Appendices: Proofs

A. Basic Lemmas

The following basic lemma is useful to our proofs, which will be used several times. Its proof follows from the convexity of $V(y,\cdot)$ and the fact that $V'_-(y,a)$ is bounded.

Lemma 1. *Under Assumption 1, for any $k \in \mathbb{N}$ and $w \in \mathcal{F}$, we have*

$$\|w_{k+1} - w\|^2 \leq \|w_k - w\|^2 + (a_0\kappa)^2\eta_k^2 + 2\eta_k \left[V(y_{j_k}, \langle w, \Phi(x_{j_k})\rangle) - V(y_{j_k}, \langle w_k, \Phi(x_{j_k})\rangle)\right]. \tag{12}$$

Proof. Since w_{k+1} is given by (3), by expanding the inner product, we have

$$\|w_{k+1} - w\|^2 = \|w_k - w\|^2 + \eta_k^2 \|V'_-(y_{j_k}, \langle w_k, \Phi(x_{j_k})\rangle)\Phi(x_{j_k})\|^2 + 2\eta_k V'_-(y_{j_k}, \langle w_k, \Phi(x_{j_k})\rangle) \langle w - w_k, \Phi(x_{j_k})\rangle.$$

The bounded assumption (4) implies that $\|\Phi(x_{j_k})\| \leq \kappa$ and by (5), $|V'_-(y_{j_k}, \langle w_k, \Phi(x_{j_k})\rangle)| \leq a_0$. We thus have

$$\|w_{k+1} - w\|^2 \leq \|w_k - w\|^2 + (a_0\kappa)^2\eta_k^2 + 2\eta_k V'_-(y_{j_k}, \langle w_k, \Phi(x_{j_k})\rangle)[\langle w, \Phi(x_{j_k})\rangle - \langle w_k, \Phi(x_{j_k})\rangle].$$

Using the convexity of $V(y_{j_k},\cdot)$ which tells us that

$$V'_-(y_{j_k}, a)(b-a) \leq V(y_{j_k}, b) - V(y_{j_k}, a), \qquad \forall a, b \in \mathbb{R},$$

we reach the desired bound. The proof is complete. □

Taking the expectation of (12) with respect to the random variable j_k, and noting that w_k is independent from j_k given $\mathbf{z}$, one can get the following result.

Lemma 2. *Under Assumption 1, for any fixed $k \in \mathbb{N}$, given any $\mathbf{z}$, assume that $w \in \mathcal{F}$ is independent of the random variable j_k. Then we have*

$$\mathbb{E}_{j_k}[\|w_{k+1} - w\|^2] \leq \|w_k - w\|^2 + (a_0\kappa)^2\eta_k^2 + 2\eta_k \left(\mathcal{E}_{\mathbf{z}}(w) - \mathcal{E}_{\mathbf{z}}(w_k)\right). \tag{13}$$

B. Sample Errors

Note that our goal is to bound the excess generalization error $\mathbb{E}[\mathcal{E}(w_T) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)]$, whereas the left-hand side of (13) is related to an empirical error. The difference between the generalization and empirical errors is a so-called sample error. To estimate this sample error, we introduce the following lemma, which gives a uniformly upper bound for sample errors over a ball $B_R = \{w \in \mathcal{F} : \|w\| \leq R\}$. Its proof is based on a standard symmetrization technique and Rademacher complexity, e.g. (Bartlett et al., 2005; Meir & Zhang, 2003). For completeness, we provide a proof here.

Lemma 3. *Assume (4) and (5). For any $R > 0$, we have*

$$\left|\mathbb{E}_{\mathbf{z}}\left[\sup_{w\in B_R} \left(\mathcal{E}(w) - \mathcal{E}_{\mathbf{z}}(w)\right)\right]\right| \leq \frac{2a_0\kappa R}{\sqrt{m}}.$$

Proof. Let $\mathbf{z}' = \{z'_i = (x'_i, y'_i)\}_{i=1}^m$ be another training sample from ρ, and assume that it is independent from $\mathbf{z}$. We have

$$\mathbb{E}_{\mathbf{z}}\left[\sup_{w\in B_R} \left(\mathcal{E}(w) - \mathcal{E}_{\mathbf{z}}(w)\right)\right] = \mathbb{E}_{\mathbf{z}}\left[\sup_{w\in B_R} \mathbb{E}_{\mathbf{z}'}[\mathcal{E}_{\mathbf{z}'}(w) - \mathcal{E}_{\mathbf{z}}(w)]\right] \leq \mathbb{E}_{\mathbf{z},\mathbf{z}'}\left[\sup_{w\in B_R} \left(\mathcal{E}_{\mathbf{z}'}(w) - \mathcal{E}_{\mathbf{z}}(w)\right)\right].$$

Let $\sigma_1, \sigma_2, \ldots, \sigma_m$ be independent random variables drawn from the Rademacher distribution, i.e. $\Pr(\sigma_i = +1) = \Pr(\sigma_i = -1) = 1/2$ for $i = 1, 2, \ldots, m$. Using a standard symmetrization technique, for example in (Meir & Zhang, 2003), we get

$$\begin{aligned}
\mathbb{E}_{\mathbf{z}}\left[\sup_{w\in B_R} \left(\mathcal{E}(w) - \mathcal{E}_{\mathbf{z}}(w)\right)\right] &\leq \mathbb{E}_{\mathbf{z},\mathbf{z}',\sigma}\left[\sup_{w\in B_R}\left\{\frac{1}{m}\sum_{i=1}^m \sigma_i(V(y'_i, \langle w, \Phi(x'_i)\rangle) - V(y_i, \langle w, \Phi(x_i)\rangle))\right\}\right] \\
&\leq 2\mathbb{E}_{\mathbf{z},\sigma}\left[\sup_{w\in B_R}\left\{\frac{1}{m}\sum_{i=1}^m \sigma_i V(y_i, \langle w, \Phi(x_i)\rangle)\right\}\right].
\end{aligned}$$

With (5), by applying Talagrand's contraction lemma, see e.g. (Bartlett et al., 2005), we derive

$$\mathbb{E}_{\mathbf{z}}\left[\sup_{w\in B_R}\left(\mathcal{E}(w)-\mathcal{E}_{\mathbf{z}}(w)\right)\right] \leq 2a_0\mathbb{E}_{\mathbf{z},\sigma}\left[\sup_{w\in B_R}\frac{1}{m}\sum_{i=1}^m \sigma_i\langle w,\Phi(x_i)\rangle\right] = 2a_0\mathbb{E}_{\mathbf{z},\sigma}\left[\sup_{w\in B_R}\left\langle w,\frac{1}{m}\sum_{i=1}^m\sigma_i\Phi(x_i)\right\rangle\right].$$

Using Cauchy-Schwartz inequality, we reach

$$\mathbb{E}_{\mathbf{z}}\left[\sup_{w\in B_R}\left(\mathcal{E}(w)-\mathcal{E}_{\mathbf{z}}(w)\right)\right] \leq 2a_0\mathbb{E}_{\mathbf{z},\sigma}\left[\sup_{w\in B_R}\|w\|\left\|\frac{1}{m}\sum_{i=1}^m\sigma_i\Phi(x_i)\right\|\right] \leq 2a_0R\mathbb{E}_{\mathbf{z},\sigma}\left[\left\|\frac{1}{m}\sum_{i=1}^m\sigma_i\Phi(x_i)\right\|\right].$$

