A Appendix: Experiments (Continue)

A.1 Further Details for the Experiment Settings

For the data partitioning, we have the Mini-ImageNet and CIFAR-100 data sets divided into the partitions $64 : 16 : 20$, which correspond to the training set, validation set and the testing set respectively. Each class is corresponding to a task. Then, for the Drug data set, we partition the tasks into $4100 : 76 : 100$ representing the training set, validation set and the testing set.

For our BASS, we apply two 2-layer FC networks for $f_1(\cdot; \boldsymbol{\theta}_1), f_2(\cdot; \boldsymbol{\theta}_2)$ respectively, and set network width $m = 200$. For deriving approximated arm rewards, we let $|\Omega_k^{\text{valid}}| = 5$. Recall that we apply the approximation approach mentioned in **Remark** 3 to reduce the space complexity and time complexity in practice for the experiments. Here, we tune the pooling step such that the inputs of $f_1(\cdot; \boldsymbol{\theta}_1), f_1(\cdot; \boldsymbol{\theta}_1)$ are approximately 50 and 20 respectively. For the learning rate, we find the learning rate for BASS with grid search from $\{0.01, 0.001, 0.0001\}$, and choose the learning rates for the meta-model $\eta_1 = 0.01, \eta_2 = 0.001$. The meta-model architecture as well as its learning rates will stay the same for all the baselines and our proposed BASS. For the CIFAR-100 and Mini-ImageNet data sets, we use the the meta-model with four convolutional blocks where the network width of each block is 32, followed by an FC layer as the output layer. For the Drug data set, we apply a meta-model with two FC layers, where the network width is 500. All the experiments are performed on a Linux machine with Intel Xeon CPU, 128GB RAM, and Tesla V100 GPU. Code will be made available at `https://github.com/yunzhe0306/Bandit_Task_Scheduler`.

A.2 Effect of the Task Noise Magnitude

We conduct the experiments to show the effects of the noise magnitude factor ϵ on the Drug and CIFAR-100 data sets. The experiment results are shown in **Table** 4.

Table 4: Comparison with baselines with different noise magnitude [data set (noise magnitude ϵ) ; final results ± standard deviation].

Algo. \ Data	Drug (0.3)	Drug (0.5)	CIFAR100 (0.3)	CIFAR100 (0.5)
Uniform	0.218±0.007	0.220±0.001	0.655±0.009	0.526±0.011
SPL	0.243±0.008	0.236±0.004	0.625±0.017	0.367±0.039
FOCAL	0.224±0.019	0.223±0.003	0.638±0.010	0.485±0.006
DAML	0.182±0.025	0.177±0.003	0.543±0.017	0.414±0.025
GCP	N/A	N/A	0.653±0.005	0.508±0.009
PAML	0.186±0.006	0.205±0.009	0.537±0.009	0.316±0.022
ATS	0.239±0.011	0.237±0.014	0.651±0.001	0.505±0.015
BASS (Ours)	**0.258±0.003**	**0.245±0.006**	**0.657±0.005**	**0.553±0.008**

With increasing noise magnitude ϵ, the performances of the meta-model trained by baselines and our BASS tend to drop, which is intuitive. In particular, for the CIFAR-100 data set, when we increase ϵ, the performance difference between BASS and the other baselines tends to increase. This can be the reason that the greedy baselines with no exploration strategies can be more susceptible to the task noise perturbation, which can lead to the sub-optimal performances of the meta-model.

Table 5: Experiment results of noise-free settings on three real data sets (5-way, 5-shot).

Data \ Algo.	Uniform	SPL	FOCAL-LOSS	DAML	GCP	PAML	ATS	BASS
Drug	0.206±0.012	0.234±0.006	0.240±0.003	0.190±0.002	N/A	0.220±0.010	0.233±0.001	**0.256±0.003**
M-ImageNet	0.576±0.016	0.554±0.004	0.582±0.005	0.437±0.015	0.564±0.002	0.467±0.007	0.561±0.004	**0.586±0.008**
CIFAR	0.681±0.010	0.681±0.008	0.692±0.023	0.662±0.027	0.681±0.016	0.640±0.011	0.695±0.035	**0.697±0.029**

From the **Table** 5, we can see that when there is no noise, the overall performance does not differ significantly across different methods. The reason could be that since the meta-learning backbone remains the same for all the methods, the meta-model performance upper bound can be similar for different scheduling algorithms, without the presence of other confounding factors (e.g., noise, task distribution skewness). In the practical application scenarios with noisy data, BASS-guided meta-models tend to perform well in presence of task noise and skewness compared with baselines, as presented by our experiments in the main body.

A.3 Parameter Study for Exploration Coefficient

As in **Eq.** 7 and **Eq.** 9, BASS involves an exploration coefficient α to balance the exploitation-exploration and the two exploration objectives. Here, we conduct the parameter study for the exploration coefficient α, and include the results with no exploration (i.e., removing f_2).

Table 6: Comparison among different α values [dataset (shot) ; final results ± standard deviation].

Algo. \ Data	Drug (1)	Drug (5)	CIFAR100 (1)	CIFAR100 (5)
No Exploration	0.234±0.003	0.239±0.012	0.256±0.027	0.537±0.012
$\alpha = 0.1$	0.231±0.005	0.233±0.013	0.264±0.051	0.522±0.024
$\alpha = 0.3$	0.228±0.013	0.231±0.008	0.268±0.047	0.528±0.014
$\alpha = 0.5$	0.236±0.004	**0.245±0.006**	**0.272±0.025**	**0.553±0.008**
$\alpha = 0.7$	**0.242±0.012**	0.227±0.006	0.241±0.005	0.543±0.021
$\alpha = 1.0$	0.236±0.002	0.235±0.013	0.266±0.006	0.537±0.005

From the results in **Table** 6, we see that the exploration module can indeed improve the performance of BASS compared with the performance with no exploration. This also fits our initial argument that the greedy algorithm alone can lead to sub-optimal performances of meta-model. By properly choosing the α value, we will be able to achieve a good balance between exploitation and exploration, as well as between the two exploration objectives. Here, setting $\alpha \in [0.5, 0.7]$ will be good enough to achieve satisfactory performances. Meanwhile, we also note that even with no exploration, our BASS still achieves good performances by directly learning the correlation between the adapted meta-parameter and the generalization score, and refining the scheduling strategy based on the status of the meta-model.

A.4 Running Time Comparison

In Figure 5, we include the running time comparison with baselines. We can see that BASS can achieve significant improvement in terms of the running time, and can take as little as 50% of ATS's running time. The intuition is that our proposed BASS only needs one round of the optimization process to update the meta-model and BASS. On the other hand, from Algorithm 1 of the ATS paper [50], we see that ATS requires two optimization rounds for each meta-training iteration to (1) update the scheduler with the temporal meta-model, and (2) update the actual meta-model respectively. Based on the figure on the RHS, we also see that BASS can achieve a relatively good balance between computational cost and performance.

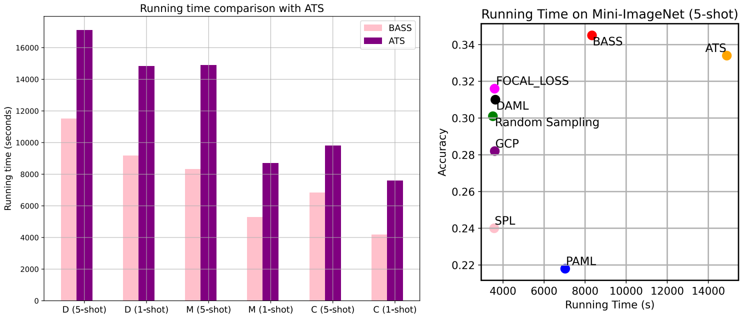

Figure 5: Running time results (including training both the scheduler and the meta-model). "D", "M" and "C" refer to the "Drug", "Mini-ImageNet", "CIFAR-100" data sets respectively. BASS can take as little as approximately 50% of ATS's running time. On the RHS, we have the scatter plot in terms of running time vs. performance on the Mini-ImageNet dataset.

A.5 Performances with Different Task Skewness Settings

In Table 7, we include the experiments with different levels of skewness. Here, we see that with less skewness levels (the skewness level reduces from Setting 1 to Setting 3), the accuracy of BASS as well as the baselines will continue to improve, while BASS still maintains decent performances.

Skewness Setting \ Algo.	Uniform	ATS	**BASS**
Skewness Setting 1	0.375±0.009	0.382±0.007	0.408±0.008
Skewness Setting 2	0.429±0.012	0.448±0.006	0.460±0.013
Skewness Setting 3	0.497±0.008	0.502±0.010	0.539±0.009

Table 7: Results for different skewness levels on CIFAR-100 data set (5-shot). (1) Setting 1 is the original setting in paper Subsec. 5.2. (2) For Setting 2, we assign 5 tasks with 8% sampling probability, 5 tasks with 3%, and the rest of the tasks equally share the 45% probability. (3) For Setting 3, we assign 5 tasks with 5%, 5 tasks with 2%, while the rest of the tasks equally share the 65% probability.

A.6 Performances with Different Batch Size

With Table 8, we include additional experiments with different batch sizes B, in comparison with the ATS and the uniform sampling approach. Here, we see that with larger B values, the accuracy of BASS as well as the baselines will generally improve.

B (batch size) \ Algo.	Uniform	ATS	**BASS**
1	0.459±0.009	0.449±0.010	0.472±0.012
2	0.526±0.011	0.515±0.015	0.553±0.008
3	0.570±0.012	0.563±0.007	0.588±0.010
5	0.581±0.005	0.571±0.007	0.586±0.009

Table 8: Results for different B values (batch sizes) on CIFAR-100 data set (5-shot).

A.7 Performances with Different Embedding Approaches of Arm Contexts

In Table 9, we include additional experiments with different levels of average pooling, such that after the average pooling, the dimensionality of the pooled vector representation will fall into $\{20, 100, 500\}$. We see that overly small dimensionality of the average-pooled vector representation (e.g., 20) can lead to sub-optimal performance of the BASS framework. In addition, we see that setting the dimensionality to 50 can generally lead to good enough performance.

Dimensionality	20	50	100	500
Accuracy	0.541±0.008	0.553±0.008	0.558±0.006	0.555±0.010

Table 9: With CIFAR-100 (5-shot), different dimensionality of the average-pooled vector representation (Remark 3) of the meta-parameters.

Here, we also include additional experimental results using MLP to map the original context into the lower dimensional space instead of using our proposed average pooling (Remark 3). Results are shown in Table 10. Here, we use the one-layer MLP with the ReLU activation to embed the original meta-parameters to the low-dimensional vector representations. We can see that the MLP-based method can indeed lead to some performance improvement. But in general, the performance difference between MLP-based embedding and the average-pooling vector representation is subtle. We also note that the MLP-based mapping approach is more time consuming compared with the average pooling approach, since we also need to train the additional embedding layer, which has a considerable number of trainable parameters.

Dimensionality	Original avg-pooled (50)	50	100	200
Accuracy	0.553±0.008	0.558±0.013	0.560±0.012	0.553±0.015

Table 10: With CIFAR-100 (5-shot), different dimensionality of the one-layer MLP(with ReLU)-embedded vector representation of the meta-parameters. "original avg-pooled (50)" refers to the average-pooled vector representation (Remark 3) with dimensionality of 50.

A.8 Additional Experiments on the "DomainNet" data set

In Table 11, we include additional experiments on the new "DomainNet" data set [36]. Within the "real" domain, we filter 100 classes that have at least 600 images. In this way, with each class being a task with 600 images, we will have a total of 100 tasks. Compared with image data sets in our paper (Mini-ImageNet and CIFAR-100), we increase the image resolution of "DomainNet" by resizing its images to 128×128 pixels. Following the settings in our paper, we divide tasks into 64 : 16 : 20 portions that correspond to the training set, validation set and the test set respectively. For the few-shot settings, we formulate the problem to be 5-shot, 5-way / 7-way with uniform sampling and ATS as baselines. With a higher image resolution of the "DomainNet" data set, BASS can still maintain the good performance compared with the baselines.

Setting \ Algo.	Uniform	ATS	**BASS**
5-way	0.475±0.002	0.483±0.006	0.511±0.012
7-way	0.411±0.005	0.372±0.009	0.435±0.008

Table 11: Results for the "DomainNet" data set (noise level 0.5, 5-shot settings).

B Appendix: Additional Discussion on the Necessity of Assumption 5.1

We would like to mention that in order to finish the convergence and generalization analysis for the neural Contextual Bandit works (e.g., [53, 2, 9]), the separateness assumption of the arm context is the minimum requirement of the data set. This is because the training data needs to be non-degenerate (i.e., every pairs of samples are distinct) to ensure that the neural network can consistently converge, as indicated by Assumption 2.1 in [3]. Therefore, our Assumption 5.1 regarding the arm separateness aims to ensure that the BASS is able to adequately learn the underlying reward mapping function with sufficient information. Comparing with the existing works, in the convergence analysis works on meta-learning [46, 47], they measure the arm separateness in terms of the minimum eigenvalue λ_0 (with $\lambda_0 > 0$) of the Neural Tangent Kernel (NTK) [22] matrix, which is comparable with our Euclidean separateness ρ. For existing neural bandit works, Assumption 5.1 in [9] is similar to our separateness assumption. Meanwhile, Assumption 4.2 in [53] and Assumption 3.4 from [52] also imply that no two arms are the same in terms of the minimum NTK matrix eigenvalue $\lambda_0 > 0$.

C Appendix: Limitation

One potential limitation of BASS is that its improvement over baselines may not be significant when dealing with noise-free settings and non-skewed task distributions (**Table** 5). Meanwhile, the non-adaptive FOCAL-LOSS [29] tends to achieve a similar performance comparing with the adaptive method ATS [50], while enjoying an advantage in terms of the computational cost. In practical terms, although BASS can generally achieve the decent performance and enjoys a smaller computational cost than ATS, the practitioner still needs to consider whether their task distribution is noisy or skewed in order to strike a good balance between the computational resource needed and the meta-model performance, as BASS can achieve a more significant advantage over baselines given the noisy or skewed task distribution.

D Appendix: Theoretical Analysis

In this section, we present the proof for **Theorem** 5.2. Here, instead of directly going for the batch setting where we adopt training task batch Ω_k for each iteration $k \in [K]$ ($|\Omega_k| = |\Omega_k^*| = B$), we first introduce the results of the single-task setting (Subsec. D.1), i.e., $|\Omega_k| = |\Omega_k^*| = 1$. Then, the results will be extended to the batch settings as in Subsec. D.2. Recall that for the meta-model, we first consider it to be a $L_{\mathcal{F}}$-layer fully-connected (FC) network (of width $m_{\mathcal{F}}$ for the theoretical analysis (lines 237-239). In particular, we follow the settings in [3] for the Gaussian initialization of weight matrices. For the weight matrix elements in meta-model's first $(L_{\mathcal{F}} - 1)$ layers, we draw each of them from the Gaussian distribution $\mathcal{N}(0, 2/m_{\mathcal{F}})$. Then, for the weight matrix elements of the last layer ($L_{\mathcal{F}}$-th layer), we draw each of them from the Gaussian distribution $\mathcal{N}(0, 1)$.

D.1 Single-task settings

For the brevity of notation, we denote the scheduler output $f(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_{k,i}];\boldsymbol{\theta}^{(k-1)}) = f_1(\boldsymbol{\chi}^q_{k,i};\boldsymbol{\theta}_1^{(k-1)}) + f_2\big([\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}^s_{k,i});\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}^q_{k,i})];\boldsymbol{\theta}_2^{(k-1)}\big)$, which corresponds to the definition in **Eq. 7**. In this case, $\mathcal{T}(K)=\{\mathcal{T}_1,\dots,\mathcal{T}_K\}$ refer to the chosen tasks and $\mathcal{T}^*(K)=\{\mathcal{T}_1^*,\dots,\mathcal{T}_K^*\}$ are the optimal ones. Based on the problem definition, we will have

$$\begin{aligned}
R_{\text{single}}(K) &= \mathbb{E}_{\mathcal{T}\sim\mathcal{P}(\mathcal{T}),\boldsymbol{x}\sim\mathcal{D}_{\mathcal{T}}}\Big[\mathcal{L}\big(\boldsymbol{x};\mathcal{I}(\mathcal{T},\boldsymbol{\Theta}^{(K)})\big)\Big] - \mathbb{E}_{\mathcal{T}\sim\mathcal{P}(\mathcal{T}),\boldsymbol{x}\sim\mathcal{D}_{\mathcal{T}}}\Big[\mathcal{L}\big(\boldsymbol{x};\mathcal{I}(\mathcal{T},\boldsymbol{\Theta}^{(K),*})\big)\Big] \\
&= \mathbb{E}_{\mathcal{T}\sim\mathcal{P}(\mathcal{T}),\boldsymbol{x}\sim\mathcal{D}_{\mathcal{T}}}\Big[\mathcal{L}\big(\boldsymbol{x};\mathcal{I}(\mathcal{T},\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K])\big)\Big] - \mathbb{E}_{\mathcal{T}\sim\mathcal{P}(\mathcal{T}),\boldsymbol{x}\sim\mathcal{D}_{\mathcal{T}}}\Big[\mathcal{L}\big(\boldsymbol{x};\mathcal{I}(\mathcal{T},\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*])\big)\Big] \\
&= h(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*]) - h(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K]) \\
&= h(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*]) - f(\boldsymbol{\chi}_K^*;\tilde{\boldsymbol{\theta}}^{(K-1)}) + f(\boldsymbol{\chi}_K^*;\tilde{\boldsymbol{\theta}}^{(K-1)}) - f(\boldsymbol{\chi}_K;\boldsymbol{\theta}^{(K-1)}) \\
&\qquad + f(\boldsymbol{\chi}_K;\boldsymbol{\theta}^{(K-1)}) - h(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K]) \\
&\leq h(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*]) - f(\boldsymbol{\chi}_K^*;\tilde{\boldsymbol{\theta}}^{(K-1)}) + f(\boldsymbol{\chi}_K^*;\tilde{\boldsymbol{\theta}}^{(K-1)}) - f(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K^*];\boldsymbol{\theta}^{(K-1)}) \\
&\qquad + f(\boldsymbol{\chi}_K;\boldsymbol{\theta}^{(K-1)}) - h(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K]) \\
&= h(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*]) - f(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*];\tilde{\boldsymbol{\theta}}^{(K-1)}) + f(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*];\tilde{\boldsymbol{\theta}}^{(K-1)}) - f(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K^*];\boldsymbol{\theta}^{(K-1)}) \\
&\qquad + f(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K];\boldsymbol{\theta}^{(K-1)}) - h(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K]) \\
&\leq |h(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*]) - f(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*];\tilde{\boldsymbol{\theta}}^{(K-1)})| + \underbrace{|f(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*];\tilde{\boldsymbol{\theta}}^{(K-1)}) - f(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K^*];\boldsymbol{\theta}^{(K-1)})|}_{I_0} \\
&\qquad + |f(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K];\boldsymbol{\theta}^{(K-1)}) - h(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K])|
\end{aligned}$$

