Never Go Full Batch: Stochastic TMLE for Large-Scale Debiased Inference

Diyang Li
Cornell University
diyang01@cs.cornell.edu

Fei Wang
Cornell University
few2001@med.cornell.edu

Kyra Gan
Cornell University
kyragan@cornell.edu

Abstract

Targeted Minimum Loss-based Estimation (TMLE) is a classical plug-in debiasing methodology that delivers doubly robust estimation and asymptotically efficient inference. Despite its statistical guarantees, each iteration of standard TMLE requires solving a minimization subproblem over the entire dataset, resulting in prohibitive computational and memory costs that render the method impractical for large-scale or real-time settings. To ease the bottlenecks, we introduce stochastic TMLE, a randomized targeting procedure that replaces each expensive full-batch fluctuation fit with mini-batch alternatives. Our theoretical analysis establishes that the stochastic iterates converge to a neighborhood (noise ball) centered around the target solution, and crucially, only a small number of subsequent full-sample TMLE iterations suffice to reach an empirical efficient influence function root. Consequently, our stochastic variant inherits the same guarantees and attractive properties as classical TMLE while substantially reducing per-iteration complexity. We further propose two computational variants with provable guarantees that broaden the algorithmic design space, offering flexibility for future developments in scalable targeted learning. Extensive experiments across diverse regimes demonstrate substantial acceleration over standard TMLE without sacrificing inferential quality.

1 Introduction

Targeted Minimum Loss-based (or Maximum Likelihood) Estimation, abbreviated TMLE [1, 2], is a powerful plug-in debiasing methodology that achieves doubly robust estimation and asymptotically efficient inference for a broad class of semiparametric target parameters. The method is widely used in problems where flexible machine learning is essential for estimating nuisance components, yet valid inference for a low-dimensional scientific estimand remains the primary goal, including treatment effect evaluation [3], missing data problems [4], policy values [5], and regression functionals [6]. While conceptually simple, naive plug-in estimators often inherit substantial bias from the initial estimator, compromising both point estimation and downstream inference [7, 8].

At a high level, TMLE provides an elegant remedy to this bias problem. Rather than accepting the plug-in estimate as given, TMLE starts from an initial estimate of the data generating distribution and then performs a sequence of targeting updates. Each update is defined through a parametric fluctuation submodel and an associated loss, and the fluctuation parameter is chosen by minimizing the current empirical loss. This targeting step is designed so that the score of the fluctuation at the origin spans the efficient influence function (EIF), which drives the updated plug-in estimator toward solving an empirical EIF estimating-equation [1]. TMLE is particularly attractive because it produces a targeted plug-in estimator that corrects first-order bias while preserving valid probabilistic structure, yielding double robustness and efficiency under standard regularity conditions [9, 10]. Owing to these properties, TMLE has become a workhorse framework in causal inference [11], personalized medicine [12], and decision-making [13].

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

Table 1: Overview of the major algorithms and guarantees in this work.

Name	Algorithm	Stochastic Sample	Stochastic Dimension	Convergence	Reference
TMLE	Algorithm 1	✗	✗	✓	[1]
STMLE	Algorithm 2	✓	✗	Noise Ball	**Ours**
S^+TMLE	Algorithm 3	✓	✗	✓	**Ours**
S^2TMLE	Algorithm 4	✓	✓	Noise Ball	**Ours**

Despite its theoretical appeal, the computational cost of standard TMLE is increasingly misaligned with modern large-scale learning practice [14, 13]. Technically, each iteration requires solving an empirical risk minimization problem for the fluctuation parameter using the *entire* dataset. When the sample sizes are large, nuisance learners are expensive, or coupled with modern workflows (e.g., cross-fitting, high-dimensional representations, or complicated loss structures), this repeated full batch training carries a nontrivial cost and dominates runtime and memory usage [15]. This bottleneck is especially acute in settings that demand low latency or limited storage, where repeatedly revisiting the whole dataset is not feasible [16].

The tension between statistical efficiency and computational scalability motivates a natural question: can we preserve the debiasing and inferential guarantees of TMLE while substantially reducing the training cost? In the optimization community, the transition from gradient descent to stochastic gradient descent (SGD) revolutionized large-scale machine learning by replacing full-batch gradient evaluations with inexpensive mini-batch approximations [17–19]. The resulting algorithms trade a modest increase in iterate variance for dramatic reductions in per-iteration cost, and under appropriate conditions converge to neighborhoods of optimal solutions [20, 21]. The SGD paradigm now underpins the training of virtually all modern machine learning systems [22, 23], and its philosophy motivates a stochastic analogue of the targeting mechanism that preserves the properties of TMLE while reducing the optimizing burden.

Technical Challenges. However, naively transplanting stochastic optimization ideas into TMLE is not straightforward. In classical stochastic gradient analysis the iterate is a finite-dimensional vector and the objective remains fixed. In contrast, the TMLE's iterate represents an estimated distribution and the targeting objective depends on nuisance fits that change after every update. Moreover, a mini-batch fluctuation fit produces the minimizer of a subsampled empirical risk, which is not an unbiased proxy for the full sample minimizer. The induced update is therefore not a standard martingale perturbation of the batch algorithm. Its noise is state-dependent and it may be correlated across iterations. Finally, stochastic targeting introduces a persistent optimization error that does not vanish just by iterating longer. One must show that this error concentrates into a small neighborhood whose radius matches the statistical rate. These intertwined difficulties explain why stochastic targeting has been rarely explored in the TMLE literature.

Our Contributions. We formally introduce Stochastic TMLE (STMLE), a randomized extension of classical TMLE that performs fluctuation updates using only a small stochastically sampled subset of observations per iteration. Despite its conceptual simplicity, STMLE proves surprisingly effective. Our theoretical analysis shows that the stochastic iterates converge to a noise ball centered around the classical TMLE solution, mirroring the well-known behavior of SGD under step size regularities [20, 21]. Empirically, the algorithm oscillates within a neighborhood of the optimal solution and requires only a few subsequent full-sample TMLE iterations to achieve exact convergence. In other words, STMLE ultimately attains a root of the EIF estimating-equation in practice, thereby inheriting all the attractive properties of standard TMLE including double robustness and asymptotic efficiency.

Building on STMLE, we further develop two variants that broaden the algorithmic design space. The first variant is adaptive STMLE (S^+TMLE), which abandons the fixed batch size in favor of an adaptive schedule inspired by self-paced learning (SPL) [24], a well-established optimization paradigm in machine learning. By gradually increasing the batch size over iterations, S^+TMLE leverages the benefits of learning-to-optimize [25] and may exhibit improved robustness in high-noise regimes from SPL. This adaptive mechanism makes the method gradually approach standard TMLE on its own (i.e. self-paced), and eliminates the need for an explicit switch to standard TMLE for final refinement. The second variant is doubly stochastic TMLE (S^2TMLE), which adopts a doubly stochastic scheme inspired by doubly stochastic gradient methods [26, 27]. In addition to STMLE, S^2TMLE subsamples coordinate directions of the fluctuation parameter, which is particularly advantageous when the target parameter is high-dimensional since the subproblem variable shares

the equivalent dimensionality by construction. Such random subspace update can also reduce the memory complexity in that regime. A comparison of all algorithms is provided in Table 1.

1.1 Related Work

The works most closely related to ours are those of Bertail et al. [28] and Chambaz et al. [29], which also address the challenge of applying TMLE to large-scale settings. Both studies investigate a survey sampling strategy that selects a representative subset from the full dataset according to carefully designed sampling rules *before* the optimization procedure begins. Standard or weighted TMLE is then executed on this fixed subset, and valid confidence intervals for the target parameter are derived under appropriate sampling assumptions. While effective in reducing computational burden, this pre-selection paradigm differs fundamentally from our work. Our approach instead accelerates the targeting iterations themselves and it repeatedly resamples observations across steps so that information from the entire dataset can be loaded over the run.

