al., 2022) freezes the backbone and learns low-rank task adaptations, reducing the cost of full-model fine-tuning.

LoRA-based agent adaptation broadly follows two designs. The first decomposes agent behavior into distinct capabilities and assigns them to **multiple adapters**. This decomposition separates reasoning, execution, and summarization (Han et al., 2025), decouples reasoning from tool use (Li et al., 2026c), or divides tool calling into tool selection and per-tool argument generation (Kadekodi et al., 2026). Beyond these predefined divisions, dedicated adapters can also be trained to address capability gaps identified from interaction trajectories (Kang et al., 2026). Such specialization requires coordination during inference and can increase serving costs when multiple agents maintain separate attention caches (Jeon et al., 2026). The second learns agent capabilities jointly with a **single adapter**. The adapter is fine-tuned on interaction trajectories, with supervision from different interaction steps updating the same low-rank parameters (Chen et al., 2023; Qian et al., 2025; Lai et al., 2025). More generally, single-adapter LoRA variants make better use of the trainable parameter budget by improving how the adapter is parameterized and optimized (Li et al., 2025b; Zhang & Pilanci, 2024; Zheng & Wu, 2026).

However, behavioral redundancy among trajectory updates can limit how effectively a single fixed-rank adapter learns from diverse supervision. Behavioral redundancy arises when different trajectory updates induce similar changes in decision representations at the same reference histories. Trajectory synthesis broadens interaction coverage (Chen et al., 2026a), while trajectory curation improves supervision quality (Wang et al., 2026). Training utility is also assessed through model-dependent signals (Fang et al., 2026a; Xia et al., 2024; Kwon et al., 2024) and critical-step analysis (Chen et al.,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

2025; Li et al., 2026a). These methods identify useful supervision but do not compare the behavioral effects of different trajectory updates. Behavioral equivalence can occur despite substantial differences in trajectory text and parameter gradients. Under a fixed rank budget, repeated equivalent updates can disproportionately shape the learned adapter and leave updates with distinct effects on these reference decisions underrepresented. This gap raises a central question. ***How can a single fixed-rank LoRA learn diverse agent behaviors?***

Addressing this question presents two challenges. Firstly, *how to identify behaviorally equivalent trajectory updates?* (**CH1**) Recorded trajectories alone do not reveal whether their training updates are behaviorally equivalent. Different interaction traces and parameter gradients may still induce similar changes in decision representations. TopoCurate (Yang et al., 2026b) estimates semantic equivalence from similarities between recorded tool actions and environment observations. However, this criterion does not compare how trajectory updates change decision representations. Identifying behavioral equivalence therefore requires comparing these changes at the same reference histories. Secondly, *how to preserve changes in decision representations under a fixed LoRA rank constraint?* (**CH2**) Aggregating diverse trajectory directions can produce a target adaptation matrix whose rank exceeds this constraint. Representing this target with a single fixed-rank adapter necessarily introduces approximation error. Matrix approximation error alone does not capture how compression distorts these changes. A suitable fixed-rank representative must preserve the effects of the target on decision representations and remain invariant to equivalent LoRA factorizations.

To address these challenges, we propose `BQ-LoRA`, a framework that learns diverse agent behaviors through a single fixed-rank LoRA adapter. Our key idea is to use a shared geometry of decision behavior to guide both trajectory weighting and low-rank compression. For **CH1**, we develop behavior quotient balancing (BQB), which compares trajectory update directions through their first-order effects on normalized decision representations at fixed interaction histories. This comparison identifies locally equivalent behavioral changes while disregarding parameter differences that are invisible to the reference decisions. BQB then reweights the directions according to their local concentration, reducing repeated contributions and giving greater relative influence to distinct behavioral changes. For **CH2**, we develop decision preserving compression (DPC), which projects the balanced direction onto the tangent space of the current fixed-rank adapter. The resulting target depends on the current adaptation matrices and remains invariant to equivalent LoRA factorizations. DPC compresses this target by jointly controlling matrix approximation error and first-order decision distortion in the same behavior geometry. Together, the modules balance behavioral coverage and preserve decision effects under the rank constraint, producing one deployable adapter.

We summarize our contributions as follows: (1) We formalize behavioral equivalence among updates through a local quotient geometry and introduce BQB to balance trajectory contributions according to their effects on agent decisions. (2) We develop DPC, which combines a factorization-invariant tangent projection with compression that accounts for both matrix approximation error and first-order decision distortion under the original LoRA rank budget. (3) We evaluate `BQ-LoRA` on AppWorld and BrowseComp-Plus with two backbone sizes, comparing performance and interaction counts across adaptation methods and trajectory curation.

2 Related Work

2.1 Low-Rank Adaptation

Low-rank adaptation (LoRA) freezes pretrained parameters and represents task-specific adaptation matrices using trainable low-rank factors (Hu et al., 2022). Existing LoRA variants improve adaptation through informative initialization (Meng et al., 2024; Wang et al., 2024; Zhang et al., 2025b;c; Feng et al., 2026; Zhang et al., 2026b), adaptive rank allocation (Zhang et al., 2023; he et al., 2025; Paischer et al., 2025; Cui et al., 2026; Liu et al., 2026a; Ye et al., 2026), modified training procedures (Hayou et al., 2024; Wang et al., 2025; Chen et al., 2026b), and input-conditioned utilization of the low-rank subspace (Borse et al., 2024; Li et al., 2025a; Yang et al., 2026a; Zou et al., 2025). Geometry-aware methods further improve LoRA optimization through Riemannian or transformation-invariant preconditioning (Zhang & Pilanci, 2024; Yen et al., 2025), factor refactorization and balancing (Zhang et al., 2025a; Castin et al., 2026), and manifold-based formulations (Zheng & Wu, 2026; Bogachev et al., 2026; Li et al., 2025b; Lion et al., 2025). Multi-LoRA approaches use multiple low-rank parameter sets. General-purpose research studies the routing and

composition of LoRA experts for multi-task adaptation and adapter reuse (Huang et al., 2024; wu et al., 2024; Dou et al., 2024; Liu & Luo, 2024; Liao et al., 2025; Xu et al., 2025; Li et al., 2026b; Zhuang et al., 2026). Recent work on agentic systems uses multiple specialized adapters to model heterogeneous capabilities and coordinate their activation across agent trajectories (Han et al., 2025; Liu et al., 2026b; Kadekodi et al., 2026; Li et al., 2026c; Kang et al., 2026).

2.2 Trajectory Redundancy

Agent trajectories with different interaction traces may provide overlapping training signals, making trajectory redundancy an important consideration in agent post-training. Existing work has primarily addressed this issue as a data-quality problem, using diversity-aware synthesis (Chen et al., 2026a) and failure-driven curation (Wang et al., 2026) to improve the composition of training data before optimization. Moving from data composition to training effects, prior work estimates sample utility from optimization signals (Fang et al., 2026a; Xia et al., 2024; Kwon et al., 2024) and localizes useful supervision through critical-step selection (Chen et al., 2025; Li et al., 2026a). These criteria identify useful supervision at different levels, but they do not determine whether distinct trajectory-induced updates produce equivalent changes in agent behavior. TopoCurate (Yang et al., 2026b) examines this equivalence through a semantic quotient topology based on embedding similarities between tool actions and environment observations. Because this relation is defined over recorded interaction content, TopoCurate does not determine whether different training updates induce the same behavioral change. Interface perturbation studies further show that successful task execution does not necessarily reflect semantic tool-use competence (Gu et al., 2026).

3 Preliminaries

A quotient manifold is obtained by identifying equivalent points in a smooth total space (Boumal, 2023). Let $\overline{\mathcal{M}}$ be a smooth manifold equipped with an equivalence relation $\sim$. For $p \in \overline{\mathcal{M}}$, its equivalence class and the resulting quotient set are defined as:

$$[p] = \left\{ q \in \overline{\mathcal{M}} : q \sim p \right\}, \qquad \mathcal{M} = \overline{\mathcal{M}}/\sim = \left\{ [p] : p \in \overline{\mathcal{M}} \right\}. \tag{1}$$

The canonical projection $\pi : \overline{\mathcal{M}} \to \mathcal{M}$ is given by $\pi(p) = [p]$. Equipped with a smooth structure, $\mathcal{M}$ is a quotient manifold of $\overline{\mathcal{M}}$ if π is a smooth submersion, i.e., $\operatorname{rank} \mathrm{D}\pi(p) = \dim \mathcal{M}$ for every $p \in \overline{\mathcal{M}}$, where $\mathrm{D}\pi(p) : T_p\overline{\mathcal{M}} \to T_{[p]}\mathcal{M}$ denotes the differential of π at p. The fiber through p is $\mathcal{F}_p = \pi^{-1}(\pi(p))$, which is an embedded submanifold of $\overline{\mathcal{M}}$. Its tangent space at p defines the vertical space $\mathcal{V}_p = T_p\mathcal{F}_p = \ker \mathrm{D}\pi(p)$. Consequently, two tangent vectors $\boldsymbol{u}, \boldsymbol{v} \in T_p\overline{\mathcal{M}}$ represent the same tangent vector in $T_{[p]}\mathcal{M}$ if and only if their difference is vertical:

$$\mathrm{D}\pi(p)[\boldsymbol{u}] = \mathrm{D}\pi(p)[\boldsymbol{v}] \quad \Longleftrightarrow \quad \boldsymbol{u} - \boldsymbol{v} \in \mathcal{V}_p. \tag{2}$$

Additional concepts on quotient manifolds are provided in Appendix A.

4 Method

4.1 Problem Statement

We consider an LLM-based agent with LoRA applied to L parameter matrices. For each layer l, LoRA adds a low-rank adaptation matrix to the frozen pretrained matrix $\boldsymbol{W}_l^0$:

$$\boldsymbol{W}_l = \boldsymbol{W}_l^0 + \boldsymbol{X}_l, \qquad \boldsymbol{X}_l = s_l \boldsymbol{B}_l \boldsymbol{A}_l. \tag{3}$$

The factors $\boldsymbol{A}_l \in \mathbb{R}^{r_l \times d_l^{\text{in}}}$ and $\boldsymbol{B}_l \in \mathbb{R}^{d_l^{\text{out}} \times r_l}$ are trainable, and $s_l > 0$ is a scaling coefficient. Their inner dimension r_l defines the fixed rank budget and ensures $\operatorname{rank}(\boldsymbol{X}_l) \le r_l$.

An agent trajectory is denoted by $\tau_i = \left(x_i, (h_{ik}, y_{ik}, o_{ik})_{k=1}^{T_i}\right)$, where x_i is the task, h_{ik} is the history before step k, y_{ik} is the agent output, and o_{ik} is the environment feedback. Each output y_{ik} consists of intermediate reasoning and a decision a_{ik}. The decision either advances the interaction or serves as the final task output. We collect the adaptation matrices as $\boldsymbol{X} = (\boldsymbol{X}_l)_{l=1}^{L}$. Given N

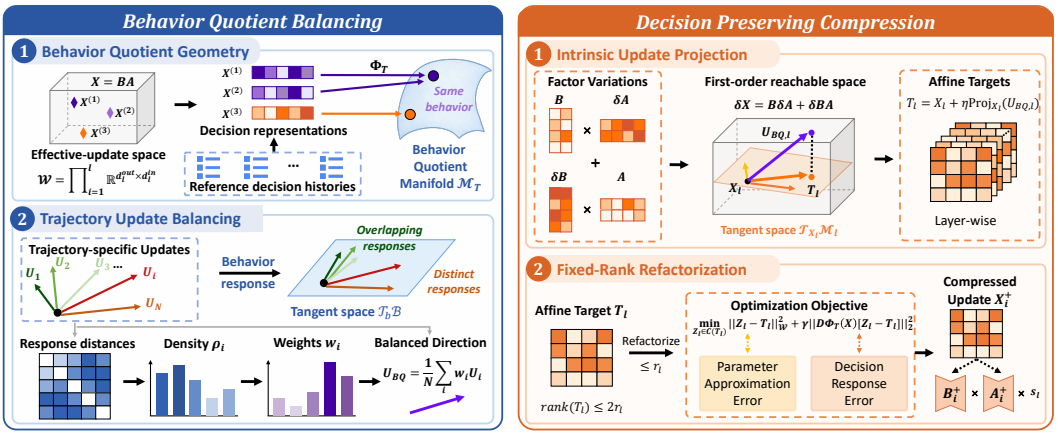


Figure 1: Overview of BQ-LoRA. BQB constructs a behavior quotient manifold and balances trajectory updates using quotient distances. DPC projects the balanced direction onto the fixed rank LoRA tangent space and refactorizes the resulting target while preserving decision changes.

trajectories $\mathcal{T} = \{\tau_i\}_{i=1}^N$, let $\mathcal{L}_i(\boldsymbol{X})$ denote the task loss associated with τ_i. Standard LoRA training minimizes the empirical task loss:

$$\mathcal{L}_{\text{task}}(\boldsymbol{X}) = \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}_i(\boldsymbol{X}). \tag{4}$$

4.2 Framework Overview

The overall framework of BQ-LoRA is illustrated in Figure 1, showing how trajectory updates are balanced and incorporated into a single LoRA adapter. It consists of two modules, i.e., behavior quotient balancing (BQB) and decision preserving compression (DPC). First, BQB constructs normalized decision representations at fixed interaction histories and groups values of $\boldsymbol{X}$ that produce the same representations, forming a local behavior quotient geometry. Within this geometry, trajectory update directions are compared through their first-order effects on the reference decision representations. Their local concentration determines the trajectory weights, reducing the dominance of repeatedly represented behavioral changes and increasing the contribution of less represented ones. The weighted trajectory directions are then aggregated into a shared update direction for $\boldsymbol{X}$. Second, DPC projects the balanced direction onto the fixed rank tangent space induced by the current LoRA factors and takes a step along it. The resulting target is compressed under the original rank budget by jointly minimizing approximation error and the distortion of first-order decision responses measured using the same behavior geometry. The compressed target is refactorized into updated LoRA factors for the next training step. These two stages are repeated during training to balance trajectory contributions while controlling the behavioral distortion introduced by the rank constraint.

4.3 Behavior Quotient Balancing

Behavior Quotient Geometry. Different trajectories can produce distinct gradients with similar effects on agent behavior. To compare these effects, we define a behavior map from $\boldsymbol{X}$ to decision representations at fixed reference histories. We equip the matrix space $\mathcal{W} = \prod_{l=1}^{L}\mathbb{R}^{d_l^{\text{out}}\times d_l^{\text{in}}}$ with the product Frobenius inner product. For a history h, let $f_{\boldsymbol{X}}(h) \in \mathbb{R}^{d_{\text{hid}}}$ denote the final representation passed to the frozen output head. For each reference decision a_{ik}, let $|a_{ik}|$ denote its number of tokens. For $1 \le t \le |a_{ik}|$, the prefix $h_{ik,t}$ extends h_{ik} with the recorded reasoning in y_{ik} followed by the first $t-1$ decision tokens. With these prefixes fixed, the behavior map is:

$$[\Phi_{\mathcal{T}}(\boldsymbol{X})]_{ik,t} = \frac{f_{\boldsymbol{X}}(h_{ik,t})}{\sqrt{NT_i|a_{ik}|}}. \tag{5}$$

The normalization averages squared representation responses within each decision. It then averages over the decisions in each trajectory before taking the mean across trajectories. Local changes in these representations serve as a surrogate for decision changes conditioned on the recorded reasoning and interaction histories. Using this map, we define behavioral equivalence on the reference collection:

$$\boldsymbol{X}^{(1)} \sim_{\mathcal{T}} \boldsymbol{X}^{(2)} \quad \Longleftrightarrow \quad \Phi_{\mathcal{T}}(\boldsymbol{X}^{(1)}) = \Phi_{\mathcal{T}}(\boldsymbol{X}^{(2)}). \tag{6}$$

To compare update directions through their behavioral effects, we describe local changes between these equivalence classes. We assume that $\Phi_{\mathcal{T}}$ is smooth with constant differential rank on a sufficiently small open neighborhood $\mathcal{O}$ of the current iterate $\boldsymbol{X}$ (Amir et al., 2023; Lukashchuk et al., 2025). The equivalence classes then form the behavior quotient manifold $\mathcal{M}_{\mathcal{T}} = \mathcal{O}/\sim_{\mathcal{T}}$, with quotient map $\pi_{\mathcal{T}}(\boldsymbol{X}) = [\boldsymbol{X}]_{\mathcal{T}}$. The differential $D\pi_{\mathcal{T}}(\boldsymbol{X})$ maps each parameter direction to a quotient tangent. At $\boldsymbol{X}$, two parameter directions represent the same quotient tangent exactly when their difference lies in $\ker D\Phi_{\mathcal{T}}(\boldsymbol{X})$. They therefore induce the same change in the reference decision representations to first order. The Euclidean inner product on the corresponding representation changes induces a nondegenerate Riemannian metric g on $\mathcal{M}_{\mathcal{T}}$. Quotient tangent distances therefore compare local behavioral changes without contributions from this kernel. Appendix B establishes the local quotient structure and derives the induced metric.