where the first inequality is due to the arm pulling mechanism, i.e., $f(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K^*];\boldsymbol{\theta}^{(K-1)}) \leq f(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K];\boldsymbol{\theta}^{(K-1)})$. Here, $f(\cdot;\tilde{\boldsymbol{\theta}}^{(K-1)})$ is defined as the "shadow" bandit model that are trained on optimal tasks $\{\mathcal{T}_1^*,\mathcal{T}_2^*,\dots,\mathcal{T}_{K-1}^*\}$ and the corresponding meta-model parameters. Here, denote $\boldsymbol{\chi}_K = \boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K] \in \mathbb{R}^p$ as the arm context given the arm $\mathcal{T}_K$ and the meta-model parameter $\boldsymbol{\Theta}^{(K-1)}$; similarly, we have $\boldsymbol{\chi}_K^* = \boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*] \in \mathbb{R}^p$ being the arm context given the arm $\mathcal{T}_K^*$ and the meta-model parameter $\boldsymbol{\Theta}^{(K-1),*}$. Thus, for the term I_0 on the RHS, we have

$$\begin{aligned}
I_0 &= |f(\chi^*;\tilde{\boldsymbol{\theta}}^{(K-1)}) - f(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K^*];\boldsymbol{\theta}^{(K-1)})| \\
&= |f(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*];\tilde{\boldsymbol{\theta}}^{(K-1)}) - f(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*];\boldsymbol{\theta}^{(K-1)}) \\
&\qquad\qquad + f(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*];\boldsymbol{\theta}^{(K-1)}) - f(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K^*];\boldsymbol{\theta}^{(K-1)})| \\
&\leq |f(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*];\tilde{\boldsymbol{\theta}}^{(K-1)}) - f(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*];\boldsymbol{\theta}^{(K-1)})| \\
&\qquad\qquad + |f(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*];\boldsymbol{\theta}^{(K-1)}) - f(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K^*];\boldsymbol{\theta}^{(K-1)})|.
\end{aligned}$$

Then, inserting the inequality will lead to

$$\begin{aligned}
R(K) \leq{} & \underbrace{|h(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K]) - f(\chi_K;\boldsymbol{\theta}^{(K-1)})|}_{I_1} + \underbrace{|f(\chi_K^*;\tilde{\boldsymbol{\theta}}^{(K-1)}) - h(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*])|}_{I_2} \\
& + \underbrace{|f(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*];\tilde{\boldsymbol{\theta}}^{(K-1)}) - f(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*];\boldsymbol{\theta}^{(K-1)})|}_{I_3} \\
& + \underbrace{|f(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*];\boldsymbol{\theta}^{(K-1)}) - f(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K^*];\boldsymbol{\theta}^{(K-1)})|}_{I_4}.
\end{aligned}$$

Here, the terms I_1, I_2 refer to the approximation error for the two bandit models (our possessed model $f(\cdot;\boldsymbol{\theta}^{(K-1)})$ and the pseudo model $f(\cdot;\tilde{\boldsymbol{\theta}}^{(K-1)})$). Then, the third term I_3 bounds the output difference when given the same input $\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K]$ to two separate bandit models, and the final term I_4 refers to the difference of the meta-model parameters when adapted to the same task with two individual sets of parameters. Here, the terms I_1, I_2 can be bounded by **Lemma** D.1, **Corollary** D.2. Then, the point is to bound the difference term I_4 when given different inputs to the bandit model.

D.1.1 Bounding error terms and assembling the regret bound

[Bounding term I_3] For error term I_3, it focuses on bounding the output difference between two bandit models $f(\cdot; \tilde{\boldsymbol{\theta}}^{(K-1)}), f(\cdot; \boldsymbol{\theta}^{(K-1)})$ given the same input $\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*]$, and we have

$$\begin{aligned}
I_3 &= |f(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*]; \tilde{\boldsymbol{\theta}}^{(K-1)}) - f(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*]; \boldsymbol{\theta}^{(K-1)})| = |f(\boldsymbol{\chi}_K^*; \tilde{\boldsymbol{\theta}}^{(K-1)}) - f(\boldsymbol{\chi}_K^*; \boldsymbol{\theta}^{(K-1)})| \\
&\leq \underbrace{|f_1(\boldsymbol{\chi}_K^*; \tilde{\boldsymbol{\theta}}_1^{(K-1)}) - f_1(\boldsymbol{\chi}_K^*; \boldsymbol{\theta}_1^{(K-1)})|}_{I_{3.1}} \\
&\quad + \underbrace{|f_2\Big([\nabla_{\tilde{\boldsymbol{\theta}}} f_1(\boldsymbol{\chi}_K^{s,*});\ \nabla_{\tilde{\boldsymbol{\theta}}} f_1(\boldsymbol{\chi}_K^{q,*})]; \tilde{\boldsymbol{\theta}}_2^{(K-1)}\Big) - f_2\Big([\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^{s,*});\ \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^{q,*})]; \boldsymbol{\theta}_2^{(K-1)}\Big)|}_{I_{3.2}}.
\end{aligned}$$

With the defined ξ_L, applying **Lemma** D.11 as well as **Corollary** D.12, we will have

$$I_{3.1} \leq \Big(1 + \mathcal{O}(\frac{KL^3 \log^{5/6}(m)}{\rho^{1/3} m^{1/6}})\Big) \cdot \mathcal{O}(\frac{K^3 L}{\rho\sqrt{m}} \log(m)) + \mathcal{O}\Big(\frac{K^4 L^2 \log^{11/6}(m)}{\rho^{4/3} m^{1/6}}\Big)$$

Then, for term $I_{3.2}$, we have

$$\begin{aligned}
I_{3.2} &= |f_2\Big([\nabla_{\tilde{\boldsymbol{\theta}}} f_1(\boldsymbol{\chi}_K^{s,*});\ \nabla_{\tilde{\boldsymbol{\theta}}} f_1(\boldsymbol{\chi}_K^{q,*})]; \tilde{\boldsymbol{\theta}}_2^{(K-1)}\Big) - f_2\Big([\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^{s,*});\ \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^{q,*})]; \boldsymbol{\theta}_2^{(K-1)}\Big)| \\
&\leq |f_2\Big([\nabla_{\tilde{\boldsymbol{\theta}}} f_1(\boldsymbol{\chi}_K^{s,*});\ \nabla_{\tilde{\boldsymbol{\theta}}} f_1(\boldsymbol{\chi}_K^{q,*})]; \tilde{\boldsymbol{\theta}}_2^{(K-1)}\Big) - f_2\Big([\nabla_{\tilde{\boldsymbol{\theta}}} f_1(\boldsymbol{\chi}_K^{s,*});\ \nabla_{\tilde{\boldsymbol{\theta}}} f_1(\boldsymbol{\chi}_K^{q,*})]; \boldsymbol{\theta}_2^{(K-1)}\Big)| \\
&\quad + |f_2\Big([\nabla_{\tilde{\boldsymbol{\theta}}} f_1(\boldsymbol{\chi}_K^{s,*});\ \nabla_{\tilde{\boldsymbol{\theta}}} f_1(\boldsymbol{\chi}_K^{q,*})]; \boldsymbol{\theta}_2^{(K-1)}\Big) - f_2\Big([\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^{s,*});\ \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^{q,*})]; \boldsymbol{\theta}_2^{(K-1)}\Big)|.
\end{aligned}$$

Here, for the first term on the RHS, we apply **Lemma** D.11 as well as **Corollary** D.12 to bound.

Then, for the second term, with Gaussian initialization of weight matrices, for the over-parameterized FC network f with Lipschitz-smooth activation functions (e.g., Sigmoid), we can have $|f(\boldsymbol{x}) - f(\boldsymbol{x}')|, \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{x}')\| \leq \xi \cdot \|\boldsymbol{x} - \boldsymbol{x}'\|$ due to its Lipschitz continuity / smoothness property [46, 17]. Meanwhile, we also have the Lipschitz continuity property for over-parameterized FC network f' with ReLU activation [3], such that $|f'(\boldsymbol{x}) - f'(\boldsymbol{x}')| \leq \xi' \cdot \|\boldsymbol{x} - \boldsymbol{x}'\|$. By the Gaussian initialization of BASS's weight matrices and the properties of over-parameterized neural networks [3, 46, 17], we have $\xi_L = \max\{\xi, \xi'\} \leq \mathcal{O}(c_\xi^L)$ being the Lipschitz constant for our f_1, f_2, where $c_\xi > 1$ is a small constant. Applying the conclusion above, we will have

$$\begin{aligned}
&|f_2\Big([\nabla_{\tilde{\boldsymbol{\theta}}} f_1(\boldsymbol{\chi}_K^{s,*});\ \nabla_{\tilde{\boldsymbol{\theta}}} f_1(\boldsymbol{\chi}_K^{q,*})]; \boldsymbol{\theta}_2^{(K-1)}\Big) - f_2\Big([\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^{s,*});\ \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^{q,*})]; \boldsymbol{\theta}_2^{(K-1)}\Big)| \\
&\quad \leq \xi_L \cdot \big\|[\nabla_{\tilde{\boldsymbol{\theta}}} f_1(\boldsymbol{\chi}_K^{s,*});\ \nabla_{\tilde{\boldsymbol{\theta}}} f_1(\boldsymbol{\chi}_K^{q,*})] - \xi_L \cdot [\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^{s,*});\ \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^{q,*})]\big\| \\
&\quad \leq \xi_L \cdot \big\|\nabla_{\tilde{\boldsymbol{\theta}}} f_1(\boldsymbol{\chi}_K^{s,*}) - \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^{s,*})\big\| + \xi_L \cdot \big\|\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^{q,*}) - \nabla_{\tilde{\boldsymbol{\theta}}} f_1(\boldsymbol{\chi}_K^{q,*})\big\| \\
&\quad \leq \xi_L \cdot \frac{KL^4 \log^{5/6}(m)}{\rho^{1/3} m^{1/6}}
\end{aligned}$$

where the last inequality is by Theorem 5 in [3] and the proof of **Lemma** D.11. With the above results, it will give us

$$I_3 \leq \Big(1 + \mathcal{O}(\frac{KL^3 \log^{5/6}(m)}{\rho^{1/3} m^{1/6}})\Big)\mathcal{O}(\frac{K^3 L}{\rho\sqrt{m}} \log(m)) + \mathcal{O}\Big(\frac{K^4 L^2 \log^{11/6}(m)}{\rho^{4/3} m^{1/6}}\Big) + \frac{\xi_L K L^4 \log^{5/6}(m)}{\rho^{1/3} m^{1/6}}$$

[Bounding term I_4] On the other hand, applying the analogous procedure for term I_4, denoting $\boldsymbol{\chi}_K^* = \boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*]$ and $\bar{\boldsymbol{\chi}}_K^* = \boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K^*]$ for the brevity of notation, we can have

$$\begin{aligned}
I_4 &= |f(\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*]; \boldsymbol{\theta}^{(K-1)}) - f(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K^*]; \boldsymbol{\theta}^{(K-1)})| \\
&\leq \underbrace{\xi_L \cdot \|\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*] - \boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K^*]\|_2}_{I_{4.1}} \\
&\qquad + \underbrace{\left|f_2\bigg([\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^{s,*});\ \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^{q,*})]; \boldsymbol{\theta}_2^{(K-1)}\bigg) - f_2\bigg([\nabla_{\boldsymbol{\theta}} f_1(\bar{\boldsymbol{\chi}}_K^{s,*});\ \nabla_{\boldsymbol{\theta}} f_1(\bar{\boldsymbol{\chi}}_K^{q,*})]; \boldsymbol{\theta}_2^{(K-1)}\bigg)\right|}_{I_{4.2}}.
\end{aligned}$$

where $\boldsymbol{\chi}_K^{s,*}, \boldsymbol{\chi}_K^{q,*}$ respectively represents the support set and query set for task $\mathcal{T}_K^*$ and the meta-parameters $\boldsymbol{\Theta}^{(K-1),*}$. Similar notation also applies to $\bar{\boldsymbol{\chi}}_K^* = \boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K^*]$. And the inequality is due to the fact that ReLU networks are naturally Lipschitz continuous w.r.t. some coefficient ξ_L when they are wide enough [3], as we have discussed above.

[Bounding term $I_{4.1}$] Based on the meta-optimization procedure (inner-loop optimization + outer-loop optimization), we have

$$\begin{aligned}
I_{4.1} &= \xi_L \cdot \|\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*] - \boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K^*]\|_2 \\
&= \xi_L \cdot \bigg\|\bigg(\boldsymbol{\Theta}^{(K-1),*} - \eta_2 \cdot \nabla_{\boldsymbol{\Theta}} \mathcal{L}(D_K^{q,*}; \boldsymbol{\Theta}_K^{(J),*})\bigg) - \bigg(\boldsymbol{\Theta}^{(K-1)} - \eta_2 \cdot \nabla_{\boldsymbol{\Theta}} \mathcal{L}(D_K^{q,*}; \boldsymbol{\Theta}_K^{(J)})\bigg)\bigg\|_2
\end{aligned}$$

where $\boldsymbol{\Theta}_K^{(J),*}$ is the task-specific parameter of $\mathcal{T}_K^*$ after adapting on $\boldsymbol{\Theta}^{(K-1),*}$ with inner-loop optimization, and the $\boldsymbol{\Theta}_K^{(J)}$ is the similar parameter after adapting on $\boldsymbol{\Theta}^{(K-1)}$. Here, we simplify the formula by representing the gradient derivation (inner-loop + outer-loop) with the mapping $H : \mathcal{T} \times \boldsymbol{\Theta} \mapsto \mathbb{R}^p$, which leads to

$$\begin{aligned}
&\|\boldsymbol{\Theta}^{(K-1),*}[\mathcal{T}_K^*] - \boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_K^*]\|_2 \\
&\qquad = \bigg\|\bigg(\boldsymbol{\Theta}^{(K-1),*} - \eta_2 \cdot \nabla_{\boldsymbol{\Theta}} \mathcal{L}(D^q; \boldsymbol{\Theta}_K^{(J),*})\bigg) - \bigg(\boldsymbol{\Theta}^{(K-1)} - \eta_2 \cdot \nabla_{\boldsymbol{\Theta}} \mathcal{L}(D^q; \boldsymbol{\Theta}_K^{(J)})\bigg)\bigg\|_2 \\
&\qquad = \|(\boldsymbol{\Theta}^{(K-1),*} - \boldsymbol{\Theta}^{(K-1)}) - \eta_2 \cdot \bigg(H(\mathcal{T}_K^*, \boldsymbol{\Theta}^{(K-1),*}) - H(\mathcal{T}_K^*, \boldsymbol{\Theta}^{(K-1)})\bigg)\|_2 \\
&\qquad = \|(\boldsymbol{\Theta}^{(K-2),*} - \boldsymbol{\Theta}^{(K-2)}) - \eta_2 \cdot \bigg(H(\mathcal{T}_K^*, \boldsymbol{\Theta}^{(K-1),*}) - H(\mathcal{T}_K^*, \boldsymbol{\Theta}^{(K-1)})\bigg) \\
&\qquad\qquad - \eta_2 \cdot \bigg(H(\mathcal{T}_{K-1}^*, \boldsymbol{\Theta}^{(K-2),*}) - H(\mathcal{T}_{K-1}^*, \boldsymbol{\Theta}^{(K-2)})\bigg)\|_2 \\
&\qquad \leq \sum_{k \in [K]} \eta_2 \cdot \big\|H(\mathcal{T}_k^*, \boldsymbol{\Theta}^{(k-1),*}) - H(\mathcal{T}_k^*, \boldsymbol{\Theta}^{(k-1)})\big\|_2
\end{aligned}$$

Recall that the past arms, including the actual chosen arms $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_K\}$ as well as the optimal ones $\{\mathcal{T}_1^*, \mathcal{T}_2^*, \ldots, \mathcal{T}_K^*\}$ are all from the candidate pool where each candidate arm is drawn i.i.d. from the task distribution $\mathcal{P}(\mathcal{T})$. Therefore, denoting the bound as $\|H(\mathcal{T}_K^*, \boldsymbol{\Theta}^{(K-1),*}) - H(\mathcal{T}_K^*, \boldsymbol{\Theta}^{(K-1)})\|_2 \leq S_1(K)$, we can have the upper bound as $I_{4.1} \leq \eta_2 \cdot \xi_L K \cdot S_1(K)$.