Beyond scalability considerations, a separate line of research has focused on improving the statistical properties of the TMLE itself. Notable contributions include higher-order TMLE [30, 31], kernel debiased plug-in estimation [32], and score-preserving TMLE [33]. These methods are largely orthogonal to our contribution and could potentially be combined with stochastic updates in future work. Finally, we note that van der Laan and Lendle [14] propose a method termed Online Targeted Learning. Despite the suggestive name, this approach does not constitute an online variant of vanilla TMLE [1] in the sense we consider here. Their framework targets a one-shot setting where data arrive sequentially and are processed exactly once, and the procedure does not involve iterative loss minimization over fluctuation submodels.

2 TMLE Revisited

This section briefly revisits the targeted learning setup and the classical TMLE targeting procedure.

Setup and Notation. Let $O_1, \dots, O_n$ be n independent and identically distributed observations (samples) drawn from an unknown data-generating distribution $\mathbf{P}_0$ on a measurable sample space. Throughout this work, our algorithmic analysis is conducted conditionally on the *realized* dataset $\{O_i\}_{i=1}^n$, with the sample size $n < \infty$ held fixed. Let $\mathcal{M}$ denote a statistical model, namely a collection of probability distributions known to contain the true distribution $\mathbf{P}_0$. Let $\Psi : \mathcal{M} \to \mathbb{R}^d$ be a d-dimensional Euclidean target parameter of interest. We denote by $\|\cdot\|$ the Euclidean norm.

We assume that the target functional Ψ is pathwise differentiable at every $\mathbf{P} \in \mathcal{M}$, and admits a pointwise well-defined efficient influence function (EIF; a.k.a. canonical gradient) $D^*(\mathbf{P})(o)$ for all $\mathbf{P}$ and all o. In plug-in estimation, we typically construct an estimator $\mathbf{P}$ of $\mathbf{P}_0$ and report $\Psi(\mathbf{P})$. In many practical settings, the target functional depends on $\mathbf{P}$ only through a relevant portion $\mathbf{Q}(\mathbf{P})$, provided that the EIF $D^*(\mathbf{P}) := D^*(\mathbf{Q}, \mathbf{G})$ can be written in terms of $\mathbf{Q}(\mathbf{P})$ and a nuisance component $\mathbf{G}(\mathbf{P})$ [2, 13]. For notational convenience, we write $\Psi(\mathbf{Q}(\mathbf{P})) = \Psi(\mathbf{P})$ with a slight abuse of notation.

Algorithm Overview. Given an initial estimator $\mathbf{P}_n^0$ of the data-generating distribution $\mathbf{P}_0$, the TMLE algorithm takes $\mathbf{Q}_0 := \mathbf{Q}(\mathbf{P}_n^0)$ as input and outputs a debiased component $\hat{\mathbf{Q}}$ that is then plugged into Ψ. The procedure iterates through a sequence of updates, and we denote the iterate at step k by $\mathbf{Q}_k$. By construction, TMLE relies on a proper loss function $(o, \mathbf{Q}) \mapsto \mathcal{L}(\mathbf{Q})(o)$ satisfying

$$\mathbf{Q}(\mathbf{P}_0) \in \arg\min_{\mathbf{Q} \in \widetilde{\mathbf{Q}}} \mathbb{E}_{o \sim \mathbf{P}_0}\big[\mathcal{L}(\mathbf{Q})\big], \quad \widetilde{\mathbf{Q}} := \big\{\mathbf{Q}(\mathbf{P}) : \mathbf{P} \in \mathcal{M}\big\}.$$

TMLE further specifies a fluctuation submodel $\mathbf{Q}_k(\epsilon \mid \mathbf{G})$, sometimes called a parametric working model [34], indexed by a d-dimensional ϵ and anchored at the current iterate, meaning that $\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big|_{\epsilon=\mathbf{0}} = \mathbf{Q}_k$ for all k. The submodel is required to satisfy the score condition stated in Definition 2.1, ensuring that the fluctuation directions are rich enough to target the EIF of Ψ.

Definition 2.1 (Generalized Score Condition [1])**.** Consider a fluctuation submodel $\{\mathbf{Q}(\epsilon \mid \mathbf{G}) : \epsilon\}$ in $\widetilde{\mathbf{Q}}$ so that the linear span of $\frac{d}{d\epsilon}\mathcal{L}\big(\mathbf{Q}(\epsilon \mid \mathbf{G})\big)(o)\big|_{\epsilon=\mathbf{0}}$ contains the linear span of the $D^*(\mathbf{Q}, \mathbf{G})(o)$.

The fluctuation parameter ϵ is constrained to lie in the feasible set $\mathcal{E} := \big\{\epsilon \in \mathbb{R}^d : \mathbf{Q}_k(\epsilon \mid \mathbf{G}) \in \widetilde{\mathbf{Q}}\big\}$. At each iteration, TMLE chooses ϵ by minimizing the empirical loss along the submodel, and updates

$\mathbf{Q}_{k+1}$ by evaluating the submodel at that specific ϵ. The resulting batch procedure is shown in Algorithm 1, whose convergence is characterized by $\epsilon_n^k = \mathbf{0}$ as defined in the original TMLE literature [1]. In practice, iterations continue until $\|\epsilon_n^k\|$ falls below a pre-specified tolerance or $k = K$. The desired convergence guarantee of Algorithm 1 is established by [35] under standard regularity conditions. Upon the arrival of convergence, the well-studied targeted learning theory [2] guarantees that the final iterate satisfies the empirical EIF estimating-equation

$$\frac{1}{n}\sum_{i=1}^{n} D^*(\hat{\mathbf{Q}}, \hat{\mathbf{G}})(O_i) = \mathbf{0}, \qquad (1)$$

which serves as the key bridge to the desirable properties of TMLE, including double robustness and asymptotic efficiency. We refer readers to [1] and [36] for detailed discussions of the underlying assumptions.

Algorithm 1 Targeted Minimum Loss-based Estimation

Input: Dataset $\{O_i\}_{i=1}^n$, initial fit $\mathbf{Q}_0$, iterations K, target parameter Ψ.

1: **for** $k = 0, \ldots, K-1$ **do**

2: $\quad \epsilon_n^k = \arg\min_{\epsilon \in \mathcal{E}} \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i)$

3: $\quad \mathbf{Q}_{k+1} = \mathbf{Q}_k(\epsilon_n^k \mid \mathbf{G})$

4: **end for**

Output: Targeted estimator $\hat{\Psi}_n = \Psi(\mathbf{Q}_K)$.

Regularity Conditions. We adopt the following standard conditions throughout this work.

Assumption 2.2. (Differentiability [9]) The map $\epsilon \mapsto \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(o)$ is differentiable at $\epsilon = \mathbf{0}$ for all k, o, and the subproblem (Line 2 of Algorithm 1) admits a solution strictly in the interior of $\mathcal{E}$.

Assumption 2.3 (Boundedness [2])**.** There exist $L_{\min} > -\infty$ and $L_{\max} < \infty$ such that $\inf_{\mathbf{Q}\in\widetilde{\mathbf{Q}}} \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}(\mathbf{Q})(O_i) \geq L_{\min}$ for all k, and $\big|\mathcal{L}(\mathbf{Q}_k(\epsilon \mid \mathbf{G}))(o)\big| \leq L_{\max}$ for all o and $\epsilon \in \mathcal{E}$.