Trajectory Update Balancing. The quotient geometry reveals overlap between trajectory updates, but uniform weighting still counts each trajectory as an independent contribution. When many trajectories induce similar decision responses, their repeated contributions can dominate the aggregate direction. Behavioral changes represented by fewer trajectories then receive less influence, even when they provide distinct responses on the reference collection. We use the local concentration of quotient tangent directions to adjust these relative contributions. For each trajectory, let $\boldsymbol{U}_i = -\nabla_{\boldsymbol{X}} \mathcal{L}_i(\boldsymbol{X})$ denote its task direction. The differential $D\pi_{\mathcal{T}}(\boldsymbol{X})$ maps $\boldsymbol{U}_i$ to a quotient tangent at $[\boldsymbol{X}]_{\mathcal{T}}$. Its behavior response $D\Phi_{\mathcal{T}}(\boldsymbol{X})[\boldsymbol{U}_i]$ describes the first-order effect of the direction on the reference decision representations. We compare these directions under the quotient metric g:

$$d_{ij}^2 = \|D\pi_{\mathcal{T}}(\boldsymbol{X})[\boldsymbol{U}_i - \boldsymbol{U}_j]\|_g^2 = \|D\Phi_{\mathcal{T}}(\boldsymbol{X})[\boldsymbol{U}_i - \boldsymbol{U}_j]\|_2^2 . \tag{7}$$

The right-hand expression evaluates the quotient tangent distance directly from the two behavior responses. It can be computed using Jacobian-vector products without explicitly constructing quotient coordinates. Following density correction for near-duplicate data (Liu et al., 2024), we measure response overlap with Gaussian affinities and define trajectory weights:

$$\rho_i = \sum_{j=1}^{N} \exp\left(-\frac{d_{ij}^2}{2\sigma_{\mathcal{T}}^2}\right), \qquad w_i = \frac{\rho_i^{-\alpha}}{\frac{1}{N}\sum_{j=1}^{N} \rho_j^{-\alpha}}. \tag{8}$$

The data-dependent bandwidth $\sigma_{\mathcal{T}} > 0$ determines the scale of the neighborhood comparison. A larger ρ_i means that more trajectory directions induce responses close to that of $\boldsymbol{U}_i$. The exponent $\alpha \in [0, 1]$ controls how strongly this overlap reduces the relative weight of a trajectory. The normalization keeps the average trajectory weight at one, so the coefficients in the weighted objective sum to one. Setting $\alpha = 0$ gives equal weight to every trajectory. For groups of similar responses with sufficiently small overlap between groups, $\alpha = 1$ approximately equalizes their total trajectory weights. Holding w_i fixed during the task update, we combine the trajectory directions:

$$\boldsymbol{U}_{\mathrm{BQ}} = \frac{1}{N} \sum_{i=1}^{N} w_i \boldsymbol{U}_i. \tag{9}$$

By linearity, the aggregate induces the weighted mean of the trajectory quotient tangents. For $\alpha > 0$, this reduces the relative contribution of densely represented behavioral changes and increases that of less represented changes. Each task loss determines an update direction, and $w_i > 0$ scales its contribution to the shared adapter. BQB adjusts behavioral coverage while retaining task supervision, without predefined capability partitions.

4.4 Decision Preserving Compression

Intrinsic Update Projection. BQB aggregates trajectory directions in $\mathcal{W}$, whereas LoRA represents each adaptation matrix through low-rank factors. The balanced direction may consequently

contain components that these factors cannot realize to first order at the current iterate. We therefore project the balanced direction onto the tangent space of changes attainable through the LoRA factors. The projection is defined by $\boldsymbol{X}$, so equivalent factorizations give the same target. For each layer l, let $\mathcal{R}_{r_l}$ denote the manifold of matrices in $\mathbb{R}^{d_l^{\text{out}} \times d_l^{\text{in}}}$ with rank r_l. Starting from the standard zero initialization of $\boldsymbol{B}_l$, we first perform a short warmup on the LoRA factors. The geometric updates below require $\operatorname{rank}(\boldsymbol{X}_l) = r_l$ at every step after warmup. At an iterate with $\operatorname{rank}(\boldsymbol{X}_l) = r_l$, differentiating $\boldsymbol{X}_l = s_l \boldsymbol{B}_l \boldsymbol{A}_l$ characterizes the tangent space (Wang et al., 2025):

$$T_{\boldsymbol{X}_l}\mathcal{R}_{r_l} = \left\{ s_l \left(\boldsymbol{B}_l \Delta \boldsymbol{A}_l + \Delta \boldsymbol{B}_l \boldsymbol{A}_l \right) : \Delta \boldsymbol{A}_l \in \mathbb{R}^{r_l \times d_l^{\text{in}}},\ \Delta \boldsymbol{B}_l \in \mathbb{R}^{d_l^{\text{out}} \times r_l} \right\}. \tag{10}$$

The two terms describe the changes to $\boldsymbol{X}_l$ obtained by varying either factor, and their sum contains all first-order directions attainable at $\boldsymbol{X}_l$. Let $\boldsymbol{P}_l \in \mathbb{R}^{d_l^{\text{out}} \times r_l}$ and $\boldsymbol{Q}_l \in \mathbb{R}^{d_l^{\text{in}} \times r_l}$ have orthonormal columns spanning $\operatorname{col}(\boldsymbol{X}_l)$ and $\operatorname{col}(\boldsymbol{X}_l^\top)$, respectively. We denote the Frobenius orthogonal projection onto $T_{\boldsymbol{X}_l}\mathcal{R}_{r_l}$ by $\operatorname{Proj}_{\boldsymbol{X}_l}$. The projected layerwise direction is (Bian et al., 2025):

$$\operatorname{Proj}_{\boldsymbol{X}_l}(\boldsymbol{U}_{\text{BQ},l}) = \boldsymbol{P}_l \boldsymbol{P}_l^\top \boldsymbol{U}_{\text{BQ},l} + \boldsymbol{U}_{\text{BQ},l} \boldsymbol{Q}_l \boldsymbol{Q}_l^\top - \boldsymbol{P}_l \boldsymbol{P}_l^\top \boldsymbol{U}_{\text{BQ},l} \boldsymbol{Q}_l \boldsymbol{Q}_l^\top. \tag{11}$$

The first two terms retain the components accessible through the current column or row space, while the last term subtracts their shared component. In the implementation, BQB applies this projection to each trajectory direction before computing weights. Writing $\operatorname{Proj}_{\boldsymbol{X}}$ for the layerwise projection, we use $\widetilde{\boldsymbol{U}}_i = \operatorname{Proj}_{\boldsymbol{X}}(\boldsymbol{U}_i)$ in place of $\boldsymbol{U}_i$ when evaluating the distances in equation 7 and the resulting weights in equation 8. Linearity gives $\operatorname{Proj}_{\boldsymbol{X}}(\boldsymbol{U}_{\text{BQ}}) = \frac{1}{N}\sum_i w_i \widetilde{\boldsymbol{U}}_i$, so projecting the balanced direction is equivalent to aggregating the projected directions with these weights. With a learning rate $\eta > 0$, the projected direction defines the layerwise affine target:

$$\boldsymbol{T}_l = \boldsymbol{X}_l + \eta \operatorname{Proj}_{\boldsymbol{X}_l}(\boldsymbol{U}_{\text{BQ},l}). \tag{12}$$

We collect these layerwise targets as $\boldsymbol{T} = (\boldsymbol{T}_l)_{l=1}^L$. The Frobenius projection gives the closest locally attainable direction to $\boldsymbol{U}_{\text{BQ},l}$, minimizing the alteration of the balanced direction within the tangent constraint. Because the projectors $\boldsymbol{P}_l \boldsymbol{P}_l^\top$ and $\boldsymbol{Q}_l \boldsymbol{Q}_l^\top$ depend only on the column and row spaces of $\boldsymbol{X}_l$, the projected direction and the resulting target are invariant to equivalent LoRA factorizations. The first-order decision change induced by $\boldsymbol{T} - \boldsymbol{X}$ is the reachable target effect that the subsequent refactorization aims to preserve.

Fixed-Rank Refactorization. The tangent projection ensures that the update direction is attainable to first order, but a finite step along that direction need not remain on the fixed-rank manifold. In particular, the affine target $\boldsymbol{T}_l$ can be expressed as the sum of two matrices of rank at most r_l, giving $\operatorname{rank}(\boldsymbol{T}_l) \le 2r_l$. When $\operatorname{rank}(\boldsymbol{T}_l) > r_l$, the target cannot be represented under the original LoRA rank budget and must be compressed. Approximation residuals with the same Frobenius norm can cause different changes in the reference decisions because the behavior map is more sensitive to some residual directions than others. Minimizing the Frobenius error alone therefore does not distinguish which parts of the target matter most for its decision effect. We use the behavior geometry from BQB to guide the selection of a rank-constrained representative. Let $\mathcal{C}(\boldsymbol{T}) \subseteq \mathcal{W}$ contain candidates $\boldsymbol{Z} = (\boldsymbol{Z}_l)_{l=1}^L$ satisfying $\operatorname{rank}(\boldsymbol{Z}_l) \le r_l$ for every layer. Each candidate is further constrained by $\operatorname{col}(\boldsymbol{Z}_l) \subseteq \operatorname{col}(\boldsymbol{T}_l)$ and $\operatorname{col}(\boldsymbol{Z}_l^\top) \subseteq \operatorname{col}(\boldsymbol{T}_l^\top)$. These constraints restrict the refactorization to the column and row spaces already present in the projected target. We obtain $\boldsymbol{X}^+$ by minimizing a combination of approximation error and behavioral distortion over $\mathcal{C}(\boldsymbol{T})$:

$$\boldsymbol{X}^+ \in \operatorname*{arg\,min}_{\boldsymbol{Z} \in \mathcal{C}(\boldsymbol{T})} \left\{ \|\boldsymbol{Z} - \boldsymbol{T}\|_{\mathcal{W}}^2 + \gamma \left\| D\Phi_{\mathcal{T}}(\boldsymbol{X})[\boldsymbol{Z} - \boldsymbol{T}] \right\|_2^2 \right\}. \tag{13}$$

The first term measures the matrix approximation error introduced by rank reduction. For the second term, linearity gives $D\Phi_{\mathcal{T}}(\boldsymbol{X})[\boldsymbol{Z} - \boldsymbol{T}] = D\Phi_{\mathcal{T}}(\boldsymbol{X})[\boldsymbol{Z} - \boldsymbol{X}] - D\Phi_{\mathcal{T}}(\boldsymbol{X})[\boldsymbol{T} - \boldsymbol{X}]$. It thus compares the first-order representation responses of the candidate and target steps at the fixed reference prefixes. Decision preservation is measured by how closely the compressed step retains the target response, with the reference histories and differential fixed throughout refactorization. The coefficient $\gamma > 0$ controls the tradeoff between matrix approximation and decision preservation. Because the behavior response is evaluated across all adapted layers jointly, the criterion accounts for their combined effect on the reference decisions. The restricted target spaces permit optimization through cores of size at most $2r_l \times 2r_l$. This objective favors representatives that approximate the

Table 1: Task performance on AppWorld and BrowseComp-Plus.

		AppWorld				BrowseComp-Plus			
		Test-N		Test-C		BM25		Embedding	
Model	Method	TGC	SGC	TGC	SGC	Accuracy	Recall	Accuracy	Recall
Qwen3.5-4B	Base Agent	27.98 ± 1.03	10.71 ± 1.79	11.99 ± 0.48	2.16 ± 0.72	19.33 ± 0.67	22.43 ± 0.58	36.67 ± 1.15	42.18 ± 0.74
	Full	58.33 ± 2.06	39.29 ± 3.57	40.29 ± 1.68	21.58 ± 2.16	26.22 ± 2.69	26.74 ± 2.18	44.22 ± 1.68	48.16 ± 1.83
	LoRA	52.98 ± 2.38	32.14 ± 3.09	35.49 ± 1.44	17.99 ± 1.90	20.89 ± 1.68	22.28 ± 1.77	40.22 ± 2.34	45.77 ± 1.96
	LoRA-S	58.93 ± 1.57	41.07 ± 1.79	39.33 ± 0.96	19.42 ± 1.44	26.89 ± 1.02	27.07 ± 1.14	42.44 ± 1.68	47.38 ± 1.43
	MoRAgent	60.12 ± 2.73	42.86 ± 3.09	40.05 ± 1.92	21.58 ± 2.88	25.56 ± 2.34	26.48 ± 2.53	43.11 ± 3.36	48.26 ± 2.39
	DART	56.55 ± 1.03	37.50 ± 1.79	39.09 ± 1.20	20.14 ± 1.25	28.22 ± 1.68	28.86 ± 1.58	45.56 ± 1.39	49.14 ± 1.21
	TopoCurate	55.95 ± 2.06	37.50 ± 3.09	41.01 ± 1.34	23.02 ± 2.16	26.44 ± 1.39	27.64 ± 1.73	43.56 ± 1.68	48.82 ± 1.62
	BQ-LoRA	**60.91** ± 1.24	**44.64** ± 1.79	**44.04** ± 1.08	**24.46** ± 1.25	**29.11** ± 1.02	**29.34** ± 0.82	**46.44** ± 1.68	**50.16** ± 0.83
Qwen3.5-9B	Base Agent	36.31 ± 0.60	19.64 ± 1.79	23.02 ± 0.24	10.79 ± 0.72	22.67 ± 0.67	24.16 ± 0.46	40.67 ± 0.67	44.36 ± 0.62
	Full	69.64 ± 2.38	53.57 ± 3.09	49.64 ± 1.44	27.34 ± 1.90	30.22 ± 2.34	31.48 ± 2.08	49.56 ± 2.69	52.36 ± 2.17
	LoRA	60.71 ± 2.06	42.86 ± 3.57	45.08 ± 1.20	20.86 ± 2.16	26.22 ± 1.68	27.14 ± 1.81	46.22 ± 2.69	48.34 ± 1.64
	LoRA-S	66.67 ± 1.19	51.79 ± 1.79	47.72 ± 1.10	25.18 ± 1.25	29.11 ± 1.39	30.26 ± 1.13	48.22 ± 1.02	50.48 ± 1.47
	MoRAgent	69.05 ± 2.98	55.36 ± 3.57	49.16 ± 1.68	28.06 ± 2.59	29.56 ± 3.36	30.92 ± 2.31	49.11 ± 2.34	51.76 ± 2.28
	DART	66.67 ± 1.57	48.21 ± 3.09	48.20 ± 0.96	25.90 ± 1.44	31.56 ± 1.02	32.84 ± 1.37	50.89 ± 1.68	53.28 ± 1.58
	TopoCurate	64.88 ± 1.79	48.21 ± 1.79	50.12 ± 1.20	29.50 ± 1.90	29.78 ± 1.68	31.72 ± 1.68	49.56 ± 1.39	52.64 ± 1.29
	BQ-LoRA	**70.44** ± 1.50	**57.14** ± 1.79	**53.32** ± 0.97	**31.65** ± 1.25	**33.11** ± 1.68	**34.26** ± 0.85	**51.78** ± 1.02	**54.38** ± 0.79

target matrices while retaining the decision changes identified by the same geometry used for trajectory balancing. The rank constraints guarantee factors $\boldsymbol{B}_l^+$ and $\boldsymbol{A}_l^+$ satisfying $\boldsymbol{X}_l^+ = s_l \boldsymbol{B}_l^+ \boldsymbol{A}_l^+$. The compressed target can be stored as a single LoRA adapter under the original rank budgets, with decision distortion included in the compression criterion.

4.5 Theoretical Analysis

We bound the representation error relative to $\boldsymbol{T}_\star$, which combines the mean direction of each behavioral group equally.

Theorem 1 (Preservation of balanced trajectory contributions)**.** *Under the assumptions and notation in Appendix C, let $\boldsymbol{Y}$ be the truncated SVD of $\boldsymbol{T}$ under the rank budgets. For $\gamma > 0$, any feasible output with objective decrease $\Delta_\gamma = F_\gamma(\boldsymbol{Y}) - F_\gamma(\boldsymbol{X}^+) \geq 0$ satisfies:*

$$\left\|\Phi_{\mathcal{T}}(\boldsymbol{X}^+) - \Phi_{\mathcal{T}}(\boldsymbol{T}_\star)\right\|_2 \leq \eta \,\mathrm{diam}_\Phi \tanh\left(\frac{1}{4}\log\frac{\max_i \kappa_i}{\min_i \kappa_i}\right) + \sqrt{\|J[\boldsymbol{Y}-\boldsymbol{T}]\|_2^2 - \frac{\Delta_\gamma}{\gamma}} + \epsilon_\gamma. \quad (14)$$

Appendix C provides the proof and explicit remainder.