Then, for the term $S_1(K)$, by definition we have $\|H(\mathcal{T}_K^*, \boldsymbol{\Theta}^{(K-1),*}) - H(\mathcal{T}_K^*, \boldsymbol{\Theta}^{(K-1)})\| \leq S_1(K)$, applying mean-reduction for the sample loss will further leads to

$$\begin{aligned}
\|H(\mathcal{T}_K^*, \boldsymbol{\Theta}^{(K-1),*}) - H(\mathcal{T}_K^*, \boldsymbol{\Theta}^{(K-1)})\| &= \|\nabla_{\boldsymbol{\Theta}} \mathcal{L}(D_K^{q,*}; \boldsymbol{\Theta}_K^{(J),*}) - \nabla_{\boldsymbol{\Theta}} \mathcal{L}(D_K^{q,*}; \boldsymbol{\Theta}_K^{(J)})\|_2 \\
&= \|\frac{1}{|D_K^{q,*}|} \sum_{\boldsymbol{x} \in D_K^{q,*}} \nabla_{\boldsymbol{\Theta}} \mathcal{L}(\boldsymbol{x}; \boldsymbol{\Theta}_K^{(J),*}) - \frac{1}{|D_K^{q,*}|} \sum_{\boldsymbol{x} \in D_K^{q,*}} \nabla_{\boldsymbol{\Theta}} \mathcal{L}(\boldsymbol{x}; \boldsymbol{\Theta}_K^{(J)})\|_2 \\
&\leq \frac{1}{|D_K^{q,*}|} \sum_{\boldsymbol{x} \in D_K^{q,*}} \|\nabla_{\boldsymbol{\Theta}} \mathcal{L}(\boldsymbol{x}; \boldsymbol{\Theta}_K^{(J),*}) - \nabla_{\boldsymbol{\Theta}} \mathcal{L}(\boldsymbol{x}; \boldsymbol{\Theta}_K^{(J)})\|_2.
\end{aligned}$$

This inequality essentially bound the gradient difference when given the same input task $\mathcal{T}_K^*$ w.r.t. different sets of model parameters. Based on the conclusion from Lemma 9 of [46] and Lemma

B.3 of [10], we have $\|\nabla_{\mathbf{\Theta}_l} f(x; \mathbf{\Theta}_K)\|_F, \|\nabla_{\mathbf{\Theta}_l} \mathcal{L}(x; \mathbf{\Theta}_K)\|_F \leq \mathcal{O}(\sqrt{m_{\mathcal{F}}}), \forall l \in [L_{\mathcal{F}}]$ for any set of parameters within the sphere $\mathbf{\Theta}_K \in \mathcal{B}(\mathbf{\Theta}_0, \omega)$ where $\mathbf{\Theta}_0$ is the center and ω is the corresponding radius (which is a small value). With a total of $L_{\mathcal{F}}$ layers for the meta-model and each layer of $m_{\mathcal{F}}$ hidden units, this will give us $\|\nabla_{\mathbf{\Theta}} \mathcal{L}(\boldsymbol{x}; \mathbf{\Theta}_K^{(J),*})\|_2, \|\nabla_{\mathbf{\Theta}} \mathcal{L}(\boldsymbol{x}; \mathbf{\Theta}_K^{(J)})\|_2 \leq \mathcal{O}(\sqrt{m_{\mathcal{F}} L_{\mathcal{F}}})$ (**Lemma** D.15). And this makes $S_1(K) \leq \mathcal{O}(\sqrt{m_{\mathcal{F}} L_{\mathcal{F}}})$. Since we have $\eta_1, \eta_2 \leq \mathcal{O}(\frac{1}{m_{\mathcal{F}}})$, summarizing the results above, the upper bound can then be derived.

[Bounding term $I_{4.2}$] Next, we proceed to bound $I_{4.2}$, which will be

$$
\begin{aligned}
I_{4.2} &= |f_2\Big([\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^{s,*});\ \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^{q,*})]; \boldsymbol{\theta}_2^{(K-1)} \Big) - f_2\Big([\nabla_{\boldsymbol{\theta}} f_1(\bar{\boldsymbol{\chi}}_K^{s,*});\ \nabla_{\boldsymbol{\theta}} f_1(\bar{\boldsymbol{\chi}}_K^{q,*})]; \boldsymbol{\theta}_2^{(K-1)} \Big)| \\
&\leq \xi_L \cdot \big\| [\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^{s,*});\ \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^{q,*})] - [\nabla_{\boldsymbol{\theta}} f_1(\bar{\boldsymbol{\chi}}_K^{s,*});\ \nabla_{\boldsymbol{\theta}} f_1(\bar{\boldsymbol{\chi}}_K^{q,*})] \big\|_2 \\
&\leq \xi_L \cdot \big\| \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^{s,*}) - \nabla_{\boldsymbol{\theta}} f_1(\bar{\boldsymbol{\chi}}_K^{s,*}) \big\|_2 + \xi_L \cdot \big\| \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^{q,*}) - \nabla_{\boldsymbol{\theta}} f_1(\bar{\boldsymbol{\chi}}_K^{q,*}) \big\|_2 \\
&\leq \xi_L^2 \cdot \big\| \boldsymbol{\chi}_K^{s,*} - \bar{\boldsymbol{\chi}}_K^{s,*} \big\|_2 + \xi_L^2 \cdot \big\| \boldsymbol{\chi}_K^{q,*} - \bar{\boldsymbol{\chi}}_K^{q,*} \big\|_2
\end{aligned}
$$

where the inequalities are due to the Lipschitz continuity / smoothness properties of over-parameterized FC networks as we discussed above. Here, we notice that the second term on the RHS can be bounded by directly applying the proving procedure of term $I_{4.1}$. Then, for the first term on the RHS, we can following a similar procedure as for $I_{4.1}$, by

$$
\begin{aligned}
&\big\| \boldsymbol{\chi}_K^{s,*} - \bar{\boldsymbol{\chi}}_K^{s,*} \big\|_2 = \big\| \mathcal{I}(\mathcal{T}_k^*,\ \mathbf{\Theta}^{(k-1),*}) - \mathcal{I}(\mathcal{T}_k^*,\ \mathbf{\Theta}^{(k-1)}) \big\|_2 \\
&\quad = \| \Big(\mathbf{\Theta}^{(K-1),*} - \eta_1 \cdot \sum_{j \in [J]} \nabla_{\mathbf{\Theta}} \mathcal{L}(D_K^{s,*}; \mathbf{\Theta}_K^{(j),*}) \Big) - \Big(\mathbf{\Theta}^{(K-1)} - \eta_1 \cdot \sum_{j \in [J]} \nabla_{\mathbf{\Theta}} \mathcal{L}(D_K^{s,*}; \mathbf{\Theta}_K^{(j)}) \Big) \|_2 \\
&\quad = \| \big(\mathbf{\Theta}^{(K-2),*} - \mathbf{\Theta}^{(K-2)}) \big) - \big(\eta_1 \cdot \sum_{j \in [J]} \nabla_{\mathbf{\Theta}} \mathcal{L}(D_{K-1}^{s,*}; \mathbf{\Theta}_{K-1}^{(j)}) - \eta_1 \cdot \sum_{j \in [J]} \nabla_{\mathbf{\Theta}} \mathcal{L}(D_{K-1}^{s,*}; \mathbf{\Theta}_{K-1}^{(j),*}) \big) \\
&\qquad\qquad - \big(\eta_1 \cdot \sum_{j \in [J]} \nabla_{\mathbf{\Theta}} \mathcal{L}(D_K^{s,*}; \mathbf{\Theta}_K^{(j)}) - \eta_1 \cdot \sum_{j \in [J]} \nabla_{\mathbf{\Theta}} \mathcal{L}(D_K^{s,*}; \mathbf{\Theta}_K^{(j),*}) \|_2 \\
&\quad \leq \eta_1 \cdot \sum_{k \in [K]} \| \sum_{j \in [J]} \nabla_{\mathbf{\Theta}} \mathcal{L}(D_k^{s,*}; \mathbf{\Theta}_k^{(j)}) - \eta_1 \cdot \sum_{j \in [J]} \nabla_{\mathbf{\Theta}} \mathcal{L}(D_k^{s,*}; \mathbf{\Theta}_k^{(j),*}) \|_2 \\
&\quad \leq \mathcal{O}(\eta_1 \cdot KJ\sqrt{m_{\mathcal{F}} L_{\mathcal{F}}})
\end{aligned}
$$

where the last inequality is due to **Lemma** D.15 and by iterating through K meta-training iterations. Summing up the results above, we will have $I_{4.2} \leq \mathcal{O}(\eta_2 \xi_L^2 \cdot K\sqrt{m_{\mathcal{F}} L_{\mathcal{F}}}) + \mathcal{O}(\eta_1 \xi_L^2 \cdot KJ\sqrt{m_{\mathcal{F}} L_{\mathcal{F}}})$.

[Summing up the results] Then, combining all the results, we would have

$$
R_{\text{single}}(K) \leq \mathcal{O}(\frac{1}{\sqrt{K}}) \cdot \Big(\sqrt{2\xi_1} + \frac{3L}{\sqrt{2}} + (1 + 2\gamma_1)\sqrt{2\log(\frac{K}{\delta})} \Big) + \mathcal{O}(\frac{\xi_L^2 KJ\sqrt{L_{\mathcal{F}}}}{\sqrt{m_{\mathcal{F}}}}) + \gamma_m
$$

where

$$
\begin{aligned}
\gamma_1 &= 2 + \mathcal{O}\left(\frac{K^3 L}{\rho\sqrt{m}} \log m \right) + \mathcal{O}\left(\frac{L^2 K^4}{\rho^{4/3} m^{1/6}} \log^{11/6}(m) \right) \\
\gamma_m &= \Big(1 + \mathcal{O}(\frac{KL^3 \log^{5/6}(m)}{\rho^{1/3} m^{1/6}}) \Big) \mathcal{O}(\frac{K^3 L}{\rho\sqrt{m}} \log(m)) + \mathcal{O}\Big(\frac{K^4 L^2 \log^{11/6}(m)}{\rho^{4/3} m^{1/6}} \Big) + \frac{\xi_L K L^4 \log^{5/6}(m)}{\rho^{1/3} m^{1/6}}
\end{aligned}
$$

Note that the majority of the terms above can be cancelled to $\mathcal{O}(1)$ with proper networks width m indicated in **Theorem** 5.2. With increasingly large network width m, these terms will also become diminutive enough to achieve our regret bound in the main body.

D.2 Extending the result to the batch settings (Proof of Theorem 5.2)

With the results and conclusions from Subsection D.1, we proceed to provide the proof of **Theorem** 5.2 under the batch settings. Recall that in our original problem formulation and Algorithm 1, we are expected to select a batch of B arms in each meta-training iteration, denoted by $\{\Omega_k\}_{k \in [K]}$. Note that each of the candidate arms from $\Omega_{\text{task}}^{(k)}$ are drawn i.i.d. from the task distribution $\mathcal{P}(\mathcal{T})$. Meantime, we will have the corresponding optimal arm batches, denoted by $\{\Omega_k^*\}_{k\in[K]}$, which minimizes the loss objective in **Eq.** 5. Recall that we update the meta-model parameters with

$$\boldsymbol{\Theta}^{(k)} = \boldsymbol{\Theta}^{(k-1)} - \eta_2 \cdot \nabla_{\boldsymbol{\Theta}}\Big(\frac{1}{|\Omega_k|}\sum_{\mathcal{T}_{k,i}\in\Omega_k}\mathcal{L}(D^q_{k,i};\boldsymbol{\Theta}^{(J)}_{k,i})\Big) = \boldsymbol{\Theta}^{(k-1)} - \frac{\eta_2}{|\Omega_k|}\sum_{\mathcal{T}_{k,i}\in\Omega_k}\nabla_{\boldsymbol{\Theta}}\Big(\mathcal{L}(D^q_{k,i};\boldsymbol{\Theta}^{(J)}_{k,i})\Big)$$

where $\boldsymbol{\Theta}^{(J)}_{k,i}$ is the task-specific parameter for $\mathcal{T}_{k,i}$ after the inner-loop optimization for J steps.

Analogously, for the notation brevity and the sake of analysis, we denote $f(\boldsymbol{\Theta}^{(K-1)}[\Omega_K];\boldsymbol{\theta}^{(k-1)}) = f_1(\boldsymbol{\chi}^q_k;\boldsymbol{\theta}^{(k-1)}_1) + f_2\Big([\nabla_{\boldsymbol{\theta}}f_1(\boldsymbol{\chi}^s_k);\nabla_{\boldsymbol{\theta}}f_1(\boldsymbol{\chi}^q_k)];\boldsymbol{\theta}^{(k-1)}_2\Big)$ where we have $\boldsymbol{\chi}^q_k := \boldsymbol{\Theta}^{(K-1)}[\Omega_K]$ being the meta-parameters adapted to batch of tasks Ω_K, and the batch-specific parameter is defined as $\boldsymbol{\chi}^s_k := \frac{1}{|\Omega_K|}\sum_{\mathcal{T}_{k,\widehat{i}}\in\Omega_K}[\boldsymbol{\Theta}^{(J)}_{k,\widehat{i}}]$. Then, the regret under the batch setting can be denoted by

$$\begin{aligned}
R(K) = R_{\text{batch}}(K) &= \mathbb{E}_{\mathcal{T}\sim\mathcal{P}(\mathcal{T}),\boldsymbol{x}\sim\mathcal{D}_{\mathcal{T}}}\Big[\mathcal{L}\big(\boldsymbol{x};\mathcal{I}(\mathcal{T},\boldsymbol{\Theta}^{(K)})\big)\Big] - \mathbb{E}_{\mathcal{T}\sim\mathcal{P}(\mathcal{T}),\boldsymbol{x}\sim\mathcal{D}_{\mathcal{T}}}\Big[\mathcal{L}\big(\boldsymbol{x};\mathcal{I}(\mathcal{T},\boldsymbol{\Theta}^{(K),*})\big)\Big] \\
&= \mathbb{E}_{\mathcal{T}\sim\mathcal{P}(\mathcal{T}),\boldsymbol{x}\sim\mathcal{D}_{\mathcal{T}}}\Big[\mathcal{L}\big(\boldsymbol{x};\mathcal{I}(\mathcal{T},\boldsymbol{\Theta}^{(K-1)}[\Omega_K])\big)\Big] - \mathbb{E}_{\mathcal{T}\sim\mathcal{P}(\mathcal{T}),\boldsymbol{x}\sim\mathcal{D}_{\mathcal{T}}}\Big[\mathcal{L}\big(\boldsymbol{x};\mathcal{I}(\mathcal{T},\boldsymbol{\Theta}^{(K-1),*}[\Omega^*_K])\big)\Big] \\
&= h(\boldsymbol{\Theta}^{(K-1),*}[\Omega^*_K]) - h(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) \\
&= h(\boldsymbol{\Theta}^{(K-1),*}[\Omega^*_K]) - f(\boldsymbol{\Theta}^{(K-1),*}[\Omega^*_K];\tilde{\boldsymbol{\theta}}^{(K-1)}) + f(\boldsymbol{\Theta}^{(K-1),*}[\Omega^*_K];\tilde{\boldsymbol{\theta}}^{(K-1)}) - f(\boldsymbol{\Theta}^{(K-1)}[\Omega_K];\boldsymbol{\theta}^{(K-1)}) \\
&\qquad + f(\boldsymbol{\Theta}^{(K-1)}[\Omega_K];\boldsymbol{\theta}^{(K-1)}) - h(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]),
\end{aligned}$$

and after applying properties of the arm pulling mechanism, it is equivalent to

$$\begin{aligned}
R(K) &\leq h(\boldsymbol{\Theta}^{(K-1),*}[\Omega^*_K]) - f(\boldsymbol{\Theta}^{(K-1),*}[\Omega^*_K];\tilde{\boldsymbol{\theta}}^{(K-1)}) \\
&\quad + f(\boldsymbol{\Theta}^{(K-1),*}[\Omega^*_K];\tilde{\boldsymbol{\theta}}^{(K-1)}) - \hat{f}(\boldsymbol{\Theta}^{(K-1),*}[\Omega^*_K];\tilde{\boldsymbol{\theta}}^{(K-1)}) \\
&\quad + \hat{f}(\boldsymbol{\Theta}^{(K-1),*}[\Omega^*_K];\tilde{\boldsymbol{\theta}}^{(K-1)}) - \hat{f}(\boldsymbol{\Theta}^{(K-1)}[\Omega^*_K];\boldsymbol{\theta}^{(K-1)}) \\
&\quad + \hat{f}(\boldsymbol{\Theta}^{(K-1)}[\Omega_K];\boldsymbol{\theta}^{(K-1)}) - f(\boldsymbol{\Theta}^{(K-1)}[\Omega_K];\boldsymbol{\theta}^{(K-1)}) \; + f(\boldsymbol{\Theta}^{(K-1)}[\Omega_K];\boldsymbol{\theta}^{(K-1)}) - h(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) \\
&\leq \underbrace{|h(\boldsymbol{\Theta}^{(K-1),*}[\Omega^*_K]) - f(\boldsymbol{\Theta}^{(K-1),*}[\Omega^*_K];\tilde{\boldsymbol{\theta}}^{(K-1)})|}_{I_5} + \underbrace{|f(\boldsymbol{\Theta}^{(K-1),*}[\Omega^*_K];\tilde{\boldsymbol{\theta}}^{(K-1)}) - \hat{f}(\boldsymbol{\Theta}^{(K-1),*}[\Omega^*_K];\tilde{\boldsymbol{\theta}}^{(K-1)})|}_{I_6} \\
&\quad + \underbrace{|\hat{f}(\boldsymbol{\Theta}^{(K-1),*}[\Omega^*_K];\tilde{\boldsymbol{\theta}}^{(K-1)}) - \hat{f}(\boldsymbol{\Theta}^{(K-1)}[\Omega^*_K];\boldsymbol{\theta}^{(K-1)})|}_{I_7} + \underbrace{|\hat{f}(\boldsymbol{\Theta}^{(K-1)}[\Omega_K];\boldsymbol{\theta}^{(K-1)}) - f(\boldsymbol{\Theta}^{(K-1)}[\Omega_K];\boldsymbol{\theta}^{(K-1)})|}_{I_8} \\
&\qquad + \underbrace{|f(\boldsymbol{\Theta}^{(K-1)}[\Omega_K];\boldsymbol{\theta}^{(K-1)}) - h(\boldsymbol{\Theta}^{(K-1)}[\Omega_K])|}_{I_9}
\end{aligned}$$

where the average value of estimated benefit scores for individual tasks $\mathcal{T}_{K,i}\in\Omega_K$ is represented as $\widehat{f}(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) = \frac{1}{|\Omega_K|}\cdot\sum_{\mathcal{T}_{K,i}\in\Omega_K}f(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_{K,i}]) = \frac{1}{|\Omega_K|}\cdot\sum_{\mathcal{T}_{K,i}\in\Omega_K}f\big(\boldsymbol{\Theta}^{(k-1)} - \eta_2\cdot\nabla_{\boldsymbol{\Theta}}\mathcal{L}(D^q_{K,i};\boldsymbol{\Theta}^{(J)}_{K,i});\boldsymbol{\theta}^{(K-1)}\big)$, and the inequality is due to the pulling mechanism of BASS. Here, I_5, I_9 individually correspond to I_1, I_2 in the single-task setting and can be bounded by **Lemma** D.3, **Corollary** D.4. Term I_7 can be upper bounded by $I_3 + I_4$ from the single-task setting above. Then, for the rest terms I_6, I_8, we proceed to bound them separately.