These conditions ensure that the empirical loss remains bounded away from $\pm\infty$ along the entire optimization trajectory. For our analysis of the stochastic variants, we impose the following additional weak assumptions. In practice, these conditions are very mild and are satisfied naturally in essentially all standard TMLE implementations on finite datasets.

Assumption 2.4 (Compact Search Set)**.** The constraint set $\mathcal{E} \subset \mathbb{R}^d$ is compact with finite diameter $R_{\mathcal{E}} := \sup_{\epsilon,\epsilon' \in \mathcal{E}} \|\epsilon - \epsilon'\| < \infty$.

Assumption 2.5 (Lipschitzness Along Submodel [35])**.** There exists a constant $L_\epsilon < \infty$ such that for all k, all o, and all $\epsilon, \epsilon' \in \mathcal{E}$, $\big|\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(o) - \mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(o)\big| \leq L_\epsilon \|\epsilon - \epsilon'\|$.

3 Stochastic TMLE Algorithm

We now present Stochastic TMLE (STMLE), our proposed (application-agnostic) randomized extension of the classical TMLE algorithm. An important practical consideration motivates our approach. Standard TMLE (Algorithm 1) targets the initial plug-in estimator by repeatedly solving a *full-sample* empirical risk minimization problem over the parameter space $\mathcal{E}$. Such targeting stage is the computational bottleneck: updating the $\mathbf{Q}_{k+1}$ and evaluating the final plug-in $\Psi(\hat{\mathbf{Q}})$ are typically cheap, while repeatedly fitting subproblems against the entire dataset is expensive and memory intensive. This asymmetry between cheap inference and expensive optimization brings about our STMLE procedure, where the key modification is conceptually simple.

At each iteration k, STMLE samples a fresh mini-batch $\mathcal{B}^k$ (subset of indices) and computes the fluctuation parameter ϵ_k by minimizing the mini-batch empirical loss rather than the full-sample loss. It then performs the same anchored fluctuation update as TMLE. The complete pseudocode is described in Algorithm 2, which is applicable to virtually *any* statistical model (functional) where the vanilla TMLE is valid. One might also note that our approach differs fundamentally from naively applying SGD to solve the subproblem, both in theory and in implementation. The modification we propose operates at the level of algorithmic construction itself, and is agnostic to the choice of specific subproblem solver, which is merely an implementation detail. Furthermore, for the sake of rigor, we adopt a deterministic tie-breaking in Line 3 of Algorithm 2 (e.g., lexicographically smallest ϵ).

Algorithm 2 Stochastic TMLE (STMLE)

Input: Dataset $\{O_i\}_{i=1}^n$, initial fit $\mathbf{Q}_0$, iterations K, target parameter Ψ, batch size B.

1: **for** $k = 0, \ldots, K-1$ **do**

2: $\quad$ Sample a mini-batch $\mathcal{B}^k \subset \{1, \ldots, n\}$

3: $\quad \epsilon_k = \arg\min_{\epsilon \in \mathcal{E}} \frac{1}{B}\sum_{i\in\mathcal{B}^k} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i)$

4: $\quad \mathbf{Q}_{k+1} = \mathbf{Q}_k(\epsilon_k \mid \mathbf{G})$

5: **end for**

Output: Targeted estimator $\hat{\Psi}_n = \Psi(\mathbf{Q}_K)$.

3.1 Convergence

In deterministic TMLE, the convergence criterion $\epsilon_n^k \rightsquigarrow \mathbf{0}$ is equivalent to the condition that the empirical loss along the chosen fluctuation submodel is minimized at $\epsilon = \mathbf{0}$. Motivated by this characterization, we consider the following measure of *distance to convergence* as

$$\Delta(q^k) \triangleq \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}(\mathbf{Q}_k)(O_i) - \inf_{\epsilon\in\mathcal{E}}\frac{1}{n}\sum_{j=1}^{n}\mathcal{L}(\mathbf{Q}_k(\epsilon \mid \mathbf{G}))(O_j),$$

where $\mathcal{E} \subset \mathbb{R}^d$ is the optimization domain introduced in Section 2. This quantity measures the gap between the current empirical loss and the minimum achievable loss along the fluctuation submodel. In standard TMLE, exact convergence implies $\epsilon_n^k = \mathbf{0}$, hence $\Delta(q^k) = 0$.

In STMLE, even if the iterate is already close to a full-sample stationary point, a fresh mini-batch can suggest a nonzero update. Analogous to stochastic optimization, one should not expect $\Delta(q^k) \rightsquigarrow 0$ by iterating longer; instead, the stochastic iterates typically oscillate in a neighborhood of the deterministic solution. We formalize this 'noise ball' behavior by showing that the long-run average of $\mathbb{E}[\Delta(q^k)]$ is bounded above.

Theorem 3.1 (Ergodic Expected Convergence)**.** *If $\mathcal{B}^k$ is sampled uniformly with replacement,*

$$\limsup_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}\left[\Delta(q^k)\right] \le 2\frac{L_{\max}}{\sqrt{B}}\left(3+\sqrt{2\left(\log 2 + d\log\left(1+\frac{R_{\mathcal{E}}L_{\epsilon}\sqrt{B}}{L_{\max}}\right)\right)}\right). \tag{2}$$

Suppose that the $\mathcal{B}^k$ is sampled uniformly without replacement. Then we have

$$\limsup_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}\left[\Delta(q^k)\right] \le 48\left(\sqrt{\log 2}+\frac{\sqrt{\pi}}{2}\right)L_{\epsilon}R_{\mathcal{E}}\sqrt{\frac{\rho_{B,n}d}{B}} + 2L_{\max}\sqrt{\frac{2\rho_{B,n}}{B}}, \tag{3}$$

where

$$\rho_{B,n} = \begin{cases} 1-\frac{B-1}{n}, & B \le \frac{n}{2}, \\ \left(1-\frac{B}{n}\right)\left(1+\frac{1}{B}\right), & B > \frac{n}{2}. \end{cases} \tag{4}$$

This ergodic-convergence result parallels the classical guarantees established for the vanilla TMLE procedure [9, 35]. Theorem 3.1 shows that Algorithm 2 admits a convergence guarantee to a ball whose radius shrinks as B grows, with the order of $\mathcal{O}\left(\frac{1+\sqrt{d\log B}}{\sqrt{B}}\right)$ for sampling with replacement, and the order of $\mathcal{O}\left((1+\sqrt{d})\sqrt{\frac{\rho_{B,n}}{B}}\right)$ for sampling without replacement. One may observe that the bound in (3) vanishes when $B = n$. However, this does not automatically ensure exact convergence of the algorithm(s). For further details, we refer the reader to Section 4 of [35].

From Noise Ball to Exact Convergence. When implementing, we invest most computational effort in fast, stochastic iterations that rapidly approach the solution manifold of the EIF estimating-equation, and then require only a small number of final full-batch *refinement steps* to attain an exact root of (1). This hybrid strategy combines the efficiency with exact convergence of standard TMLE.

3.2 Escaping Suboptimality: A Case Study

Beyond computational acceleration, stochastic fluctuation updates offer a potential statistical advantage analogous to the ability of SGD to escape saddle points and poor local minima. In the semi-parametric context, the estimating-equation (1) may admit multiple roots [37], and different roots can correspond to different values of loss even though they all satisfy the same EIF moment condition. Standard TMLE is deterministic and converges to a fixed reachable root determined by the $\mathbf{P}_n^0$. In contrast, the randomness in STMLE can enable exploration of the solution manifold and potentially lead to solutions with better empirical fit.