By Jensen's inequality, we get

$$\mathbb{E}_{\mathbf{z}}\left[\sup_{w\in B_R}\left(\mathcal{E}(w)-\mathcal{E}_{\mathbf{z}}(w)\right)\right] \leq 2a_0R\left[\mathbb{E}_{\mathbf{z},\sigma}\left\|\frac{1}{m}\sum_{i=1}^m\sigma_i\Phi(x_i)\right\|^2\right]^{1/2} = 2a_0R\left[\frac{1}{m^2}\mathbb{E}_{\mathbf{z},\sigma}\sum_{i=1}^m\|\Phi(x_i)\|^2\right]^{1/2}.$$

The desired result thus follows by introducing (4) to the above. Note that the above procedure also applies if we replace $\mathcal{E}(w)-\mathcal{E}_{\mathbf{z}}(w)$ with $\mathcal{E}_{\mathbf{z}}(w)-\mathcal{E}(w)$. The proof is complete. □

The following lemma gives upper bounds on the iterated sequence.

Lemma 4. *Under Assumption 1. Then for any $t\in\mathbb{N}$, we have*

$$\|w_{t+1}\| \leq \sqrt{(a_0\kappa)^2\sum_{k=1}^t\eta_k^2 + 2|V|_0\sum_{k=1}^t\eta_k}.$$

Proof. Using Lemma 1 with $w=0$, we have

$$\|w_{k+1}\|^2 \leq \|w_k\|^2 + (a_0\kappa)^2\eta_k^2 + 2\eta_k\left[V(y_{j_k},0) - V(y_{j_k},\langle w_k,\Phi(x_{j_k})\rangle)\right].$$

Noting that $V(y,a)\geq 0$ and $V(y_{j_k},0)\leq |V|_0$, we thus get

$$\|w_{k+1}\|^2 \leq \|w_k\|^2 + (a_0\kappa)^2\eta_k^2 + 2\eta_k|V|_0.$$

Applying this inequality iteratively for $k=1,\cdots,t$, and introducing with $w_1=0$, one can get that

$$\|w_{t+1}\|^2 \leq (a_0\kappa)^2\sum_{k=1}^t\eta_k^2 + 2|V|_0\sum_{k=1}^t\eta_k,$$

which leads to the desired result by taking square root on both sides. □

According to the above two lemmas, we can bound the sample errors as follows.

Lemma 5. *Assume (4) and (5). Then, for any $k\in\mathbb{N}$,*

$$|\mathbb{E}_{\mathbf{z},J}[\mathcal{E}_{\mathbf{z}}(w_k)-\mathcal{E}(w_k)]| \leq \frac{2a_0\kappa R_k}{\sqrt{m}},$$

where

$$R_k = \sqrt{(a_0\kappa)^2\sum_{k=1}^t\eta_k^2 + 2|V|_0\sum_{k=1}^t\eta_k}. \tag{14}$$

When the loss function is smooth, by Theorems 2.2 and 3.9 from (Hardt et al., 2016), we can control the sample errors as follows.

Lemma 6. *Under Assumptions 1 and 3, let $\eta_t \leq 2/(\kappa^2 L)$ for all $k \in [T]$,*

$$|\mathbb{E}_{\mathbf{z},J}[\mathcal{E}_{\mathbf{z}}(w_k) - \mathcal{E}(w_k)]| \leq \frac{2(a_0\kappa)^2 \sum_{i=1}^k \eta_i}{m}.$$

Proof. Note that by (4), Assumption 3 and (2), for all $(x,y) \in \mathbf{z}$, $w, w' \in \mathcal{F}$,

$$\begin{aligned}
&\|V'(y, \langle w, \Phi(x)\rangle)\Phi(x) - V'(y, \langle w', \Phi(x)\rangle)\Phi(x)\| \leq \kappa |V'(y, \langle w, \Phi(x)\rangle) - V'(y, \langle w', \Phi(x)\rangle)| \\
\leq& \kappa L |\langle w, \Phi(x)\rangle - \langle w', \Phi(x)\rangle| = \kappa L |\langle w - w', \Phi(x)\rangle| \leq \kappa L \|w - w'\| \|\Phi(x)\| \\
\leq& \kappa^2 L \|w - w'\|,
\end{aligned}$$

and

$$\|V'(y, \langle w, \Phi(x)\rangle)\Phi(x)\| \leq \kappa a_0.$$

That is, for every $(x,y) \in \mathbf{z}$, $V(y, \langle \cdot, \Phi(x)\rangle)$ is $(\kappa^2 L)$-smooth and (κa_0)-Lipschitz. Now the results follow directly by using Theorems 2.2 and 3.8 from (Hardt et al., 2016). □

C. Excess Errors for Weighted Averages

Lemma 7. *Under Assumption 1, assume that there exists a non-decreasing sequence $\{b_k > 0\}_k$ such that*

$$|\mathbb{E}_{\mathbf{z},J}[\mathcal{E}_{\mathbf{z}}(w_k) - \mathcal{E}(w_k)]| \leq b_k, \quad \forall k \in [T]. \tag{15}$$

Then for any $t \in [T]$ and any fixed $w \in \mathcal{F}$,

$$\sum_{k=1}^t 2\eta_k \mathbb{E}_{\mathbf{z},J}\left[\mathcal{E}(w_k)\right] \leq b_t \sum_{k=1}^t 2\eta_k + (a_0\kappa)^2 \sum_{k=1}^t \eta_k^2 + \sum_{k=1}^t 2\eta_k \mathcal{E}(w) + \|w\|^2. \tag{16}$$

Proof. By Lemma 2, we have (13). Rewriting $-\mathcal{E}_{\mathbf{z}}(w_k)$ as

$$-\mathcal{E}_{\mathbf{z}}(w_k) + \mathcal{E}(w_k) - \mathcal{E}(w_k),$$

taking the expectation with respect to $J(T)$ and $\mathbf{z}$ on both sides, noting that w is independent of J and $\mathbf{z}$, and applying Condition (15), we derive

$$\mathbb{E}_{\mathbf{z},J}[\|w_{k+1} - w\|^2] \leq \mathbb{E}_{\mathbf{z},J}[\|w_k - w\|^2] + (a_0\kappa)^2\eta_k^2 + 2\eta_k\left(\mathcal{E}(w) - \mathbb{E}_{\mathbf{z},J}\left[\mathcal{E}(w_k)\right]\right) + 2\eta_k b_k,$$

which is equivalent to

$$2\eta_k \mathbb{E}_{\mathbf{z},J}\left[\mathcal{E}(w_k)\right] \leq 2\eta_k \mathcal{E}(w) + \mathbb{E}_{\mathbf{z},J}[\|w_k - w\|^2 - \|w_{k+1} - w\|^2] + (a_0\kappa)^2\eta_k^2 + 2\eta_k b_k.$$

Summing up over $k = 1, \cdots, t$, and introducing with $w_1 = 0$,

$$\sum_{k=1}^t 2\eta_k \mathbb{E}_{\mathbf{z},J}\left[\mathcal{E}(w_k)\right] \leq \sum_{k=1}^t 2\eta_k \mathcal{E}(w) + \|w\|^2 + (a_0\kappa)^2 \sum_{k=1}^t \eta_k^2 + \sum_{k=1}^t 2\eta_k b_k.$$

The proof can be finished by noting that b_k is non-decreasing. □

Now, we are in a position to prove Theorem 1.