D.2.1 Bounding error terms and assembling the regret bound

We begin with the term I_8, and then proceed to I_6. For the chosen batch of tasks Ω_K in the round K, we will have $f_1(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) = f_1\big(\boldsymbol{\Theta}^{(K-1)} - \eta_2\cdot\nabla_{\boldsymbol{\Theta}}\big(\frac{1}{|\Omega_K|}\sum_{\mathcal{T}_{K,i}\in\Omega_K}\mathcal{L}(D^q_{K,i};\boldsymbol{\Theta}^{(J)}_{K,i})\big);\boldsymbol{\theta}^{(K-1)}_1\big)$, In this case, the average value of estimation sampling probabilities for tasks $\mathcal{T}_{K,i}\in\Omega_K$ is

$$\begin{aligned}
&\widehat{f}(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) = \widehat{f}_1(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) + \widehat{f}_2(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) \\
&\quad = \frac{1}{|\Omega_K|}\sum_{\mathcal{T}_{K,i}\in\Omega_K}\Big[f_1(\boldsymbol{\Theta}^{(K-1)}[\mathcal{T}_{K,i}];\boldsymbol{\theta}^{(K-1)}_1) + f_2\big([\nabla_{\boldsymbol{\theta}}f_1(\boldsymbol{\chi}^s_{K,i});\ \nabla_{\boldsymbol{\theta}}f_1(\boldsymbol{\chi}^q_{K,i})];\boldsymbol{\theta}^{(K-1)}_2\big)\Big]
\end{aligned}$$

[Bounding the f_1 output difference] Next, let us first proceed to bound the output difference with respect to the exploitation module f_1, where we can transform this term into

$$
\begin{aligned}
&f_1(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) - \widehat{f}_1(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) \\
&\quad= f_1\Bigg(\boldsymbol{\Theta}_1^{(k-1)} - \eta_2 \cdot \nabla_{\boldsymbol{\Theta}}\big(\frac{1}{|\Omega_K|}\sum_{\mathcal{T}_{K,i}\in\Omega_K}\mathcal{L}(D^q_{K,i};\boldsymbol{\Theta}^{(J)}_{K,i})\big);\boldsymbol{\theta}_1^{(K-1)}\Bigg) \\
&\qquad\qquad - \frac{1}{|\Omega_K|}\cdot\sum_{\mathcal{T}_{K,j}\in\Omega_K} f_1\Bigg(\boldsymbol{\Theta}^{(k-1)} - \eta_2\cdot\nabla_{\boldsymbol{\Theta}}\mathcal{L}(D^q_{K,j};\boldsymbol{\Theta}^{(J)}_{K,j});\boldsymbol{\theta}_1^{(K-1)}\Bigg) \\
&\quad= \frac{1}{|\Omega_K|}\cdot\sum_{\mathcal{T}_{K,j}\in\Omega_K}\Bigg(f_1\big(\boldsymbol{\Theta}^{(k-1)} - \eta_2\cdot\nabla_{\boldsymbol{\Theta}}\big(\frac{1}{|\Omega_K|}\sum_{\mathcal{T}_{K,i}\in\Omega_K}\mathcal{L}(D^q_{K,i};\boldsymbol{\Theta}^{(J)}_{K,i})\big);\boldsymbol{\theta}_1^{(K-1)}\big) \\
&\qquad\qquad - f_1\big(\boldsymbol{\Theta}^{(k-1)} - \eta_2\cdot\nabla_{\boldsymbol{\Theta}}\mathcal{L}(D^q_{K,j};\boldsymbol{\Theta}^{(J)}_{K,j});\boldsymbol{\theta}_1^{(K-1)}\big)\Bigg).
\end{aligned}
$$

Then, applying the Lipschitz continuity property will lead to

$$
\begin{aligned}
&f_1(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) - \widehat{f}_1(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) \\
&\qquad\leq \frac{\eta_2\cdot\xi_L}{|\Omega_K|}\cdot\sum_{\mathcal{T}_{K,i}\in\Omega_K}\|\nabla_{\boldsymbol{\Theta}}\big(\frac{1}{|\Omega_K|}\sum_{\mathcal{T}_{K,j}\in\Omega_K}\mathcal{L}(D^q_{K,j};\boldsymbol{\Theta}^{(J)}_{K,j})\big) - \nabla_{\boldsymbol{\Theta}}\mathcal{L}(D^q_{K,i};\boldsymbol{\Theta}^{(J)}_{K,i})\|_2.
\end{aligned}
$$

Here, by the definition of the outer-loop optimization of first-order meta-learning, we will have an alternative form the inequality, denoted by

$$
\begin{aligned}
&f_1(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) - \widehat{f}_1(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) \leq \\
&\qquad\frac{\eta_2\cdot\xi_L}{|\Omega_K|}\cdot\Bigg(\sum_{\mathcal{T}_{K,i}\in\Omega_K}\|\frac{1}{|\Omega_K|}\sum_{\mathcal{T}_{K,j}\in\Omega_K}\nabla_{\boldsymbol{\Theta}}\big(\mathcal{L}(D^q_{K,j};\boldsymbol{\Theta}^{(J)}_{K,j})\big) - \nabla_{\boldsymbol{\Theta}}\mathcal{L}(D^q_{K,i};\boldsymbol{\Theta}^{(J)}_{K,i})\|_2\Bigg).
\end{aligned}
$$

For the term in the parentheses on the RHS, substituting the backward operation with the $H(\cdot,\cdot)$ mapping function, we have

$$
\begin{aligned}
&\sum_{\mathcal{T}_{K,i}\in\Omega_K}\|\frac{1}{|\Omega_K|}\sum_{\mathcal{T}_{K,j}\in\Omega_K}\nabla_{\boldsymbol{\Theta}}\big(\mathcal{L}(D^q_{K,j};\boldsymbol{\Theta}^{(J)}_{K,j})\big) - \nabla_{\boldsymbol{\Theta}}\mathcal{L}(D^q_{K,i};\boldsymbol{\Theta}^{(J)}_{K,i})\|_2 \\
&\quad= \sum_{\mathcal{T}_{K,i}\in\Omega_K}\|\frac{1}{|\Omega_K|}\sum_{\mathcal{T}_{K,j}\in\Omega_K}\nabla_{\boldsymbol{\Theta}}\big(\mathcal{L}(D^q_{K,j};\boldsymbol{\Theta}^{(J)}_{K,j})\big) - \frac{1}{|\Omega_K|}\sum_{\mathcal{T}_{K,j}\in\Omega_K}\nabla_{\boldsymbol{\Theta}}\mathcal{L}(D^q_{K,i};\boldsymbol{\Theta}^{(J)}_{K,i})\|_2 \\
&\quad\leq \frac{1}{|\Omega_K|}\sum_{\mathcal{T}_{K,i}\in\Omega_K}\sum_{\mathcal{T}_{K,j}\in\Omega_K}\|\nabla_{\boldsymbol{\Theta}}\mathcal{L}(D^q_{K,j};\boldsymbol{\Theta}^{(J)}_{K,j}) - \nabla_{\boldsymbol{\Theta}}\mathcal{L}(D^q_{K,i};\boldsymbol{\Theta}^{(J)}_{K,i})\|_2 \\
&\quad= \frac{1}{|\Omega_K|}\sum_{\mathcal{T}_{K,i}\in\Omega_K}\sum_{\mathcal{T}_{K,j}\in\Omega_K}\|H(\mathcal{T}_{K,i},\boldsymbol{\Theta}^{(K-1)}) - H(\mathcal{T}_{K,j},\boldsymbol{\Theta}^{(K-1)})\|_2 \\
&\quad\leq |\Omega_K|\cdot S_1(K).
\end{aligned}
$$

with $|\Omega_K| = B$. Therefore, the f_1 part of error term I_8 could be bounded by $f_1(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) - \widehat{f}_1(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) \leq \eta_2\cdot\xi_L\cdot B\cdot S_1(K)$. where the upper bound $S_1(K) \leq \mathcal{O}(\sqrt{m_{\mathcal{F}}L_{\mathcal{F}}})$ can be found in the procedure bounding term $I_{4.1}$.

[Bounding the f_2 output difference] Then, with $\boldsymbol{\chi}_K = \boldsymbol{\Theta}^{(K-1)}[\Omega_K]$, we proceed to bound the output difference with respect to the exploration module, which is represented by

$$
\begin{aligned}
&f_2(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) - \widehat{f}_2(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) \\
&\quad= f_2\big([\nabla_{\boldsymbol{\theta}}f_1(\boldsymbol{\chi}^s_K);\ \nabla_{\boldsymbol{\theta}}f_1(\boldsymbol{\chi}^q_K)];\boldsymbol{\theta}_2^{(K-1)}\big) - \frac{1}{|\Omega_K|}\cdot\sum_{\mathcal{T}_{K,j}\in\Omega_K}f_2\big([\nabla_{\boldsymbol{\theta}}f_1(\boldsymbol{\chi}^s_{K,j});\ \nabla_{\boldsymbol{\theta}}f_1(\boldsymbol{\chi}^q_{K,j})];\boldsymbol{\theta}_2^{(K-1)}\big) \\
&\quad= \frac{1}{|\Omega_K|}\cdot\sum_{\mathcal{T}_{K,j}\in\Omega_K}\Bigg(f_2\big([\nabla_{\boldsymbol{\theta}}f_1(\boldsymbol{\chi}^s_K);\ \nabla_{\boldsymbol{\theta}}f_1(\boldsymbol{\chi}^q_K)];\boldsymbol{\theta}_2^{(K-1)}\big) - f_2\big([\nabla_{\boldsymbol{\theta}}f_1(\boldsymbol{\chi}^s_{K,j});\ \nabla_{\boldsymbol{\theta}}f_1(\boldsymbol{\chi}^q_{K,j})];\boldsymbol{\theta}_2^{(K-1)}\big)\Bigg)
\end{aligned}
$$

By adopting the Lipschitz continuity property of f_2, we will have

$$\begin{aligned}
&f_2(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) - \widehat{f}_2(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) \\
&\leq \frac{\xi_L}{|\Omega_K|} \cdot \sum_{\mathcal{T}_{K,j} \in \Omega_K} \big\| [\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^s);\ \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^q)] - [\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_{K,j}^s);\ \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_{K,j}^q)] \big\|_2 \\
&\leq \frac{\xi_L}{|\Omega_K|} \cdot \sum_{\mathcal{T}_{K,j} \in \Omega_K} \big\| \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^s) - \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_{K,j}^s) \big\|_2 + \big\| \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_K^q) - \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_{K,j}^q) \big\|_2 \\
&\leq \frac{\xi_L^2}{|\Omega_K|} \cdot \sum_{\mathcal{T}_{K,j} \in \Omega_K} \big\| \boldsymbol{\chi}_K^s - \boldsymbol{\chi}_{K,j}^s \big\|_2 + \big\| \boldsymbol{\chi}_K^q - \boldsymbol{\chi}_{K,j}^q \big\|_2 \\
&= \frac{\xi_L^2 \cdot \eta_2}{|\Omega_K|} \cdot \sum_{\mathcal{T}_{K,j} \in \Omega_K} \| \nabla_{\boldsymbol{\Theta}} \big(\frac{1}{|\Omega_K|} \sum_{\mathcal{T}_{K,i} \in \Omega_K} \mathcal{L}(D_{K,i}^q; \boldsymbol{\Theta}_{K,i}^{(J)}) \big) - \nabla_{\boldsymbol{\Theta}} \mathcal{L}(D_{K,j}^q; \boldsymbol{\Theta}_{K,j}^{(J)}) \|_2 \\
&\qquad + \frac{\xi_L^2}{|\Omega_K|} \cdot \sum_{\mathcal{T}_{K,j} \in \Omega_K} \big\| \boldsymbol{\chi}_K^s - \boldsymbol{\chi}_{K,j}^s \big\|_2 \\
&\leq \eta_2 \cdot \xi_L^2 B \cdot S_1(K) + \frac{\xi_L^2 \cdot \eta_1}{|\Omega_K|} \cdot \sum_{\mathcal{T}_{K,j} \in \Omega_K} \| \frac{1}{|\Omega_K|} \sum_{\mathcal{T}_{K,i} \in \Omega_K} \boldsymbol{\Theta}_{K,i}^{(J)} - \frac{1}{|\Omega_K|} \sum_{\mathcal{T}_{K,i} \in \Omega_K} \boldsymbol{\Theta}_{K,j}^{(J)} \|_2
\end{aligned}$$

where the last inequality is by applying the conclusion when bounding the output difference w.r.t. the exploitation module f_1. Then, for the second term on the RHS,

$$\begin{aligned}
&\sum_{\mathcal{T}_{K,j} \in \Omega_K} \| \frac{1}{|\Omega_K|} \sum_{\mathcal{T}_{K,i} \in \Omega_K} \boldsymbol{\Theta}_{K,i}^{(J)} - \frac{1}{|\Omega_K|} \sum_{\mathcal{T}_{K,i} \in \Omega_K} \boldsymbol{\Theta}_{K,j}^{(J)} \|_2 \leq \frac{1}{|\Omega_K|} \sum_{\mathcal{T}_{K,j} \in \Omega_K} \sum_{\mathcal{T}_{K,i} \in \Omega_K} \| \boldsymbol{\Theta}_{K,i}^{(J)} - \boldsymbol{\Theta}_{K,j}^{(J)} \|_2 \\
&\leq \frac{1}{|\Omega_K|} \sum_{\mathcal{T}_{K,j} \in \Omega_K} \sum_{\mathcal{T}_{K,i} \in \Omega_K} \| \big(\boldsymbol{\Theta}^{K-1} - \sum_{\tau \in [\tau]} \nabla_{\boldsymbol{\Theta}} \mathcal{L}(D_{K,i}^s; \boldsymbol{\Theta}_{K,i}^{(\tau)}) \big) - \big(\boldsymbol{\Theta}^{K-1} - \sum_{\tau \in [J]} \nabla_{\boldsymbol{\Theta}} \mathcal{L}(D_{K,j}^s; \boldsymbol{\Theta}_{K,j}^{(\tau)}) \big) \|_2 \\
&= \frac{1}{|\Omega_K|} \sum_{\mathcal{T}_{K,j} \in \Omega_K} \sum_{\mathcal{T}_{K,i} \in \Omega_K} \| \sum_{\tau \in [J]} \nabla_{\boldsymbol{\Theta}} \mathcal{L}(D_{K,i}^s; \boldsymbol{\Theta}_{K,i}^{(\tau)}) - \sum_{\tau \in [J]} \nabla_{\boldsymbol{\Theta}} \mathcal{L}(D_{K,j}^s; \boldsymbol{\Theta}_{K,j}^{(\tau)}) \|_2 \\
&\leq |\Omega_K| J \cdot \mathcal{O}(\sqrt{m_{\mathcal{F}} L_{\mathcal{F}}})
\end{aligned}$$

where the last inequality is due to **Lemma** D.15. Summing up all the results above will give us the upper bound for f_2 output difference $f_2(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) - \widehat{f}_2(\boldsymbol{\Theta}^{(K-1)}[\Omega_K]) \leq \mathcal{O}(\eta_2 \cdot \xi_L^2 B \cdot \sqrt{m_{\mathcal{F}} L_{\mathcal{F}}} + \eta_1 \cdot BJ \cdot \sqrt{m_{\mathcal{F}} L_{\mathcal{F}}})$.

[Similar procedure for term I_6] Analogously, we can apply the same derivation for the error term I_6, which leads to

$$\begin{aligned}
I_6 &= f(\boldsymbol{\Theta}^{(K-1),*}[\Omega_K^*]; \tilde{\boldsymbol{\theta}}^{(K-1)}) - \hat{f}(\boldsymbol{\Theta}^{(K-1),*}[\Omega_K^*]; \tilde{\boldsymbol{\theta}}^{(K-1)}) \\
&= f_1 \Bigg(\boldsymbol{\Theta}^{(k-1),*} - \eta_2 \nabla_{\boldsymbol{\Theta}} \big(\frac{1}{|\Omega_K^*|} \sum_{\mathcal{T}_{i^*} \in \Omega_K^*} \mathcal{L}(D_{i^*}^q; \boldsymbol{\Theta}_{i^*}^{(J)}) \big); \tilde{\boldsymbol{\theta}}_1^{(K-1)} \Bigg) \\
&\qquad\qquad - \frac{1}{|\Omega_K^*|} \cdot \sum_{\mathcal{T}_{i^*} \in \Omega_K^*} f_1 \Bigg(\boldsymbol{\Theta}^{(k-1),*} - \eta_2 \nabla_{\boldsymbol{\Theta}} \mathcal{L}(D_{i^*}^q; \boldsymbol{\Theta}_{i^*}^{(J)}); \tilde{\boldsymbol{\theta}}_1^{(K-1)} \Bigg).
\end{aligned}$$

Following a similar procedure as that of term I_8 will give us a similar bound as

$$\begin{aligned}
I_6 = f(\boldsymbol{\Theta}^{(K-1),*}[\Omega_K^*]; \tilde{\boldsymbol{\theta}}^{(K-1)}) &- \hat{f}(\boldsymbol{\Theta}^{(K-1),*}[\Omega_K^*]; \tilde{\boldsymbol{\theta}}^{(K-1)}) \\
&\leq \mathcal{O}(\eta_2 \cdot \xi_L^2 B \cdot \sqrt{m_{\mathcal{F}} L_{\mathcal{F}}} + BJ \cdot \sqrt{m_{\mathcal{F}} L_{\mathcal{F}}} \eta_1).
\end{aligned}$$

where the learning rate $\eta_1, \eta_2 \leq \mathcal{O}(\frac{1}{m_{\mathcal{F}}})$ is a small value. Then, the upper bounds for error terms I_6, I_8 are given as desired.