5 Experiments

5.1 Experimental Setups

Datasets. We conduct experiments on AppWorld (Trivedi et al., 2024) and BrowseComp-Plus (Chen et al., 2026c). **AppWorld** evaluates agents on multi-step tasks in a stateful world of simulated applications. Given a natural language request, the agent invokes application APIs and responds to environment feedback to complete the task. Programmatic tests verify task completion and the resulting application states. **BrowseComp-Plus** evaluates deep search agents on difficult questions that require iterative retrieval and evidence synthesis. Agents search a fixed document corpus containing human-verified supporting evidence. For both benchmarks, we perform supervised fine-tuning on collected successful trajectories. Details of the datasets are provided in Appendix E.1.

Baselines. We compare `BQ-LoRA` with seven baselines. Standard adaptation baselines include: (1) Base, which evaluates the pretrained model without adaptation, (2) Full FT, which updates all model parameters, and (3) LoRA (Hu et al., 2022), which introduces a low-rank adaptation matrix into each adapted layer. We then compare two multi-LoRA methods designed for agent tuning: (4) MoRAgent (Han et al., 2025), which assigns separate LoRA groups to the reasoner, executor, and

Table 2: Mean interaction counts on AppWorld (steps) and BrowseComp-Plus (search calls).

Model	Method	AppWorld		BrowseComp-Plus	
		Test-N	Test-C	BM25	Embedding
Qwen3.5-4B	Base Agent	31.28 ± 1.24	36.64 ± 1.58	16.95 ± 0.86	15.54 ± 0.72
	Full	17.75 ± 0.64	26.10 ± 1.12	16.70 ± 0.78	11.65 ± 0.46
	LoRA	18.26 ± 0.82	27.12 ± 1.36	17.73 ± 0.94	14.35 ± 0.68
	LoRA-S	18.65 ± 0.76	27.20 ± 1.18	17.20 ± 0.83	13.25 ± 0.62
	MoRAgent	17.10 ± 0.91	25.35 ± 1.42	16.85 ± 0.88	12.35 ± 0.74
	DART	18.00 ± 0.68	26.25 ± 1.06	16.15 ± 0.72	11.90 ± 0.48
	TopoCurate	17.35 ± 0.73	24.90 ± 1.21	15.75 ± 0.81	12.05 ± 0.57
	BQ-LoRA	**16.70** ± 0.61	**24.10** ± 1.14	**15.35** ± 0.76	**11.35** ± 0.52
Qwen3.5-9B	Base Agent	26.20 ± 1.08	29.92 ± 1.46	15.65 ± 0.78	13.16 ± 0.64
	Full	17.45 ± 0.58	24.25 ± 0.96	15.50 ± 0.66	11.10 ± 0.42
	LoRA	17.16 ± 0.72	24.96 ± 1.18	17.92 ± 0.91	13.48 ± 0.63
	LoRA-S	17.65 ± 0.64	25.75 ± 1.09	16.95 ± 0.74	12.65 ± 0.56
	MoRAgent	16.10 ± 0.79	23.60 ± 1.26	16.30 ± 0.82	11.80 ± 0.68
	DART	16.90 ± 0.56	24.20 ± 0.88	15.75 ± 0.63	11.55 ± 0.44
	TopoCurate	16.65 ± 0.67	23.15 ± 1.04	15.30 ± 0.71	11.65 ± 0.53
	BQ-LoRA	**15.95** ± 0.59	**22.70** ± 0.94	**15.05** ± 0.68	**10.85** ± 0.47

summarizer roles, and (5) DART (Li et al., 2026c), which separates reasoning and tool use updates across distinct low-rank modules. For single-LoRA optimization, we include (6) LoRA-S (Zheng & Wu, 2026), which optimizes low-rank factors on a quotient manifold through a Sylvester equation. For trajectory curation, we include (7) TopoCurate (Yang et al., 2026b), which uses a semantic quotient topology of multi-trial interactions to select informative trajectories.

Implementation Details. We use Qwen3.5-4B and Qwen3.5-9B as the backbone models (Team et al., 2026). All training runs use a global batch size of 64. For all LoRA-based methods, we set the rank to $r_l = 32$ and the LoRA scaling parameter to $\alpha_{\text{LoRA}} = 32$, which gives $s_l = \alpha_{\text{LoRA}}/r_l = 1$.

5.2 Performance Evaluation

Main Results. Tables 1 and 2 summarize task performance and interaction counts, respectively. We report mean task performance and standard deviation over three random seeds. (1) Standard LoRA improves task completion and answer accuracy over the base agent but trails stronger baselines. Full fine-tuning also falls short of the best low-rank methods. Its flexibility may favor fitting trajectory-specific patterns, while low-rank constraints can limit overfitting. (2) LoRA-S consistently outperforms LoRA, supporting optimization that removes factorization redundancy. However, equivalent factorizations describe the same adaptation matrix. They do not capture redundancy between behavioral changes across trajectories. (3) Multiple adapters benefit from capability specialization. MoRAgent leads the low-rank baselines on Test-N, while DART leads on BrowseComp-Plus. Both require separate adapters with predefined capability partitions. (4) TopoCurate leads the baselines on Test-C through trajectory curation. However, similar recorded interactions need not induce similar behavioral changes. Its selection criterion therefore does not directly address redundancy in training effects. (5) BQ-LoRA achieves the highest mean scores on all eight metrics. Its average exceeds the highest baseline average by 2.64 and 2.78 percentage points for the 4B and 9B backbones, respectively. It also achieves the lowest mean interaction counts in every setting. These results support balancing updates and preserving decision effects within one fixed-rank adapter.

Ablation Study. Table 3 evaluates the contribution of each module by removing it from the full method. BQ-LoRA-w/o-BQB uses uniform trajectory weights and retains DPC, while BQ-LoRA-w/o-DPC uses BQB alone. Removing BQB reduces Test-C TGC by 3.27 and 3.44 percentage points for the 4B and 9B backbones, respectively, and lowers BrowseComp-Plus accuracy with both retrievers. The larger drops on Test-C suggest that balancing helps generalization to more challenging tasks. Removing DPC reduces all four metrics in Table 3, with average drops of 2.61

Table 3: Ablation results on AppWorld (TGC) and BrowseComp-Plus (accuracy)

Model	Method	AppWorld		BrowseComp-Plus	
		Test-N	Test-C	BM25	Embedding
Qwen3.5-4B	LoRA	52.98 ± 2.38	35.49 ± 1.44	20.89 ± 1.68	40.22 ± 2.34
	BQ-LoRA-w/o-BQB	59.72 ± 1.82	40.77 ± 1.10	27.11 ± 1.68	43.56 ± 1.39
	BQ-LoRA-w/o-DPC	56.94 ± 2.09	41.33 ± 0.97	27.56 ± 2.04	44.22 ± 2.14
	BQ-LoRA	**60.91** ± 1.24	**44.04** ± 1.08	**29.11** ± 1.02	**46.44** ± 1.68
Qwen3.5-9B	LoRA	60.71 ± 2.06	45.08 ± 1.20	26.22 ± 1.68	46.22 ± 2.69
	BQ-LoRA-w/o-BQB	68.65 ± 2.09	49.88 ± 1.20	30.67 ± 2.40	49.78 ± 1.39
	BQ-LoRA-w/o-DPC	64.68 ± 1.82	50.20 ± 0.97	31.56 ± 1.68	50.22 ± 2.52
	BQ-LoRA	**70.44** ± 1.50	**53.32** ± 0.97	**33.11** ± 1.68	**51.78** ± 1.02

and 3.00 percentage points for the two backbones. Retaining DPC performs better on Test-N, while retaining BQB performs better on Test-C and BrowseComp-Plus accuracy for both backbones. Both ablations outperform LoRA, and their combination achieves the highest means in every setting, supporting complementary benefits. Appendix E.4 provides further ablations.

Hyperparameter Sensitivity. We vary the balancing exponent α from 0 to 1 and the compression coefficient γ from 0.1 to 10 at the values shown in Figure 2. The curves average the eight performance metrics for each backbone. (1) Performance peaks at $\alpha = 0.75$ for 4B and $\alpha = 1$ for 9B. The gains over uniform weighting at $\alpha = 0$ support reducing the influence of densely represented responses. The 4B curve flattens at larger α, suggesting diminishing returns from stronger balancing. (2) Both backbones perform best at $\gamma = 0.8$. Small γ emphasizes matrix approximation, whereas large γ penalizes decision distortion more strongly. The decline toward either end of the tested range supports an intermediate tradeoff between weight approximation and decision preservation.

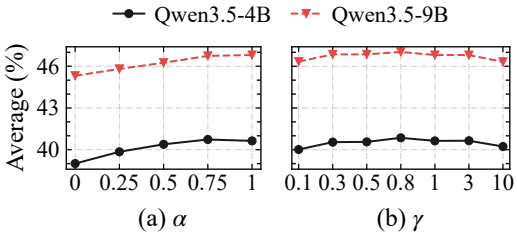
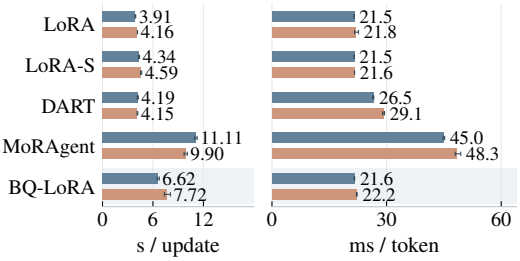


Figure 2: Hyperparameter sensitivity.

Computational Complexity. Figure 3 compares training time and decoding latency. The additional training cost of BQ-LoRA comes from reference decision derivatives for trajectory balancing and rank-constrained refactorization. Gradient reuse and distributed geometric computation reduce this overhead, keeping training faster than MoRAgent. Decoding latency remains close to ordinary LoRA because both modules operate during training and retain a single adapter for inference. Appendix E.3 gives further resource measurements.



(a) Training time (b) Decoding latency

Figure 3: Qwen3.5-4B costs on AppWorld (blue) and BrowseComp-Plus (orange).

6 Conclusion

We presented BQ-LoRA for adapting LLM agents with a single fixed-rank LoRA. BQB balances trajectory updates in a local behavior quotient geometry, reducing repeated contributions from updates with equivalent effects on decision representations. DPC projects the balanced update and compresses the resulting target while controlling decision distortion. Together, they preserve the original rank budget and require no capability-specific routing. Experiments on AppWorld and BrowseComp-Plus show strong task performance and reduced interaction counts.

AI USE STATEMENT

We used generative AI tools for code development, writing refinement, and table and figure formatting.

ETHICS STATEMENT

We do not foresee specific ethical concerns arising from this work. The research conducted in this paper conforms to the ICLR Code of Ethics.

REPRODUCIBILITY STATEMENT

The experimental settings and implementation details are provided in Section 5.1 and Appendix E.2. We will release the code upon acceptance to facilitate reproducibility.

REFERENCES

Tal Amir, Steven Gortler, Ilai Avni, Ravina Ravina, and Nadav Dym. Neural injective functions for multisets, measures and graphs via a finite witness theorem. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 42516–42551. Curran Associates, Inc., 2023. doi: 10.52202/075280-1845. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/84b686f7cc7b7751e9aaac0da74f755a-Paper-Conference.pdf.

Fengmiao Bian, Jinyang ZHENG, Ziyun Liu, Jianzhou Luo, and Jian-Feng CAI. Finding low-rank matrix weights in dnns via riemannian optimization: Radagrad and radamw. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 59916–59950. Curran Associates, Inc., 2025. doi: 10.52202/085713-2004. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/5679173c400b332796426e443ab5ea0d-Paper-Conference.pdf.

Vladimir Bogachev, Vladimir Aletov, Alexander Molozhavenko, Denis Bobkov, Vera Soboleva, Aibek Alanov, and Maxim Rakhuba. Lora meets riemannion: Muon optimizer for parametrization-independent low-rank adapters. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 82190–82215, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/8505a7e1da40320bf5692d7edec5aa45-Paper-Conference.pdf.

Shubhankar Borse, Shreya Kadambi, Nilesh Prasad Pandey, Kartikeya Bhardwaj, Viswanath Ganapathy, Sweta Priyadarshi, Risheek Garrepalli, Rafael Esteves, Munawar Hayat, and Fatih Porikli. Foura: Fourier low-rank adaptation. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 71504–71539. Curran Associates, Inc., 2024. doi: 10.52202/079017-2284. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/83960718b4d12f799985206f1b1cf00f-Paper-Conference.pdf.

Nicolas Boumal. *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, 2023. doi: 10.1017/9781009166164.

Valérie Castin, Kimia Nadjahi, Pierre Ablin, and Gabriel Peyré. Balanced loRA: Removing parameter invariance to accelerate convergence. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=kHInw3cjCP.

Baian Chen, Chang Shu, Ehsan Shareghi, Nigel Collier, Karthik Narasimhan, and Shunyu Yao. Fireact: Toward language agent fine-tuning, 2023. URL https://arxiv.org/abs/2310.05915.

Guhong Chen, Chenghao Sun, Cheng Fu, Qiyao Wang, Zhihong Huang, ChaoPeng Wei, Guangxu Chen, Feiteng Fang, Ahmadreza Argha, Bing Zhao, Xander Xu, Qi Han, Hamid Alinejad-Rokny, Qiang Qu, Binhua Li, Shiwen Ni, Min Yang, HU Wei, and Yongbin Li. Beyond quantity: Trajectory diversity scaling for code agents. In Maria Liakata, Viviane P. Moreira, Jiajun Zhang, and David Jurgens (eds.), *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 15676–15691, San Diego, California, United States, July 2026a. Association for Computational Linguistics. ISBN 979-8-89176-395-1. doi: 10.18653/v1/2026.findings-acl.768. URL https://aclanthology.org/2026.findings-acl.768/.

Nan Chen, Soledad Villar, and Soufiane Hayou. Learning rate scaling across loRA ranks and transfer to full finetuning. In *Forty-third International Conference on Machine Learning*, 2026b. URL https://openreview.net/forum?id=vCAIwEJZQ6.

Zhixun Chen, Ming Li, Yuxuan Huang, Yali Du, Meng Fang, and Tianyi Zhou. ATLAS: Agent tuning via learning critical steps. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (eds.), *Findings of the Association for Computational Linguistics: ACL 2025*, pp. 25334–25349, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-256-5. doi: 10.18653/v1/2025.findings-acl.1299. URL https://aclanthology.org/2025.findings-acl.1299/.

Zijian Chen, Xueguang Ma, Shengyao Zhuang, Ping Nie, Kai Zou, Sahel Sharifymoghaddam, Andrew Liu, Joshua Green, Kshama Patel, Ruoxi Meng, Mingyi Su, Yanxi Li, Haoran Hong, Xinyu Shi, Xuye Liu, Hosna Oyarhoseini, Nandan Thakur, Crystina Zhang, Luyu Gao, Wenhu Chen, and Jimmy Lin. BrowseComp-plus: A fair and disentangled evaluation benchmark for deep search agents. In Maria Liakata, Viviane P. Moreira, Jiajun Zhang, and David Jurgens (eds.), *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 22349–22370, San Diego, California, United States, July 2026c. Association for Computational Linguistics. ISBN 979-8-89176-390-6. doi: 10.18653/v1/2026.acl-long.1023. URL https://aclanthology.org/2026.acl-long.1023/.

Xuan Cui, HuiYue Li, Run Zeng, Yunfei Zhao, Jinrui Qian, Wei Duan, Bo Liu, and Zhanpeng Zhou. IGU-loRA: Adaptive rank allocation via integrated gradients and uncertainty-aware scoring. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=MnToYQx9My.

Shihan Dou, Enyu Zhou, Yan Liu, Songyang Gao, Wei Shen, Limao Xiong, Yuhao Zhou, Xiao Wang, Zhiheng Xi, Xiaoran Fan, Shiliang Pu, Jiang Zhu, Rui Zheng, Tao Gui, Qi Zhang, and Xuanjing Huang. LoRAMoE: Alleviating world knowledge forgetting in large language models via MoE-style plugin. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar (eds.), *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1932–1945, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.acl-long.106. URL https://aclanthology.org/2024.acl-long.106/.

Ruishan Fang, Siyuan Lu, Chenyi Zhuang, and Tao Lin. Rods: Reward-driven online data synthesis for multi-turn tool-use agents, 2026a. URL https://arxiv.org/abs/2606.19047.