Proof of Theorem 1. According to Lemma 5, Condition (15) is satisfied for

$$b_t = \frac{2a_0\kappa\sqrt{\sum_{k=1}^t (a_0\kappa\eta_k)^2 + 2|V|_0 \sum_{k=1}^t \eta_k}}{\sqrt{m}}.$$

By Lemma 7, we thus have (16). Dividing both sides by $\sum_{k=1}^t 2\eta_k$, and using the convexity of $V(y,\cdot)$ which implies

$$\frac{\sum_{k=1}^t \eta_k \mathcal{E}(w_k)}{\sum_{k=1}^t \eta_k} \geq \mathcal{E}(\frac{\sum_{k=1}^t \eta_k w_k}{\sum_{k=1}^t \eta_k}) = \mathcal{E}(\overline{w}_t), \tag{17}$$

we get that

$$\mathbb{E}_{\mathbf{z},J}\left[\mathcal{E}(\overline{w}_t)\right] \leq b_t + \frac{(a_0\kappa)^2}{2}\frac{\sum_{k=1}^t \eta_k^2}{\sum_{k=1}^t \eta_k} + \mathcal{E}(w) + \frac{\|w\|^2}{2\sum_{k=1}^t \eta_k}.$$

For any fixed $\epsilon > 0$, we know that there exists a $w_\epsilon \in \mathcal{F}$, such that $\mathcal{E}(w_\epsilon) \leq \inf_{w\in\mathcal{F}} \mathcal{E}(w) + \epsilon$. Letting $t = t^*(m)$, and $w = w_\epsilon$, we have

$$\mathbb{E}_{\mathbf{z},J}\left[\mathcal{E}(w_{t^*(m),w})\right] \leq b_{t^*(m)} + \frac{(a_0\kappa)^2}{2}\frac{\sum_{k=1}^{t^*(m)} \eta_k^2}{\sum_{k=1}^{t^*(m)} \eta_k} + \inf_{w\in\mathcal{F}} \mathcal{E}(w) + \epsilon + \frac{\|w_\epsilon\|^2}{2\sum_{k=1}^{t^*(m)} \eta_k}.$$

Letting $m \to \infty$, and using Conditions (A) and (B) which imply

$$\lim_{m\to\infty} \frac{1}{\sum_{k=1}^{t^*(m)} \eta_k} = 0, \quad \lim_{m\to\infty} \frac{\sum_{k=1}^{t^*(m)} \eta_k^2}{\sum_{k=1}^{t^*(m)} \eta_k} = 0, \quad \text{and } \lim_{m\to\infty} \frac{\sum_{k=1}^{t^*(m)} \eta_k^2}{m} = \lim_{m\to\infty} \frac{\sum_{k=1}^{t^*(m)} \eta_k^2}{\sum_{k=1}^{t^*(m)} \eta_k}\frac{\sum_{k=1}^{t^*(m)} \eta_k}{m} = 0,$$

we reach

$$\lim_{m\to\infty} \mathbb{E}_{\mathbf{z},J}\left[\mathcal{E}(w_{t^*(m),w})\right] \leq \inf_{w\in\mathcal{F}} \mathcal{E}(w) + \epsilon.$$

Since $\epsilon > 0$ is arbitrary, the desired result thus follows. The proof is complete. □

Lemma 8. *Under the assumptions of Lemma 7, let Assumption 2 hold. Then for any* $t \in [T]$,

$$\sum_{k=1}^t 2\eta_k \mathbb{E}_{\mathbf{z},J}\left[\mathcal{E}(w_k) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)\right] \leq b_t \sum_{k=1}^t 2\eta_k + (a_0\kappa)^2 \sum_{k=1}^t \eta_k^2 + 2c_\beta \left(\sum_{k=1}^t \eta_k\right)^{1-\beta}. \tag{18}$$

Proof. By Lemma 7, we have (16). Subtracting $\sum_{k=1}^t 2\eta_k \inf_{w\in\mathcal{F}} \mathcal{E}(w)$ from both sides,

$$\sum_{k=1}^t 2\eta_k \mathbb{E}_{\mathbf{z},J}\left[\mathcal{E}(w_k) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)\right] \leq b_t \sum_{k=1}^t 2\eta_k + (a_0\kappa)^2 \sum_{k=1}^t \eta_k^2 + \sum_{k=1}^t 2\eta_k\left[\mathcal{E}(w) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)\right] + \|w\|^2.$$

Taking the infimum over $w \in \mathcal{F}$, recalling that $\mathcal{D}(\lambda)$ is defined by (6), we have

$$\sum_{k=1}^t 2\eta_k \mathbb{E}_{\mathbf{z},J}\left[\mathcal{E}(w_k) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)\right] \leq b_t \sum_{k=1}^t 2\eta_k + (a_0\kappa)^2 \sum_{k=1}^t \eta_k^2 + \sum_{k=1}^t 2\eta_k \mathcal{D}\left(\frac{1}{\sum_{k=1}^t \eta_k}\right).$$

Using Assumption 2 to the above, we get the desired result. The proof is complete. □

Collecting some of the above analysis, we get the following result.

Proposition 1. *Under the assumptions of Lemma 8, we have*

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(\overline{w}_t)] - \inf_{w\in\mathcal{F}} \mathcal{E}(w) \leq b_t + \frac{(a_0\kappa)^2}{2}\frac{\sum_{k=1}^t \eta_k^2}{\sum_{k=1}^t \eta_k} + c_\beta\left(\frac{1}{\sum_{k=1}^t \eta_k}\right)^\beta. \tag{19}$$

Proof. By Lemma 8, we have (18). Dividing both sides by $\sum_{k=1}^t 2\eta_k$, and using (17), we get the desired bound. □

D. From Weighted Averages to the Last Iterate

A basic tool for studying the convergence for iterates is the following decomposition, as often done in (Shamir & Zhang, 2013) for classical online learning or subgradient descent algorithms (Lin et al., 2016). It enables us to study the weighted excess generalization error $2\eta_t \mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)]$ in terms of "weighted averages" and moving weighted averages. In what follows, we will write $\mathbb{E}_{\mathbf{z},J}$ as $\mathbb{E}$ for short.