[Assembling the results] Then, combining all the results, we would have

$$R(K) \leq \mathcal{O}(\frac{1}{\sqrt{K}}) \cdot \Bigg(\sqrt{2\xi_1} + \frac{3L}{\sqrt{2}} + (1 + 2\gamma_1) \sqrt{2 \log(\frac{K}{\delta})} \Bigg) + \mathcal{O}(\frac{\xi_L^2 KBJ \sqrt{L_{\mathcal{F}}}}{\sqrt{m_{\mathcal{F}}}}) + \gamma_m$$

where

$$\gamma_1 = 2 + \mathcal{O}\left(\frac{K^3 L}{\rho\sqrt{m}}\log m\right) + \mathcal{O}\left(\frac{L^2 K^4}{\rho^{4/3} m^{1/6}}\log^{11/6}(m)\right)$$

$$\gamma_m = \left(1 + \mathcal{O}(\frac{K L^3 \log^{5/6}(m)}{\rho^{1/3} m^{1/6}})\right)\mathcal{O}(\frac{K^3 L}{\rho\sqrt{m}}\log(m)) + \mathcal{O}\left(\frac{K^4 L^2 \log^{11/6}(m)}{\rho^{4/3} m^{1/6}}\right) + \frac{\xi_L K L^4 \log^{5/6}(m)}{\rho^{1/3} m^{1/6}}$$

Similarly, with proper networks width m as in **Theorem** 5.2, the majority of the terms above can be cancelled to $\mathcal{O}(1)$. With increasingly large network width m under the over-parameterization settings, γ_1, γ_m will also become diminutive.

D.3 Performance Guarantee for the Exploitation and Exploration Modules

In this subsection, we would like to give the performance guarantee for the proposed BASS framework, and the corresponding performance bound can be applied to derive an upper bound for the error terms I_1, I_2 for the single-task settings and I_5, I_9 under the batch settings. Up to meta-training iteration $k \in [K]$ (before updating the meta-parameters and BASS), we denote all the past records received as $\mathcal{P}_{k-1}$. Before presenting the main lemmas, we first introduce the following operator. Inspired by [3], with two arbitrary vectors $\tilde{\boldsymbol{\chi}}, \boldsymbol{\chi}$ such that $\|\tilde{\boldsymbol{\chi}}\|_2 \leq 1, \|\boldsymbol{\chi}\|_2 = 1$, we have the following operator

$$\phi(\tilde{\boldsymbol{\chi}}, \boldsymbol{\chi}) = (\frac{\tilde{\boldsymbol{\chi}}}{\sqrt{2}}, \frac{\boldsymbol{\chi}}{2}, c) \tag{11}$$

as the concatenation of the two vectors $\frac{\tilde{\boldsymbol{\chi}}}{\sqrt{2}}, \frac{\boldsymbol{\chi}}{2}$ and one constant c, where $c = \sqrt{\frac{3}{4} - (\frac{\|\tilde{\boldsymbol{\chi}}\|_2}{\sqrt{2}})^2} \geq \frac{1}{2}$. And this operator transforms the transformed vector into unit norm, $\|\phi(\tilde{\boldsymbol{\chi}}, \boldsymbol{\chi})\|_2 = 1$. The idea of this operator is to make the gradients $\nabla_{\boldsymbol{\theta}} f_1(\cdot; \boldsymbol{\theta}_1)$ of the exploitation model, which is the input of the exploration model $f_2(\cdot; \boldsymbol{\theta}_2)$, comply with the normalization requirement and the separateness assumption (**Assumption** 5.1). For the sake of analysis, we will adopt this operation in the following proof. Note that this operator is just one possible solution, and our results could be easily generalized to other forms of input gradients under the unit-length and separateness assumption. Similar ideas are also applied in previous works [9]. We begin to bound the single-task settings with the following lemma.

Lemma D.1. *For the constants $c'_g > 0$, $0 < \rho \leq \mathcal{O}(\frac{1}{L})$ and $\xi_1 \in (0, 1)$, given the past records $\mathcal{P}_{k-1}$, we suppose m, η_1, η_2 satisfy the conditions in **Theorem** 5.2, and randomly draw the parameter $\{\boldsymbol{\theta}_1^{(k)}, \boldsymbol{\theta}_2^{(k)}\} \sim \{\widetilde{\boldsymbol{\theta}}_1^{(\tau)}, \widetilde{\boldsymbol{\theta}}_2^{(\tau)}\}_{\tau\in[k]}$. Consider the past records $\mathcal{P}_{k-1}$ up to round k are generated by a fixed policy when witness the candidate arms $\{\Omega_{task}^{(\tau)}\}_{\tau\in[k]}$. Then, with probability at least $1 - \delta$ given an arm-reward pair $(\mathcal{T}_{k,\widehat{i}}, r_{k,\widehat{i}})$, we have*

$$\mathbb{E}_{\mathcal{T}_{k,i}\sim\mathcal{P}(\mathcal{T})}\left[|f_2\left(\phi(\frac{[\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_{k,\widehat{i}}^s);\ \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_{k,\widehat{i}}^q)]}{c'_g L}, \boldsymbol{\chi}_{k,\widehat{i}}^q); \boldsymbol{\theta}_2^{(k-1)}\right) - \left(r_\tau - f_1(\boldsymbol{\chi}_{k,\widehat{i}}^q; \boldsymbol{\theta}_1^{(k-1)})\right)|\ |\Omega_{task}^{(k)}, \mathcal{P}_{k-1}\right]$$
$$\leq \frac{1}{\sqrt{k}} \cdot \left(\sqrt{2\xi_1} + \frac{3L}{\sqrt{2}} + (1 + 2\gamma_1)\sqrt{2\log(\frac{k}{\delta})}\right)$$

where

$$\gamma_1 = 2 + \mathcal{O}\left(\frac{k^3 L}{\rho\sqrt{m}}\log m\right) + \mathcal{O}\left(\frac{L^2 k^4}{\rho^{4/3} m^{1/6}}\log^{11/6}(m)\right).$$

Proof. The proof of this lemma is inspired by Lemma C.1 from [9]. First, we can derive the output upper bound

$$\left|f_2\left(\phi(\frac{[\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_{k,\widehat{i}}^s);\ \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_{k,\widehat{i}}^q)]}{c'_g L}, \boldsymbol{\chi}_{k,\widehat{i}}^q); \boldsymbol{\theta}_2^{(k-1)}\right) - \left(r_k - f_1(\boldsymbol{\chi}_{k,\widehat{i}}^q; \boldsymbol{\theta}_1^{(k-1)})\right)\right|$$
$$\leq \left|f_2\left(\phi(\frac{[\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_{k,\widehat{i}}^s);\ \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_{k,\widehat{i}}^q)]}{c'_g L}, \boldsymbol{\chi}_{k,\widehat{i}}^q); \boldsymbol{\theta}_2^{(k-1)}\right)\right| + \left|f_1(\boldsymbol{\chi}_{k,\widehat{i}}^q; \boldsymbol{\theta}_1^{(k-1)})\right| + 1$$
$$\leq 1 + 2\gamma_1$$

by triangle inequality and applying the generalization result of FC networks (**Lemma** D.5) on $f_1(\cdot;\boldsymbol{\theta}_1), f_2(\cdot;\boldsymbol{\theta}_2)$.

For the brevity of notation, we use $\nabla f_1(\mathcal{T}_{k,\widehat{i}})$ to denote $\phi(\frac{[\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}^s_{k,\widehat{i}});\ \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}^q_{k,\widehat{i}})]}{c'_g L}, \boldsymbol{\chi}^q_{k,\widehat{i}})$ and apply $(\boldsymbol{\chi}_k, r_k)$ as $(\boldsymbol{\chi}^q_{k,\widehat{i}}, r_{k,\widehat{i}})$ for the following proof. Define the difference sequence as

$$
\begin{aligned}
V_\tau^{(1)} = \mathbb{E}\bigg[\bigg| f_2\Big(\nabla f_1(\mathcal{T}_{\tau,\widehat{i}}); \boldsymbol{\theta}_2^{(\tau-1)}\Big) - \Big(r_\tau - f_1(\boldsymbol{\chi}_\tau; \boldsymbol{\theta}_1^{(\tau-1)})\Big)\bigg|\bigg] \\
- \bigg| f_2\Big(\nabla f_1(\mathcal{T}_{\tau,\widehat{i}}); \boldsymbol{\theta}_2^{(\tau-1)}\Big) - \Big(r_\tau - f_1(\boldsymbol{\chi}_\tau; \boldsymbol{\theta}_1^{(\tau-1)})\Big)\bigg|.
\end{aligned}
$$

Since the past rewards and the received arm-reward pairs $(\boldsymbol{\chi}_\tau, r_\tau)$ are generated by the same reward mapping function, we have the expectation

$$
\begin{aligned}
\mathbb{E}[V_\tau^{(1)} | F_\tau] =& \mathbb{E}\bigg[\bigg| f_2\Big(\nabla f_1(\mathcal{T}_{\tau,\widehat{i}}); \boldsymbol{\theta}_2^{(\tau-1)}\Big) - \Big(r_\tau - f_1(\boldsymbol{\chi}_\tau; \boldsymbol{\theta}_1^{(\tau-1)})\Big)\bigg|\bigg] \\
&- \mathbb{E}\bigg[\bigg| f_2\Big(\nabla f_1(\mathcal{T}_{\tau,\widehat{i}}); \boldsymbol{\theta}_2^{(\tau-1)}\Big) - \Big(r_\tau - f_1(\boldsymbol{\chi}_\tau; \boldsymbol{\theta}_1^{(\tau-1)})\Big)\bigg| \Big| F_\tau\bigg] = 0.
\end{aligned}
$$

where F_τ denotes the filtration given the past records $\mathcal{P}_\tau$, up to round $\tau \in [k]$. This also gives the fact that $V_\tau^{(1)}$ is a martingale difference sequence. Then, after applying the martingale difference sequence over $[k]$, we have

$$
\begin{aligned}
\frac{1}{k}\sum_{\tau \in [k]} V_\tau^{(1)} =& \frac{1}{k}\sum_{\tau \in [k]} \mathbb{E}\bigg[\bigg| f_2\Big(\nabla f_1(\mathcal{T}_{\tau,\widehat{i}}); \boldsymbol{\theta}_2^{(\tau-1)}\Big) - \Big(r_\tau - f_1(\boldsymbol{\chi}_\tau; \boldsymbol{\theta}_1^{(\tau-1)})\Big)\bigg|\bigg] \\
&- \frac{1}{k}\sum_{\tau \in [k]} \bigg| f_2\Big(\nabla f_1(\mathcal{T}_{\tau,\widehat{i}}); \boldsymbol{\theta}_2^{(\tau-1)}\Big) - \Big(r_\tau - f_1(\boldsymbol{\chi}_\tau; \boldsymbol{\theta}_1^{(\tau-1)})\Big)\bigg|.
\end{aligned}
$$

Then, by applying the Azuma-Hoeffding inequality, it leads to

$$
\mathbb{P}\bigg[\frac{1}{k}\sum_{\tau \in [k]} V_\tau^{(1)} - \frac{1}{k}\sum_{\tau \in [k]} \mathbb{E}[V_\tau^{(1)}] \geq (1 + 2\gamma_1)\sqrt{\frac{2\log(1/\delta)}{k}}\bigg] \leq \delta
$$

Since the expectation of $V_\tau^{(1)}$ is zero, with the probability at least $1 - \delta$ and an existing set of parameters $\boldsymbol{\theta}_2$ s.t. $\|\boldsymbol{\theta}_2 - \boldsymbol{\theta}_2^{(0)}\| \leq \mathcal{O}\left(\frac{k^3}{\rho\sqrt{m}}\log m\right)$, the above inequality implies

$$
\begin{aligned}
&\frac{1}{k}\sum_{\tau \in [k]} V_\tau^{(1)} \leq (1 + 2\gamma_1)\sqrt{\frac{2\log(1/\delta)}{k}} \implies \\
&\quad \mathbb{E}_{\mathcal{T}_{k,i} \sim \mathcal{P}(\mathcal{T})} \mathbb{E}_{\{\boldsymbol{\theta}_1^{(k-1)}, \boldsymbol{\theta}_2^{(k-1)}\}} \bigg[\bigg| f_2\Big(\nabla f_1(\mathcal{T}_{\tau,\widehat{i}}); \boldsymbol{\theta}_2^{(k-1)}\Big) - \Big(r_\tau - f_1(\boldsymbol{\chi}_\tau; \boldsymbol{\theta}_1^{(k-1)})\Big)\bigg|\bigg] \\
&\quad = \frac{1}{k}\sum_{\tau \in [k]} \mathbb{E}\bigg[\bigg| f_2\Big(\nabla f_1(\mathcal{T}_{\tau,\widehat{i}}); \boldsymbol{\theta}_2^{(\tau-1)}\Big) - \Big(r_k - f_1(\boldsymbol{\chi}_\tau; \boldsymbol{\theta}_1^{(\tau-1)})\Big)\bigg|\bigg] \\
&\quad \leq \frac{1}{k}\sum_{\tau \in [k]} \bigg| f_2\Big(\nabla f_1(\mathcal{T}_{\tau,\widehat{i}}); \boldsymbol{\theta}_2^{(\tau-1)}\Big) - \Big(r_\tau - f_1(\boldsymbol{\chi}_\tau; \boldsymbol{\theta}_1^{(\tau-1)})\Big)\bigg| + (1 + 2\gamma_1)\sqrt{\frac{2\log(1/\delta)}{k}} \\
&\quad \underset{\text{(i)}}{\leq} \frac{1}{k}\sum_{\tau \in [k]} \bigg| f_2\Big(\nabla f_1(\mathcal{T}_{\tau,\widehat{i}}); \boldsymbol{\theta}_2\Big) - \Big(r_\tau - f_1(\boldsymbol{\chi}_\tau; \boldsymbol{\theta}_1^{(\tau-1)})\Big)\bigg| + \frac{3L}{\sqrt{2k}} + (1 + 2\gamma_1)\sqrt{\frac{2\log(1/\delta)}{k}} \\
&\quad \leq \frac{1}{\sqrt{k}}\sqrt{\sum_{\tau \in [k]} \bigg| f_2\Big(\nabla f_1(\mathcal{T}_{\tau,\widehat{i}}); \boldsymbol{\theta}_2\Big) - \Big(r_\tau - f_1(\boldsymbol{\chi}_\tau; \boldsymbol{\theta}_1^{(\tau-1)})\Big)\bigg|^2} + \frac{3L}{\sqrt{2k}} + (1 + 2\gamma_1)\sqrt{\frac{2\log(1/\delta)}{k}} \\
&\quad \underset{\text{(ii)}}{\leq} \sqrt{\frac{2\xi_1}{k}} + \frac{3L}{\sqrt{2k}} + (1 + 2\gamma_1)\sqrt{\frac{2\log(1/\delta)}{k}}.
\end{aligned}
$$

where the first equality is due to the sampling of candidate tasks and the model parameters. Here, the upper bound (i) is derived by applying the conclusions of **Lemma** D.6 and **Lemma** D.10, and the inequality (ii) is derived by adopting **Lemma** D.6 while defining the empirical loss to be $\frac{1}{2}\sum_{\tau\in[k]}\left|f_2\left(\nabla f_1(\mathcal{T}_{\tau,\hat{i}});\boldsymbol{\theta}_2\right)-\left(r_\tau-f_1(\boldsymbol{\chi}_\tau;\boldsymbol{\theta}_1^{(\tau-1)})\right)\right|^2\leq\xi_1$. Finally, applying the union bound would give the aforementioned results.

$\square$

Here, analogous to the trained parameters, we consider the shadow parameters as $\{\boldsymbol{\theta}_1^{(k),*},\boldsymbol{\theta}_2^{(k),*}\}\sim\{\widetilde{\boldsymbol{\theta}}_1^{(\tau),*},\widetilde{\boldsymbol{\theta}}_2^{(\tau),*}\}_{\tau\in[k]}$. Similarly, each pair $\{\widetilde{\boldsymbol{\theta}}_1^{(\tau),*},\widetilde{\boldsymbol{\theta}}_2^{(\tau),*}\}$ is separately trained on past received rewards of the optimal arm(s) $\{r_{\tau',i^*}\}_{\tau'\in[\tau],\mathcal{T}_{\tau',i^*}\in\Omega_k^*}$ and past exploration scores of the optimal arm(s) $\{e_{\tau',i^*}\}_{\tau'\in[\tau],\mathcal{T}_{\tau',i^*}\in\Omega_k^*}$ with $J_{\boldsymbol{\theta}}$-iteration GD, starting from the random initialization $\{\boldsymbol{\theta}_1^{(0)},\boldsymbol{\theta}_2^{(0)}\}$.