Runnan Fang, Shihao Cai, Baixuan Li, Jialong Wu, Guangyu Li, Wenbiao Yin, Xinyu Wang, Xiaobin Wang, Liangcai Su, Zhen Zhang, Shibin Wu, Zhengwei Tao, Yong Jiang, Pengjun Xie, Ningyu Zhang, Fei Huang, Wentao Zhang, and Jingren Zhou. Towards general agentic intelligence via environment scaling. In Maria Liakata, Viviane P. Moreira, Jiajun Zhang, and David Jurgens (eds.), *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 17610–17621, San Diego, California, United States, July 2026b. Association for Computational Linguistics. ISBN 979-8-89176-395-1. doi: 10.18653/v1/2026.findings-acl.872. URL https://aclanthology.org/2026.findings-acl.872/.

Zhi-Quan Feng, Ying-Jia Lin, and Hung-Yu Kao. Learning in the fisher subspace: A guided initialization for loRA fine-tuning. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=P76Q8Pi1E9.

Weizheng Gu, Chengze Li, Zhuohao Yu, Mengyuan Sun, Zhibang Yang, Wei Wang, Hongrui Jia, Shikun Zhang, and Wei Ye. What do agents learn from trajectory-SFT: Semantics or interfaces? In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=u1i3XoSXQG.

Jing Han, Binwei Yan, Tianyu Guo, Zheyuan Bai, Mengyu Zheng, Hanting Chen, and Ying Nie. MoRAgent: Parameter efficient agent tuning with mixture-of-roles. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 21916–21931. PMLR, 13–19 Jul 2025.

Soufiane Hayou, Nikhil Ghosh, and Bin Yu. LoRA+: Efficient low rank adaptation of large models. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 17783–17806. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/hayou24a.html.

haonan he, Peng Ye, Yuchen Ren, yuan yuan, LuyangZhou, ShucunJu, and lei chen. Gora: Gradient-driven adaptive low rank adaptation. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, pp. 114074–114102. Curran Associates, Inc., 2025. doi: 10.52202/085713-3805. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/a5e4907a40c0dcb8433a35c714ba9d79-Paper-Conference.pdf.

Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

Chengsong Huang, Qian Liu, Bill Yuchen Lin, Tianyu Pang, Chao Du, and Min Lin. Lorahub: Efficient cross-task generalization via dynamic loRA composition. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=TrloAXEJ2B.

Hyesung Jeon, Hyeongju Ha, and Jae-Joon Kim. LRAgent: Efficient KV cache sharing for multi-loRA LLM agents. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=ooRukY3zjQ.

Rohan Kadekodi, Zhan Jin, Keisuke Kamahori, Yile Gu, Sean Khatiri, Noah H. Bayindirli, Sergey Gorbunov, and Baris Kasikci. Dualtune: Decoupled fine-tuning for on-device agentic systems, 2026. URL https://openreview.net/forum?id=oYwPx0y0RE.

Hangoo Kang, Tarun Suresh, Jon Saad-Falcon, and Azalia Mirhoseini. TRACE: Capability-targeted agentic training. In *Second Workshop on Agents in the Wild: Safety, Security, and Beyond*, 2026. URL https://openreview.net/forum?id=p37UqCmcxG.

Yongchan Kwon, Eric Wu, Kevin Wu, and James Y Zou. Datainf: Efficiently estimating data influence in lora-tuned llms and diffusion models. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 21921–21942, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/5e84a0f233611a1dc8fb794dc52415a3-Paper-Conference.pdf.

Hanyu Lai, Junjie Gao, Xiao Liu, Yifan Xu, Shudan Zhang, Yuxiao Dong, and Jie Tang. AndroidGen: Building an android language agent under data scarcity. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2727–2749, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-251-0. doi: 10.18653/v1/2025.acl-long.138. URL https://aclanthology.org/2025.acl-long.138/.

Mukai Li, Qingcheng Zeng, Tianqing Fang, Zhenwen Liang, Linfeng Song, Qi Liu, Haitao Mi, and Dong Yu. Verified critical step optimization for LLM agents. In Maria Liakata, Viviane P. Moreira,

Jiajun Zhang, and David Jurgens (eds.), *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 39627–39639, San Diego, California, United States, July 2026a. Association for Computational Linguistics. ISBN 979-8-89176-395-1. doi: 10.18653/v1/2026.findings-acl.1974. URL https://aclanthology.org/2026.findings-acl.1974/.

Shiwei Li, Xiandi Luo, Haozhao Wang, Xing Tang, Ziqiang Cui, Dugang Liu, Yuhua Li, Xiuqiang He, and Ruixuan Li. Beyond higher rank: Token-wise input-output projections for efficient low-rank adaptation. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 136414–136435. Curran Associates, Inc., 2025a. doi: 10.52202/085713-4552. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/c74429a930838b5bac03c681f0a0ce11-Paper-Conference.pdf.

Wenbing Li, Zikai Song, Hang Zhou, Junqing Yu, Yunyao Zhang, and Wei Yang. Loramixer: Coordinate modular lora experts through serial attention routing. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 14694–14716, 2026b. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/18610cbcd1da57854aa05ebdc5cd3168-Paper-Conference.pdf.

Yu Li, Mingyang Yi, Xiuyu Li, Ju Fan, Fuxin Jiang, Binbin Chen, Peng Li, Jie Song, and Tieying Zhang. Reasoning and tool-use compete in agentic rl:from quantifying interference to disentangled tuning, 2026c. URL https://arxiv.org/abs/2602.00994.

Zhizhong Li, Sina Sajadmanesh, Jingtao Li, and Lingjuan Lyu. Stella: Subspace learning in low-rank adaptation using stiefel manifold. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 75066–75092. Curran Associates, Inc., 2025b. doi: 10.52202/085713-2521. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/6cb0c6e7d50d5d65613f0456ca85e2db-Paper-Conference.pdf.

Mengqi Liao, Wei Chen, Junfeng Shen, Shengnan Guo, and Huaiyu Wan. Hmora: Making llms more effective with hierarchical mixture of lora experts. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 91309–91330, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/e43a33994a28f746dcfd53eb51ed3c2d-Paper-Conference.pdf.

Kai Lion, Liang Zhang, Bingcong Li, and Niao He. Polar: Polar-decomposed low-rank adapter representation. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 149313–149358. Curran Associates, Inc., 2025. doi: 10.52202/085713-4996. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/db79b6c50430a52fbab9d63efb8433ca-Paper-Conference.pdf.

Muqing Liu, Chongjie Si, and Yuheng Jia. FlexloRA: Entropy-guided flexible low-rank adaptation. In *The Fourteenth International Conference on Learning Representations*, 2026a. URL https://openreview.net/forum?id=tqnkbdYWWm.

Ye Liu, Kevin Qinghong Lin, Chang-Wen Chen, and Mike Zheng Shou. Videomind: A chain-of-lora agent for temporal-grounded video reasoning. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 57481–57506, 2026b. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/5e84e4413268b713f0d4a1b23a9dae57-Paper-Conference.pdf.

Zefang Liu and Jiahua Luo. AdamoLE: Fine-tuning large language models with adaptive mixture of low-rank adaptation experts. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=ndY9qFf9Sa.

Zifan Liu, Amin Karbasi, and Theodoros Rekatsinas. Tsds: Data selection for task-specific model finetuning. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 10117–10147. Curran Associates, Inc., 2024. doi: 10.52202/

079017-0325. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/13848b5893119ff772b69812c95914fa-Paper-Conference.pdf.

Mykola Lukashchuk, Raphaël Trésor, Wouter Nuijten, Ismail Senoz, and Bert Vries. The quotient bayesian learning rule. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 8788–8827. Curran Associates, Inc., 2025. doi: 10.52202/085713-0298. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/0ce1eb87dbb03fdfa872a93d15cfe333-Paper-Conference.pdf.

Fanxu Meng, Zhaohui Wang, and Muhan Zhang. Pissa: Principal singular values and singular vectors adaptation of large language models. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 121038–121072. Curran Associates, Inc., 2024. doi: 10.52202/079017-3846. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/db36f4d603cc9e3a2a5e10b93e6428f2-Paper-Conference.pdf.

Fabian Paischer, Lukas Hauzenberger, Thomas Schmied, Benedikt Alkin, Marc Deisenroth, and Sepp Hochreiter. Parameter efficient fine-tuning via explained variance adaptation. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, pp. 46215–46267. Curran Associates, Inc., 2025. doi: 10.52202/085713-1539. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/41d33bd41fd44bd9dba0e092047cf213-Paper-Conference.pdf.

Cheng Qian, Emre Can Acikgoz, Hongru Wang, Xiusi Chen, Avirup Sil, Dilek Hakkani-Tür, Gokhan Tur, and Heng Ji. SMART: Self-aware agent for tool overuse mitigation. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (eds.), *Findings of the Association for Computational Linguistics: ACL 2025*, pp. 4604–4621, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-256-5. doi: 10.18653/v1/2025.findings-acl.239. URL https://aclanthology.org/2025.findings-acl.239/.

Weiwei Sun, Miao Lu, Zhan Ling, Kang Liu, Xuesong Yao, Yiming Yang, and Jiecao Chen. Scaling long-horizon agent via context folding. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=lNRgWoGfYg.

Qwen Team et al. Qwen3. 5: Towards native multimodal agents, february 2026. *URL https://qwen.ai/blog*, 2026.

Harsh Trivedi, Tushar Khot, Mareike Hartmann, Ruskin Manku, Vinty Dong, Edward Li, Shashank Gupta, Ashish Sabharwal, and Niranjan Balasubramanian. AppWorld: A controllable world of apps and people for benchmarking interactive coding agents. In *ACL*, 2024.

Loring W. Tu. *An Introduction to Manifolds*. Universitext. Springer, 2 edition, 2011. doi: 10.1007/978-1-4419-7400-6.

Dingzirui Wang, Xuanliang Zhang, Keyan Xu, Qingfu Zhu, and Wanxiang Che. Curateevo: Data-curation evolving for agentic post-training, 2026. URL https://arxiv.org/abs/2607.06140.

Shaowen Wang, Linxi Yu, and Jian Li. Lora-ga: Low-rank adaptation with gradient approximation. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 54905–54931. Curran Associates, Inc., 2024. doi: 10.52202/079017-1741. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/62c4718cc334f6a0a62fb81c4a2095a1-Paper-Conference.pdf.

Zhengbo Wang, Jian Liang, Ran He, Zilei Wang, and Tieniu Tan. LoRA-pro: Are low-rank adapters properly optimized? In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=gTwRMU3lJ5.

xun wu, Shaohan Huang, and Furu Wei. Mixture of lora experts. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 47302–47318, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/ce806d8b4bf38cd92d483e5a0490d983-Paper-Conference.pdf.

Mengzhou Xia, Sadhika Malladi, Suchin Gururangan, Sanjeev Arora, and Danqi Chen. LESS: Selecting influential data for targeted instruction tuning. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 54104–54132. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/xia24c.html.

Jingwei Xu, Junyu Lai, and Yunpeng Huang. Meteora: Multiple-tasks embedded lora for large language models. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 55049–55076, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/8951bbdcf234132bcce680825e7cb354-Paper-Conference.pdf.

Guang Yang, Changhao Guan, Chao Huang, Yufeng Chen, and Kaiyu Huang. Adaptive utilization of low-rank adaptation via conditioned gating. In *Forty-third International Conference on Machine Learning*, 2026a. URL https://openreview.net/forum?id=cA7SvKStDh.

Jinluan Yang, Yuxin Liu, Zhengyu Chen, Chengcheng Han, Yueqing Sun, Qi Gu, Hui Su, Xunliang Cai, Fei Wu, and Kun Kuang. Topocurate:modeling interaction topology for tool-use agent training, 2026b. URL https://arxiv.org/abs/2603.01714.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=WE_vluYUL-X.

Jingqi Ye, haonan he, Minglei Li, Fujun Han, Tao Chen, and Peng Ye. Gradient intrinsic dimensionality alignment: Narrowing the gap between low-rank adaptation and full fine-tuning. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=kObvnQ6pUx.

Jui-Nan Yen, Si Si, Zhao Meng, Felix Yu, Venkata Sai Surya Subramanyam Duvvuri, Inderjit Dhillon, Cho-Jui Hsieh, and Sanjiv Kumar. Lora done rite: Robust invariant transformation equilibration for lora optimization. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 75815–75834, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/bcbc0f660d2dde42f9d1d0ecb14a6f9a-Paper-Conference.pdf.

Aohan Zeng, Mingdao Liu, Rui Lu, Bowen Wang, Xiao Liu, Yuxiao Dong, and Jie Tang. AgentTuning: Enabling generalized agent abilities for LLMs. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar (eds.), *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 3053–3077, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-acl.181. URL https://aclanthology.org/2024.findings-acl.181/.

Fangzhao Zhang and Mert Pilanci. Riemannian preconditioned LoRA for fine-tuning foundation models. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 59641–59669. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/zhang24ax.html.

Minxing Zhang, Yi Yang, Roy Xie, Bhuwan Dhingra, Shuyan Zhou, and Jian Pei. Generalizability of large language model-based agents: A comprehensive survey. *ACM Computing Surveys*, 58 (10):1–44, 2026a.

Qingru Zhang, Minshuo Chen, Alexander Bukharin, Pengcheng He, Yu Cheng, Weizhu Chen, and Tuo Zhao. Adaptive budget allocation for parameter-efficient fine-tuning. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=lq62uWRJjiY.

Qingyue Zhang, Chang Chu, Tianren Peng, Qi Li, Xiangyang Luo, Zhihao Jiang, and Shao-Lun Huang. LoRA-DA: Data-aware initialization for low-rank adaptation via asymptotic analysis. In *Forty-third International Conference on Machine Learning*, 2026b. URL https://openreview.net/forum?id=LvKSVUbXc9.

Yilang Zhang, Bingcong Li, and Georgios Giannakis. Reflora: Refactored low-rank adaptation for efficient fine-tuning of large models. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 92847–92878. Curran Associates, Inc., 2025a. doi: 10.52202/085713-3106. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/85c1ee0b3cdf49b84f14b163b01f2711-Paper-Conference.pdf.

Yuanhe Zhang, Fanghui Liu, and Yudong Chen. LoRA-one: One-step full gradient could suffice for fine-tuning large language models, provably and efficiently. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 75513–75574. PMLR, 13–19 Jul 2025b. URL https://proceedings.mlr.press/v267/zhang25ax.html.

Zicheng Zhang, Haoran Li, Yifeng Zhang, Guoqiang Gong, Jiaxing Wang, Junxing Hu, Pengzhang Liu, and Qixia Jiang. The primacy of magnitude in low-rank adaptation. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, pp. 39–69. Curran Associates, Inc., 2025c. doi: 10.52202/085713-0002. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/0010665e949927b74faf6e3ada6d7f72-Paper-Conference.pdf.

Jinyang Zheng and Tong Wu. Lora-s: An efficient low rank adaptation scheme via sylvester equation. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 101788–101811, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/a6a90bcc2aa470c3871b2d39a67d26e8-Paper-Conference.pdf.

Yuan Zhuang, Yi Shen, Yuexin Bian, Qing Su, Shihao Ji, Yuanyuan Shi, and Fei Miao. Ld-mole: Learnable dynamic routing for mixture of lora experts. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 64840–64858, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/6a39cf3b666f8bdb2223f253981f3869-Paper-Conference.pdf.

Heming Zou, Yunliang Zang, Wutong Xu, Yao Zhu, and Xiangyang Ji. Flylora: Boosting task decoupling and parameter efficiency via implicit rank-wise mixture-of-experts. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 10386–10419. Curran Associates, Inc., 2025. doi: 10.52202/085713-0349. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/0f234479d4be0102b6a9ec5ebf75e881-Paper-Conference.pdf.

A Additional Concepts on Quotient Manifolds

Quotient Set. Let $\overline{\mathcal{M}}$ be a smooth manifold equipped with an equivalence relation $\sim$. The equivalence class of $p \in \overline{\mathcal{M}}$ is defined as (Boumal, 2023; Tu, 2011):

$$[p] = \left\{ q \in \overline{\mathcal{M}} : q \sim p \right\}.$$

The quotient set collects these classes:

$$\mathcal{M} = \overline{\mathcal{M}} / \sim = \left\{ [p] : p \in \overline{\mathcal{M}} \right\}.$$

The canonical projection $\pi : \overline{\mathcal{M}} \to \mathcal{M}$ maps each point to its equivalence class, $\pi(p) = [p]$.