Let $\mathbf{Q}_{\mathrm{TMLE}}$ denote the deterministic limit obtained by running vanilla TMLE from the initial estimator $\mathbf{Q}(\mathbf{P}_n^0)$. For each specific sequence of mini-batches $(\mathcal{B}^0,\ldots,\mathcal{B}^{K-1})$, let $\mathbf{Q}_{\mathrm{ST}}(\mathcal{B}^0,\ldots,\mathcal{B}^{K-1})$ denote the final deterministic output obtained by first running STMLE for K iterations using exactly $\mathcal{B}^0,\ldots,\mathcal{B}^{K-1}$, then running vanilla TMLE from the resulting iterate $\mathbf{Q}_K$ until convergence. We measure the *quality* of an estimate $\mathbf{Q}$ via the empirical cost (training fit) $\mathbf{J}(\mathbf{Q}) := \sum_{i=1}^{n}\mathcal{L}(\mathbf{Q})(O_i)$, whose gradient coincides with the empirical EIF residual and therefore serves as a transparent

diagnostic for semiparametric efficiency. We emphasize that $\mathbf{J}(\mathbf{Q})$ is used here only as a transparent proxy for comparing solutions; in applications, one could replace it by e.g., cross-validated criterion. Define the best achievable cost as

$$\mathbf{J}_K^\star := \min_{\substack{\mathcal{B}^0,\dots,\mathcal{B}^{K-1}\subset\{1,\dots,n\} \\ |\mathcal{B}^0|=\cdots=|\mathcal{B}^{K-1}|=B}} \mathbf{J}\big(\mathbf{Q}_{\mathrm{ST}}(\mathcal{B}^0,\dots,\mathcal{B}^{K-1})\big)\,. \tag{5}$$

Lemma 3.2 (Unconditional Gain)**.** *We have*

$$\mathbb{E}\big[\min\big\{\mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}),\mathbf{J}(\mathbf{Q}_{\mathrm{ST}})\big\}\big] \le \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \big(\mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbf{J}_K^\star\big)\tbinom{n}{B}^{-K},$$

where $\binom{n}{B}$ *denotes the binomial coefficient.*

Lemma 3.2 shows that the set of stochastic trajectories contains a strictly improving candidate with nonzero probability.[1] To obtain more insightful guarantees, we consider a stylized but useful regime where only finitely many empirical EIF roots are reachable by semi-parametric TMLE procedure, and the last stochastic mini-batch step effectively selects among their basins.

Assumption 3.3 (Finite Reachable EIF-root Set)**.** There exists a data-dependent subset $\big\{\mathbf{Q}_{(1)},\dots,\mathbf{Q}_{(m)}\big\} \subset \widetilde{\mathbf{Q}}$ with $m \ge 2$ such that the deterministic TMLE limit can take only values in this set, and each $\mathbf{Q}_{(j)}$ satisfies equation (1).

Assumption 3.4 (Basin-selection Consistency)**.** If $\mathcal{B}$ is the last mini-batch used to produce $\mathbf{Q}_K$, then the TMLE limit of $\mathbf{Q}_K$ is determined by which reachable root has the smallest mini-batch empirical loss on that same $\mathcal{B}$.

In words, Assumption 3.4 captures the intuition that mini-batch randomness can redirect the optimization trajectory toward different basins of attraction. Let $j^\star \in \arg\min_{1\le j\le m} \mathbf{J}\big(\mathbf{Q}_{(j)}\big)$ and $\Omega \triangleq \min_{j\ne j^\star}\big[\mathbf{J}\big(\mathbf{Q}_{(j)}\big) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big)\big]/n$. Based on the Lemma 3.2, we can have a improvement bound as follows.

Theorem 3.5. *Under Assumption 3.3, 3.4 we have*

$$\mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbb{E}\left[\mathbf{J}(\mathbf{Q}_{\mathrm{ST}})\right] \ge \big(\mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbf{J}(\mathbf{Q}_{(j^\star)})\big) - 2nL_{\max}\min\big\{1,\,(m-1)\,e^{-B\Omega^2/(8L_{\max}^2)}\big\}.$$

Remark 3.6. Under reasonable choice of the batch size (see (8) in Appendix A), we can obtain the strict inequality $\mathbb{E}\left[\mathbf{J}\big(\mathbf{Q}_{\mathrm{ST}}\big)\right] < \mathbf{J}\big(\mathbf{Q}_{\mathrm{TMLE}}\big)$. Theorem 3.5 states that if vanilla TMLE converges to a suboptimal root, then the hybrid refinement achieves strictly better expected loss under appropriate magnitude of B. We leave a more in-depth analysis of this escape phenomenon to future work.

4 Algorithmic Variants

In this section, we present two computational variants of STMLE, namely S$^+$TMLE and S^{2}TMLE, which expand the algorithmic design space and pave the way for future developments.

4.1 S$^+$TMLE: Adaptive Batch Sizes

STMLE with a fixed batch size enjoys the noise-ball convergence guarantees, but the residual error $\Delta(q^k)$ cannot vanish asymptotically. While one can always append a short run of full-sample TMLE refinement to obtain exact $\hat{\mathbf{Q}}$, fixed-batch STMLE requires an explicit switching rule and leaves the choice of B to the user. In this subsection, we seek a *data-driven* approach that allows the algorithm to learn appropriate batch sizes automatically, and gradually transitioning from stochastic targeting to deterministic targeting on its own. Our design is inspired by self-paced learning (SPL) [24], a learning-to-optimize paradigm that mimics the human process of mastering easy patterns first and incorporating harder examples later. It has demonstrated success across diverse applications including computer vision, natural language processing [38, 39]. We adapt this principle to STMLE framework, yielding adaptive STMLE (S$^+$TMLE).

SPL Formulations. We briefly recap SPL and emphasize the aspects most relevant to our adaptive design. Consider a standard machine learning objective of the form $\min_{\mathbf{w}} \sum_{i=1}^n \ell\big(\mathbf{w}; O_i\big)$, where O_i denotes the i-th training sample, $\ell(\cdot)$ is the loss function, and $\mathbf{w}$ represents the model parameters. SPL augments the primal learning objective by introducing a latent weight vector $\mathbf{v} = [v_1,\dots,v_n]^\top$ with each weight $v_i \in [0,1]$ serves as a learnable importance score (or inclusion probability) attached to the i-th sample.

[1] In Section 3.2 and the remainder of the main body, random subsets like $\mathcal{B}^k$ are sampled with replacement across iterations.

Given an age parameter $\lambda \geq 0$ and a self-paced regularizer (SP-regularizer) $f(v_i, \lambda)$, SPL solves

$$\min_{\mathbf{w},\mathbf{v}} \sum_{i=1}^{n} \Big(v_i \cdot \ell\big(\mathbf{w}; O_i\big) + f(v_i, \lambda) \Big), \quad (6)$$

where the form of $f(v_i, \lambda)$ governs the sample selection mechanism. The joint optimization is typically performed via an alternative optimization strategy (AOS). Let $\ell_i := \ell(\mathbf{w}; O_i)$ denote the loss of sample i at the current parameters. The AOS updates of (6) take the form

$$\begin{cases} v_i^\star(\ell_i, \lambda) = \arg\min_{v_i \in [0,1]} \{ v_i \ell_i + f(v_i, \lambda) \}, \\ \mathbf{w}^\star = \arg\min_{\mathbf{w}} \sum_{i=1}^{n} v_i^\star \, \ell\big(\mathbf{w}; O_i\big). \end{cases}$$

Algorithm 3 Adaptive STMLE (S^+TMLE)

Input: Dataset $\{O_i\}_{i=1}^n$, initial fit $\mathbf{Q}_0$, iterations K, target parameter Ψ.