Lemma 9. *We have*

$$2\eta_t \mathbb{E}\left\{\mathcal{E}(w_t) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)\right\} \leq \frac{1}{t}\sum_{k=1}^{t} 2\eta_k \mathbb{E}\left\{\mathcal{E}(w_k) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)\right\} + \sum_{k=1}^{t-1}\frac{1}{k(k+1)}\sum_{i=t-k+1}^{t} 2\eta_i \mathbb{E}\left\{\mathcal{E}(w_i) - \mathcal{E}(w_{t-k})\right\}. \tag{20}$$

Proof. Let $\{u_t\}_t$ be a real-valued sequence. For $k = 1, \cdots, t-1$,

$$\frac{1}{k}\sum_{i=t-k+1}^{t} u_i - \frac{1}{k+1}\sum_{i=t-k}^{t} u_i = \frac{1}{k(k+1)}\left\{(k+1)\sum_{i=t-k+1}^{t} u_i - k\sum_{i=t-k}^{t} u_i\right\} = \frac{1}{k(k+1)}\sum_{i=t-k+1}^{t}(u_i - u_{t-k}).$$

Summing over $k = 1, \cdots, t-1$, and rearranging terms, we get

$$u_t = \frac{1}{t}\sum_{i=1}^{t} u_i + \sum_{k=1}^{t-1}\frac{1}{k(k+1)}\sum_{i=t-k+1}^{t}(u_i - u_{t-k}).$$

Choosing $u_t = 2\eta_t \mathbb{E}\left\{\mathcal{E}(w_t) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)\right\}$ in the above, we get

$$\begin{aligned} 2\eta_t \mathbb{E}\left\{\mathcal{E}(w_t) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)\right\} =& \frac{1}{t}\sum_{i=1}^{t} 2\eta_i \mathbb{E}\left\{\mathcal{E}(w_i) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)\right\} \\ &+ \sum_{k=1}^{t-1}\frac{1}{k(k+1)}\sum_{i=t-k+1}^{t}\left(2\eta_i \mathbb{E}\left\{\mathcal{E}(w_i) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)\right\} - 2\eta_{t-k}\mathbb{E}\left\{\mathcal{E}(w_{t-k}) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)\right\}\right), \end{aligned}$$

which can be rewritten as

$$\begin{aligned} 2\eta_t \mathbb{E}\left\{\mathcal{E}(w_t) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)\right\} =& \frac{1}{t}\sum_{k=1}^{t} 2\eta_k \mathbb{E}\left\{\mathcal{E}(w_k) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)\right\} + \sum_{k=1}^{t-1}\frac{1}{k(k+1)}\sum_{i=t-k+1}^{t} 2\eta_i \mathbb{E}\left\{\mathcal{E}(w_i) - \mathcal{E}(w_{t-k})\right\} \\ &+ \sum_{k=1}^{t-1}\frac{1}{k+1}\left[\frac{1}{k}\sum_{i=t-k+1}^{t} 2\eta_i - 2\eta_{t-k}\right]\mathbb{E}\left\{\mathcal{E}(w_{t-k}) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)\right\}. \end{aligned}$$

Since, $\mathcal{E}(w_{t-k}) - \inf_{w\in\mathcal{F}} \mathcal{E}(w) \geq 0$ and that $\{\eta_t\}_{t\in\mathbb{N}}$ is a non-increasing sequence, we know that the last term of the above inequality is at most zero. Therefore, we get the desired result. The proof is complete. □

The first term of the right-hand side of (20) is the weighted excess generalization error, and it can be estimated easily by (18), while the second term (sum of moving averages) can be estimated by the following lemma.

Lemma 10. *Under the assumptions of Lemma 7, we have*

$$\sum_{k=1}^{t-1}\frac{1}{k(k+1)}\sum_{i=t-k+1}^{t} 2\eta_i \mathbb{E}\left\{\mathcal{E}(w_i) - \mathcal{E}(w_{t-k})\right\} \leq \sum_{i=1}^{t-1}\frac{(a_0\kappa\eta_i)^2 + 4b_t\eta_i}{t-i} - \frac{1}{t}\sum_{k=1}^{t}(a_0\kappa\eta_k)^2 + 4b_t\eta_k) + (a_0\kappa\eta_t)^2 + 4b_t\eta_t. \tag{21}$$

Proof. Given any sample $\mathbf{z}$, note that w_{t-k} is depending only on $j_1, j_2, \cdots, j_{t-k-1}$, and thus is independent from j_{i+1} for any $t \geq i \geq t-k$. Following from Lemma 2, for any $i \geq t-k$,

$$\mathbb{E}_{j_{i+1}}[\|w_{i+1} - w_{t-k}\|^2] \leq \|w_i - w_{t-k}\|^2 + (a_0\kappa)^2\eta_i^2 + 2\eta_i\left(\mathcal{E}_{\mathbf{z}}(w_{t-k}) - \mathcal{E}_{\mathbf{z}}(w_i)\right).$$

Taking the expectation on both sides, and bounding $\mathbb{E}[\mathcal{E}_{\mathbf{z}}(w_{t-k}) - \mathcal{E}_{\mathbf{z}}(w_i)]$ as

$$= \mathbb{E}[\mathcal{E}_{\mathbf{z}}(w_{t-k}) - \mathcal{E}(w_{t-k}) + \mathcal{E}(w_i) - \mathcal{E}_{\mathbf{z}}(w_i) + \mathcal{E}(w_{t-k}) - \mathcal{E}(w_i)] \leq 2b_t + \mathbb{E}[\mathcal{E}(w_{t-k}) - \mathcal{E}(w_i)]$$

by Condition (15), and rearranging terms, we get

$$2\eta_i \mathbb{E}\left[\mathcal{E}(w_i) - \mathcal{E}(w_{t-k})\right] \leq \mathbb{E}[\|w_i - w_{t-k}\|^2 - \|w_{i+1} - w_{t-k}\|^2] + (a_0\kappa)^2\eta_i^2 + 4\eta_i b_t.$$

Summing up over $i = t-k, \cdots, t$, we get

$$\sum_{i=t-k}^{t} 2\eta_i \mathbb{E}\left[\mathcal{E}(w_i) - \mathcal{E}(w_{t-k})\right] \leq (a_0\kappa)^2 \sum_{i=t-k}^{t} \eta_i^2 + 4b_t \sum_{i=t-k}^{t} \eta_i.$$

The left-hand side is exactly $\sum_{i=t-k+1}^{t} 2\eta_i \mathbb{E}\left[\mathcal{E}(w_i) - \mathcal{E}(w_{t-k})\right]$. Thus, dividing both sides by $k(k+1)$, and then summing up over $k = 1, \cdots, t-1$,

$$\sum_{k=1}^{t-1} \frac{1}{k(k+1)} \sum_{i=t-k+1}^{t} 2\eta_i \mathbb{E}\left\{\mathcal{E}(w_i) - \mathcal{E}(w_{t-k})\right\} \leq \sum_{k=1}^{t-1} \frac{1}{k(k+1)} \sum_{i=t-k}^{t} ((a_0\kappa\eta_i)^2 + 4b_t\eta_i).$$

Exchanging the order in the sum, and setting $\xi_i = (a_0\kappa\eta_i)^2 + 4b_t\eta_i$ for all $i \in [t]$, we obtain

$$\begin{aligned}
\sum_{k=1}^{t-1} \frac{1}{k(k+1)} \sum_{i=t-k+1}^{t} 2\eta_i \mathbb{E}\left\{\mathcal{E}(w_i) - \mathcal{E}(w_{t-k})\right\} &\leq \sum_{i=1}^{t-1} \sum_{k=t-i}^{t-1} \frac{1}{k(k+1)} \xi_i + \sum_{k=1}^{t-1} \frac{1}{k(k+1)} \xi_t \\
&= \sum_{i=1}^{t-1} \left(\frac{1}{t-i} - \frac{1}{t}\right) \xi_i + \left(1 - \frac{1}{t}\right) \xi_t \\
&= \sum_{i=1}^{t-1} \frac{1}{t-i} \xi_i + \xi_t - \frac{1}{t} \sum_{k=1}^{t} \xi_k.
\end{aligned}$$