Quotient Space. For a subset $\mathcal{U} \subseteq \mathcal{M}$, the preimage under π collects all points in the equivalence classes belonging to $\mathcal{U}$:

$$\pi^{-1}(\mathcal{U}) = \left\{ p \in \overline{\mathcal{M}} : [p] \in \mathcal{U} \right\}. \tag{15}$$

The *quotient topology* declares $\mathcal{U}$ open exactly when this preimage is open in the total space:

$$\mathcal{U} \text{ is open in } \mathcal{M} \quad \Longleftrightarrow \quad \pi^{-1}(\mathcal{U}) \text{ is open in } \overline{\mathcal{M}}. \tag{16}$$

Equipped with these open sets, $\mathcal{M}$ is a quotient space, and π is continuous.

Quotient Manifold. A quotient space is a quotient manifold if it admits a smooth manifold structure with the quotient topology and the canonical projection π is a smooth submersion. The latter means that π is smooth and its differential $D\pi(p)$ satisfies:

$$\operatorname{rank} D\pi(p) = \dim \mathcal{M}, \qquad \forall\, p \in \overline{\mathcal{M}}. \tag{17}$$

To establish this structure, it suffices to find a smooth map $F : \overline{\mathcal{M}} \to \mathcal{N}$ onto a known smooth manifold $\mathcal{N}$ satisfying, for all $p, q \in \overline{\mathcal{M}}$:

$$p \sim q \quad \Longleftrightarrow \quad F(p) = F(q).$$

The differential must have rank $\dim \mathcal{N}$ throughout $\overline{\mathcal{M}}$. The surjective smooth submersion F is open, so the bijection $[p] \mapsto F(p)$ is a homeomorphism from $\mathcal{M}$ with its quotient topology onto $\mathcal{N}$. Transporting the smooth structure through this identification makes π a smooth submersion and equips $\mathcal{M}$ with a unique smooth quotient structure. The identification gives $\dim \mathcal{M} = \dim \mathcal{N}$. The fiber $\mathcal{F}_p = \pi^{-1}([p])$ therefore has dimension:

$$\dim \mathcal{F}_p = \dim \overline{\mathcal{M}} - \dim \mathcal{N}.$$

The vertical space $\mathcal{V}_p$ is the tangent space $T_p\mathcal{F}_p$ to the fiber:

$$\mathcal{V}_p = T_p\mathcal{F}_p = \ker D\pi(p) = \ker DF(p). \tag{18}$$

B Local Behavior Quotient Geometry

We explain the local quotient manifold construction for the behavior classes of LoRA adaptation matrices. For fixed reference agent trajectories $\mathcal{T}$, the behavior map $\Phi_{\mathcal{T}}$ in equation 5 collects decision representations at the recorded prefixes. Let $\mathcal{O} \subseteq \mathcal{W}$ be an open neighborhood of the current adaptation on which $\Phi_{\mathcal{T}}$ is smooth with constant differential rank. After shrinking $\mathcal{O}$, the constant rank theorem makes the image $\Phi_{\mathcal{T}}(\mathcal{O})$ an embedded manifold (Tu, 2011):

$$\operatorname{rank} D\Phi_{\mathcal{T}}(\boldsymbol{X}) = \dim \Phi_{\mathcal{T}}(\mathcal{O}), \qquad \forall\, \boldsymbol{X} \in \mathcal{O}. \tag{19}$$

Within $\mathcal{O}$, let $[\boldsymbol{X}]_{\mathcal{T}}$ denote the class of $\boldsymbol{X}$ under the behavioral equivalence relation $\sim_{\mathcal{T}}$ in equation 6. These classes are exactly the fibers of $\Phi_{\mathcal{T}}$, so the criterion in Appendix A equips the quotient $\mathcal{M}_{\mathcal{T}} = \mathcal{O}/\sim_{\mathcal{T}}$ with a smooth structure. The induced diffeomorphism is $\widetilde{\Phi}_{\mathcal{T}}([\boldsymbol{X}]_{\mathcal{T}}) = \Phi_{\mathcal{T}}(\boldsymbol{X})$. The canonical projection $\pi_{\mathcal{T}}(\boldsymbol{X}) = [\boldsymbol{X}]_{\mathcal{T}}$ therefore factors as:

$$\pi_{\mathcal{T}} = \widetilde{\Phi}_{\mathcal{T}}^{-1} \circ \Phi_{\mathcal{T}}.$$

Since the differential of $\widetilde{\Phi}_{\mathcal{T}}$ is invertible, the projection has full quotient rank:

$$\operatorname{rank} D\pi_{\mathcal{T}}(\boldsymbol{X}) = \dim \Phi_{\mathcal{T}}(\mathcal{O}) = \dim \mathcal{M}_{\mathcal{T}}. \tag{20}$$

Thus $\pi_{\mathcal{T}}$ is a smooth submersion, establishing $\mathcal{M}_{\mathcal{T}}$ as a local quotient manifold.

Induced Riemannian Metric. Differentiating $\Phi_{\mathcal{T}} = \widetilde{\Phi}_{\mathcal{T}} \circ \pi_{\mathcal{T}}$ identifies the quotient tangent space $T_{[\boldsymbol{X}]_{\mathcal{T}}}\mathcal{M}_{\mathcal{T}}$ with the image of the behavior differential:

$$D\Phi_{\mathcal{T}}(\boldsymbol{X}) = D\widetilde{\Phi}_{\mathcal{T}}\left([\boldsymbol{X}]_{\mathcal{T}}\right) \circ D\pi_{\mathcal{T}}(\boldsymbol{X}).$$

Invertibility of the first factor gives $\ker D\pi_{\mathcal{T}}(\boldsymbol{X}) = \ker D\Phi_{\mathcal{T}}(\boldsymbol{X})$. The resulting tangent identification is:

$$T_{[\boldsymbol{X}]_{\mathcal{T}}}\mathcal{M}_{\mathcal{T}} \simeq \mathcal{W}/\ker D\Phi_{\mathcal{T}}(\boldsymbol{X}) \simeq \operatorname{im} D\Phi_{\mathcal{T}}(\boldsymbol{X}). \tag{21}$$

We define g by pulling back the Euclidean inner product $\langle\cdot,\cdot\rangle_2$ on the agent decision representations through $\widetilde{\Phi}_{\mathcal{T}}$ (Boumal, 2023). Writing * for metric pullback, the factorization of $\Phi_{\mathcal{T}}$ gives:

$$\pi_{\mathcal{T}}^* g = \pi_{\mathcal{T}}^* \widetilde{\Phi}_{\mathcal{T}}^* \langle\cdot,\cdot\rangle_2 = \Phi_{\mathcal{T}}^* \langle\cdot,\cdot\rangle_2. \tag{22}$$

This identity determines g uniquely because $\pi_{\mathcal{T}}$ is a surjective submersion. For directions $\boldsymbol{U}, \boldsymbol{V} \in \mathcal{W}$ of the adaptation matrices, the chain rule gives its explicit form:

$$\begin{aligned} &g_{[\boldsymbol{X}]_{\mathcal{T}}}\left(D\pi_{\mathcal{T}}(\boldsymbol{X})[\boldsymbol{U}], D\pi_{\mathcal{T}}(\boldsymbol{X})[\boldsymbol{V}]\right) \\ &\quad = \left\langle D\Phi_{\mathcal{T}}(\boldsymbol{X})[\boldsymbol{U}], D\Phi_{\mathcal{T}}(\boldsymbol{X})[\boldsymbol{V}]\right\rangle_2. \end{aligned} \tag{23}$$

The definition through $\widetilde{\Phi}_{\mathcal{T}}$ makes g smooth and independent of the parameter representative. The kernel identity ensures positive definiteness on the quotient tangent space:

$$\begin{aligned} g_{[\boldsymbol{X}]_{\mathcal{T}}}\left(D\pi_{\mathcal{T}}(\boldsymbol{X})[\boldsymbol{U}], D\pi_{\mathcal{T}}(\boldsymbol{X})[\boldsymbol{U}]\right) = 0 \quad &\Longleftrightarrow \quad D\Phi_{\mathcal{T}}(\boldsymbol{X})[\boldsymbol{U}] = 0 \\ &\Longleftrightarrow \quad D\pi_{\mathcal{T}}(\boldsymbol{X})[\boldsymbol{U}] = 0. \end{aligned} \tag{24}$$

Thus $D\widetilde{\Phi}_{\mathcal{T}}([\boldsymbol{X}]_{\mathcal{T}})$ is an isometry onto $\operatorname{im} D\Phi_{\mathcal{T}}(\boldsymbol{X})$, preserving inner products between quotient tangents and their decision responses. Let $\|\cdot\|_g$ denote the norm induced by g. For the projected trajectory directions $\widetilde{\boldsymbol{U}}_i$ in BQB, the quotient tangent distance in equation 7 is:

$$d_{ij}^2 = \left\| D\pi_{\mathcal{T}}(\boldsymbol{X})[\widetilde{\boldsymbol{U}}_i - \widetilde{\boldsymbol{U}}_j] \right\|_g^2 = \left\| D\Phi_{\mathcal{T}}(\boldsymbol{X})[\widetilde{\boldsymbol{U}}_i - \widetilde{\boldsymbol{U}}_j] \right\|_2^2. \tag{25}$$

Consequently $d_{ij} = 0$ exactly when $\widetilde{\boldsymbol{U}}_i - \widetilde{\boldsymbol{U}}_j \in \ker D\Phi_{\mathcal{T}}(\boldsymbol{X})$. Thus d_{ij} is a pseudometric on projected parameter directions and a metric on their quotient tangents. The geometry depends on LoRA factors only through the adaptation matrices, so equivalent factorizations induce the same metric.

C Joint Preservation of Balanced Behavior

We bound the behavioral error after balancing and compression relative to an equal-group reference target. The reference histories remain fixed throughout. Write $J = D\Phi_{\mathcal{T}}(\boldsymbol{X})$ and let $\operatorname{Proj}_{\boldsymbol{X}}$ apply the projections in equation 11 to all layers. Operator norms $\|\cdot\|_{\mathrm{op}}$ and adjoints * use the product Frobenius and Euclidean inner products.

C.1 Density Compensation and the Reference Target

Fix any $\alpha \in [0,1]$. Let $\{\mathcal{I}_a\}_{a=1}^K$ be an analysis partition of the trajectories into K nonempty groups of sizes $n_a = |\mathcal{I}_a|$. The partition defines the equal-group reference target used in the analysis. For $i \in \mathcal{I}_a$, the equal-group coefficient is $p_i^\star = 1/(Kn_a)$. The method uses $p_i = w_i/N = \rho_i^{-\alpha}/\sum_j \rho_j^{-\alpha}$, with the positive densities used in equation 8. The reference target is:

$$\boldsymbol{T}_\star = \boldsymbol{X} + \eta \operatorname{Proj}_{\boldsymbol{X}}\left(\sum_i p_i^\star \boldsymbol{U}_i\right). \tag{26}$$

The actual balanced target uses the same LoRA tangent projection:

$$\boldsymbol{T} = \boldsymbol{X} + \eta \operatorname{Proj}_{\boldsymbol{X}}\left(\sum_i p_i \boldsymbol{U}_i\right).$$

For $i \in \mathcal{I}_a$, define $\kappa_i = \rho_i^\alpha/n_a > 0$ to measure the remaining imbalance after density weighting. Let $u_{\max} = \max_i \|\operatorname{Proj}_{\boldsymbol{X}}(\boldsymbol{U}_i)\|_{\mathcal{W}}$ bound the projected directions. Their diameter is $\operatorname{diam}_{\mathcal{W}} = \max_{i,j} \|\operatorname{Proj}_{\boldsymbol{X}}(\boldsymbol{U}_i - \boldsymbol{U}_j)\|_{\mathcal{W}}$. The corresponding response diameter is $\operatorname{diam}_\Phi = \max_{i,j} \|J[\operatorname{Proj}_{\boldsymbol{X}}(\boldsymbol{U}_i - \boldsymbol{U}_j)]\|_2$. Assume that $\Phi_{\mathcal{T}}$ is twice continuously differentiable on a convex neighborhood containing $\boldsymbol{X}$ and both targets, with the bilinear operator norm of $D^2\Phi_{\mathcal{T}}$ bounded by $L_\Phi \geq 0$. Define the density compensation allowance:

$$\delta = \eta\left(\operatorname{diam}_\Phi + \eta L_\Phi u_{\max} \operatorname{diam}_{\mathcal{W}}\right) \tanh\left(\frac{1}{4}\log\frac{\max_i \kappa_i}{\min_i \kappa_i}\right). \tag{27}$$

Proposition 1 (Finite target error from density compensation)**.** *Under the preceding assumptions, the actual target satisfies:*

$$\|\Phi_{\mathcal{T}}(\boldsymbol{T}) - \Phi_{\mathcal{T}}(\boldsymbol{T}_\star)\|_2 \le \delta. \tag{28}$$

Proof. Put $z_i = \kappa_i^{-1}$ and $\mu = \sum_i p_i^\star z_i$. The trajectory coefficients satisfy:

$$\begin{aligned} p_i &= \frac{\rho_i^{-\alpha}}{\sum_j \rho_j^{-\alpha}} \\ &= \frac{n_a^{-1}\kappa_i^{-1}}{\sum_b \sum_{j\in\mathcal{I}_b} n_b^{-1}\kappa_j^{-1}} \\ &= \frac{p_i^\star z_i}{\mu}, \qquad i \in \mathcal{I}_a. \end{aligned}$$

Let $z_- = \min_i z_i$ and $z_+ = \max_i z_i$. Write $d_{\mathrm{TV}} = \frac{1}{2}\sum_i |p_i - p_i^\star|$ for the total variation between the coefficient vectors. For $z_- < z_+$, the chord bound for the function $z \mapsto |z - \mu|$ on $[z_-, z_+]$ gives:

$$\begin{aligned} d_{\mathrm{TV}} &= \frac{1}{2\mu}\sum_i p_i^\star |z_i - \mu| \\ &\le \frac{(z_+ - \mu)(\mu - z_-)}{(z_+ - z_-)\mu} \\ &= \frac{z_+ + z_- - \mu - z_+ z_-/\mu}{z_+ - z_-} \\ &\le \frac{\sqrt{z_+} - \sqrt{z_-}}{\sqrt{z_+} + \sqrt{z_-}} \\ &= \tanh\left(\frac{1}{4}\log\frac{\max_i \kappa_i}{\min_i \kappa_i}\right). \end{aligned} \tag{29}$$

The penultimate step uses $\mu + z_+ z_-/\mu \ge 2\sqrt{z_+ z_-}$. When $z_- = z_+$, both sides of equation 29 vanish. The positive and negative parts of $p_i - p_i^\star$ have equal mass. Their projected response means are at most diam_Φ apart, so:

$$\|J[\boldsymbol{T} - \boldsymbol{T}_\star]\|_2 = \eta \left\| \sum_i (p_i - p_i^\star) J[\mathrm{Proj}_{\boldsymbol{X}}(\boldsymbol{U}_i)] \right\|_2 \le \eta \,\mathrm{diam}_\Phi \, d_{\mathrm{TV}}.$$

The same argument in matrix space gives $\|\boldsymbol{T} - \boldsymbol{T}_\star\|_{\mathcal{W}} \le \eta \,\mathrm{diam}_{\mathcal{W}}\, d_{\mathrm{TV}}$. Both targets lie within distance $\eta u_{\max}$ of $\boldsymbol{X}$, as does their connecting segment. Integration along this segment gives:

$$\begin{aligned} \|\Phi_{\mathcal{T}}(\boldsymbol{T}) - \Phi_{\mathcal{T}}(\boldsymbol{T}_\star)\|_2 &\le \|J[\boldsymbol{T} - \boldsymbol{T}_\star]\|_2 \\ &\quad + \int_0^1 L_\Phi \|(1-t)\boldsymbol{T}_\star + t\boldsymbol{T} - \boldsymbol{X}\|_{\mathcal{W}} \|\boldsymbol{T} - \boldsymbol{T}_\star\|_{\mathcal{W}} \, dt \\ &\le \eta \,\mathrm{diam}_\Phi \, d_{\mathrm{TV}} + \eta^2 L_\Phi u_{\max} \,\mathrm{diam}_{\mathcal{W}}\, d_{\mathrm{TV}} \\ &\le \delta. \end{aligned} \tag{30}$$

If all κ_i agree, then $p_i = p_i^\star$ and $\boldsymbol{T} = \boldsymbol{T}_\star$. □

The density estimate holds for any positive ρ_i, including those computed from the projected trajectory responses in Algorithm 1.

If ρ_i / n_a is constant across trajectories, the total coefficient of group $\mathcal{I}_a$ satisfies:

$$\sum_{i\in\mathcal{I}_a} p_i = \frac{n_a^{1-\alpha}}{\sum_{b=1}^K n_b^{1-\alpha}}. \tag{31}$$

For unequal group sizes, increasing α reduces the ratio between the largest and smallest group contributions, reaching equal contributions at $\alpha = 1$. The variation of κ_i retains this remaining dependence on group size in equation 29.