1: **for** $k = 0, \ldots, K-1$ **do**

2: $\quad \mathbf{v}^k = \arg\min_{\mathbf{0} \preceq \mathbf{v} \preceq \mathbf{1}} \sum_{i=1}^{n} v_i \mathcal{L}(\mathbf{Q}_k)(O_i) + f(\mathbf{v}, \lambda^k)$

3: $\quad \epsilon_k = \arg\min_{\epsilon \in \mathcal{E}} \frac{1}{|\mathcal{B}^k|} \sum_{i=1}^{n} v_i^k \mathcal{L}(\mathbf{Q}_k(\epsilon \mid \mathbf{G}))(O_i)$

4: $\quad \mathbf{Q}_{k+1} = \mathbf{Q}_k(\epsilon_k \mid \mathbf{G})$

5: $\quad$ Set $\lambda^{k+1} \geq \lambda^k$ by the user-specified or predefined age schedule.

6: **end for**

Output: Targeted estimator $\hat{\Psi}_n = \Psi(\mathbf{Q}_K)$.

By gradually increasing λ over iterations, the algorithm progressively incorporates more samples into the effective training set. In SPL practice, the weight update admits a closed-form solution for most standard choices of the SP-regularizer. As a canonical example, under the `Hard` SP-regularizer $f^H(v, \lambda) := -\lambda v$, one obtains $v^\star(\ell, \lambda) = 1$ if $\ell < \lambda$ and $v^\star = 0$ otherwise [24]. Under this SP-regularizer, samples with loss below the threshold λ are included in training while others are excluded.

Algorithm & Guarantees. Inspired by the philosophy of SPL, we propose S^+TMLE, an adaptive variant that replaces the fixed batch size in STMLE with a learned sample weighting scheme. At each iteration k, S^+TMLE first solves the SPL subproblem to obtain weights $\mathbf{v}^k$ based on the current losses $\mathcal{L}(\mathbf{Q}_k)(O_i)$ and the age parameter λ^k. The fluctuation parameter ϵ_k is then computed by minimizing the weighted empirical loss, where the effective batch size $|\mathcal{B}^k|$ equals $\|\mathbf{v}^k\|_1 = \sum_{i=1}^n v_i^k$. The complete procedure is presented in Algorithm 3.

Remark 4.1 (Age Schedule). The practical choices for $\{\lambda^k\}_k$ has been extensively explored in the SPL literature and applications, including linear or logarithmic growth, and data-adaptive rules based on quantiles of the current loss distribution [40, 41].

Lemma 4.2 (Well-posedness)**.** *Fix any iteration index k and any λ. The solution set of $\mathbf{v}^k \in \arg\min_{\mathbf{0} \preceq \mathbf{v} \preceq \mathbf{1}} \{ \sum_{i=1}^n v_i \mathcal{L}(\mathbf{Q}_k)(O_i) + f(\mathbf{v}, \lambda) \}$ is nonempty, i.e., at least one minimizer exists.*

Lemma 4.2 implies that the iterates of Algorithm 3 are always well-defined, regardless of the current model maturity. Theorem 4.3 shows that S^+TMLE automatically transitions from stochastic to full-batch updates within finite iterations.

Theorem 4.3 (Finite-time Recovery)**.** *Suppose $\{\lambda_k\}_{k \geq 0} \subset [0, \infty)$ is monotonically nondecreasing with $\lim_{k \to \infty} \lambda_k = \infty$, and let f be one of the mainstream SP-regularizers* `Hard`*,* `Mixture`*, or* `LOG` *(defined in Appendix A.3). Then there exists a finite index K_0 such that for all $k > K_0$, the (unique) optimal weight vector equals $\mathbf{v}^k = \mathbf{1}$.*

Once $\mathbf{v}^k \equiv \mathbf{1}$, Algorithm 3 reduces to standard TMLE and inherits all of its convergence guarantees. This self-annealing property eliminates the need for switching between stochastic and deterministic phases, providing a principled approach to scalable targeted learning.

4.2 S^2TMLE: Doubly Stochastic Targeting

The STMLE algorithm reduces computational cost by subsampling observations, but the per-iteration complexity still scales linearly with the dimension d of the target parameter Ψ, which motivates a second layer of stochasticity that reduces the dimension of each subproblem. In optimization community, the doubly stochastic gradient methods have proven highly effective in large-scale machine learning environments [26, 27]. We adapt this principle to the targeted learning setting, proposing doubly stochastic TMLE (S^2TMLE).

At iteration k, S^{2}TMLE uniformly samples a mini-batch of indices $\mathcal{B}^k \subseteq \{1,\dots,n\}$ of size B, and a coordinate subset $\mathcal{S}^k \subseteq \{1,\dots,d\}$ of size S. We use the notation $\epsilon_{\mathcal{S}}$ to denote the subvector of ϵ indexed by $\mathcal{S}$, and $D^*_{\mathcal{S}}$ to denote the corresponding components of the EIF. To streamline notation, we assume that for any $\epsilon \in \mathcal{E}$ and any coordinate subset $\mathcal{S}$, the vector obtained by setting coordinates outside $\mathcal{S}$ to zero also belongs to $\mathcal{E}$. Let $\bar{\mathcal{S}}^k := \{1,\dots,d\} \setminus \mathcal{S}^k$ denote the complement. The fluctuation parameter is then computed by minimizing the mini-batch loss over $\epsilon_{\mathcal{S}^k}$, while setting $\epsilon_{\bar{\mathcal{S}}^k} \equiv \mathbf{0}_{\bar{\mathcal{S}}^k}$. The corresponding targeted update is then applied through the same fluctuation mechanism. The pseudocode is shown in Algorithm 4. When $S = d$, the S^{2}TMLE reduces to STMLE, and when additionally $B = n$, it recovers the vanilla TMLE in Algorithm 1.

Algorithm 4 Doubly Stochastic TMLE (S^{2}TMLE)

Input: Dataset $\{O_i\}_{i=1}^n$, initial fit $\mathbf{Q}_0$, iterations K, target parameter Ψ, batch size B, dimension size S.

1: **for** $k = 0,\dots,K-1$ **do**
2: Sample $\mathcal{B}^k \subset \{1,\dots,n\}$, $\mathcal{S}^k \subset \{1,\dots,d\}$.
3: $\epsilon^k_{\mathcal{S}^k} = \arg\min_{\epsilon_{\mathcal{S}^k}} \frac{1}{B} \sum_{i \in \mathcal{B}^k} \mathcal{L}(\mathbf{Q}_k(\epsilon_{\mathcal{S}^k}, D^*_{\mathcal{S}^k} \mid \mathbf{G}))(O_i)$
4: $\epsilon^k_{\bar{\mathcal{S}}^k} = \mathbf{0}_{\bar{\mathcal{S}}^k}$, $\mathbf{Q}_{k+1} = \mathbf{Q}_k(\epsilon^k, D^* \mid \mathbf{G})$
5: **end for**

Output: Targeted estimator $\hat{\Psi}_n = \Psi(\mathbf{Q}_K)$.

Theorem 4.4 (Ergodic Expected Convergence)**.** *We have*

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{E}\left[\Delta(q^k)\right] \le \frac{4L_\epsilon R_{\mathcal{E}}}{\sqrt{B}} + L_\epsilon R_{\mathcal{E}} \sqrt{1 - \frac{S}{d}} + 4L_{\max} \sqrt{\frac{2\left(\log 2 + d\log(1+2\sqrt{B})\right)}{B}}. \quad (7)$$

Inequality (7) separates the interpretable effects of the sample and dimension subsampling. The first and third terms capture the effect of observation subsampling and scale as $\mathcal{O}(B^{-1/2})$ and $\mathcal{O}(\sqrt{d\log B/B})$ respectively. The second term $L_\epsilon R_{\mathcal{E}}\sqrt{1-S/d}$ captures the approximation error introduced by coordinate subsampling, which is monotonically decreasing in S and vanishes when $S = d$. Conversely, when S is small relative to d, this term dominates and the algorithm converges to a larger noise ball.