Corollary D.2. *For the constants $0<\rho\leq\mathcal{O}(1/L)$ and $\xi_1\in(0,1)$, given the past records $\mathcal{P}_{k-1}$, we suppose m,η_1,J satisfy the conditions in **Theorem** 5.2, and randomly draw the parameters $\{\boldsymbol{\theta}_1^{(k),*},\boldsymbol{\theta}_2^{(k),*}\}\sim\{\widetilde{\boldsymbol{\theta}}_1^{(\tau),*},\widetilde{\boldsymbol{\theta}}_2^{(\tau),*}\}_{\tau\in[k]}$. For the optimal arm $\mathcal{T}_{k,i^*}\in\Omega_{task}^k$, consider its union set with the the collection of past optimal arms $\mathcal{P}_{k-1}^*\cup\{\mathcal{T}_{k,i^*},r_{k,i^*}\}$ are generated by a fixed policy when witness the candidate arms $\{\Omega_{task}^{(\tau)}\}_{\tau\in[k]}$, with $\mathcal{P}_{k-1}^*$ being the collection chosen by this policy. Then, with probability at least $1-\delta$, we have*

$$\mathbb{E}_{\mathcal{T}_{k,i}\sim\mathcal{P}(\mathcal{T})}\left[|f_2\left(\phi(\frac{[\nabla_{\boldsymbol{\theta}}f_1(\boldsymbol{\chi}_k^{s,*});\ \nabla_{\boldsymbol{\theta}}f_1(\boldsymbol{\chi}_k^{q,*})]}{c_g'L},\boldsymbol{\chi}_k^{q,*});\boldsymbol{\theta}_2^{(k-1),*}\right)-\left(r_\tau-f_1(\boldsymbol{\chi}_k^{q,*};\boldsymbol{\theta}_1^{(k-1),*})\right)|\ \middle|\Omega_{task}^{(k)},\mathcal{P}_{k-1}^*\right]$$
$$\leq\frac{1}{\sqrt{k}}\cdot\left(\sqrt{2\xi_1}+\frac{3L}{\sqrt{2}}+(1+\gamma_1)\sqrt{2\log(\frac{k}{\delta})}\right)+\Gamma_k$$

where r_{τ,i^*} is the corresponding reward generated by the mapping function given an arm $\boldsymbol{\chi}_{\tau,i^*}$, and

$$\Gamma_k=\left(1+\mathcal{O}(\frac{kL^3\log^{5/6}(m)}{\rho^{1/3}m^{1/6}})\right)\cdot\mathcal{O}(\frac{k^4L}{\rho\sqrt{m}}\log(m))+\mathcal{O}\left(\frac{k^5L^2\log^{11/6}(m)}{\rho^{4/3}m^{1/6}}\right).$$

Proof. This corollary is the direct application of Lemma D.1 by following a similar proof procedure. First, suppose the shadow models $f_1(\cdot;\boldsymbol{\theta}_2),f_2(\cdot;\boldsymbol{\theta}_2)$ are trained on the alternative trajectory $\mathcal{P}_{k-1}^*$. Analogous to the proof of Lemma D.1, we can define the following martingale difference sequence with regard to the previous records $\mathcal{P}_{k-1}^*$ up to round $\tau\in[t]$ as

$$\begin{aligned}V_\tau^{(1),*}=\mathbb{E}\left[\left|f_2\left(\nabla f_1(\mathcal{T}_{\tau,i^*});\boldsymbol{\theta}_2^{(\tau-1),*}\right)-\left(r_\tau^*-f_1(\boldsymbol{\chi}_\tau^*;\boldsymbol{\theta}_1^{(\tau-1),*})\right)\right|\right]\\-\left|f_2\left(\nabla f_1(\mathcal{T}_{\tau,i^*});\boldsymbol{\theta}_2^{(\tau-1),*}\right)-\left(r_\tau^*-f_1(\boldsymbol{\chi}_\tau^*;\boldsymbol{\theta}_1^{(\tau-1),*})\right)\right|.\end{aligned}$$

Since the records in set $\mathcal{P}_{k-1}^*$ are sharing the same reward mapping function, we have the expectation

$$\begin{aligned}\mathbb{E}[V_\tau^{(1),*}|F_\tau^*]=&\mathbb{E}\left[\left|f_2\left(\nabla f_1(\mathcal{T}_{\tau,i^*});\boldsymbol{\theta}_2^{(\tau-1),*}\right)-\left(r_\tau^*-f_1(\boldsymbol{\chi}_\tau^*;\boldsymbol{\theta}_1^{(\tau-1),*})\right)\right|\right]\\&-\mathbb{E}\left[\left|f_2\left(\nabla f_1(\mathcal{T}_{\tau,i^*});\boldsymbol{\theta}_2^{(\tau-1),*}\right)-\left(r_\tau^*-f_1(\boldsymbol{\chi}_\tau^*;\boldsymbol{\theta}_1^{(\tau-1),*})\right)\right|\middle|F_\tau^*\right]=0.\end{aligned}$$

where F_τ^* denotes the filtration given the past records $\mathcal{P}_{k-1}^*$. The mean value of $V_\tau^{(1),*}$ across different time steps will be

$$\begin{aligned}\frac{1}{k}\sum_{\tau\in[k]}V_\tau^{(1),*}=&\frac{1}{k}\sum_{\tau\in[k]}\mathbb{E}\left[\left|f_2\left(\nabla f_1(\mathcal{T}_{\tau,i^*});\boldsymbol{\theta}_2^{(\tau-1),*}\right)-\left(r_\tau^*-f_1(\boldsymbol{\chi}_\tau^*;\boldsymbol{\theta}_1^{(\tau-1),*})\right)\right|\right]\\&-\frac{1}{k}\sum_{\tau\in[k]}\left|f_2\left(\nabla f_1(\mathcal{T}_{\tau,i^*});\boldsymbol{\theta}_2^{(\tau-1),*}\right)-\left(r_\tau^*-f_1(\boldsymbol{\chi}_\tau^*;\boldsymbol{\theta}_1^{(\tau-1),*})\right)\right|.\end{aligned}$$

with the expectation of zero. Afterwards, applying the Azuma-Hoeffding inequality, with a constant $\delta \in (0, 1)$, it leads to

$$\mathbb{P}\bigg[\frac{1}{k}\sum_{\tau \in [k]} V_{\tau}^{(1),*} - \frac{1}{k}\sum_{\tau \in [k]} \mathbb{E}[V_{\tau}^{(1),*}] \geq (1 + 2\gamma_1)\sqrt{\frac{2\log(1/\delta)}{k}}\bigg] \leq \delta$$

To bound the output difference between the shadow model $f_1(\cdot; \boldsymbol{\theta}_1^{(k-1),*}), f_2(\cdot; \boldsymbol{\theta}_2^{(k-1),*})$ and the model we trained based on received records $f_1(\cdot; \boldsymbol{\theta}_1^{(k-1)}), f_2(\cdot; \boldsymbol{\theta}_2^{(k-1)})$, we apply the conclusion from **Lemma** D.11, which leads to that given arbitrary input vectors $\boldsymbol{x}, \boldsymbol{x}'$, we have

$$\begin{aligned}&|f_1(\boldsymbol{x}; \boldsymbol{\theta}_1^{(k-1),*}) - f_1(\boldsymbol{x}; \boldsymbol{\theta}_1^{(k-1)})|, \ |f_2(\boldsymbol{x}'; \boldsymbol{\theta}_2^{(k-1),*}) - f_2(\boldsymbol{x}'; \boldsymbol{\theta}_2^{(k-1)})| \leq \\ &\qquad \bigg(1 + \mathcal{O}(\frac{kL^3\log^{5/6}(m)}{\rho^{1/3}m^{1/6}})\bigg) \cdot \mathcal{O}(\frac{k^3 L}{\rho\sqrt{m}}\log(m)) + \mathcal{O}\bigg(\frac{k^4 L^2 \log^{11/6}(m)}{\rho^{4/3}m^{1/6}}\bigg).\end{aligned}$$

Finally, combining all the results will finish the proof.

□

We will also be able to have the performance guarantee under the batch settings. Recall that given a batch of chosen tasks $\Omega_k \subset \Omega_{\text{task}}^{(k)}, k \in [K]$, we have the meta-parameters adapted to this batch of tasks being $\boldsymbol{\Theta}^{(K-1)}[\Omega_K]$, which we consider as the input for the $f_1(\cdot; \boldsymbol{\theta}_1)$ model, where the tasks within each collection are sampled from the task distribution. Thus, chosen task batches from different iterations are also independent from each other. Intuitively, we can also define the corresponding reward for arm batch Ω_k as $r_k = h(\boldsymbol{\Theta}^{(k-1)}[\Omega_k])$. Then, we bound the batch settings with the following lemma and corollary.

Lemma D.3. *For the constants $c_g' > 0$, $\rho \in (0, \mathcal{O}(\frac{1}{L}))$ and $\xi_1 \in (0, 1)$, given the past records $\mathcal{P}_{k-1}$, we suppose m, η_1, η_2 satisfy the conditions in **Theorem** 5.2, and randomly draw the parameter $\{\boldsymbol{\theta}_1^{(k)}, \boldsymbol{\theta}_2^{(k)}\} \sim \{\widetilde{\boldsymbol{\theta}}_1^{(\tau)}, \widetilde{\boldsymbol{\theta}}_2^{(\tau)}\}_{\tau \in [k]}$. Consider the past records $\mathcal{P}_{k-1}$ up to round k are generated by a fixed policy when witness the candidate arms $\{\Omega_{task}^{(\tau)}\}_{\tau \in [k]}$. Then, with probability at least $1 - \delta$ given the pair of chosen arm batch and the reward (Ω_k, r_k) in round k, we have*

$$\begin{aligned}&\mathbb{E}_{\mathcal{T}_{k,i} \sim \mathcal{P}(\mathcal{T})}\bigg[|f_2\bigg(\phi(\frac{[\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_k^s); \ \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_k^q)]}{c_g' L}, \boldsymbol{\chi}_k^q); \boldsymbol{\theta}_2^{(k-1)}\bigg) - \bigg(r_\tau - f_1(\boldsymbol{\chi}_k^q; \boldsymbol{\theta}_1^{(k-1)})\bigg)| \ |\Omega_{task}^{(k)}, \mathcal{P}_{k-1}\bigg] \\ &\qquad \leq \frac{1}{\sqrt{k}} \cdot \bigg(\sqrt{2\xi_1} + \frac{3L}{\sqrt{2}} + (1 + 2\gamma_1)\sqrt{2\log(\frac{k}{\delta})}\bigg)\end{aligned}$$

where

$$\gamma_1 = 2 + \mathcal{O}\left(\frac{k^3 L}{\rho\sqrt{m}}\log m\right) + \mathcal{O}\left(\frac{L^2 k^4}{\rho^{4/3}m^{1/6}}\log^{11/6}(m)\right).$$

Proof. The proof of this lemma is analogous to the proof of **Lemma** D.1. First, we can derive the output upper bound

$$\begin{aligned}&\bigg|f_2\bigg(\phi(\frac{[\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_k^s); \ \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_k^q)]}{c_g' L}, \boldsymbol{\chi}_k^q); \boldsymbol{\theta}_2^{(k-1)}\bigg) - \bigg(r_k - f_1(\boldsymbol{\chi}_k^q; \boldsymbol{\theta}_1^{(k-1)})\bigg)\bigg| \\ &\qquad \leq \bigg|f_2\bigg(\phi(\frac{[\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_k^s); \ \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_k^q)]}{c_g' L}, \boldsymbol{\chi}_k^q); \boldsymbol{\theta}_2^{(k-1)}\bigg)\bigg| + \bigg|f_1(\boldsymbol{\chi}_k^q; \boldsymbol{\theta}_1^{(k-1)})\bigg| + 1 \\ &\qquad \leq 1 + 2\gamma_1\end{aligned}$$

by triangle inequality and applying the generalization result of FC networks (**Lemma** D.5) on $f_1(\cdot; \boldsymbol{\theta}_1), f_2(\cdot; \boldsymbol{\theta}_2)$, where $c_g' > 0$ is a positive number to scale the concatenated gradient vector.

For the brevity of notation, we use $\nabla f_1(\Omega_k)$ to denote $\phi(\frac{[\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_k^s); \ \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_k^q)]}{c_g' L}, \boldsymbol{\chi}_k^q)$ and apply $(\boldsymbol{\chi}_k, r_k)$ as $(\boldsymbol{\chi}_k^q, r_k)$ for the following proof. Define the difference sequence as

$$\begin{aligned}V_\tau^{(2)} = \mathbb{E}\bigg[\bigg|f_2\bigg(\nabla f_1(\Omega_\tau); \boldsymbol{\theta}_2^{(\tau-1)}\bigg) - \bigg(r_\tau - f_1(\boldsymbol{\chi}_\tau; \boldsymbol{\theta}_1^{(\tau-1)})\bigg)\bigg|\bigg] \\ - \bigg|f_2\bigg(\nabla f_1(\Omega_\tau); \boldsymbol{\theta}_2^{(\tau-1)}\bigg) - \bigg(r_\tau - f_1(\boldsymbol{\chi}_\tau; \boldsymbol{\theta}_1^{(\tau-1)})\bigg)\bigg|.\end{aligned}$$

Since the past rewards and the received arm batch-reward pairs $(\boldsymbol{\chi}_\tau, r_\tau)$ are generated by the same reward mapping function, we have the expectation

$$\begin{aligned}\mathbb{E}[V_\tau^{(2)}|F_\tau] =& \mathbb{E}\left[\left|f_2\left(\nabla f_1(\Omega_\tau);\boldsymbol{\theta}_2^{(\tau-1)}\right) - \left(r_\tau - f_1(\boldsymbol{\chi}_\tau;\boldsymbol{\theta}_1^{(\tau-1)})\right)\right|\right] \\ &- \mathbb{E}\left[\left|f_2\left(\nabla f_1(\Omega_\tau);\boldsymbol{\theta}_2^{(\tau-1)}\right) - \left(r_\tau - f_1(\boldsymbol{\chi}_\tau;\boldsymbol{\theta}_1^{(\tau-1)})\right)\right| \Big| F_\tau\right] = 0.\end{aligned}$$

where F_τ denotes the filtration given the past records $\mathcal{P}_\tau$, up to round $\tau \in [k]$. This also gives the fact that $V_\tau^{(2)}$ is a martingale difference sequence. Then, after applying the martingale difference sequence over $[k]$, we have

$$\begin{aligned}\frac{1}{k}\sum_{\tau\in[k]} V_\tau^{(2)} =& \frac{1}{k}\sum_{\tau\in[k]} \mathbb{E}\left[\left|f_2\left(\nabla f_1(\Omega_\tau);\boldsymbol{\theta}_2^{(\tau-1)}\right) - \left(r_\tau - f_1(\boldsymbol{\chi}_\tau;\boldsymbol{\theta}_1^{(\tau-1)})\right)\right|\right] \\ &- \frac{1}{k}\sum_{\tau\in[k]} \left|f_2\left(\nabla f_1(\Omega_\tau);\boldsymbol{\theta}_2^{(\tau-1)}\right) - \left(r_\tau - f_1(\boldsymbol{\chi}_\tau;\boldsymbol{\theta}_1^{(\tau-1)})\right)\right|.\end{aligned}$$

By the Azuma-Hoeffding inequality, it leads to $\mathbb{P}\left[\frac{1}{k}\sum_{\tau\in[k]} V_\tau^{(2)} - \frac{1}{k}\sum_{\tau\in[k]} \mathbb{E}[V_\tau^{(2)}] \geq (1 + 2\gamma_1)\sqrt{\frac{2\log(1/\delta)}{k}}\right] \leq \delta$. As we have discussed, the tasks within each collection are sampled from the task distribution, which makes chosen task batches from different iterations $\Omega_k, k \in [K]$ are also independent from each other. Since the expectation of $V_\tau^{(2)}$ is zero, with the probability at least $1-\delta$ and an existing set of parameters $\boldsymbol{\theta}_2$ s.t. $\|\boldsymbol{\theta}_2 - \boldsymbol{\theta}_2^{(0)}\| \leq \mathcal{O}\left(\frac{k^3}{\rho\sqrt{m}}\log m\right)$, the above inequality implies

$$\begin{aligned}&\frac{1}{k}\sum_{\tau\in[k]} V_\tau^{(1)} \leq (1+2\gamma_1)\sqrt{\frac{2\log(1/\delta)}{k}} \implies \\ &\quad \mathbb{E}_{\mathcal{T}_{k,i}\sim\mathcal{P}(\mathcal{T})}\mathbb{E}_{\{\boldsymbol{\theta}_1^{(k-1)},\boldsymbol{\theta}_2^{(k-1)}\}}\left[\left|f_2\left(\nabla f_1(\Omega);\boldsymbol{\theta}_2^{(k-1)}\right) - \left(r - f_1(\boldsymbol{\chi};\boldsymbol{\theta}_1^{(k-1)})\right)\right|\right] \\ &\quad = \frac{1}{k}\sum_{\tau\in[k]} \mathbb{E}\left[\left|f_2\left(\nabla f_1(\Omega_\tau);\boldsymbol{\theta}_2^{(\tau-1)}\right) - \left(r_k - f_1(\boldsymbol{\chi}_\tau;\boldsymbol{\theta}_1^{(\tau-1)})\right)\right|\right] \\ &\quad \leq \frac{1}{k}\sum_{\tau\in[k]} \left|f_2\left(\nabla f_1(\Omega_\tau);\boldsymbol{\theta}_2^{(\tau-1)}\right) - \left(r_\tau - f_1(\boldsymbol{\chi}_\tau;\boldsymbol{\theta}_1^{(\tau-1)})\right)\right| + (1+2\gamma_1)\sqrt{\frac{2\log(1/\delta)}{k}} \\ &\quad \underset{\text{(i)}}{\leq} \frac{1}{k}\sum_{\tau\in[k]} \left|f_2\left(\nabla f_1(\Omega_\tau);\boldsymbol{\theta}_2\right) - \left(r_\tau - f_1(\boldsymbol{\chi}_\tau;\boldsymbol{\theta}_1^{(\tau-1)})\right)\right| + \frac{3L}{\sqrt{2k}} + (1+2\gamma_1)\sqrt{\frac{2\log(1/\delta)}{k}} \\ &\quad \leq \frac{1}{\sqrt{k}}\sqrt{\sum_{\tau\in[k]} \left|f_2\left(\nabla f_1(\Omega_\tau);\boldsymbol{\theta}_2\right) - \left(r_\tau - f_1(\boldsymbol{\chi}_\tau;\boldsymbol{\theta}_1^{(\tau-1)})\right)\right|^2} + \frac{3L}{\sqrt{2k}} + (1+2\gamma_1)\sqrt{\frac{2\log(1/\delta)}{k}} \\ &\quad \underset{\text{(ii)}}{\leq} \sqrt{\frac{2\xi_2}{k}} + \frac{3L}{\sqrt{2k}} + (1+2\gamma_1)\sqrt{\frac{2\log(1/\delta)}{k}}.\end{aligned}$$

where the first equality is due to the sampling of candidate tasks and the model parameters. Here, the upper bound (i) is derived by applying the conclusions of **Lemma** D.6 and **Lemma** D.10, and the inequality (ii) is derived by adopting **Lemma** D.6 while defining the empirical loss to be $\frac{1}{2}\sum_{\tau\in[k]}\left|f_2\left(\nabla f_1(\Omega_\tau);\boldsymbol{\theta}_2\right) - \left(r_\tau - f_1(\boldsymbol{\chi}_\tau;\boldsymbol{\theta}_1^{(\tau-1)})\right)\right|^2 \leq \xi_2$. Finally, applying the union bound would give the aforementioned results.