Density accuracy from response separation. The Gaussian affinities allow the density allowance to be related to the separation between projected responses. Suppose distances within each group are at most $d_{\text{in}} \geq 0$, while distances between groups are at least $d_{\text{out}} \geq 0$. The corresponding affinity bounds are $q_{\text{in}} = \exp(-d_{\text{in}}^2/(2\sigma_{\mathcal{T}}^2))$ and $q_{\text{out}} = \exp(-d_{\text{out}}^2/(2\sigma_{\mathcal{T}}^2))$. Let $n_{\min} = \min_a n_a$ and $n_{\max} = \max_a n_a$.

Proposition 2 (Density compensation under response separation)**.** *Under these distance bounds, the imbalance factor satisfies:*

$$\frac{\max_i \kappa_i}{\min_i \kappa_i} \leq \left(\frac{n_{\max}}{n_{\min}}\right)^{1-\alpha} \left(\frac{1 + (N/n_{\min} - 1)q_{\text{out}}}{q_{\text{in}}}\right)^{\alpha}. \tag{32}$$

Proof. For $i \in \mathcal{I}_a$, the within-group terms contribute at least $n_a q_{\text{in}}$, and each term from another group is at most q_{out}. Since each affinity is at most one, the density satisfies:

$$n_a q_{\text{in}} \leq \rho_i \leq n_a + (N - n_a)q_{\text{out}}. \tag{33}$$

Using $\kappa_i = n_a^{\alpha-1}(\rho_i/n_a)^{\alpha}$ and $0 \leq \alpha \leq 1$ gives:

$$n_{\max}^{\alpha-1} q_{\text{in}}^{\alpha} \leq \kappa_i \leq n_{\min}^{\alpha-1} \left(1 + (N/n_{\min} - 1)q_{\text{out}}\right)^{\alpha}.$$

Dividing the upper bound by the lower bound proves equation 32. □

At $\alpha = 1$, density compensation approaches equal group contributions when q_{in} approaches one and $(N/n_{\min} - 1)q_{\text{out}}$ approaches zero. The second condition accounts for the cumulative contribution from other groups, including when a group contains few trajectories. The factors q_{in} and q_{out} quantify response separation relative to the bandwidth in Appendix E.2. Proposition 1 bounds the target error using the resulting densities.

C.2 Compression within the Rank Budget

Let $\boldsymbol{Y}$ be the layerwise rank-r_l truncated SVD approximation of $\boldsymbol{T}$, with residual $\boldsymbol{R}_0 = \boldsymbol{Y} - \boldsymbol{T}$. Consider any feasible output $\boldsymbol{X}^+ \in \mathcal{C}(\boldsymbol{T})$ satisfying $F_\gamma(\boldsymbol{X}^+) \leq F_\gamma(\boldsymbol{Y})$ for the objective in equation 13. The achieved objective decrease is:

$$\Delta_\gamma = F_\gamma(\boldsymbol{Y}) - F_\gamma(\boldsymbol{X}^+) \geq 0. \tag{34}$$

For $\gamma \geq 0$, define the residual radius:

$$\beta_\gamma^2 = \|\boldsymbol{R}_0\|_{\mathcal{W}}^2 + \gamma \|J\boldsymbol{R}_0\|_2^2. \tag{35}$$

In particular, $\beta_0 = \|\boldsymbol{R}_0\|_{\mathcal{W}}$.

Proposition 3 (Compression after a finite number of updates)**.** *For $\gamma > 0$, every such feasible output satisfies:*

$$\|J[\boldsymbol{X}^+ - \boldsymbol{T}]\|_2^2 \leq \|J\boldsymbol{R}_0\|_2^2 - \frac{\Delta_\gamma}{\gamma}. \tag{36}$$

Proof. Write $\boldsymbol{R}_\gamma = \boldsymbol{X}^+ - \boldsymbol{T}$. Optimality of truncated SVD for the Frobenius error gives $\|\boldsymbol{R}_\gamma\|_{\mathcal{W}}^2 \geq \|\boldsymbol{R}_0\|_{\mathcal{W}}^2$. Expanding the objective decrease therefore gives:

$$\begin{aligned} \gamma\left(\|J\boldsymbol{R}_0\|_2^2 - \|J\boldsymbol{R}_\gamma\|_2^2\right) &= \Delta_\gamma + \|\boldsymbol{R}_\gamma\|_{\mathcal{W}}^2 - \|\boldsymbol{R}_0\|_{\mathcal{W}}^2 \\ &\geq \Delta_\gamma. \end{aligned} \tag{37}$$

Dividing by γ proves equation 36 and shows that its right-hand side is nonnegative. □

The proposition controls the response error at each retained compression iterate. Every strict objective decrease reduces the first-order response error relative to the SVD initialization for the same target. The objective bound also controls the matrix residual:

$$\|\boldsymbol{X}^+ - \boldsymbol{T}\|_{\mathcal{W}}^2 \leq F_\gamma(\boldsymbol{X}^+) = \beta_\gamma^2 - \Delta_\gamma \leq \beta_\gamma^2. \tag{38}$$

Local spectral refinement. We next describe the response components that compression can attenuate at a stationary output. Let $\mathcal{E}$ be the linear space satisfying the target column and row restrictions in DPC, without the rank constraints. Assume that the smallest retained singular value $\sigma_Y = \min_l \sigma_{r_l}(\boldsymbol{Y}_l)$ is positive. Within $\mathcal{E}$, let $\mathcal{S}$ be the fixed-rank tangent space at $\boldsymbol{Y}$, with normal space $\mathcal{S}^\perp$. Let $J_{\mathcal{S}}$ denote the restriction of the behavior differential J to the tangent space $\mathcal{S}$. Its restriction to the normal space $\mathcal{S}^\perp$ is denoted by $J_{\mathcal{S}^\perp}$. Identities act on the indicated spaces. The compression response operator is:

$$\mathcal{F}_\gamma = (I + \gamma J_{\mathcal{S}} J_{\mathcal{S}}^*)^{-1}. \tag{39}$$

Let $\mathcal{G}_{\gamma,\mathcal{S}} = I_{\mathcal{S}} + \gamma J_{\mathcal{S}}^* J_{\mathcal{S}}$ be the quadratic operator on $\mathcal{S}$. For $\gamma > 0$, define the curvature coefficient using the restricted response operators:

$$C_\gamma = \frac{4\|\mathcal{F}_\gamma J_{\mathcal{S}^\perp}\|_{\mathrm{op}} + 2\sqrt{5}\,\left\|J_{\mathcal{S}}\mathcal{G}_{\gamma,\mathcal{S}}^{-1}\right\|_{\mathrm{op}}\sqrt{1+\gamma\|J_{\mathcal{S}^\perp}\|_{\mathrm{op}}^2}}{\sigma_Y}. \tag{40}$$

Proposition 4 (Compression response under the rank constraint)**.** *In addition to $F_\gamma(\boldsymbol{X}^+) \le F_\gamma(\boldsymbol{Y})$, assume $\gamma > 0$ and $\beta_\gamma \le \sigma_Y/4$. If $\boldsymbol{X}^+$ is stationary for F_γ on the fixed-rank manifold within $\mathcal{E}$, then:*

$$\|J[\boldsymbol{X}^+ - \boldsymbol{T}] - \mathcal{F}_\gamma J\boldsymbol{R}_0\|_2 \le C_\gamma \beta_\gamma^2. \tag{41}$$

Proof. Define $\mathcal{G}_\gamma = I_{\mathcal{E}} + \gamma (J|_{\mathcal{E}})^*(J|_{\mathcal{E}})$ and let $\|\cdot\|_{\mathcal{G}_\gamma}$ be its induced norm. For $\boldsymbol{R}_\gamma = \boldsymbol{X}^+ - \boldsymbol{T}$, comparison with the feasible SVD approximation gives:

$$\|\boldsymbol{R}_\gamma\|_{\mathcal{W}}^2 \le \|\boldsymbol{R}_\gamma\|_{\mathcal{G}_\gamma}^2 \le \|\boldsymbol{R}_0\|_{\mathcal{G}_\gamma}^2 = \beta_\gamma^2.$$

The output consequently remains within the fixed-rank neighborhood:

$$\|\boldsymbol{X}^+ - \boldsymbol{Y}\|_{\mathcal{W}} \le \|\boldsymbol{R}_\gamma\|_{\mathcal{W}} + \|\boldsymbol{R}_0\|_{\mathcal{W}} \le 2\beta_\gamma \le \sigma_Y/2.$$

Every layer of $\boldsymbol{X}^+$ therefore has rank r_l. In singular-vector coordinates within the target spaces, write $\boldsymbol{Y}_l = \mathrm{diag}(\boldsymbol{\Sigma}_{Y,l}, 0)$, where $\boldsymbol{\Sigma}_{Y,l}$ contains the retained singular values. The Schur complement expresses a nearby rank-r_l matrix through tangent blocks $\boldsymbol{H}_l^{11}, \boldsymbol{H}_l^{12}, \boldsymbol{H}_l^{21}$ (Boumal, 2023, Section 7.5):

$$\boldsymbol{X}_l^+ = \begin{pmatrix} \boldsymbol{\Sigma}_{Y,l} + \boldsymbol{H}_l^{11} & \boldsymbol{H}_l^{12} \\ \boldsymbol{H}_l^{21} & \boldsymbol{H}_l^{21}(\boldsymbol{\Sigma}_{Y,l} + \boldsymbol{H}_l^{11})^{-1}\boldsymbol{H}_l^{12} \end{pmatrix}. \tag{42}$$

Let $\boldsymbol{H} \in \mathcal{S}$ collect the tangent blocks and let $\boldsymbol{N}(\boldsymbol{H}) \in \mathcal{S}^\perp$ collect the lower right blocks, so $\boldsymbol{X}^+ - \boldsymbol{Y} = \boldsymbol{H} + \boldsymbol{N}(\boldsymbol{H})$. For $h = \|\boldsymbol{H}\|_{\mathcal{W}} \le 2\beta_\gamma \le \sigma_Y/2$, the normal component satisfies:

$$\begin{aligned} \|\boldsymbol{N}(\boldsymbol{H})\|_{\mathcal{W}} &\le \frac{\left(\sum_l \|\boldsymbol{H}_l^{21}\|_F^2 \|\boldsymbol{H}_l^{12}\|_F^2\right)^{1/2}}{\sigma_Y - h} \\ &\le \frac{\sum_l \left(\|\boldsymbol{H}_l^{21}\|_F^2 + \|\boldsymbol{H}_l^{12}\|_F^2\right)}{2(\sigma_Y - h)} \\ &\le \frac{h^2}{2(\sigma_Y - h)} \le \frac{4\beta_\gamma^2}{\sigma_Y}. \end{aligned} \tag{43}$$

For a tangent variation $\boldsymbol{V} \in \mathcal{S}$, put $\boldsymbol{L}_l = (\boldsymbol{\Sigma}_{Y,l} + \boldsymbol{H}_l^{11})^{-1}$. Only the lower right block of the derivative is nonzero:

$$[D\boldsymbol{N}(\boldsymbol{H})[\boldsymbol{V}]]_l^{22} = \boldsymbol{V}_l^{21}\boldsymbol{L}_l\boldsymbol{H}_l^{12} + \boldsymbol{H}_l^{21}\boldsymbol{L}_l\boldsymbol{V}_l^{12} - \boldsymbol{H}_l^{21}\boldsymbol{L}_l\boldsymbol{V}_l^{11}\boldsymbol{L}_l\boldsymbol{H}_l^{12}.$$

Applying the Cauchy Schwarz inequality to the tangent blocks gives:

$$\begin{aligned} \|D\boldsymbol{N}(\boldsymbol{H})\|_{\mathrm{op}}^2 &\le \max_l \left\{ \frac{\|\boldsymbol{H}_l^{12}\|_F^2 + \|\boldsymbol{H}_l^{21}\|_F^2}{(\sigma_Y - h)^2} + \frac{\|\boldsymbol{H}_l^{12}\|_F^2\|\boldsymbol{H}_l^{21}\|_F^2}{(\sigma_Y - h)^4} \right\} \\ &\le \frac{h^2}{(\sigma_Y - h)^2} + \frac{h^4}{4(\sigma_Y - h)^4} \le \frac{5h^2}{\sigma_Y^2} \le \frac{20\beta_\gamma^2}{\sigma_Y^2}. \end{aligned} \tag{44}$$

Let $\Pi_{\mathcal{S}}$ and $\Pi_{\mathcal{S}^\perp}$ project orthogonally onto $\mathcal{S}$ and $\mathcal{S}^\perp$ within $\mathcal{E}$. Truncated SVD gives $\boldsymbol{R}_0 \in \mathcal{S}^\perp$, and $\boldsymbol{R}_\gamma = \boldsymbol{R}_0 + \boldsymbol{H} + \boldsymbol{N}(\boldsymbol{H})$. Stationarity in the chart implies:

$$0 = \Pi_{\mathcal{S}}\mathcal{G}_\gamma\boldsymbol{R}_\gamma + D\boldsymbol{N}(\boldsymbol{H})^*\Pi_{\mathcal{S}^\perp}\mathcal{G}_\gamma\boldsymbol{R}_\gamma.$$

Since $\boldsymbol{R}_0$ and $\boldsymbol{N}(\boldsymbol{H})$ lie in $\mathcal{S}^\perp$, this equation isolates the tangent component:

$$\mathcal{G}_{\gamma,\mathcal{S}}\boldsymbol{H} = -\gamma J_{\mathcal{S}}^* J[\boldsymbol{R}_0 + \boldsymbol{N}(\boldsymbol{H})] - D\boldsymbol{N}(\boldsymbol{H})^* \Pi_{\mathcal{S}^\perp} \mathcal{G}_\gamma \boldsymbol{R}_\gamma.$$

Substitution and the identity $I - \gamma J_{\mathcal{S}} \mathcal{G}_{\gamma,\mathcal{S}}^{-1} J_{\mathcal{S}}^* = \mathcal{F}_\gamma$ give:

$$\begin{aligned} J\boldsymbol{R}_\gamma - \mathcal{F}_\gamma J\boldsymbol{R}_0 &= J[\boldsymbol{R}_0 + \boldsymbol{N}(\boldsymbol{H})] + J_{\mathcal{S}}\boldsymbol{H} - \mathcal{F}_\gamma J\boldsymbol{R}_0 \\ &= \mathcal{F}_\gamma J_{\mathcal{S}^\perp}\boldsymbol{N}(\boldsymbol{H}) - J_{\mathcal{S}}\mathcal{G}_{\gamma,\mathcal{S}}^{-1} D\boldsymbol{N}(\boldsymbol{H})^* \Pi_{\mathcal{S}^\perp}\mathcal{G}_\gamma \boldsymbol{R}_\gamma. \end{aligned} \tag{45}$$

The normal metric residual is bounded in the compression norm:

$$\|\Pi_{\mathcal{S}^\perp}\mathcal{G}_\gamma \boldsymbol{R}_\gamma\|_{\mathcal{W}} \le \left\|\Pi_{\mathcal{S}^\perp}\mathcal{G}_\gamma^{1/2}\right\|_{\mathrm{op}} \|\boldsymbol{R}_\gamma\|_{\mathcal{G}_\gamma} \le \sqrt{1+\gamma\|J_{\mathcal{S}^\perp}\|_{\mathrm{op}}^2}\,\beta_\gamma.$$

Keeping the restricted operator norms in equation 45 completes the estimate:

$$\begin{aligned} \|J\boldsymbol{R}_\gamma - \mathcal{F}_\gamma J\boldsymbol{R}_0\|_2 &\le \|\mathcal{F}_\gamma J_{\mathcal{S}^\perp}\|_{\mathrm{op}} \|\boldsymbol{N}(\boldsymbol{H})\|_{\mathcal{W}} \\ &\quad + \left\|J_{\mathcal{S}}\mathcal{G}_{\gamma,\mathcal{S}}^{-1}\right\|_{\mathrm{op}} \sqrt{1+\gamma\|J_{\mathcal{S}^\perp}\|_{\mathrm{op}}^2}\, \|D\boldsymbol{N}(\boldsymbol{H})\|_{\mathrm{op}}\beta_\gamma \\ &\le \frac{4\|\mathcal{F}_\gamma J_{\mathcal{S}^\perp}\|_{\mathrm{op}} + 2\sqrt{5}\,\left\|J_{\mathcal{S}}\mathcal{G}_{\gamma,\mathcal{S}}^{-1}\right\|_{\mathrm{op}} \sqrt{1+\gamma\|J_{\mathcal{S}^\perp}\|_{\mathrm{op}}^2}}{\sigma_Y}\,\beta_\gamma^2 \\ &= C_\gamma \beta_\gamma^2. \end{aligned}$$

□

Let λ_k be the eigenvalues of $J_{\mathcal{S}}J_{\mathcal{S}}^*$. Write $\boldsymbol{c} = J\boldsymbol{R}_0$ and let c_k be its coordinates in a corresponding orthonormal eigenbasis. The leading compression response has the spectral representation:

$$\|\mathcal{F}_\gamma \boldsymbol{c}\|_2^2 = \sum_k \frac{|c_k|^2}{(1+\gamma\lambda_k)^2}. \tag{46}$$

Compression attenuates each response component with $\lambda_k > 0$ and leaves components orthogonal to the reachable response space unchanged at this order.