From the above analysis, we can conclude the proof. □

Proposition 2. *Under the assumptions of Lemma 8, we have*

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w \in \mathcal{F}} \mathcal{E}(w)] \leq b_t \left(1 + \sum_{k=1}^{t-1} \frac{2\eta_k}{\eta_t(t-k)}\right) + \sum_{k=1}^{t-1} \frac{(a_0\kappa\eta_k)^2}{2\eta_t(t-k)} + \frac{(a_0\kappa)^2\eta_t}{2} + \frac{c_\beta}{\eta_t t} \left(\sum_{k=1}^{t} \eta_k\right)^{1-\beta} \tag{22}$$

Proof. Plugging (18) and (21) into (20), by a direct calculation, we get

$$2\eta_t \mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w \in \mathcal{F}} \mathcal{E}(w)] \leq \frac{2c_\beta}{t} \left(\sum_{k=1}^{t} \eta_k\right)^{1-\beta} + \sum_{k=1}^{t-1} \frac{(a_0\kappa\eta_k)^2 + 4b_t\eta_k}{t-k} - \frac{2b_t}{t} \sum_{k=1}^{t} \eta_k + (a_0\kappa\eta_t)^2 + 4b_t\eta_t.$$

Since $\{\eta_t\}_t$ is non-increasing, $\frac{2b_t}{t} \sum_{k=1}^{t} \eta_k \geq 2b_t\eta_t$. Thus,

$$2\eta_t \mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w \in \mathcal{F}} \mathcal{E}(w)] \leq \frac{2c_\beta}{t} \left(\sum_{k=1}^{t} \eta_k\right)^{1-\beta} + 2\eta_t b_t + \sum_{k=1}^{t-1} \frac{(a_0\kappa\eta_k)^2 + 4b_t\eta_k}{t-k} + (a_0\kappa\eta_t)^2.$$

Dividing both sides with $2\eta_t$, and rearranging terms, one can conclude the proof. □

Now, we are ready to prove Theorems 2 and 3.

Proof of Theorem 2. By Lemma 6, the condition (15) is satisfied with $b_k = 2(a_0\kappa)^2 \sum_{i=1}^k \eta_i/m$. It thus follows from Propositions 1 and 2 that

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(\overline{w}_t)] - \inf_{w\in\mathcal{F}} \mathcal{E}(w) \leq 2(a_0\kappa)^2 \frac{\sum_{k=1}^t \eta_k}{m} + \frac{(a_0\kappa)^2}{2} \frac{\sum_{k=1}^t \eta_k^2}{\sum_{k=1}^t \eta_k} + c_\beta \left(\frac{1}{\sum_{k=1}^t \eta_k} \right)^\beta, \tag{23}$$

and

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)] \leq 2(a_0\kappa)^2 \frac{\sum_{k=1}^t \eta_k}{m} \left(1 + \sum_{k=1}^{t-1} \frac{2\eta_k}{\eta_t(t-k)}\right) + \frac{(a_0\kappa)^2}{2} \sum_{k=1}^{t-1} \frac{\eta_k^2}{\eta_t(t-k)} + \frac{(a_0\kappa)^2}{2}\eta_t + c_\beta \frac{\left(\sum_{k=1}^t \eta_k\right)^{1-\beta}}{\eta_t t}. \tag{24}$$

By noting that $1 \leq \eta_{t-1}/\eta_t \leq \sum_{k=1}^{t-1} \eta_k/(\eta_t(t-k))$,

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)] \leq 6(a_0\kappa)^2 \frac{\sum_{k=1}^t \eta_t}{m} \sum_{k=1}^{t-1} \frac{\eta_k}{\eta_t(t-k)} + \frac{(a_0\kappa)^2}{2} \sum_{k=1}^{t-1} \frac{\eta_k^2}{\eta_t(t-k)} + \frac{(a_0\kappa)^2}{2}\eta_t + c_\beta \frac{\left(\sum_{k=1}^t \eta_k\right)^{1-\beta}}{\eta_t t}. \tag{25}$$

The proof is complete. □

Proof of Theorems 3. By Propositions 1 and 2, we have (19) and (22). Also, by Lemma 5, we have $b_t \leq \frac{2a_0\kappa R_t}{\sqrt{m}}$. Then

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(\overline{w}_t)] - \inf_{w\in\mathcal{F}} \mathcal{E}(w) \leq 2a_0\kappa \frac{R_t}{\sqrt{m}} + \frac{(a_0\kappa)^2}{2} \frac{\sum_{k=1}^t \eta_k^2}{\sum_{k=1}^t \eta_k} + c_\beta \left(\frac{1}{\sum_{k=1}^t \eta_k} \right)^\beta, \tag{26}$$

and

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)] \leq 2a_0\kappa \frac{R_t}{\sqrt{m}} \left(1 + \sum_{k=1}^{t-1} \frac{2\eta_k}{\eta_t(t-k)}\right) + \frac{(a_0\kappa)^2}{2} \sum_{k=1}^{t-1} \frac{\eta_k^2}{\eta_t(t-k)} + \frac{(a_0\kappa)^2}{2}\eta_t + c_\beta \frac{\left(\sum_{k=1}^t \eta_k\right)^{1-\beta}}{\eta_t t}. \tag{27}$$

Note that $1 \leq \eta_{t-1}/\eta_t$ since η_t is non-increasing. Thus,

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)] \leq 6a_0\kappa \frac{R_t}{\sqrt{m}} \sum_{k=1}^{t-1} \frac{\eta_k}{\eta_t(t-k)} + \frac{(a_0\kappa)^2}{2} \sum_{k=1}^{t-1} \frac{\eta_k^2}{\eta_t(t-k)} + \frac{(a_0\kappa)^2}{2}\eta_t + c_\beta \frac{\left(\sum_{k=1}^t \eta_k\right)^{1-\beta}}{\eta_t t}. \tag{28}$$

Recall that R_t is given by (14) and that η_k is non-increasing, we thus have

$$R_t \leq \sqrt{(a_0\kappa)^2\eta_1 + 2|V|_0} \sqrt{\sum_{k=1}^t \eta_k}. \tag{29}$$

Introducing the above into (26) and (28),

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(\overline{w}_t)] - \inf_{w\in\mathcal{F}} \mathcal{E}(w) \leq 2a_0\kappa \sqrt{(a_0\kappa)^2\eta_1 + 2|V|_0} \sqrt{\frac{\sum_{k=1}^t \eta_k}{m}} + \frac{(a_0\kappa)^2}{2} \frac{\sum_{k=1}^t \eta_k^2}{\sum_{k=1}^t \eta_k} + c_\beta \left(\frac{1}{\sum_{k=1}^t \eta_k} \right)^\beta, \tag{30}$$

and

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)] \leq 6a_0\kappa \sqrt{(a_0\kappa)^2\eta_1 + 2|V|_0} \sqrt{\frac{\sum_{k=1}^t \eta_k}{m}} + \frac{(a_0\kappa)^2}{2} \sum_{k=1}^{t-1} \frac{\eta_k^2}{\eta_t(t-k)} + \frac{(a_0\kappa)^2}{2}\eta_t + c_\beta \frac{\left(\sum_{k=1}^t \eta_k\right)^{1-\beta}}{\eta_t t}.$$

□

E. Explicit Convergence Rates

In this section, we prove Corollaries 1-8. We first introduce the following basic estimates.

