

# Hessian 기반 초기화를 통한 LoRA 의 초기 학습 향상에 관한 연구

김윤지, 배준우, 안진현\*  
명지대학교

poonji1004@mju.ac.kr, ag660340@mju.ac.kr, \*wlsgus3396@mju.ac.kr

## A Study on Improving Early Training of LoRA via Hessian-Based Initialization

Yoonji Kim, JunWoo Bae, Jin-Hyun Ahn\*  
Myongji Univ.

### 요약

본 논문은 파라미터 효율적 미세조정(PEFT, Parameter-Efficient Fine-Tuning) 기법 중 하나인 Low-Rank Adaptation(LoRA)의 초기화 방식에 관한 연구를 다룬다. 기존 LoRA 는 저랭크 파라미터를  $A = \mathcal{N}(0, \sigma^2)$ ,  $B = 0$ 으로 초기화하여 사전학습된 모델의 출력을 보존하는 장점을 가지지만, 학습 초기 단계에서 저랭크 업데이트가 충분히 활성화되지 않는 한계를 가진다. 이에 본 연구에서는 손실 함수의 곡률 정보를 반영하기 위해 Hessian-vector product(HVP)를 활용한 LoRA 저랭크 파라미터 초기화 방법을 제안한다. 제안한 방법은 저랭크 파라미터가 모델의 중요한 방향으로 정렬되도록 초기화함으로써, 학습 초기 단계에서의 비효율적인 파라미터 탐색을 완화하는 것을 목표로 한다. 간단한 실험을 통해 제안한 초기화 전략이 기존 LoRA 대비 더 빠른 초기 수렴 특성을 보임을 확인하였다. 본 연구는 초기화 전략이 LoRA 기반 PEFT 의 학습 효율에 중요한 영향을 미친다는 점을 실험적으로 분석하였으며, 향후 다양한 대규모 언어 모델 및 PEFT 기법으로의 확장 가능성을 제시한다.

### I. 서 론

대규모 언어 모델(LLMs, Large Language Models)은 자연어 이해 및 생성 성능에서 획기적인 발전을 이루며 다양한 응용 분야에서 활용되고 있다. 특히 Transformer 구조를 기반으로 한 LLM은 방대한 데이터와 매개변수를 통해 뛰어난 표현력을 확보하였으나, 그에 따른 막대한 파라미터 수와 계산 비용은 실제 환경에서의 학습 및 배포에 큰 제약으로 작용한다. 이러한 문제로 인해, 제한된 자원 환경에서도 LLM을 효율적으로 활용하기 위한 파라미터 효율적 미세조정(PEFT, Parameter-Efficient Fine-Tuning) 기법에 대한 연구가 활발히 진행되고 있다. 이러한 흐름 속에서 제안된 Low-Rank Adaptation(LoRA) [1] 기법은 기존 사전학습된 모델의 가중치를 고정한 상태에서, 저랭크 행렬을 추가로 학습함으로써 미세조정을 수행하는 방식이다. LoRA는 전체 모델 파라미터를 업데이트하지 않고도 높은 성능을 달성할 수 있어, 메모리 사용량과 학습 비용을 크게 절감할 수 있다는 장점을 가진다. 이로 인해 LoRA는 다양한 LLM 미세조정 시나리오에서 표준적인 접근 방식으로 널리 채택되고 있다. 그러나 LoRA는 저랭크 행렬을 무작위 초기화한 후 학습을 진행하는 구조적 특성으로 인해, 학습 초기에 모델 출력에 미치는 영향이 제한적이며, 결과적으로 파라미터 업데이트가 비효율적으로 이루어

어지는 문제가 존재한다. 특히, 저랭크 공간이 모델의 중요한 방향을 충분히 반영하지 못할 경우, 학습이 느리게 진행되거나, 동일한 rank 대비 성능 저하가 발생하는 현상이 보고되고 있다. 이러한 현상은 LoRA가 갖는 일종의 lazy learning 특성으로 해석될 수 있다.