Remark 4.5. The doubly stochastic framework can be extended further by introducing additional stochasticity.

5 Experiments

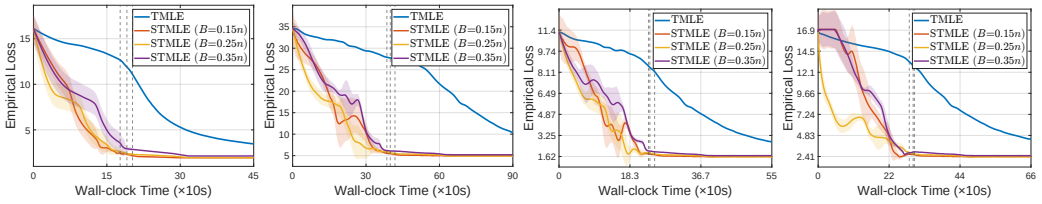
Setup & Results We generate $n = 10^6$ (million) i.i.d. observations $O_i \in \mathbb{R}^{30}$ and consider four standard plug-in target parameters. The first target $\Psi_{(1)}$ is a distributional functional defined as the squared cumulative distribution function (CDF) evaluated at 30 equally spaced quantile points, based on a scalar component of O_i. The $\Psi_{(2)}$ is a 30-dimensional median regression parameter with a linear working model. The remaining two targets adopt a causal structure with 28 covariates, a binary treatment, and a scalar outcome: $\Psi_{(3)}$ is the treatment-specific mean (TSM), and $\Psi_{(4)}$ is the variance of the conditional treatment effect (VTE). For *each* Ψ, we employ two distinct synthetic data-generating processes (denoted DGP1 and DGP2) to assess algorithms across different underlying distributions. Complete specifications of details are provided in Appendix I.

Table 2: Peak RAM usage ($\times$10 MB). The $\setminus$ indicates S^{2}TMLE degenerates to STMLE.

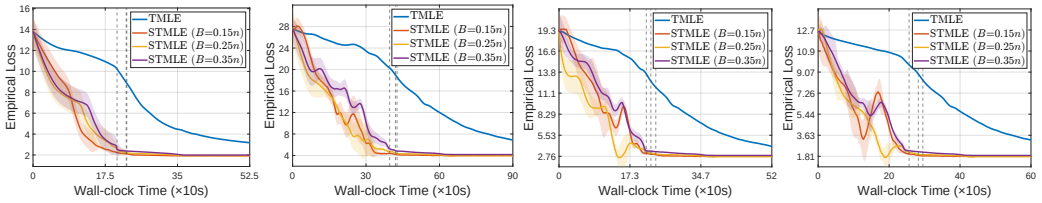
Experiment	TMLE	STMLE	S$^+$TMLE	S^{2}TMLE
$\Psi_{(1)}$ + DGP1	27.8$\pm$3.6	19.6$\pm$2.8	21.8$\pm$3.0	18.4$\pm$2.6
$\Psi_{(1)}$ + DGP2	29.2$\pm$3.9	20.3$\pm$3.0	22.6$\pm$3.2	19.1$\pm$2.8
$\Psi_{(2)}$ + DGP1	82.4$\pm$7.7	58.9$\pm$5.9	63.5$\pm$6.4	54.6$\pm$5.4
$\Psi_{(2)}$ + DGP2	86.8$\pm$8.1	60.7$\pm$6.2	65.4$\pm$6.8	56.2$\pm$5.7
$\Psi_{(3)}$ + DGP1	44.8$\pm$5.3	36.9$\pm$4.6	39.8$\pm$4.9	$\setminus$
$\Psi_{(3)}$ + DGP2	46.2$\pm$5.6	38.0$\pm$4.8	41.0$\pm$5.2	$\setminus$
$\Psi_{(4)}$ + DGP1	49.9$\pm$5.9	39.6$\pm$4.9	42.8$\pm$5.4	$\setminus$
$\Psi_{(4)}$ + DGP2	51.3$\pm$6.2	40.8$\pm$5.1	44.1$\pm$5.7	$\setminus$

Figure 1 reports the wall-clock time required for each algorithm to reach convergence across all configurations. The results demonstrate that STMLE (and its variants) achieve substantial speedups over standard TMLE. Table 2 summarizes the peak RAM usage during execution. The stochastic algorithms consistently maintain a smaller memory footprint than the full-batch baseline, as they avoid materializing the complete empirical loss over all n observations. This reduction is particularly significant in memory-constrained environments and enables deployment on commodity hardware for large-scale datasets.

Real-World Benchmarks We also evaluate our methods on two classic real-world benchmarks spanning various fields. For estimating the TSM, we use the National Supported Work (NSW)

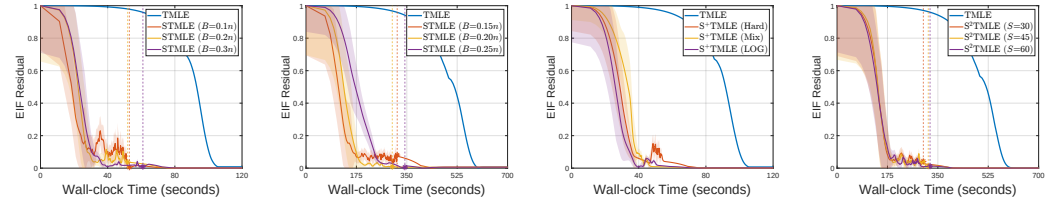


(a) $\Psi_{(1)}$ on DGP1 (b) $\Psi_{(2)}$ on DGP1 (c) $\Psi_{(3)}$ on DGP1 (d) $\Psi_{(4)}$ on DGP1



(e) $\Psi_{(1)}$ on DGP2 (f) $\Psi_{(2)}$ on DGP2 (g) $\Psi_{(3)}$ on DGP2 (h) $\Psi_{(4)}$ on DGP2

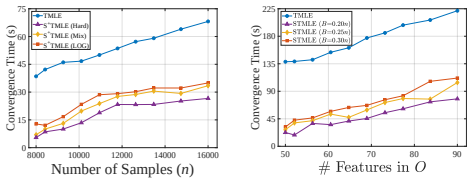
Figure 1: STMLE: empirical loss vs. training time. Results for the other two algorithms (S$^+$TMLE, S^{2}TMLE) are in Appendix J. Shaded regions indicate variance across runs.



(a) $\Psi_{(3)}$ on NSW-Job (b) $\Psi_{(2)}$ on YearPrediction (c) $\Psi_{(3)}$ on NSW-Job (d) $\Psi_{(2)}$ on YearPrediction

Figure 2: Statistical performance of the proposed algorithms. We plot th normalized EIF residual (in norm) vs. running time. The '*' marker indicates the switch to vanilla TMLE.

job-training dataset [42], a canonical benchmark in causal inference. Each observation comprises a binary treatment indicator A (participation), an 8-dimensional covariate vector W, and a scalar outcome Y (post-program earnings). We construct a combined sample of $n = 16{,}177$ observations by pooling the experimental treatment group with comparison units. For median regression, we employ the YearPredictionMSD dataset from the UCI Repository [43], where the task is to predict a song's release year from 90 real-valued audio features. It contains $n = 515{,}345$ observations, providing a setting where computational scalability is essential. Figure 2 plots the norm of the empirical EIF residual (1) against cumulative training time. All stochastic variants converge rapidly to a small neighborhood of zero that is consistent with the noise-ball guarantees established. It further reveals that the residual stochastic plateau accounts for only 5-15% of the total descent, indicating that the bulk of the targeting work is already completed by the switch time. To investigate scalability, we construct subsampled versions of the real datasets by varying the sample size n or the observation dimension. Figure 3 reports this favorable scaling, which makes the proposed methods particularly attractive for modern large-scale studies.