Lemma 11. *Let $\theta \in \mathbb{R}_+$, and $t \in \mathbb{N}$, $t \geq 3$. Then*

$$\sum_{k=1}^{t} k^{-\theta} \leq \begin{cases} t^{1-\theta}/(1-\theta), & \text{when } \theta < 1, \\ \log t + 1, & \text{when } \theta = 1, \\ \theta/(\theta-1), & \text{when } \theta > 1, \end{cases}$$

and

$$\sum_{k=1}^{t} k^{-\theta} \geq \begin{cases} \frac{1-4^{\theta-1}}{1-\theta} t^{1-\theta} & \text{when } \theta < 1, \\ \ln t & \text{when } \theta = 1. \end{cases}$$

Proof. By using

$$\sum_{k=1}^{t} k^{-\theta} = 1 + \sum_{k=2}^{t} \int_{k-1}^{k} du k^{-\theta} \leq 1 + \sum_{k=2}^{t} \int_{k-1}^{k} u^{-\theta} du = 1 + \int_{1}^{t} u^{-\theta} du,$$

which leads to the first part of the result. Similarly,

$$\sum_{k=1}^{t} k^{-\theta} = \sum_{k=1}^{t} k^{-\theta} \geq \sum_{k=1}^{t} \int_{k}^{k+1} u^{-\theta} du = \int_{1}^{t+1} u^{-\theta} du,$$

which leads to the second part of the result. The proof is complete. □

Lemma 12. *Let $q \in \mathbb{R}_+$ and $t \in \mathbb{N}$, $t \geq 3$. Then*

$$\sum_{k=1}^{t-1} \frac{1}{t-k} k^{-q} \leq \begin{cases} 2^q [2 + (1-q)^{-1}] t^{-q} \log t, & \text{when } q < 1, \\ 8t^{-1} \log t, & \text{when } q = 1, \\ (2^q + 2q)/(q-1) t^{-1}, & \text{when } q > 1, \end{cases}$$

Proof. We split the sum into two parts

$$\begin{aligned} \sum_{k=1}^{t-1} \frac{1}{t-k} k^{-q} &= \sum_{t/2 \leq k \leq t-1} \frac{1}{t-k} k^{-q} + \sum_{1 \leq k < t/2} \frac{1}{t-k} k^{-q} \\ &\leq 2^q t^{-q} \sum_{t/2 \leq k \leq t-1} \frac{1}{t-k} + 2t^{-1} \sum_{1 \leq k < t/2} k^{-q} \\ &= 2^q t^{-q} \sum_{1 \leq k \leq t/2} k^{-1} + 2t^{-1} \sum_{1 \leq k < t/2} k^{-q}. \end{aligned}$$

Applying Lemma 11, we get

$$\sum_{k=1}^{t-1} \frac{1}{t-k} k^{-q} \leq 2^q t^{-q} (\log(t/2) + 1) + \begin{cases} 2^q t^{-q}/(1-q), & \text{when } q < 1, \\ 4t^{-1} \log t, & \text{when } q = 1, \\ 2qt^{-1}/(q-1), & \text{when } q > 1, \end{cases}$$

which leads to the desired result by using $t^{-q+1} \log t \leq 1/(\mathrm{e}(q-1)) \leq 1/(2(q-1))$ when $q > 1$. □

The bounds in the above two lemmas involve constant factor $1/(1-\theta)$ or $1/(1-q)$, which tend to be infinity as $\theta \to 1$ or $q \to 1$. To avoid these, we introduce the following complement results.

Lemma 13. *Let $\theta \in \mathbb{R}_+$, and $t \in \mathbb{N}$, with $t \geq 3$. Then*

$$\sum_{k=1}^{t} k^{-\theta} \leq t^{\max(1-\theta,0)} 2 \log t.$$

Proof. Note that

$$\sum_{k=1}^{t} k^{-\theta} = \sum_{k=1}^{t} k^{-1} k^{1-\theta} \leq t^{\max(1-\theta,0)} \sum_{k=1}^{t} k^{-1}.$$

The proof can be finished by applying Lemma 11. □

Lemma 14. *Let $q \in \mathbb{R}_+$ and $t \in \mathbb{N}$, $t \geq 3$. Then*

$$\sum_{k=1}^{t-1} \frac{1}{t-k} k^{-q} \leq 4 t^{-\min(q,1)} \log t.$$

Proof. Note that

$$\sum_{k=1}^{t-1} \frac{1}{t-k} k^{-q} = \sum_{k=1}^{t-1} \frac{k^{1-q}}{(t-k)k} \leq t^{\max(1-q,0)} \sum_{k=1}^{t-1} \frac{1}{(t-k)k},$$

and that by Lemma 11,

$$\sum_{k=1}^{t-1} \frac{1}{(t-k)k} = \frac{1}{t} \sum_{k=1}^{t-1} \left(\frac{1}{t-k} + \frac{1}{k} \right) = \frac{2}{t} \sum_{k=1}^{t-1} \frac{1}{k} \leq \frac{4}{t} \log t.$$

□

With the above estimates and Theorems 2, 3, we can get the following two propositions.

Proposition 3. *Under Assumptions 1, 2 and 3, let $\eta_t = \eta t^{-\theta}$ for some positive constant $\eta \leq \frac{1}{\kappa^2 L}$ with $\theta \in [0,1)$ for all $t \in \mathbb{N}$. Then for all $t \in \mathbb{N}$,*

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(\overline{w}_t)] - \inf_{w \in \mathcal{F}} \mathcal{E}(w) \leq \frac{2(a_0\kappa)^2}{1-\theta} \frac{\eta t^{1-\theta}}{m} + \frac{(a_0\kappa)^2(1-\theta)}{1-4^{\theta-1}} \eta t^{-\min(\theta,1-\theta)} \log t + c_\beta \left(\frac{1-\theta}{1-4^{\theta-1}} \right)^\beta \left(\frac{1}{\eta t^{1-\theta}} \right)^\beta,$$

and

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w \in \mathcal{F}} \mathcal{E}(w)] \leq \frac{18(a_0\kappa)^2}{1-\theta} \frac{\eta t^{1-\theta} \log t}{m} + 3(a_0\kappa)^2 \eta t^{-\min(\theta,1-\theta)} \log t + \frac{c_\beta}{1-\theta} \left(\frac{1}{\eta t^{1-\theta}} \right)^\beta.$$

Proof. Following the proof of Theorem 2, we have (23) and (24).
We first consider the case $\overline{w}_t$. With $\eta_t = \eta t^{-\theta}$, (23) reads as