C.3 Proof of Theorem 1

The final error is measured against the equal-group target in equation 26:

$$E_{\mathrm{BQ}} = \|\Phi_{\mathcal{T}}(\boldsymbol{X}^+) - \Phi_{\mathcal{T}}(\boldsymbol{T}_\star)\|_2. \tag{47}$$

Take $\gamma > 0$ and a feasible compression output satisfying $F_\gamma(\boldsymbol{X}^+) \le F_\gamma(\boldsymbol{Y})$. Assume that $\Phi_{\mathcal{T}}$ is twice continuously differentiable on a convex neighborhood containing $\boldsymbol{X}$ and both targets, with $\|D^2\Phi_{\mathcal{T}}\|_{\mathrm{op}} \le L_\Phi$. Require this neighborhood to contain $\boldsymbol{X}^+$ as well. The residual bound in equation 38 places this output and both targets in the closed ball centered at $\boldsymbol{X}$ with radius $\eta u_{\max} + \beta_\gamma$. Define the finite-step remainder:

$$\epsilon_\gamma = \eta^2 L_\Phi u_{\max}\,\mathrm{diam}_{\mathcal{W}} \tanh\left(\frac{1}{4}\log\frac{\max_i \kappa_i}{\min_i \kappa_i}\right) + L_\Phi\left(\|\boldsymbol{T}-\boldsymbol{X}\|_{\mathcal{W}} + \frac{\beta_\gamma}{2}\right)\beta_\gamma. \tag{48}$$

Proof. Integration from the balanced target controls the nonlinear compression response:

$$\begin{aligned} \|\Phi_{\mathcal{T}}(\boldsymbol{X}^+) - \Phi_{\mathcal{T}}(\boldsymbol{T}) - J[\boldsymbol{X}^+ - \boldsymbol{T}]\|_2 &= \left\|\int_0^1 \left(D\Phi_{\mathcal{T}}(\boldsymbol{T} + t[\boldsymbol{X}^+ - \boldsymbol{T}]) - J\right)[\boldsymbol{X}^+ - \boldsymbol{T}]\,dt\right\|_2 \\ &\le \int_0^1 L_\Phi\left(\|\boldsymbol{T}-\boldsymbol{X}\|_{\mathcal{W}} + t\|\boldsymbol{X}^+ - \boldsymbol{T}\|_{\mathcal{W}}\right)\|\boldsymbol{X}^+ - \boldsymbol{T}\|_{\mathcal{W}}\,dt \\ &\le L_\Phi\left(\|\boldsymbol{T}-\boldsymbol{X}\|_{\mathcal{W}} + \frac{\beta_\gamma}{2}\right)\beta_\gamma. \end{aligned} \tag{49}$$

The triangle inequality combines the density and compression errors:

$$\begin{aligned} E_{\mathrm{BQ}} &\le \|\Phi_{\mathcal{T}}(\boldsymbol{T}) - \Phi_{\mathcal{T}}(\boldsymbol{T}_\star)\|_2 + \|J[\boldsymbol{X}^+ - \boldsymbol{T}]\|_2 \\ &\quad + \|\Phi_{\mathcal{T}}(\boldsymbol{X}^+) - \Phi_{\mathcal{T}}(\boldsymbol{T}) - J[\boldsymbol{X}^+ - \boldsymbol{T}]\|_2 \\ &\le \delta + \sqrt{\|J\boldsymbol{R}_0\|_2^2 - \frac{\Delta_\gamma}{\gamma}} + L_\Phi \left(\|\boldsymbol{T} - \boldsymbol{X}\|_{\mathcal{W}} + \frac{\beta_\gamma}{2} \right) \beta_\gamma \\ &= \eta \,\mathrm{diam}_\Phi \tanh\left(\frac{1}{4} \log \frac{\max_i \kappa_i}{\min_i \kappa_i} \right) + \sqrt{\|J\boldsymbol{R}_0\|_2^2 - \frac{\Delta_\gamma}{\gamma}} + \epsilon_\gamma. \end{aligned} \tag{50}$$

The last inequality uses Propositions 1 and 3 together with equation 49. Substituting $\boldsymbol{R}_0 = \boldsymbol{Y} - \boldsymbol{T}$ proves equation 14. □

Size of the finite-step remainder. The LoRA tangent constraint makes the rank compression residual quadratic in the step size. Let $\sigma_X = \min_l \sigma_{r_l}(\boldsymbol{X}_l) > 0$ and assume $\eta u_{\max} \le \sigma_X/2$. Write $\boldsymbol{U}_{\mathrm{proj}} = \mathrm{Proj}_{\boldsymbol{X}}(\boldsymbol{U}_{\mathrm{BQ}}) = (\boldsymbol{T} - \boldsymbol{X})/\eta$, with $\|\boldsymbol{U}_{\mathrm{proj}}\|_{\mathcal{W}} \le u_{\max}$. In singular-vector coordinates of $\boldsymbol{X}_l$, let $\boldsymbol{\Sigma}_{X,l}$ contain its nonzero singular values and let $\boldsymbol{U}^{11}_{\mathrm{proj},l}, \boldsymbol{U}^{12}_{\mathrm{proj},l}, \boldsymbol{U}^{21}_{\mathrm{proj},l}$ be the tangent blocks of $\boldsymbol{U}_{\mathrm{proj},l}$. The lower right block of $\boldsymbol{T}_l$ is zero because $\boldsymbol{U}_{\mathrm{proj},l} \in T_{\boldsymbol{X}_l}\mathcal{R}_{r_l}$. Completing this block with $\eta^2 \boldsymbol{U}^{21}_{\mathrm{proj},l}(\boldsymbol{\Sigma}_{X,l} + \eta \boldsymbol{U}^{11}_{\mathrm{proj},l})^{-1}\boldsymbol{U}^{12}_{\mathrm{proj},l}$ gives a rank-r_l comparison matrix. Optimality of truncated SVD therefore gives:

$$\begin{aligned} \beta_0 &\le \frac{\eta^2 \left(\sum_l \|\boldsymbol{U}^{21}_{\mathrm{proj},l}\|_F^2 \|\boldsymbol{U}^{12}_{\mathrm{proj},l}\|_F^2 \right)^{1/2}}{\sigma_X - \eta u_{\max}} \\ &\le \frac{\eta^2 u_{\max}^2}{2(\sigma_X - \eta u_{\max})} \le \frac{\eta^2 u_{\max}^2}{\sigma_X}. \end{aligned} \tag{51}$$

The weighted residual consequently satisfies:

$$\beta_\gamma \le \sqrt{1 + \gamma \|J\|_{\mathrm{op}}^2}\, \beta_0 \le \sqrt{1 + \gamma \|J\|_{\mathrm{op}}^2}\, \frac{\eta^2 u_{\max}^2}{\sigma_X}.$$

Let $b_\gamma = \sqrt{1 + \gamma \|J\|_{\mathrm{op}}^2}\, u_{\max}^2/\sigma_X$, so that $\beta_\gamma \le b_\gamma \eta^2$. Substituting this bound and $\|\boldsymbol{T} - \boldsymbol{X}\|_{\mathcal{W}} \le \eta u_{\max}$ into equation 48 gives:

$$\epsilon_\gamma \le \eta^2 L_\Phi u_{\max}\, \mathrm{diam}_{\mathcal{W}} \tanh\left(\frac{1}{4} \log \frac{\max_i \kappa_i}{\min_i \kappa_i} \right) + L_\Phi u_{\max} b_\gamma \eta^3 + \frac{L_\Phi b_\gamma^2}{2} \eta^4.$$

When the values $\kappa_i = \rho_i^\alpha / n_a$ agree, the density term vanishes. For equal κ_i and fixed γ, the compression term in equation 14 is at most $\|J\|_{\mathrm{op}} u_{\max}^2 \eta^2/\sigma_X$, while the remaining error is bounded by the explicit cubic and quartic terms above.

C.4 Decision Distributions at the Reference Prefixes

The frozen output head relates representation preservation to conditional decision distributions. Let $\boldsymbol{W}_{\mathrm{out}}$ denote its matrix and let $p_{\boldsymbol{Z}}(\cdot \mid h)$ be the next-token distribution at a fixed prefix h for adaptation $\boldsymbol{Z}$:

$$p_{\boldsymbol{Z}}(\cdot \mid h) = \mathrm{softmax}\left(\boldsymbol{W}_{\mathrm{out}} f_{\boldsymbol{Z}}(h) \right). \tag{52}$$

For adaptations $\boldsymbol{Z}$ and $\boldsymbol{Z}'$, average their Kullback Leibler divergence with the normalization of equation 5:

$$\mathcal{K}_{\mathcal{T}}(\boldsymbol{Z}, \boldsymbol{Z}') = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T_i} \sum_{k=1}^{T_i} \frac{1}{|a_{ik}|} \sum_{t=1}^{|a_{ik}|} \mathrm{KL}\left(p_{\boldsymbol{Z}}(\cdot \mid h_{ik,t}) \,\|\, p_{\boldsymbol{Z}'}(\cdot \mid h_{ik,t}) \right). \tag{53}$$

Proposition 5 (Conditional decision preservation)**.** *For the same fixed reference prefixes, the average divergence satisfies:*

$$\mathcal{K}_{\mathcal{T}}(\boldsymbol{Z}, \boldsymbol{Z}') \le \frac{\|\boldsymbol{W}_{\mathrm{out}}\|_{\mathrm{op}}^2}{4} \|\Phi_{\mathcal{T}}(\boldsymbol{Z}) - \Phi_{\mathcal{T}}(\boldsymbol{Z}')\|_2^2. \tag{54}$$

Proof. For a logit vector $\boldsymbol{z}$, define $A(\boldsymbol{z}) = \log \sum_v \exp(z_v)$ and write $\boldsymbol{p} = \mathrm{softmax}(\boldsymbol{z})$. The Hessian of A is $\mathrm{diag}(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^\top$. For any vector $\boldsymbol{u}$, the variance bound for a variable supported on $[\min_v u_v, \max_v u_v]$ gives:

$$\boldsymbol{u}^\top \nabla^2 A(\boldsymbol{z})\boldsymbol{u} = \mathrm{Var}_{v\sim\boldsymbol{p}}(u_v) \le \frac{(\max_v u_v - \min_v u_v)^2}{4} \le \frac{\|\boldsymbol{u}\|_2^2}{2}. \tag{55}$$

Taylor expansion of A consequently gives:

$$\begin{aligned} \mathrm{KL}\left(\mathrm{softmax}(\boldsymbol{z}) \,\|\, \mathrm{softmax}(\boldsymbol{z}')\right) &= A(\boldsymbol{z}') - A(\boldsymbol{z}) - \nabla A(\boldsymbol{z})^\top(\boldsymbol{z}' - \boldsymbol{z}) \\ &\le \frac{1}{4}\|\boldsymbol{z}' - \boldsymbol{z}\|_2^2. \end{aligned} \tag{56}$$

Applying this inequality to the logits at each reference prefix gives:

$$\mathrm{KL}\left(p_{\boldsymbol{Z}}(\cdot \mid h) \,\|\, p_{\boldsymbol{Z}'}(\cdot \mid h)\right) \le \frac{\|\boldsymbol{W}_{\mathrm{out}}\|_{\mathrm{op}}^2}{4}\|f_{\boldsymbol{Z}}(h) - f_{\boldsymbol{Z}'}(h)\|_2^2.$$

Averaging over the recorded decision prefixes with the weights in equation 53 proves equation 54. □

Taking $\boldsymbol{Z} = \boldsymbol{X}^+$ and $\boldsymbol{Z}' = \boldsymbol{T}_\star$ applies Theorem 1 to conditional decision distributions at the fixed reference prefixes.

D Training Algorithm

Algorithm 1 presents the overall training procedure of `BQ-LoRA`.

E Additional Experiments

E.1 Datasets

AppWorld. AppWorld (Trivedi et al., 2024) evaluates API-based task completion in simulated applications. We train on 642 successful trajectories generated by Qwen3.8-max across 87 training tasks. We evaluate on the full Normal and Challenge test sets, which contain 168 and 417 tasks, respectively. Task Goal Completion (TGC) measures the percentage of tasks that pass all evaluation tests. Scenario Goal Completion (SGC) measures the percentage of scenarios in which all three constituent tasks succeed, capturing consistency across task variants.

BrowseComp-Plus. BrowseComp-Plus (Chen et al., 2026c) evaluates iterative retrieval and evidence synthesis. We train on 668 trajectories generated by Qwen3.8-max and filtered for answer correctness. Following Context-Folding (Sun et al., 2026), we evaluate on a fixed subset of 150 questions disjoint from the training set. Accuracy measures the percentage of questions for which the final answer is judged semantically equivalent to the reference answer. Recall measures the fraction of annotated evidence documents retrieved across all search calls for each question, averaged over questions.

E.2 Implementation Details

The experimental server is equipped with eight NVIDIA H20 GPUs. We tune learning rates separately for each method. Full fine-tuning uses a learning rate of 5×10^{-6}, and LoRA uses 2×10^{-4}. The default training budgets are five epochs on AppWorld and three epochs on BrowseComp-Plus, with cosine learning rate decay and a warmup over the first 1% of training steps. For adapter methods, we adapt the same text attention and feed-forward projections, including the input and output projections of gated linear attention. The token embeddings and output head remain frozen. Training supervises assistant tokens, including recorded reasoning, while masking prompts and tool observations. For each trajectory τ_i, we compute $\mathcal{L}_i(\boldsymbol{X})$ by averaging its token losses over its supervised assistant tokens. BQB combines these trajectory losses with coefficients w_i/N, giving the weighted objective $\frac{1}{N}\sum_{i=1}^N w_i \mathcal{L}_i(\boldsymbol{X})$. BQB evaluates equation 7 on the projected directions $\widetilde{\boldsymbol{U}}_i$ and uses

Algorithm 1 BQ-LoRA training

Input Training trajectories $\mathcal{D}$, batch size N, and a prescribed training budget
Input Frozen matrices $\{\boldsymbol{W}_l^0\}_{l=1}^L$ and initial factors $\{\boldsymbol{A}_l, \boldsymbol{B}_l\}_{l=1}^L$
Input Rank budgets r_l, scales $s_l > 0$, learning rate $\eta > 0$, and balancing strength $\alpha \in [0, 1]$
Input Positive bandwidth rule $\sigma_{\mathcal{T}}$ and compression strength $\gamma > 0$
Output Trained LoRA factors $\{\boldsymbol{A}_l, \boldsymbol{B}_l\}_{l=1}^L$

For zero-initialized adapters, perform a short warmup on the LoRA factors within the training budget
while the training budget is not exhausted **do**
  Select $\mathcal{T} = \{\tau_i\}_{i=1}^N$ from $\mathcal{D}$ and retain its recorded reference prefixes
  $\boldsymbol{X} \leftarrow (s_l \boldsymbol{B}_l \boldsymbol{A}_l)_{l=1}^L$
  Construct $\Phi_{\mathcal{T}}$ using equation 5
  Compute the orthonormal column and row bases $\boldsymbol{P}_l, \boldsymbol{Q}_l$ for each layer
  Behavior quotient balancing
  for $i = 1, \ldots, N$ **do**
    $\boldsymbol{U}_i \leftarrow -\nabla_{\boldsymbol{X}} \mathcal{L}_i(\boldsymbol{X})$
    $\widetilde{\boldsymbol{U}}_i \leftarrow \mathrm{Proj}_{\boldsymbol{X}}(\boldsymbol{U}_i)$ using the layerwise projection in equation 11
    $\boldsymbol{v}_i \leftarrow D\Phi_{\mathcal{T}}(\boldsymbol{X})[\widetilde{\boldsymbol{U}}_i]$ using a Jacobian-vector product
  end for
  $d_{ij}^2 \leftarrow \|\boldsymbol{v}_i - \boldsymbol{v}_j\|_2^2$ for all $i, j \in \{1, \ldots, N\}$
  Choose $\sigma_{\mathcal{T}} > 0$ using the prescribed bandwidth rule
  $\rho_i \leftarrow \sum_{j=1}^N \exp\left(-d_{ij}^2/(2\sigma_{\mathcal{T}}^2)\right)$ for all i
  $w_i \leftarrow \rho_i^{-\alpha} / \left(\frac{1}{N} \sum_{j=1}^N \rho_j^{-\alpha}\right)$ for all i
  Hold w_i fixed and form $\boldsymbol{U}_{\mathrm{BQ}} \leftarrow \frac{1}{N} \sum_{i=1}^N w_i \boldsymbol{U}_i$
  Decision preserving compression
  for $l = 1, \ldots, L$ **do**
    Compute $\mathrm{Proj}_{\boldsymbol{X}_l}(\boldsymbol{U}_{\mathrm{BQ},l})$ using equation 11
    $\boldsymbol{T}_l \leftarrow \boldsymbol{X}_l + \eta\, \mathrm{Proj}_{\boldsymbol{X}_l}(\boldsymbol{U}_{\mathrm{BQ},l})$
  end for
  Collect $\boldsymbol{T} \leftarrow (\boldsymbol{T}_l)_{l=1}^L$ and hold $D\Phi_{\mathcal{T}}(\boldsymbol{X})$ fixed
  Form $\mathcal{C}(\boldsymbol{T})$ using the following constraints for each layer l:

$$\mathrm{rank}(\boldsymbol{Z}_l) \le r_l,$$
$$\mathrm{col}(\boldsymbol{Z}_l) \subseteq \mathrm{col}(\boldsymbol{T}_l),$$
$$\mathrm{col}(\boldsymbol{Z}_l^\top) \subseteq \mathrm{col}(\boldsymbol{T}_l^\top).$$