(a) $\Psi_{(3)}$ on NSW-Job (b) $\Psi_{(2)}$ on YearPrediction

Figure 3: Scalability on real-world benchmarks.

6 Conclusion & Future Work

This paper introduced a family of stochastic extensions to the classical TMLE framework, addressing the computational bottleneck inherent in full-sample debiasing. Several directions merit future investigation. Incorporating advanced optimization techniques such as momentum acceleration or averaging could further improve the convergence of TMLE algorithms. Moreover, establishing formal properties in STMLE that guarantee robustness to initialization of the estimand would further strengthen its practical reliability. Importance sampling and stratified batching are promising future directions, but introduce additional sampling costs and require new theoretical analysis.

References

[1] Mark J Van Der Laan and Daniel Rubin. Targeted maximum likelihood learning. 2006.

[2] Mark J Van der Laan, Sherri Rose, et al. *Targeted learning: causal inference for observational and experimental data*, volume 4. Springer, 2011.

[3] Megan S Schuler and Sherri Rose. Targeted maximum likelihood estimation for causal inference in observational studies. *American journal of epidemiology*, 185(1):65–73, 2017.

[4] Susan Gruber, Rachael V Phillips, Hana Lee, Martin Ho, John Concato, and Mark J van der Laan. Targeted learning: toward a future informed by real-world evidence. *Statistics in Biopharmaceutical Research*, 16(1):11–25, 2024.

[5] Aurelien Bibaut, Ivana Malenica, Nikos Vlassis, and Mark Van Der Laan. More efficient off-policy evaluation through regularized targeted learning. In *International Conference on Machine Learning*, pages 654–663. PMLR, 2019.

[6] Yi Li, David Mccoy, Nolan Gunter, Kaitlyn Lee, Alejandro Schuler, and Mark van der Laan. Targeted deep architectures: A tmle-based framework for robust causal inference in neural networks. *arXiv preprint arXiv:2507.12435*, 2025.

[7] Ildar Abdulovich Ibragimov and Rafail Zalmanovich Has' Minskii. *Statistical estimation: asymptotic theory*, volume 16. Springer Science & Business Media, 2013.

[8] Oliver Hines, Oliver Dukes, Karla Diaz-Ordaz, and Stijn Vansteelandt. Demystifying statistical learning based on efficient influence functions. *The American Statistician*, 76(3):292–304, 2022.

[9] Mark J van der Laan, Sherri Rose, and Susan Gruber. Readings in targeted maximum likelihood estimation. 2009.

[10] Mark J Van der Laan and Maya L Petersen. Targeted learning. In *Ensemble machine learning*, pages 117–156. Springer, 2012.

[11] Susan Gruber, Hana Lee, Rachael Phillips, and Mark van der Laan. Causal inference with targeted learning for producing and evaluating real-world evidence. In *Real-world evidence in medical product development*, pages 125–143. Springer, 2023.

[12] Oleg Sofrygin, Zheng Zhu, Julie A Schmittdiel, Alyce S Adams, Richard W Grant, Mark J van der Laan, and Romain Neugebauer. Targeted learning with daily ehr data. *Statistics in medicine*, 38(16):3073–3090, 2019.

[13] Mark J Van der Laan and Sherri Rose. *Targeted learning in data science*. Springer, 2018.

[14] Mark J van der Laan and Samuel D Lendle. Online targeted learning. 2014.

[15] Léon Bottou, Frank E Curtis, and Jorge Nocedal. Optimization methods for large-scale machine learning. *SIAM review*, 60(2):223–311, 2018.

[16] Shai Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2025.

[17] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951.

[18] Shun-ichi Amari. Backpropagation and stochastic gradient descent method. *Neurocomputing*, 5 (4-5):185–196, 1993.

[19] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010: 19th International Conference on Computational StatisticsParis France, August 22-27, 2010 Keynote, Invited and Contributed Papers*, pages 177–186. Springer, 2010.

[20] Vatsal Shah, Xiaoxia Wu, and Sujay Sanghavi. Choosing the sample with lowest loss makes sgd robust. In *International Conference on Artificial Intelligence and Statistics*, pages 2120–2130. PMLR, 2020.

[21] Xiang Cheng, Dong Yin, Peter Bartlett, and Michael Jordan. Stochastic gradient and langevin processes. In *International Conference on Machine Learning*, pages 1810–1819. PMLR, 2020.

[22] Idan Amir, Yair Carmon, Tomer Koren, and Roi Livni. Never go full batch (in stochastic convex optimization). *Advances in Neural Information Processing Systems*, 34:25033–25043, 2021.

[23] Edwin KP Chong, Wu-Sheng Lu, and Stanislaw H Zak. *An Introduction to Optimization: With Applications to Machine Learning*. John Wiley & Sons, 2023.

[24] M Kumar, Benjamin Packer, and Daphne Koller. Self-paced learning for latent variable models. *Advances in neural information processing systems*, 23, 2010.

[25] Tianlong Chen, Xiaohan Chen, Wuyang Chen, Howard Heaton, Jialin Liu, Zhangyang Wang, and Wotao Yin. Learning to optimize: A primer and a benchmark. *Journal of Machine Learning Research*, 23(189):1–59, 2022.

[26] Bo Dai, Bo Xie, Niao He, Yingyu Liang, Anant Raj, Maria-Florina Balcan, and Le Song. Scalable kernel methods via doubly stochastic gradients. *Advances in neural information processing systems*, 27, 2014.

[27] Michalis Titsias and Miguel Lázaro-Gredilla. Doubly stochastic variational bayes for non-conjugate inference. In *International conference on machine learning*, pages 1971–1979. PMLR, 2014.

[28] Patrice Bertail, Antoine Chambaz, and Emilien Joly. Practical targeted learning from large data sets by survey sampling. *arXiv preprint arXiv:1606.09522*, 2016.

[29] Antoine Chambaz, Emilien Joly, and Xavier Mary. Targeted learning using adaptive survey sampling. In *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*, pages 541–559. Springer, 2018.

[30] Marco Carone, Iván Díaz, and Mark J van der Laan. Higher-order targeted minimum loss-based estimation. 2014.

[31] Mark van der Laan, Zeyi Wang, and Lars van der Laan. Higher order targeted maximum likelihood estimation. *arXiv preprint arXiv:2101.06290*, 2021.

[32] Brian M Cho, Yaroslav Mukhin, Kyra Gan, and Ivana Malenica. Kernel debiased plug-in estimation: Simultaneous, automated debiasing without influence functions for many target parameters. In *International Conference on Machine Learning*, pages 8534–8555. PMLR, 2024.

[33] Noel Pimentel, Alejandro Schuler, and Mark van der Laan. Score-preserving targeted maximum likelihood estimation. *arXiv preprint arXiv:2502.00200*, 2025.

[34] Iván Díaz and Michael Rosenblum. Targeted maximum likelihood estimation using exponential families. *The international journal of biostatistics*, 11(2):233–251, 2015.

[35] Diyang Li and Kyra Gan. Targeted maximum likelihood learning: An optimization perspective. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[36] Dylan J Foster and Vasilis Syrgkanis. Orthogonal statistical learning. *The Annals of Statistics*, 51(3):879–908, 2023.

[37] Anastasios A Tsiatis. *Semiparametric theory and missing data*. Springer, 2006.

[38] Yixiao Ge, Feng Zhu, Dapeng Chen, Rui Zhao, et al. Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. *Advances in neural information processing systems*, 33:11309–11321, 2020.