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(\overline{w}_t)] - \inf_{w \in \mathcal{F}} \mathcal{E}(w) \leq 2(a_0\kappa)^2 \frac{\eta \sum_{k=1}^t k^{-\theta}}{m} + \frac{(a_0\kappa)^2}{2} \frac{\eta \sum_{k=1}^t k^{-2\theta}}{\sum_{k=1}^t k^{-\theta}} + c_\beta \left(\frac{1}{\eta \sum_{k=1}^t k^{-\theta}} \right)^\beta.$$

Lemma 11 tells us that

$$\frac{1-4^{\theta-1}}{1-\theta} t^{1-\theta} \leq \sum_{k=1}^{t} k^{-\theta} \leq \frac{t^{1-\theta}}{1-\theta}.$$

Thus,

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(\overline{w}_t)] - \inf_{w \in \mathcal{F}} \mathcal{E}(w) \leq \frac{2(a_0\kappa)^2}{1-\theta} \frac{\eta t^{1-\theta}}{m} + \frac{(a_0\kappa)^2(1-\theta)}{2(1-4^{\theta-1})} \frac{\eta \sum_{k=1}^t k^{-2\theta}}{t^{1-\theta}} + c_\beta \left(\frac{1-\theta}{1-4^{\theta-1}} \right)^\beta \left(\frac{1}{\eta t^{1-\theta}} \right)^\beta.$$

Using Lemma 13 to the above, we can get the first part of the desired results.
Now consider the case w_t. With $\eta_t = \eta t^{-\theta}$, (24) is exactly

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)] \leq 2(a_0\kappa)^2 \frac{\eta \sum_{k=1}^t k^{-\theta}}{m} \left(1 + 2t^\theta \sum_{k=1}^{t-1} \frac{k^{-\theta}}{t-k}\right) + \frac{(a_0\kappa)^2}{2} \eta t^{-\theta} \left(t^{2\theta} \sum_{k=1}^{t-1} \frac{k^{-2\theta}}{t-k} + 1\right) + c_\beta \frac{\left(\sum_{k=1}^t k^{-\theta}\right)^{1-\beta}}{\eta^\beta t^{1-\theta}}.$$

Applying Lemma 14 to bound $\sum_{k=1}^{t-1} \frac{k^{-\theta}}{t-k}$ and $\sum_{k=1}^{t-1} \frac{k^{-2\theta}}{t-k}$, by a simple calculation, we derive

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)] \leq 2(a_0\kappa)^2 \frac{\eta \sum_{k=1}^t k^{-\theta}}{m} \cdot (9\log t) + 3(a_0\kappa)^2 \eta t^{-\min(\theta,1-\theta)} \log t + c_\beta \frac{\left(\sum_{k=1}^t k^{-\theta}\right)^{1-\beta}}{\eta^\beta t^{1-\theta}}.$$

Using Lemma 11 to upper bound $\sum_{k=1}^t k^{-\theta}$, one can get the second part of the desired results. □

Proposition 4. *Under Assumptions 1 and 2, let $\eta_t = \eta t^{-\theta}$ for all $t \in \mathbb{N}$, with $0 < \eta \leq 1$ and $\theta \in [0,1)$. Then for all $t \in \mathbb{N}$,*

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(\overline{w}_t)] - \inf_{w\in\mathcal{F}} \mathcal{E}(w) \leq 2a_0\kappa \sqrt{\frac{(a_0\kappa)^2 + 2|V|_0}{1-\theta}} \sqrt{\frac{\eta t^{1-\theta}}{m}} + \frac{(a_0\kappa)^2(1-\theta)}{1-4^{\theta-1}} \eta t^{-\min(\theta,1-\theta)} \log t + c_\beta \left(\frac{1-\theta}{1-4^{\theta-1}}\right)^\beta \left(\frac{1}{\eta t^{1-\theta}}\right)^\beta,$$

and

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)] \leq 18a_0\kappa \sqrt{\frac{(a_0\kappa)^2 + 2|V|_0}{1-\theta}} \sqrt{\frac{\eta t^{1-\theta}}{m}} \log t + 3(a_0\kappa)^2 \eta t^{-\min(\theta,1-\theta)} \log t + \frac{c_\beta}{1-\theta} \left(\frac{1}{\eta t^{1-\theta}}\right)^\beta.$$

Proof. Following the proof of Theorem 3, we have (26) and (27), where R_t satisfies (29). Comparing (26), (27) with (23), (24), we find that the differences are the terms related sample errors, i.e., the term $2(a_0\kappa^2) \sum_{k=1}^t \eta_k/m$ in (23), (24), while $2a_0\kappa R_t/\sqrt{m}$ in (26), (27). Thus, following from the proof of Proposition 3, we get

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(\overline{w}_t)] - \inf_{w\in\mathcal{F}} \mathcal{E}(w) \leq \frac{2a_0\kappa R_t}{\sqrt{m}} + \frac{(a_0\kappa)^2(1-\theta)}{1-4^{\theta-1}} \eta t^{-\min(\theta,1-\theta)} \log t + c_\beta \left(\frac{1-\theta}{1-4^{\theta-1}}\right)^\beta \left(\frac{1}{\eta t^{1-\theta}}\right)^\beta,$$

and

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)] \leq \frac{2a_0\kappa R_t}{\sqrt{m}} \cdot 9\log t + 3(a_0\kappa)^2 \eta t^{-\min(\theta,1-\theta)} \log t + \frac{c_\beta}{1-\theta} \left(\frac{1}{\eta t^{1-\theta}}\right)^\beta.$$

Recall that R_t satisfies (29), with $\eta_t = \eta t^{-\theta}$, where $\eta \leq 1$, by Lemma 11, we know that

$$R_t \leq \sqrt{\frac{(a_0\kappa)^2 + 2|V|_0}{1-\theta}} \sqrt{\eta t^{1-\theta}}.$$

From the above analysis, one can conclude the proof. □

We are ready to prove Corollaries 1-8.

Proof of Corollary 1. Applying Proposition 3 with $\theta = 0$, $\eta = \eta_1/\sqrt{m}$, we derive

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(\overline{w}_t)] - \inf_{w\in\mathcal{F}} \mathcal{E}(w) \leq 2\eta_1(a_0\kappa)^2 \frac{t}{\sqrt{m^3}} + 2(a_0\kappa)^2 \eta_1 \frac{\log t}{\sqrt{m}} + \frac{2c_\beta}{\eta_1^\beta} \left(\frac{\sqrt{m}}{t}\right)^\beta,$$

and

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)] \leq 18\eta_1(a_0\kappa)^2 \frac{t\log t}{\sqrt{m^3}} + 3\eta_1(a_0\kappa)^2 \frac{\log t}{\sqrt{m}} + \frac{c_\beta}{\eta_1^\beta}\left(\frac{\sqrt{m}}{t}\right)^\beta.$$

The proof is complete. □

Proof of Corollary 2. Applying Proposition 3 with $\eta = \eta_1, \theta = 1/2$, we get