□

Analogously, we consider the shadow parameters as $\{\boldsymbol{\theta}_1^{(k),*}, \boldsymbol{\theta}_2^{(k),*}\} \sim \{\widetilde{\boldsymbol{\theta}}_1^{(\tau),*}, \widetilde{\boldsymbol{\theta}}_2^{(\tau),*}\}_{\tau \in [k]}$ where each pair $\{\widetilde{\boldsymbol{\theta}}_1^{(\tau),*}, \widetilde{\boldsymbol{\theta}}_2^{(\tau),*}\}$ is separately trained on past received rewards of the optimal arm(s) $\{r_{\tau', i^*}\}_{\tau' \in [\tau], \mathcal{T}_{\tau', i^*} \in \Omega_k^*}$ and past exploration scores of the optimal arm(s) $\{e_{\tau', i^*}\}_{\tau' \in [\tau], \mathcal{T}_{\tau', i^*} \in \Omega_k^*}$ with $J_{\boldsymbol{\theta}}$-iteration GD starting from the random initialization $\{\boldsymbol{\theta}_1^{(0)}, \boldsymbol{\theta}_2^{(0)}\}$.

Corollary D.4. *For the constants $\rho \in (0, \mathcal{O}(\frac{1}{L}))$ and $\xi_1 \in (0,1)$, given the past records $\mathcal{P}_{k-1}$, we suppose m, η_1, J satisfy the conditions in **Theorem** 5.2, and randomly draw the parameters $\{\boldsymbol{\theta}_1^{(k),*}, \boldsymbol{\theta}_2^{(k),*}\} \sim \{\widetilde{\boldsymbol{\theta}}_1^{(\tau),*}, \widetilde{\boldsymbol{\theta}}_2^{(\tau),*}\}_{\tau \in [k]}$. For the optimal arm batch $\Omega_k^* \subset \Omega_{task}^k$, consider its union set with the the collection of past optimal arms $\mathcal{P}_{k-1}^* \cup \{\Omega_k^*, r_k^*\}$ are generated by a fixed policy when witness the candidate arms $\{\Omega_{task}^{(\tau)}\}_{\tau \in [k]}$, with $\mathcal{P}_{k-1}^*$ being the collection chosen by this policy. Then, with probability at least $1-\delta$, we have*

$$\mathbb{E}_{\mathcal{T}_{k,i} \sim \mathcal{P}(\mathcal{T})}\Big[\big| f_2\Big(\phi(\frac{[\nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_k^{s,*});\ \nabla_{\boldsymbol{\theta}} f_1(\boldsymbol{\chi}_k^{q,*})]}{c_g' L}, \boldsymbol{\chi}_k^{q,*}); \boldsymbol{\theta}_2^{(k-1),*}\Big) - \Big(r_\tau - f_1(\boldsymbol{\chi}_k^{q,*}; \boldsymbol{\theta}_1^{(k-1),*})\Big)\big| \,\big|\Omega_{task}^{(k)}, \mathcal{P}_{k-1}^* \Big]$$
$$\leq \frac{1}{\sqrt{k}} \cdot \Big(\sqrt{2\xi_2} + \frac{3L}{\sqrt{2}} + (1+\gamma_1)\sqrt{2\log(\frac{k}{\delta})} \Big) + \Gamma_k$$

where r_{τ, i^*} is the corresponding reward generated by the mapping function given an arm $\boldsymbol{\chi}_{\tau, i^*}$, and

$$\Gamma_k = \Big(1 + \mathcal{O}(\frac{kL^3 \log^{5/6}(m)}{\rho^{1/3} m^{1/6}})\Big) \cdot \mathcal{O}(\frac{k^4 L}{\rho\sqrt{m}} \log(m)) + \mathcal{O}\Big(\frac{k^5 L^2 \log^{11/6}(m)}{\rho^{4/3} m^{1/6}}\Big).$$

This corollary is a directly application of **Lemma** D.3 and can be obtained with a similar proof as in **Corollary** D.2.

D.4 Ancillary Lemmas

Applying $\mathcal{P}_{k-1}$ as the training data, we have the following properties for the over-parameterized FC network $f(\cdot; \boldsymbol{\theta})$ after GD.

Lemma D.5. *For the constants $\rho \in (0, \mathcal{O}(\frac{1}{L}))$ and $\xi_1 \in (0,1)$, given the past records $\mathcal{P}_{k-1}$ up to time step k, we suppose m, η_1, J_1 satisfy the conditions in **Theorem** 5.2. Then, with probability at least $1-\delta$, given a sample-label pair $(\boldsymbol{x}, r)$, we have*

$$|f(\boldsymbol{x}; \boldsymbol{\theta}^{(k)})| \leq \gamma_1 = 2 + \mathcal{O}\left(\frac{k^3 L}{\rho\sqrt{m}} \log m\right) + \mathcal{O}\left(\frac{L^2 k^4}{\rho^{4/3} m^{1/6}} \log^{11/6}(m)\right).$$

Proof. The LHS of the inequality could be written as

$$\begin{aligned} |f(\boldsymbol{x}; \boldsymbol{\theta})| \leq & |f(\boldsymbol{x}; \boldsymbol{\theta}) - f(\boldsymbol{x}; \boldsymbol{\theta}^{(0)}) - \langle \nabla_{\boldsymbol{\theta}^{(0)}} f(\boldsymbol{x}; \boldsymbol{\theta}^{(0)}), \boldsymbol{\theta} - \boldsymbol{\theta}^{(0)} \rangle| \\ & + |f(\boldsymbol{x}; \boldsymbol{\theta}^{(0)}) + \langle \nabla_{\boldsymbol{\theta}^{(0)}} f(\boldsymbol{x}; \boldsymbol{\theta}^{(0)}), \boldsymbol{\theta} - \boldsymbol{\theta}^{(0)} \rangle|. \end{aligned}$$

Here, we could bound the first term on the RHS with **Lemma** D.7. Applying **Lemma** D.8 on the second term, and recalling $\|\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}\|_2 \leq \omega$, would give

$$\begin{aligned} |f(\boldsymbol{x}; \boldsymbol{\theta})| \leq & \ 2 + \|\nabla_{\boldsymbol{\theta}^{(0)}} f(\boldsymbol{x}; \boldsymbol{\theta}^{(0)})\|_2 \|\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}\|_2 + \\ & \mathcal{O}(\omega^{1/3} L^2 \sqrt{m \log(m)}) \cdot \|\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}\|_2 \\ \leq & \ 2 + \mathcal{O}(L) \cdot \|\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}\|_2 + \mathcal{O}(L^2 \sqrt{m \log(m)}) (\|\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}\|_2)^{\frac{4}{3}}. \end{aligned}$$

Then, applying the conclusion of **Lemma** D.6 would lead to

$$\begin{aligned} |f(\boldsymbol{x}; \boldsymbol{\theta})| & \leq 2 + \mathcal{O}(L) \cdot \mathcal{O}\left(\frac{k^3}{\rho\sqrt{m}} \log m\right) + \mathcal{O}(L^2 \sqrt{m \log(m)}) \left(\mathcal{O}(\frac{k^3}{\rho\sqrt{m}} \log m)\right)^{\frac{4}{3}} \\ & = 2 + \mathcal{O}\left(\frac{k^3 L}{\rho\sqrt{m}} \log m\right) + \mathcal{O}\left(\frac{L^2 k^4}{\rho^{4/3} m^{1/6}} \log^{11/6}(m)\right) = \gamma_1. \end{aligned}$$

□

Lemma D.6 (Theorem 1 from [3])**.** *For any $0 < \xi_1 \leq 1$, $0 < \rho \leq \mathcal{O}(\frac{1}{L})$. Given the past records $\mathcal{P}_{k-1}$, suppose m, η_1, J satisfy the conditions in* ***Theorem*** *5.2, then with probability at least $1-\delta$, we could have*

1. *$\mathcal{L}(\boldsymbol{\theta}) \leq \xi_1$ after J iterations of GD.*
2. *For any $j \in [J]$, $\|\boldsymbol{\theta}^{(j)} - \boldsymbol{\theta}^{(0)}\| \leq \mathcal{O}\left(\frac{k^3}{\rho\sqrt{m}} \log m\right)$.*

In particular, **Lemma** D.6 above provides the convergence guarantee for $f(\cdot;\boldsymbol{\theta})$ after certain rounds of GD training on the past records $\mathcal{P}_{k-1}$.

Lemma D.7 (Lemma 4.1 in [10])**.** *Assume a constant ω such that $\mathcal{O}(m^{-3/2}L^{-3/2}[\log(TnL^2/\delta)]^{3/2}) \leq \omega \leq \mathcal{O}(L^{-6}[\log m]^{-3/2})$ and n training samples. With randomly initialized $\boldsymbol{\theta}^{(0)}$, for parameters $\boldsymbol{\theta}, \boldsymbol{\theta}'$ satisfying $\|\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}\|, \|\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}\| \leq \omega$, we have*

$$|f(\boldsymbol{x};\boldsymbol{\theta}) - f(\boldsymbol{x};\boldsymbol{\theta}') - \langle \nabla_{\boldsymbol{\theta}'} f(\boldsymbol{x};\boldsymbol{\theta}'), \boldsymbol{\theta} - \boldsymbol{\theta}' \rangle| \leq \mathcal{O}(\omega^{1/3} L^2 \sqrt{m \log(m)}) \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|$$

with the probability at least $1-\delta$.

Lemma D.8. *Assume m, η_1, J satisfy the conditions in* ***Theorem*** *5.2 and $\boldsymbol{\theta}^{(0)}$ is randomly initialized. Then, with probability at least $1-\delta$ and given an arm $\|\boldsymbol{x}\|_2 = 1$, we have*

1. *$|f(\boldsymbol{x};\boldsymbol{\theta}^{(0)})| \leq 2$,*
2. *$\|\nabla_{\boldsymbol{\theta}^{(0)}} f(\boldsymbol{x};\boldsymbol{\theta}^{(0)})\|_2 \leq \mathcal{O}(L)$.*

Proof. The conclusion (1) is a direct application of Lemma 7.1 in [3]. Suppose the parameters of the L-layer FC network are $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_L\}$. For conclusion (2), applying Lemma 7.3 in [3], for each layer $\boldsymbol{\theta}_l \in \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_L\}$, we have

$$\|\nabla_{\boldsymbol{\theta}_l} f(\boldsymbol{x};\boldsymbol{\theta}^{(0)})\|_2 = \|(\boldsymbol{\theta}_L \boldsymbol{D}_{L-1} \cdots \boldsymbol{D}_{l+1}\boldsymbol{\theta}_{l+1}) \cdot (\boldsymbol{D}_{l+1}\boldsymbol{\theta}_{l+1} \cdots \boldsymbol{D}_1 \boldsymbol{\theta}_1) \cdot \boldsymbol{x}^\intercal\|_2 = \mathcal{O}(\sqrt{L}).$$

where $\boldsymbol{D}$ is the diagonal matrix corresponding to the activation function. Then, we could have the conclusion that

$$\|\nabla_{\boldsymbol{\theta}^{(0)}} f(\boldsymbol{x};\boldsymbol{\theta}^{(0)})\|_2 = \sqrt{\sum_{l\in[L]} \|\nabla_{\boldsymbol{\theta}_l} f(\boldsymbol{x};\boldsymbol{\theta}^{(0)})\|_2^2} = \mathcal{O}(L).$$

□

Lemma D.9 (Theorem 5 in [3])**.** *Assume m, η_1, J satisfy the conditions in* ***Theorem*** *5.2 and $\boldsymbol{\theta}^{(0)}$ being randomly initialized. Then, with probability at least $1-\delta$, and for all parameter $\boldsymbol{\theta}$ such that $\|\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}\|_2 \leq \omega$, we have*

$$\|\nabla_{\boldsymbol{\theta}} f(\boldsymbol{x};\boldsymbol{\theta}) - \nabla_{\boldsymbol{\theta}^{(0)}} f(\boldsymbol{x};\boldsymbol{\theta}^{(0)})\|_2 \leq \mathcal{O}(\omega^{1/3} L^3 \sqrt{\log(m)})$$

Lemma D.10. *Assume m, η_1 satisfy the condition in* ***Theorem*** *5.2. For notation brevity, suppose the training sample-label pairs are $\{\boldsymbol{x}_\tau, r_\tau\}_{\tau\in[k]}$. With the probability at least $1-\delta$, we have*

$$\sum_{\tau\in[k]} |f(\boldsymbol{x}_\tau;\boldsymbol{\theta}^{(\tau)}) - r_\tau| \leq \sum_{\tau\in[k]} |f(\boldsymbol{x}_\tau;\boldsymbol{\theta}^{(k)}) - r_\tau| + \frac{3L\sqrt{2k}}{2}$$

Proof. With the notation from Lemma 4.3 in [10], set $R = \frac{k^3 \log(m)}{\delta}$, $\nu = R^2$, and $\epsilon = \frac{LR}{\sqrt{2\nu k}}$. Then, considering the loss function to be $\mathcal{L}(\boldsymbol{\theta}) := \sum_{\tau\in[k]} |f(\boldsymbol{x}_\tau;\boldsymbol{\theta}) - r_\tau|$ would complete the proof. □

Lemma D.11. *Consider a randomly initialized L-layer ReLU fully-connected network $f(\cdot;\boldsymbol{\theta}_0)$. For any $0 < \xi_2 \leq 1$, $0 < \rho \leq \mathcal{O}(\frac{1}{L})$. Let there be two sets of training samples $\mathcal{P}_k, \mathcal{P}'_k$ with the unit-length and the ρ-separateness assumption, and let $\boldsymbol{\theta}$ be the trained parameter on $\mathcal{P}_k$ while $\boldsymbol{\theta}'$ is the trained parameter on $\mathcal{P}'_k$. Suppose the conditions in* ***Theorem*** *5.2 are satisfied. Then, with probability at least $1-\delta$, we have*

$$|f(\boldsymbol{x};\boldsymbol{\theta}) - f(\boldsymbol{x};\boldsymbol{\theta}')| \leq \left(1 + \mathcal{O}(\frac{kL^3 \log^{5/6}(m)}{\rho^{1/3} m^{1/6}})\right) \cdot \mathcal{O}(\frac{k^3 L}{\rho\sqrt{m}} \log(m)) + \mathcal{O}\left(\frac{k^4 L^2 \log^{11/6}(m)}{\rho^{4/3} m^{1/6}}\right)$$

when given a new sample $\boldsymbol{x} \in \mathbb{R}^d$.

Proof. First, based on the conclusion from Theorem 1 from [3] and regarding the t samples, the trained the parameters satisfy $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2, \|\boldsymbol{\theta}' - \boldsymbol{\theta}_0\|_2 \leq \mathcal{O}(\frac{k^3}{\rho\sqrt{m}} \log(m)) = \omega$ where $\boldsymbol{\theta}_0$ is the randomly initialized parameter. Then, we could have

$$\begin{aligned}\|\nabla_{\boldsymbol{\theta}} f(\boldsymbol{x}; \boldsymbol{\theta})\|_2 &\leq \|\nabla_{\boldsymbol{\theta}_0} f(\boldsymbol{x}; \boldsymbol{\theta}_0)\|_2 + \|\nabla_{\boldsymbol{\theta}} f(\boldsymbol{x}; \boldsymbol{\theta}) - \nabla_{\boldsymbol{\theta}_0} f(\boldsymbol{x}; \boldsymbol{\theta}_0)\|_2 \\ &\leq \left(1 + \mathcal{O}(\frac{kL^3 \log^{5/6}(m)}{\rho^{1/3} m^{1/6}})\right) \cdot \mathcal{O}(L)\end{aligned}$$

w.r.t. the conclusion from Theorem 1 and Theorem 5 of [3]. Then, regarding the Lemma 4.1 from [10], we would have

$$|f(\boldsymbol{x}; \boldsymbol{\theta}) - f(\boldsymbol{x}; \boldsymbol{\theta}') - \langle \nabla_{\boldsymbol{\theta}'} f(\boldsymbol{x}; \boldsymbol{\theta}'), \boldsymbol{\theta} - \boldsymbol{\theta}' \rangle| \leq \mathcal{O}(\omega^{1/3} L^2 \sqrt{m \log(m)}) \cdot \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_2.$$

Therefore, the our target could be reformed as

$$\begin{aligned}|f(\boldsymbol{x}; \boldsymbol{\theta}) - f(\boldsymbol{x}; \boldsymbol{\theta}')| &\leq \|\nabla_{\boldsymbol{\theta}'} f(\boldsymbol{x}; \boldsymbol{\theta}')\|_2 \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_2 + \mathcal{O}(\omega^{1/3} L^2 \sqrt{m \log(m)}) \cdot \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_2 \\ &\leq \left(1 + \mathcal{O}(\frac{kL^3 \log^{5/6}(m)}{\rho^{1/3} m^{1/6}})\right) \cdot \mathcal{O}(L) \cdot \omega + \mathcal{O}(\omega^{4/3} L^2 \sqrt{m \log(m)})\end{aligned}$$

Substituting the ω with its value would complete the proof.

□

Corollary D.12. *Following a similar settings as in* ***Lemma*** *D.11, consider a randomly initialized L-layer fully-connected network $f(\cdot; \boldsymbol{\theta}_0)$ with Sigmoid activation. For any $0 < \xi_2 \leq 1$, $0 < \rho \leq \mathcal{O}(\frac{1}{L})$. Let there be two sets of training samples $\mathcal{P}_k, \mathcal{P}'_k$ with the unit-length and the ρ-separateness assumption, and let $\boldsymbol{\theta}$ be the trained parameter on $\mathcal{P}_k$ while $\boldsymbol{\theta}'$ is the trained parameter on $\mathcal{P}'_k$. Suppose the conditions in* ***Theorem*** *5.2 are satisfied. Then, with probability at least $1 - \delta$, we have*

$$|f(\boldsymbol{x}; \boldsymbol{\theta}) - f(\boldsymbol{x}; \boldsymbol{\theta}')| \leq \left(1 + \mathcal{O}(\frac{kL^3 \log^{5/6}(m)}{\rho^{1/3} m^{1/6}})\right) \cdot \mathcal{O}(\frac{k^3 L}{\rho\sqrt{m}} \log(m)) + \mathcal{O}\left(\frac{k^4 L^2 \log^{11/6}(m)}{\rho^{4/3} m^{1/6}}\right)$$

when given a new sample $\boldsymbol{x} \in \mathbb{R}^d$.

Proof. This corollary is an intuitive extension of **Lemma** D.11. Since the result from Theorem 1 of [3] also applies to Lipschitz-smooth (i.e., Sigmoid) activation functions, combining the proof of **Lemma** D.11 and the result from Lemma 7 in [46] will give the conclusion.