  $\boldsymbol{X}^+ \in \arg\min_{\boldsymbol{Z} \in \mathcal{C}(\boldsymbol{T})} \left\{ \|\boldsymbol{Z} - \boldsymbol{T}\|_{\mathcal{W}}^2 + \gamma \|D\Phi_{\mathcal{T}}(\boldsymbol{X})[\boldsymbol{Z} - \boldsymbol{T}]\|_2^2 \right\}$
  for $l = 1, \ldots, L$ **do**
    Factorize $\boldsymbol{X}_l^+ = s_l \boldsymbol{B}_l^+ \boldsymbol{A}_l^+$ under the rank budget r_l
    $\boldsymbol{A}_l \leftarrow \boldsymbol{A}_l^+$ and $\boldsymbol{B}_l \leftarrow \boldsymbol{B}_l^+$
  end for
end while
return $\{\boldsymbol{A}_l, \boldsymbol{B}_l\}_{l=1}^L$

these distances to compute trajectory weights. DPC uses the same map to measure distortion of the target response during compression. Both modules use the same fixed reference histories within each update. We set $\sigma_{\mathcal{T}}^2$ to the median of the strictly positive squared distances d_{ij}^2 over $i < j$ in the current batch. The bandwidth is recomputed for each batch, and coincident responses receive uniform weights when all distances vanish. Trajectory weights are held fixed when differentiating the training loss. DPC jointly optimizes the layer cores with at most three Riemannian gradient steps from a truncated SVD initialization. The behavior derivative includes all adapted layers, while layerwise SVD retractions enforce their rank budgets. Backtracking permits four trials with step halving and accepts decreases exceeding 10^{-8} of the initial objective. The solver stops on a negligible projected gradient or a failed search. It also stops when an accepted decrease is below 10^{-3} of the initial objective. Derivative products reuse intermediate quantities from the fixed linearization within each update. The rank of 32 is specified per adapter, so methods with multiple adapters have larger total trainable parameter budgets.

Table 4: Component ablations on AppWorld.

Model	Method	Test-N		Test-C	
		TGC	SGC	TGC	SGC
Qwen3.5-4B	LoRA	52.98 ± 2.38	32.14 ± 3.09	35.49 ± 1.44	17.99 ± 1.90
	BQ-LoRA w/o BQB	59.72 ± 1.82	42.26 ± 2.73	40.77 ± 1.10	22.54 ± 1.50
	BQ-LoRA w/o DPC	56.94 ± 2.09	38.10 ± 2.73	41.33 ± 0.97	23.26 ± 1.50
	BQ-LoRA	$\mathbf{60.91} \pm 1.24$	$\mathbf{44.64} \pm 1.79$	$\mathbf{44.04} \pm 1.08$	$\mathbf{24.46} \pm 1.25$
Qwen3.5-9B	LoRA	60.71 ± 2.06	42.86 ± 3.57	45.08 ± 1.20	20.86 ± 2.16
	BQ-LoRA w/o BQB	68.65 ± 2.09	54.76 ± 2.73	49.88 ± 1.20	28.30 ± 1.81
	BQ-LoRA w/o DPC	64.68 ± 1.82	48.21 ± 1.79	50.20 ± 0.97	29.50 ± 1.44
	BQ-LoRA	$\mathbf{70.44} \pm 1.50$	$\mathbf{57.14} \pm 1.79$	$\mathbf{53.32} \pm 0.97$	$\mathbf{31.65} \pm 1.25$

Table 5: Component ablations on BrowseComp-Plus. Bold values indicate the best result for each metric within each model group.

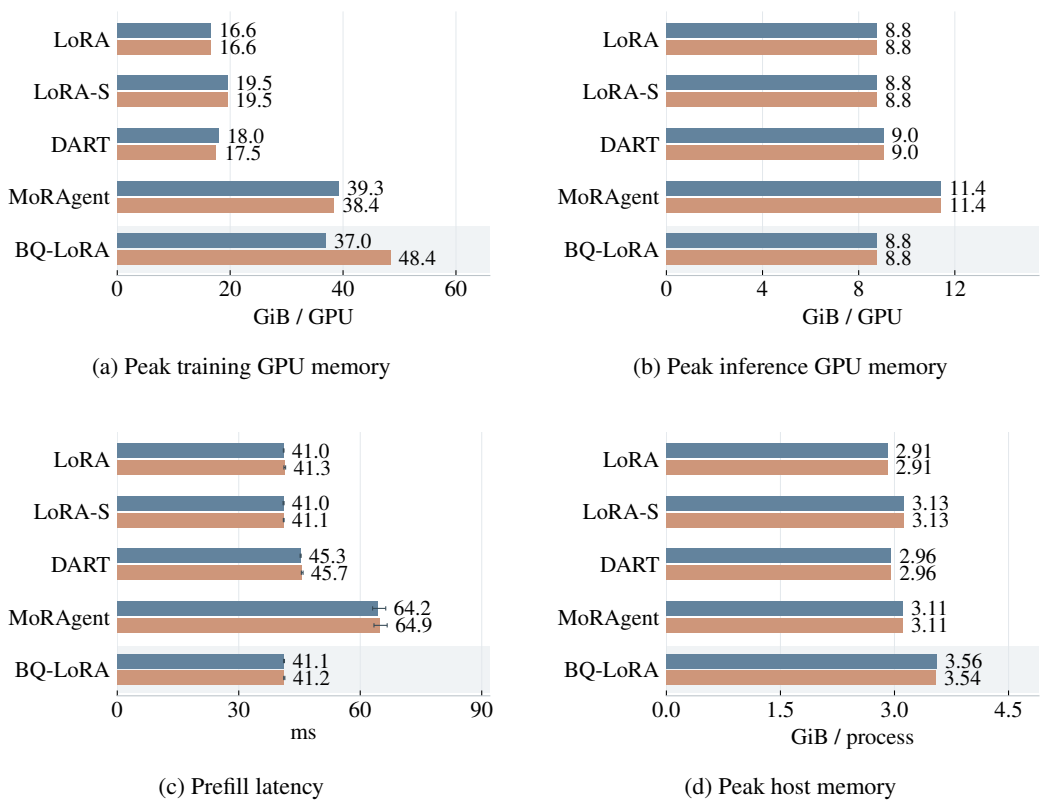
Model	Method	BM25		Embedding	
		Accuracy	Recall	Accuracy	Recall
Qwen3.5-4B	LoRA	20.89 ± 1.68	22.28 ± 1.77	40.22 ± 2.34	45.77 ± 1.96
	BQ-LoRA w/o BQB	27.11 ± 1.68	27.40 ± 1.80	43.56 ± 1.39	48.80 ± 1.20
	BQ-LoRA w/o DPC	27.56 ± 2.04	28.00 ± 2.10	44.22 ± 2.14	48.20 ± 1.80
	BQ-LoRA	$\mathbf{29.11} \pm 1.02$	$\mathbf{29.34} \pm 0.82$	$\mathbf{46.44} \pm 1.68$	$\mathbf{50.16} \pm 0.83$
Qwen3.5-9B	LoRA	26.22 ± 1.68	27.14 ± 1.81	46.22 ± 2.69	48.34 ± 1.64
	BQ-LoRA w/o BQB	30.67 ± 2.40	31.90 ± 2.10	49.78 ± 1.39	52.60 ± 1.30
	BQ-LoRA w/o DPC	31.56 ± 1.68	32.80 ± 1.50	50.22 ± 2.52	52.90 ± 2.30
	BQ-LoRA	$\mathbf{33.11} \pm 1.68$	$\mathbf{34.26} \pm 0.85$	$\mathbf{51.78} \pm 1.02$	$\mathbf{54.38} \pm 0.79$

E.3 Computational Complexity

The efficiency measurements use the behavior map in equation 5, with the same reference construction as the task performance experiments. For each training batch $\mathcal{T}$, the map includes the full hidden representation at every recorded decision token, with the normalization specified in equation 5. The response dimension is $d_\Phi = d_{\text{hid}} \sum_{i=1}^{N} \sum_{k=1}^{T_i} |a_{ik}|$. Let C_{update} denote the cost of computing trajectory gradients and constructing the projected target, including optimizer updates and basis construction. Let C_{ref} include all derivative computations and applications for this behavior map. With a fixed compression solver budget, the cost per update is:

$$C_{\text{BQ}} = O\left(C_{\text{update}} + C_{\text{ref}} + N^2 d_\Phi + \sum_{l=1}^{L} r_l^3\right). \tag{57}$$

To reduce overhead, projected trajectory directions are recovered from LoRA factor gradients and retained in low-rank form for aggregation. The target bases are constructed once per update and reused throughout compression. Response inner products are accumulated over reference blocks to compute the trajectory distances. The batch size controls the number of pairwise comparisons, with a batch of 64 producing a 64×64 affinity matrix. All d_Φ response coordinates contribute to these comparisons. Compression operates on cores of size at most $2r_l \times 2r_l$, so the cubic term depends on the adapter ranks. For rank 32, these cores are at most 64×64, and the fixed solver budget bounds the number of decompositions. During DPC, the reference prefixes and the differential $D\Phi_{\mathcal{T}}(\boldsymbol{X})$ remain fixed. We reuse intermediate quantities from the fixed linearization to apply the differential and its adjoint to core directions across all adapted layers jointly, without materializing the full derivative matrix. The associated derivative computation and reuse costs are included in C_{ref}. Figures 3 and 4 report runtime and memory, including communication. Training remains faster than MoRAgent, while decoding latency stays close to ordinary LoRA. Both modules operate during training, allowing performance gains to be pursued while retaining the inference efficiency of a single LoRA adapter.



(a) Peak training GPU memory

(b) Peak inference GPU memory

(c) Prefill latency

(d) Peak host memory

Figure 4: Resource measurements on AppWorld (blue) and BrowseComp-Plus (orange).

Table 6: Comparison of trajectory weighting methods on Qwen3.5-4B.

	AppWorld				BrowseComp-Plus			
	Test-N		Test-C		BM25		Embedding	
Method	TGC	SGC	TGC	SGC	Accuracy	Recall	Accuracy	Recall
LoRA	52.98	32.14	35.49	17.99	20.89	22.28	40.22	45.77
+ Text Weighting	54.76	33.93	37.17	17.99	24.67	26.11	41.33	45.95
+ Parameter Weighting	54.17	33.93	38.13	19.42	24.00	26.70	42.00	45.41
+ BQB	**56.94**	**38.10**	**41.33**	**23.26**	**27.56**	**28.00**	**44.22**	**48.20**

E.4 Ablation Studies

Component Ablation. Tables 4 and 5 present the complete component ablation results. They supplement the TGC and accuracy scores in Table 3 with SGC on AppWorld and recall on BrowseComp-Plus. Either module alone improves all reported mean scores over LoRA for both backbones, and combining the modules further improves every metric. Retaining DPC gives higher Test-N TGC and SGC, whereas retaining BQB gives higher Test-C scores for both backbones. On BrowseComp-Plus, BQB alone also gives higher accuracy with either retriever, while the relative recall depends on the backbone and retriever. On Test-N, the combined method increases SGC by 2.38 percentage points over the stronger ablation for both backbones. These gains indicate more consistent completion across the task variants within a scenario. The full method also achieves the highest recall under both BrowseComp-Plus retrievers, extending the improvements beyond final answer accuracy to evidence retrieval. The improvement from combining the modules is consistent across these changes in the relative ranking of the individual components.

Table 7: Comparison of compression methods on Qwen3.5-4B.

	AppWorld				BrowseComp-Plus			
	Test-N		Test-C		BM25		Embedding	
Method	TGC	SGC	TGC	SGC	Accuracy	Recall	Accuracy	Recall
LoRA	52.98	32.14	35.49	17.99	20.89	22.28	40.22	45.77
+ SVD	57.74	39.29	39.09	20.86	25.33	25.86	42.67	47.89
+ DPC with Identity Metric	57.74	39.29	39.09	20.86	25.33	25.76	42.67	47.83
+ DPC	**59.72**	**42.26**	**40.77**	**22.54**	**27.11**	**27.40**	**43.56**	**48.80**

Table 8: Rank sensitivity on AppWorld.

Rank	Test-N						Test-C					
	TGC			SGC			TGC			SGC		
	LoRA	LoRA-S	BQ-LoRA	LoRA	LoRA-S	BQ-LoRA	LoRA	LoRA-S	BQ-LoRA	LoRA	LoRA-S	BQ-LoRA
8	47.42	53.57	55.75	26.19	34.52	38.10	30.46	34.53	38.29	13.43	14.87	18.94
16	50.40	56.35	58.53	29.76	38.10	41.67	33.25	37.17	41.41	15.83	17.03	22.06
32	52.98	58.93	60.91	32.14	41.07	44.64	35.49	39.33	44.04	17.99	19.42	24.46
64	55.36	60.32	62.30	35.12	42.86	46.43	37.65	41.01	45.48	19.42	21.58	25.90
128	56.35	61.11	62.70	36.31	44.05	47.02	38.85	42.13	45.72	20.38	22.54	26.38

Behavior Quotient Balancing. Table 6 compares BQB with trajectory weighting based on text semantic distances or parameter update distances. Both alternatives improve TGC and answer accuracy over LoRA, but their relative performance varies across settings. The text criterion performs better on Test-N TGC and BM25 accuracy, while the parameter criterion performs better on Test-C TGC and embedding accuracy. BQB compares local responses of the projected trajectory directions at fixed reference histories, so the weighting reflects their effects on the same decision representations. It achieves the highest reported score on every metric. Relative to parameter weighting, BQB improves Test-C TGC and SGC by 3.20 and 3.84 percentage points, respectively. On BrowseComp-Plus, BM25 accuracy exceeds text weighting, the stronger alternative for this metric, by 2.89 percentage points. The advantage also holds with embedding retrieval, where BQB improves accuracy by 2.22 percentage points over parameter weighting and recall by 2.25 points over text weighting. These results support measuring overlap through local reference responses when balancing trajectory contributions.

Decision Preserving Compression. In Table 7, the SVD baseline applies tangent projection followed by truncated SVD under the same rank budget. Relative to LoRA, this variant already improves Test-N TGC by 4.76 percentage points. DPC with Identity Metric retains the DPC solver and iteration budget but replaces the behavioral metric with the identity metric. This comparison examines the effect of the iterative solver when compression residuals are measured by their Frobenius magnitude. The identity variant matches SVD on all AppWorld metrics and both accuracy scores, with recall differences of at most 0.10 percentage points. Full DPC improves every metric and raises Test-N SGC and BM25 accuracy over SVD by 2.97 and 1.78 percentage points, respectively, based on the reported means. The similarity between SVD and the identity variant suggests that the gains depend on the behavior metric used during compression. DPC uses the reference differential to distinguish residual directions and penalizes those that more strongly distort the target response. This supports accounting for the local response of compression residuals at the reference histories, since equal Frobenius errors can produce different representation responses.

E.5 Rank sensitivity

Table 8 shows that `BQ-LoRA` achieves comparable or better performance with a smaller rank budget. On both backbones, it matches or exceeds LoRA at rank 64 using only rank 8, and surpasses LoRA-S at rank 128 using rank 32. The smaller gains beyond rank 32 further support this default choice.