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(\overline{w}_t)] - \inf_{w\in\mathcal{F}} \mathcal{E}(w) \leq 4(a_0\kappa)^2\eta_1 \frac{\sqrt{t}}{m} + (a_0\kappa)^2\eta_1 \frac{\log t}{\sqrt{t}} + c_\beta \eta_1^{-\beta}\frac{1}{t^{\beta/2}},$$

and

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)] \leq 36(a_0\kappa)^2\eta_1 \frac{\sqrt{t}\log t}{m} + 3(a_0\kappa)^2\eta_1 \frac{\log t}{\sqrt{t}} + 2c_\beta \eta_1^{-\beta}\frac{1}{t^{\beta/2}}.$$

□

Proof of Corollary 3. Applying Proposition 3 with $\theta = 0$ and $\eta = \eta_1 m^{-\frac{\beta}{\beta+1}}$, we get

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(\overline{w}_t)] - \inf_{w\in\mathcal{F}} \mathcal{E}(w) \leq 2\eta_1(a_0\kappa)^2 m^{-\frac{\beta+2}{\beta+1}} t + 2\eta_1(a_0\kappa)^2 m^{-\frac{\beta}{\beta+1}}\log t + \frac{2c_\beta}{\eta_1^\beta} m^{\frac{\beta^2}{\beta+1}} t^{-\beta},$$

and

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)] \leq 18\eta_1(a_0\kappa)^2 m^{-\frac{\beta+2}{\beta+1}} t\log t + 3\eta_1(a_0\kappa)^2 m^{-\frac{\beta}{\beta+1}}\log t + \frac{c_\beta}{\eta_1^\beta} m^{\frac{\beta^2}{\beta+1}} t^{-\beta}.$$

The proof is complete. □

Proof of Corollary 4. Applying Proposition 3 with $\eta = \eta_1$ and $\theta = \frac{\beta}{\beta+1}$,

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(\overline{w}_t)] - \inf_{w\in\mathcal{F}} \mathcal{E}(w) \leq 4(a_0\kappa)^2\eta_1 \frac{t^{\frac{1}{\beta+1}}}{m} + 2\eta_1(a_0\kappa)^2 t^{-\frac{\beta}{\beta+1}} + 2c_\beta\eta_1^{-\beta} t^{-\frac{\beta}{\beta+1}},$$

and

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)] \leq 36(a_0\kappa)^2\eta_1 \frac{t^{\frac{1}{1+\beta}}\log t}{m} + 3\eta_1(a_0\kappa)^2 t^{-\frac{\beta}{\beta+1}}\log t + 2c_\beta\eta_1^{-\beta} t^{-\frac{\beta}{\beta+1}}.$$

For the above two inequalities, we used that $\beta \in (0,1]$, $\theta = \frac{\beta}{\beta+1} \leq 1/2$ and $4^{\theta-1} \leq 1/2$. □

Proof of Corollary 5. Applying Proposition 4 with $\eta = 1/\sqrt{m}$ and $\theta = 0$,

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(\overline{w}_t)] - \inf_{w\in\mathcal{F}} \mathcal{E}(w) \leq 2a_0\kappa\sqrt{(a_0\kappa)^2 + 2|V|_0}\frac{\sqrt{t}}{m^{3/4}} + 2(a_0\kappa)^2\frac{\log t}{\sqrt{m}} + 2c_\beta\left(\frac{\sqrt{m}}{t}\right)^\beta,$$

and

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w\in\mathcal{F}} \mathcal{E}(w)] \leq 18a_0\kappa\sqrt{(a_0\kappa)^2 + 2|V|_0}\frac{\sqrt{t}\log t}{m^{3/4}} + 3(a_0\kappa)^2\frac{\log t}{\sqrt{m}} + c_\beta\left(\frac{\sqrt{m}}{t}\right)^\beta.$$

The proof is complete. □

Proof of Corollary 6. Applying Proposition 4 with $\eta = 1$ and $\theta = 1/2$, we get

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(\overline{w}_t)] - \inf_{w \in \mathcal{F}} \mathcal{E}(w) \leq 2\sqrt{2} a_0 \kappa \sqrt{(a_0\kappa)^2 + 2|V|_0} \frac{t^{1/4}}{\sqrt{m}} + \frac{(a_0\kappa)^2 \log t}{\sqrt{t}} + \frac{c_\beta}{t^{\beta/2}}.$$

and

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w \in \mathcal{F}} \mathcal{E}(w)] \leq 18\sqrt{2} a_0 \kappa \sqrt{(a_0\kappa)^2 + 2|V|_0} \frac{t^{1/4} \log t}{\sqrt{m}} + 3(a_0\kappa)^2 \frac{\log t}{\sqrt{t}} + 2c_\beta \frac{1}{t^{\beta/2}}.$$

□

Proof of Corollary 7. Using Proposition 4 with $\eta = m^{-\frac{2\beta}{2\beta+1}}$ and $\theta = 0$, we get

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(\overline{w}_t)] - \inf_{w \in \mathcal{F}} \mathcal{E}(w) \leq 2a_0\kappa \sqrt{(a_0\kappa)^2 + 2|V|_0} m^{-\frac{4\beta+1}{4\beta+2}} \sqrt{t} + 2(a_0\kappa)^2 m^{-\frac{2\beta}{2\beta+1}} \log t + 2c_\beta m^{\frac{2\beta^2}{2\beta+1}} t^{-\beta},$$

and

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w \in \mathcal{F}} \mathcal{E}(w)] \leq 18a_0\kappa \sqrt{(a_0\kappa)^2\eta + 2|V|_0} m^{-\frac{4\beta+1}{4\beta+2}} \sqrt{t} \log t + 3(a_0\kappa)^2 m^{-\frac{2\beta}{2\beta+1}} \log t + c_\beta m^{\frac{2\beta^2}{2\beta+1}} t^{-\beta}.$$

The proof is complete. □

Proof of Corollary 8. Let $\theta = \frac{2\beta}{2\beta+1}$. Obviously, $\theta \in [0, \frac{2}{3}]$ since $\beta \in (0, 1]$. Thus, $\frac{1}{1-\theta} = 2\beta + 1 \leq 3$, $\frac{1-\theta}{1-4^{\theta-1}} \leq \frac{1}{1-4^{-1/3}} \leq 2$. Following from Proposition 4,

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(\overline{w}_t)] - \inf_{w \in \mathcal{F}} \mathcal{E}(w) \leq 2\sqrt{3} a_0\kappa \sqrt{(a_0\kappa)^2 + 2|V|_0} \frac{t^{\frac{1}{4\beta+2}}}{\sqrt{m}} + 2(a_0\kappa)^2 t^{-\frac{\min(2\beta,1)}{2\beta+1}} \log t + 2c_\beta t^{-\frac{\beta}{2\beta+1}},$$

and

$$\mathbb{E}_{\mathbf{z},J}[\mathcal{E}(w_t) - \inf_{w \in \mathcal{F}} \mathcal{E}(w)] \leq 18\sqrt{3} a_0\kappa \sqrt{(a_0\kappa)^2 + 2|V|_0} \frac{t^{\frac{1}{4\beta+2}}}{\sqrt{m}} \log t + 3(a_0\kappa)^2 t^{-\frac{\min(2\beta,1)}{2\beta+1}} \log t + 3c_\beta t^{-\frac{\beta}{2\beta+1}}.$$

□