□

D.5 Regret Bound for Uniform Sampling

Lemma D.13 (Regret Bound for the Uniform Sampling Approach)**.** *When applying the uniform sampling as in most meta-learning frameworks, we denote the corresponding sampled task series as $\Omega_u(K)$. We will have $R_u(K) = \mathbb{E}_{\mathcal{T} \sim \mathcal{P}(\mathcal{T}), \boldsymbol{x} \sim \mathcal{D}_{\mathcal{T}}} \left[\mathcal{L}(\boldsymbol{x}; \mathcal{I}(\mathcal{T}, \boldsymbol{\Theta}_u^{(K)})) - \mathcal{L}(\boldsymbol{x}; \mathcal{I}(\mathcal{T}, \boldsymbol{\Theta}^{(K),*})) \right]$. where $\boldsymbol{\Theta}_u^{(K)}$ refer to the meta-parameters trained with uniform sampling. With $\|\boldsymbol{\Theta}_u^{(K)} - \boldsymbol{\Theta}^{(K),*}\|_2 \leq \omega$, we have the regret bound for the uniform sampling as*

$$\begin{aligned}R_u(K) &= \mathbb{E}_{\mathcal{T} \sim \mathcal{P}(\mathcal{T}), \boldsymbol{x} \sim \mathcal{D}_{\mathcal{T}}} \left[\mathcal{L}(\boldsymbol{x}; \mathcal{I}(\mathcal{T}, \boldsymbol{\Theta}_u^{(K)})) - \mathcal{L}(\boldsymbol{x}; \mathcal{I}(\mathcal{T}, \boldsymbol{\Theta}^{(K),*})) \right] \\ &\leq \sqrt{m_{\mathcal{F}} L_{\mathcal{F}}} \cdot \omega + \mathcal{O}(\omega^{4/3} L_{\mathcal{F}}^3 \sqrt{m_{\mathcal{F}} \log(m_{\mathcal{F}})}) + \mathcal{O}(\sqrt{\frac{L_{\mathcal{F}}}{m_{\mathcal{F}}}}) \\ &\leq \min\left\{ \mathcal{O}\left(K L_{\mathcal{F}} + \frac{K^{4/3} L_{\mathcal{F}}^{11/3} \sqrt{\log(m_{\mathcal{F}})}}{m_{\mathcal{F}}^{1/6}} + \sqrt{\frac{L_{\mathcal{F}}}{m_{\mathcal{F}}}} \right), 1 \right\}\end{aligned}$$

Proof. Here, for the simplicity of notation, we denote $\boldsymbol{\Theta} = \mathcal{I}(\mathcal{T}, \boldsymbol{\Theta})$, and neglect the expectation terms. Note that the difference between adapted meta-parameters and the original meta-parameters is small enough and can be well-bounded. We will then have

$$\begin{aligned}
R_u(K) &= \mathbb{E}_{\mathcal{T}\sim\mathcal{P}(\mathcal{T}),\boldsymbol{x}\sim\mathcal{D}_{\mathcal{T}}}\Big[\mathcal{L}(\boldsymbol{x};\mathcal{I}(\mathcal{T},\boldsymbol{\Theta}_u^{(K)})) - \mathcal{L}(\boldsymbol{x};\mathcal{I}(\mathcal{T},\boldsymbol{\Theta}^{(K),*}))\Big] \\
&= \widetilde{\mathcal{L}}(\widetilde{\boldsymbol{\Theta}}_u^{(K)}) - \widetilde{\mathcal{L}}(\widetilde{\boldsymbol{\Theta}}^{(K),*})
\end{aligned}$$

where the two sets of meta-parameters are trained with uniformly sampled tasks and the optimal tasks, and $\widetilde{\boldsymbol{\Theta}}$ is used to denote the adapted meta-parameters $\mathcal{I}(\mathcal{T},\boldsymbol{\Theta})$ for simplicity. With any convex loss function (e.g., L_2 loss or cross-entropy loss) under the over-parameterization settings, we will have the generalization loss being almost convex w.r.t. the meta-parameters as in **Lemma** D.14, which leads to

$$\begin{aligned}
&\widetilde{\mathcal{L}}(\widetilde{\boldsymbol{\Theta}}_u^{(K)}) - \widetilde{\mathcal{L}}(\widetilde{\boldsymbol{\Theta}}^{(K),*}) \leq \langle \nabla_{\widetilde{\boldsymbol{\Theta}}}\widetilde{\mathcal{L}}(\widetilde{\boldsymbol{\Theta}}_u^{(K)}), \widetilde{\boldsymbol{\Theta}}_u^{(K)} - \widetilde{\boldsymbol{\Theta}}^{(K),*}\rangle + \epsilon \\
&\quad \leq \|\nabla_{\widetilde{\boldsymbol{\Theta}}}\widetilde{\mathcal{L}}(\widetilde{\boldsymbol{\Theta}}_u^{(K)})\|_2\|\widetilde{\boldsymbol{\Theta}}_u^{(K)} - \widetilde{\boldsymbol{\Theta}}^{(K),*}\|_2 + \epsilon \\
&\quad \leq \|\nabla_{\widetilde{\boldsymbol{\Theta}}}\widetilde{\mathcal{L}}(\widetilde{\boldsymbol{\Theta}}_u^{(K)})\|_2\|\boldsymbol{\Theta}_u^{(K)} - \boldsymbol{\Theta}^{(K),*}\|_2 + \eta_1\cdot\mathcal{O}(\sqrt{m_{\mathcal{F}}L_{\mathcal{F}}}) + \epsilon \\
&\quad \overset{\text{(i)}}{\leq} \sqrt{m_{\mathcal{F}}L_{\mathcal{F}}}\cdot\omega + \mathcal{O}(\omega^{4/3}L_{\mathcal{F}}^3\sqrt{m_{\mathcal{F}}\log(m_{\mathcal{F}})}) + \mathcal{O}(\sqrt{\frac{L_{\mathcal{F}}}{m_{\mathcal{F}}}}) \\
&\quad \overset{\text{(ii)}}{\leq} \mathcal{O}\bigg(KL_{\mathcal{F}} + \frac{K^{4/3}L_{\mathcal{F}}^{11/3}\sqrt{\log(m_{\mathcal{F}})}}{m_{\mathcal{F}}^{1/6}} + \sqrt{\frac{L_{\mathcal{F}}}{m_{\mathcal{F}}}}\bigg) \\
&\overset{\text{(iii)}}{\Longrightarrow} \widetilde{\mathcal{L}}(\widetilde{\boldsymbol{\Theta}}_u^{(K)}) - \widetilde{\mathcal{L}}(\widetilde{\boldsymbol{\Theta}}^{(K),*}) \leq \min\bigg\{\mathcal{O}\bigg(KL_{\mathcal{F}} + \frac{K^{4/3}L_{\mathcal{F}}^{11/3}\sqrt{\log(m_{\mathcal{F}})}}{m_{\mathcal{F}}^{1/6}} + \sqrt{\frac{L_{\mathcal{F}}}{m_{\mathcal{F}}}}\bigg),\ 1\bigg\}
\end{aligned}$$

where $\epsilon = \mathcal{O}(\omega^{4/3}L_{\mathcal{F}}^3\sqrt{m_{\mathcal{F}}\log(m_{\mathcal{F}})}) > 0$, and $\|\boldsymbol{\Theta}_u^{(K),*} - \boldsymbol{\Theta}_u^{(K)}\|_2 \leq \omega$. Here, the first inequality is due to **Lemma** D.14 and the convexity of the loss function. The third inequality is due to the upper bound for meta-model gradients (**Lemma** D.15). The (i) is due to **Lemma** D.16 and sufficiently small learning rate $\eta_1 \leq \mathcal{O}(\frac{1}{m_{\mathcal{F}}})$. Based on **Lemma** D.15, we will have $\|\nabla_{\boldsymbol{\Theta}}\mathcal{L}(\boldsymbol{x};\boldsymbol{\Theta}_K^{(J),*})\|_2, \|\nabla_{\boldsymbol{\Theta}}\mathcal{L}(\boldsymbol{x};\boldsymbol{\Theta}_K^{(J)})\|_2 \leq \mathcal{O}(\sqrt{m_{\mathcal{F}}L_{\mathcal{F}}})$. Since we have $\eta_1, \eta_2 \leq \mathcal{O}(\frac{1}{m})$, starting from randomly initialized $\boldsymbol{\Theta}^{(0)}$, the parameter shift caused by GD can be upper bounded by $\|\boldsymbol{\Theta}_u^{(K),*} - \boldsymbol{\Theta}_u^{(K)}\|_2 \leq \omega = \mathcal{O}(K\cdot\sqrt{\frac{L_{\mathcal{F}}}{m_{\mathcal{F}}}})$. The implication (iii) is because the loss function $\mathcal{L}(\cdot;\cdot)$ has the value range [0, 1]. ☐

Here, we notice that the RHS of the regret bound in **Lemma** D.13 has two terms. Although the second term can be reduced to $\mathcal{O}(1)$ with sufficiently large meta-model width $m_{\mathcal{F}} > \mathcal{O}(\text{Poly}(K, L, \rho^{-1}))$, the first term tends to grow along with more iterations K and the larger meta-model width $m_{\mathcal{F}}$. The reason is that the radius for the parameter shift during meta-training ω can be as large as $\mathcal{O}(\frac{1}{\sqrt{m_{\mathcal{F}}}})$, which means that it cannot cancel out the effects of gradient norms, which have the order of $\mathcal{O}(\sqrt{m_{\mathcal{F}}})$. In this case, we will not able to include a $m_{\mathcal{F}}$ term to the denominator to scale down the regret with $m_{\mathcal{F}}$, and make the upper bound narrower than 1.

Lemma D.14. *Given an arbitrary sample $\boldsymbol{x}$ and its label, let $\widetilde{\mathcal{L}}(\boldsymbol{\Theta}) = \mathcal{L}(\boldsymbol{x};\boldsymbol{\Theta})$. Suppose $m_{\mathcal{F}}, \eta_1, \eta_2$ satisfy the conditions in Theorem 5.2. With probability at least $1 - \mathcal{O}(KL_{\mathcal{F}}^2)\cdot\exp[-\Omega(m_{\mathcal{F}}\omega^{2/3}L_{\mathcal{F}})]$ over randomness of $\boldsymbol{\Theta}^{(0)}$, for all $k \in [K]$, and $\boldsymbol{\Theta}, \boldsymbol{\Theta}'$ satisfying $\|\boldsymbol{\Theta} - \boldsymbol{\Theta}^{(0)}\|_2 \leq \omega$ and $\|\boldsymbol{\Theta}' - \boldsymbol{\Theta}^{(0)}\|_2 \leq \omega$ with $\omega \leq \mathcal{O}(L_{\mathcal{F}}^{-6}[\log m_{\mathcal{F}}]^{-3/2})$, it holds uniformly that*

$$\widetilde{\mathcal{L}}(\boldsymbol{\Theta}) - \widetilde{\mathcal{L}}(\boldsymbol{\Theta}') \leq \langle \nabla_{\boldsymbol{\Theta}}\widetilde{\mathcal{L}}(\boldsymbol{\Theta}), \boldsymbol{\Theta} - \boldsymbol{\Theta}'\rangle + \epsilon.$$

with $\epsilon = \mathcal{O}(\omega^{4/3}L_{\mathcal{F}}^3\sqrt{\log m_{\mathcal{F}}})$) being a small constant.

proof. This proof follows an analogous approach as the proof of Lemma 4.2 in [10]. Let $\nabla_{\mathcal{F}}\widetilde{\mathcal{L}}(\boldsymbol{\Theta}')$ be the derivative of $\widetilde{\mathcal{L}}$ with respective to $\mathcal{F}(\boldsymbol{x};\boldsymbol{\Theta})$. Then, it holds that $|\nabla_{\mathcal{F}}\widetilde{\mathcal{L}}(\boldsymbol{\Theta}')| \leq \mathcal{O}(1)$ based on

Lemma D.15. Then, by convexity of $\widetilde{\mathcal{L}}$, we have

$$
\begin{aligned}
&\widetilde{\mathcal{L}}(\boldsymbol{\Theta}')-\widetilde{\mathcal{L}}(\boldsymbol{\Theta}) \\
&\qquad \overset{\text{(i)}}{\geq} \nabla_{\mathcal{F}}\widetilde{\mathcal{L}}(\boldsymbol{\Theta}) \cdot (\mathcal{F}(\boldsymbol{x};\boldsymbol{\Theta}') - \mathcal{F}(\boldsymbol{x};\boldsymbol{\Theta})) \\
&\qquad \overset{\text{(ii)}}{\geq} \nabla_{\mathcal{F}}\widetilde{\mathcal{L}}(\boldsymbol{\Theta}') \cdot \langle \nabla_{\boldsymbol{\Theta}}\mathcal{F}(\boldsymbol{x};\boldsymbol{\Theta}), \boldsymbol{\Theta}' - \boldsymbol{\Theta}\rangle \\
&\qquad\qquad - |\nabla_{\mathcal{F}}\widetilde{\mathcal{L}}(\boldsymbol{\Theta}')| \cdot |\mathcal{F}(\boldsymbol{x};\boldsymbol{\Theta}') - \mathcal{F}(\boldsymbol{x};\boldsymbol{\Theta}) - \langle \nabla\mathcal{F}(\boldsymbol{x};\boldsymbol{\Theta}), \boldsymbol{\Theta}' - \boldsymbol{\Theta}\rangle| \\
&\qquad \geq \langle \nabla_{\boldsymbol{\Theta}}\widetilde{\mathcal{L}}(\boldsymbol{\Theta}), \boldsymbol{\Theta}' - \boldsymbol{\Theta}\rangle - |\nabla_{\mathcal{F}}\widetilde{\mathcal{L}}(\boldsymbol{\Theta}')| \cdot |\mathcal{F}(\boldsymbol{x};\boldsymbol{\Theta}') - \mathcal{F}(\boldsymbol{x};\boldsymbol{\Theta}) - \langle \nabla\mathcal{F}(\boldsymbol{x};\boldsymbol{\Theta}), \boldsymbol{\Theta}' - \boldsymbol{\Theta}\rangle| \\
&\qquad \overset{\text{(iii)}}{\geq} \langle \nabla_{\boldsymbol{\Theta}}\widetilde{\mathcal{L}}(\boldsymbol{\Theta}), \boldsymbol{\Theta}' - \boldsymbol{\Theta}\rangle - \mathcal{O}(\omega^{4/3}L_{\mathcal{F}}^{3}\sqrt{m_{\mathcal{F}}\log(m_{\mathcal{F}})}) \\
&\qquad \geq \langle \nabla_{\boldsymbol{\Theta}}\widetilde{\mathcal{L}}(\boldsymbol{\Theta}), \boldsymbol{\Theta}' - \boldsymbol{\Theta}\rangle - \epsilon
\end{aligned}
$$

where (i) is due to the convexity of the loss function $\mathcal{L}$, (ii) is an application of triangle inequality, and (iii) is the application of and **Lemma** D.16. Finally, denoting $\epsilon = \mathcal{O}(\omega^{4/3}L_{\mathcal{F}}^{3}\sqrt{m_{\mathcal{F}}\log m_{\mathcal{F}}})$ will complete the proof.

□

Lemma D.15. *Suppose $m_{\mathcal{F}}, \eta_1, \eta_2$ satisfy the conditions in Theorem 5.2. With probability at least $1-\mathcal{O}(KL_{\mathcal{F}})\cdot\exp(-\Omega(m_{\mathcal{F}}\omega^{2/3}L_{\mathcal{F}}))$ over the random initialization, $\boldsymbol{\Theta}$ satisfying $\|\boldsymbol{\Theta}-\boldsymbol{\Theta}^{(0)}\|_2 \leq \omega$ with $\omega \leq \mathcal{O}(L_{\mathcal{F}}^{-9/2}[\log m_{\mathcal{F}}]^{-3})$, it holds uniformly that*

$$
\begin{aligned}
\|\nabla_{\boldsymbol{\Theta}}\mathcal{F}(\boldsymbol{x};\boldsymbol{\Theta})\|_2 &\leq \mathcal{O}(\sqrt{m_{\mathcal{F}}L_{\mathcal{F}}}), \\
\|\nabla_{\boldsymbol{\Theta}}\mathcal{L}(\boldsymbol{x};\boldsymbol{\Theta})\|_2 &\leq \mathcal{O}(\sqrt{m_{\mathcal{F}}L_{\mathcal{F}}}).
\end{aligned}
$$

Proof. This lemma is a direct application of Lemma 9 of [46] and Lemma B.2, B.3 of [10].

□

Lemma D.16. *Suppose $m_{\mathcal{F}}, \eta_1, \eta_2$ satisfy the conditions in Theorem 5.2. With probability at least $1-\mathcal{O}(KL_{\mathcal{F}})\cdot\exp(-\Omega(m_{\mathcal{F}}\omega^{2/3}L_{\mathcal{F}}))$, for all $t \in [T], i \in [k]$, $\boldsymbol{\Theta}, \boldsymbol{\Theta}'$ (or Θ, Θ') satisfying $\|\boldsymbol{\Theta}-\boldsymbol{\Theta}^{(0)}\|_2, \|\boldsymbol{\Theta}'-\boldsymbol{\Theta}^{(0)}\|_2 \leq \omega$ with $\omega \leq \mathcal{O}(L_{\mathcal{F}}^{-9/2}[\log m_{\mathcal{F}}]^{-3})$, it holds uniformly that*

$$
|\mathcal{F}(\boldsymbol{x};\boldsymbol{\Theta}) - \mathcal{F}(\boldsymbol{x};\boldsymbol{\Theta}') - \langle \nabla_{\boldsymbol{\Theta}'}\mathcal{F}(\boldsymbol{x};\boldsymbol{\Theta}'), \boldsymbol{\Theta} - \boldsymbol{\Theta}'\rangle| \leq \mathcal{O}(w^{1/3}L_{\mathcal{F}}^{2}\sqrt{m_{\mathcal{F}}\log(m_{\mathcal{F}})})\|\boldsymbol{\Theta} - \boldsymbol{\Theta}'\|_2.
$$

Proof. The proof for this lemma directly follows the proof of Lemma 4.1 in [10] and Lemma 7 in [46]. □