본 논문에서는 이러한 LoRA의 한계를 극복하기 위해, Hessian-vector product(HVP) [2] 정보를 활용한 새로운 LoRA 초기화 방법을 제안한다. 제안하는 방법은 모델 파라미터 공간에서 손실 함수의 곡률 정보를 반영하여, 저랭크 행렬이 보다 의미 있는 방향으로 초기화되도록 유도한다. 이를 통해 LoRA가 학습 초기 단계부터 효과적으로 작동하도록 하며, 동일한 rank 조건에서도 더 빠른 수렴과 향상된 성능을 달성할 수 있음을 보인다. 실험 결과를 통해 제안 기법이 기존 LoRA 대비 효율적인 학습 특성을 가지며, LLM 미세조정에서의 실용성을 향상시킬 것을 검증한다.

### II. 관련 연구

#### A. LoRA 기반 미세조정 기법

LoRA는 사전학습된 대규모 언어 모델의 가중치를 고정한 상태에서, 저랭크 행렬을 추가적으로 학습함으로써 효율적인 미세조정을 가능하게 하는 기법이다. LoRA는 전체 파라미터를 업데이트하지 않

고도 경쟁력 있는 성능을 달성할 수 있어, 대규모 모델 환경에서 널리 활용되고 있다. 그러나 저랭크 행렬을 무작위로 초기화하는 기존 LoRA 방식은 학습 초기에 모델 출력에 미치는 영향이 제한적이며, 이로 인해 학습 효율이 저하되는 문제가 지적되어 왔다.

### B. LoRA 개선 및 초기화 전략 연구

이러한 한계를 완화하기 위해 LoRA의 학습 효율을 개선하고자 하는 다양한 연구들이 제안되었다.

[3], [4]

PiSSA [5]는 사전학습된 가중치 행렬에 대해 특이값 분해(SVD, Singular Value Decomposition)를 수행하고, 상위 특이값 및 특이벡터를 LoRA의 저랭크 파라미터 초기화에 직접 활용하는 방식을 제안하였다. 이 방법은 저랭크 공간이 사전학습 가중치의 지배적인 선형 구조를 반영하도록 유도함으로써, 무작위 초기화 대비 초기 학습 안정성과 수렴 특성을 개선하는 데 초점을 둔다.

LoRA-One [6]은 LoRA 파라미터의 초기화 구조로 인해 학습 초기 단계에서 업데이트가 충분히 활성화되지 않는 문제를 지적하고, 이를 완화하기 위해 학습 과정에서 관측되는 그래디언트 정보를 기반으로 저랭크 파라미터를 재초기화하는 전략을 제안하였다. 해당 접근은 학습 도중 특정 시점에서 LoRA 파라미터를 다시 정렬함으로써, 제한된 rank에서도 학습 효율을 향상시키는 것을 목표로 한다.

## III. 제안 방법

### A. 연구 개요

본 연구에서는 LoRA의 초기화 방식으로 인해 발생하는 학습 초기 비효율성 문제를 완화하기 위해, 손실 함수의 곡률 정보를 활용한 LoRA 저랭크 파라미터 초기화 방법을 제안한다. 일반적으로  $k$  번째 가중치 행렬  $W_k$ 에 대해 LoRA는 다음과 같은 저랭크 업데이트를 적용한다.

$$A_k \in R^{r \times d_2}, B_k \in R^{d_1 \times r}$$

기존 LoRA에서는  $A_k = \mathcal{N}(0, \sigma^2)$ 으로 초기화하고,  $B_k = 0$ 으로 초기화함으로써 학습 초기에  $\Delta W_k = 0$ 이 되도록 설계한다. 이러한 초기화 방식은 사전학습된 모델의 출력을 보존하는 장점이 있으나, 학습 초기 단계에서 저랭크 업데이트가 손실 함수에 미치는 영향이 제한적이라는 한계를 가진다.

### B. HVP 기반 곡률 정보

손실 함수  $L$ 에 대한 Hessian 행렬  $H \triangleq \nabla^2 L$ 은 파라미터 공간에서의 국소 곡률을 나타내며, 특정 방향으로의 파라미터 변화가 손실 값에 미치는 민감도를 정량적으로 표현한다. 그러나 대규모 언어 모델 환경에서 Hessian 행렬을 명시적으로 계산하는 것은 계산 비용 측면에서 현실적으로 어렵다. 본 연구에서는 이러한 문제를 해결하기 위해 HVP를 활용하여 곡률 정보를 효율적으로 반영한다. 임의의 벡터  $v$ 에 대한 HVP는 다음과 같이 정의된다.

$$Hv(v) = Hv = \nabla^2 L \cdot v$$

HVP는 Hessian 행렬을 직접 구성하지 않고도 특정 방향에 대한 2차 곡률 정보를 계산할 수 있다는 장점을 가진다.

### C. 곡률 정보 기반 LoRA 초기화

본 연구의 핵심 아이디어는 LoRA의 저랭크 파라미터가 곡률 정보가 큰 방향으로 정렬되도록 초기화함으로써, 학습 초기 단계부터 저랭크 업데이트가 손실 함수에 실질적인 영향을 미치도록 하는 것이다. 이를 위해, HVP를 통해 추출된 곡률 정보를 기반으로,  $A_k, B_k$ 의 초기 방향을 설정한다. 이러한 초기화 전략은 LoRA의 학습 과정이나 파라미터 수를 변경하지 않으며, 초기화 단계에서만 추가 정보를 활용한다는 점에서 기존 LoRA 방식과 차별성을 가진다.

## IV. 실험

Table 1. Early training convergence comparison

Method	$\mathcal{L} < 6.0$ (Epoch)	Final Train $\mathcal{L}$	Peak Grad Norm	Grad < 2 (Epoch)
LoRA-One	0.32	5.55	25.3	0.35
Ours	<b>0.14</b>	<b>4.87</b>	<b>14.2</b>	<b>0.18</b>

Table 1은 학습 과정에서의 손실 감소 속도와 gradient 안정성을 정량적으로 비교한 결과를 나타낸다. 제안한 방법은 초기 학습 단계에서 더 빠른 손실 감소와 낮은 gradient 변동성을 보이며, 곡률 정보를 반영한 초기화의 효과를 확인할 수 있다.

## V. 결론

본 논문에서는 LoRA의 초기화 구조로 인해 발생하는 학습 초기 비효율성 문제를 분석하고, HVP를 활용한 곡률 정보 기반 LoRA 초기화 방법을 제안하였다. 제안한 방법은 저랭크 파라미터가 학습 초기 단계부터 의미 있는 방향으로 활성화되도록 유도한다. 간단한 실험을 통해 제안한 초기화 전략이 기존 LoRA 대비 더 빠른 초기 수렴 특성을 보임을 확인하였다. 향후 연구에서는 곡률 정보의 효율적인 근사 방법과 다양한 모델로의 확장을 진행할 예정이다.

### ACKNOWLEDGMENT

본 과제(결과물)는 2025년도 교육부 및 경기도의 재원으로 경기 RISE 센터의 지원을 받아 수행된 지역혁신중심 대학지원체계(RISE)의 결과입니다. (2025-RISE-09-A15)

### 참 고 문 헌

- [1] Hu, Edward J., et al., "LoRA: Low-Rank Adaptation of Large Language Models," ICLR, 2022.
- [2] Pearlmutter, B. A., "Fast Exact Multiplication by the Hessian," Neural Computation, 1994.
- [3] Liu, Shih-Yang, et al., "DoRA: Weight-Decomposed Low-Rank Adaptation of Large Language Models," ICML, 2024.
- [4] Zhang, Qingyue, et al., "LoRA-DA: Data-Aware Initialization for Low-Rank Adaptation via Asymptotic Analysis," arXiv, 2025.
- [5] Meng, Fanxu, et al., "PiSSA: Principal Singular Values and Singular Vectors Adaptation of Large Language Models," NeurIPS, 2024.
- [6] Zhang, Yuanhe, et al., "LoRA-One: One-Step Full Gradient Could Suffice for Fine-Tuning LargeLanguage Models, Provably and Efficiently," ICML, 2025.