[39] Petru Soviany, Radu Tudor Ionescu, Paolo Rota, and Nicu Sebe. Curriculum learning: A survey. *International Journal of Computer Vision*, 130(6):1526–1565, 2022.

[40] Deyu Meng, Qian Zhao, and Lu Jiang. A theoretical understanding of self-paced learning. *Information Sciences*, 414:319–328, 2017.

[41] Wei Zheng, Xiaofeng Zhu, Guoqiu Wen, Yonghua Zhu, Hao Yu, and Jiangzhang Gan. Unsupervised feature selection by self-paced learning regularization. *Pattern recognition letters*, 132: 4–11, 2020.

[42] Rajeev H Dehejia and Sadek Wahba. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American statistical Association*, 94(448): 1053–1062, 1999.

[43] Arthur Asuncion, David Newman, et al. Uci machine learning repository, 2007.

[44] Deyu Meng, Qian Zhao, and Lu Jiang. What objective does self-paced learning indeed optimize? *arXiv preprint arXiv:1511.06049*, 2015.

[45] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[46] Yurii D Burago and Viktor A Zalgaller. *Geometric inequalities*, volume 285. Springer Science & Business Media, 2013.

[47] Rémi Bardenet and Odalric-Ambrym Maillard. Concentration inequalities for sampling without replacement. *Bernoulli*, 21(3):1361–1385, 2015.

[48] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

[49] Peter S Bullen, Dragoslav S Mitrinovic, and Means Vasic. *Means and their inequalities*, volume 31. Springer Science & Business Media, 2013.

[50] Walter Rudin. Principles of mathematical analysis. *3rd ed.*, 1976.

[51] Stephen Abbott. *Understanding analysis*. Springer, 2015.

[52] Mark J. van der Laan, Eric C. Polley, and Alan E. Hubbard. Super learner. *Statistical Applications in Genetics and Molecular Biology*, 6(1):Article 25, 2007. doi: 10.2202/1544-6115.1309.

Appendix

The appendix is organized as follows.

A Technical Background **14**

A.1 Algorithm Roadmap 14

A.2 Batch Size Condition 14

A.3 Self-Paced Regularizer 14

B Extra Discussions **15**

B.1 Stochastic-to-Deterministic Transition 15

B.2 On the Choice of B and S 15

B.3 Limitations 15

B.4 Survey-Weighted TMLE 15

C Proof of Theorem 3.1 **16**

C.1 Case I 16

C.2 Case II 21

D Proof of Lemma 3.2 **28**

E Proof of Theorem 3.5 **29**

F Proof of Lemma 4.2 **33**

G Proof of Theorem 4.3 **35**

H Proof of Theorem 4.4 **37**

I Implementation Details **44**

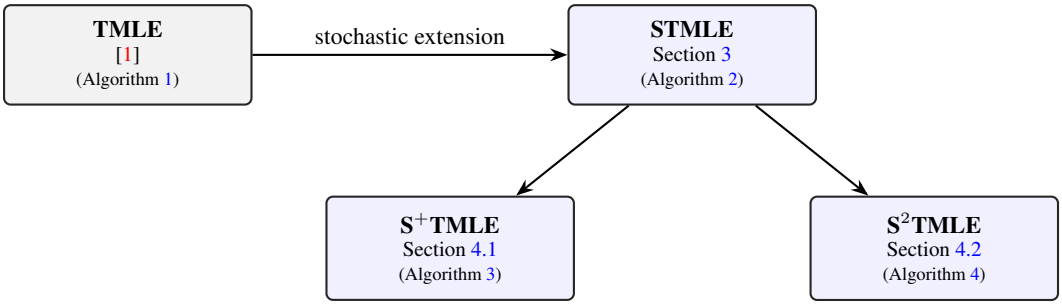
I.1 Target Parameters 44

I.2 Synthetic Data-Generating Distributions 45

J Additional Results **46**

A Technical Background

A.1 Algorithm Roadmap



A.2 Batch Size Condition

We present the detailed specification of batch size B that guarantees the strict inequality in Remark 3.6.

$$B \in \begin{cases} \varnothing, & \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) = \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big), \\ \{1,2,\ldots,n\}, & \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big) > 2nL_{\max}, \\ \{1,2,\ldots,n\}, & 0 < \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big) \le 2nL_{\max} \text{ and } m = 1, \\ \left\{ \left\lfloor \dfrac{8L_{\max}^2}{\Omega^2} \ln \dfrac{2nL_{\max}(m-1)}{\mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big)} \right\rfloor + 1, \ldots, n \right\}, \\ \quad 0 < \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big) \le 2nL_{\max} \text{ and } m > 1. \end{cases} \tag{8}$$

A.3 Self-Paced Regularizer

In this part, we provide additional background on the self-paced regularizers used in S$^+$TMLE (Algorithm 3) and present the closed-form solutions for the optimal sample weights under each regularizer.

Hard regularizer [24]. The `Hard` SP-regularizer is defined as $f^H(v,\lambda) := -\lambda v$, which is the simplest choice, corresponding to a linear penalty that encourages large weights. The optimal weight admits

$$v^\star(\ell,\lambda) = \begin{cases} 1, & \text{if } \ell < \lambda, \\ 0, & \text{if } \ell \ge \lambda. \end{cases}$$

Mixture regularizer [44]. The `Mixture` SP-regularizer is defined as

$$f^M(v,\lambda,\gamma) := \frac{\gamma^2}{v + \gamma/\lambda}, \tag{9}$$

where $\gamma > \sqrt{L_{\max}}$ is an additional hyperparameter. In the context of Theorem 4.3, we require $\gamma > \sqrt{L_{\max}}$ to ensure the finite-time recovery property. The optimal weight is given by

$$v^\star(\ell,\lambda,\gamma) = \begin{cases} 1, & \text{if } \ell \le \left(\dfrac{\lambda\gamma}{\lambda+\gamma}\right)^2, \\ \gamma\left(\dfrac{1}{\sqrt{\ell}} - \dfrac{1}{\lambda}\right), & \text{if } \left(\dfrac{\lambda\gamma}{\lambda+\gamma}\right)^2 < \ell < \lambda^2, \\ 0, & \text{if } \ell \ge \lambda^2. \end{cases}$$

LOG regularizer [40]. The `LOG` SP-regularizer is defined as

$$f^L(v,\lambda,\alpha) := \frac{1}{\alpha}\mathrm{KL}\big(1+\alpha\lambda, v\big) = \frac{1}{\alpha}\left((1+\alpha\lambda)\log\frac{1+\alpha\lambda}{v} - (1+\alpha\lambda) + v\right),$$

where $\mathrm{KL}(\cdot,\cdot)$ denotes the Kullback-Leibler (KL) distance between two variables and $\alpha > 0$ is a hyperparameter. The optimal weight is given by

$$v^\star(\ell,\lambda,\alpha) = \begin{cases} 1, & \text{if } \ell \le \lambda, \\ \dfrac{1+\alpha\lambda}{1+\alpha\ell}, & \text{if } \ell > \lambda. \end{cases}$$

Remark A.1 (Sparsity of S$^+$TMLE). Although S$^+$TMLE can be written as a weighted objective with auxiliary variables $\{v_i\}_{i=1}^n$, it does not in general reduce to a fixed reweighting of the full batch. The SP-regularizer encompasses a broad, *potentially infinite* family of functions, and many concrete designs yield genuinely sparse weights by construction. For instance, the Mixture Regularizer used in our experiments assigns $v_i = 0$ to all samples with $\ell \ge \lambda^2$, recovering a true subsampling effect rather than soft reweighting.

B Extra Discussions

B.1 Stochastic-to-Deterministic Transition

The hybrid strategy described in Section 3 alternates a stochastic phase with a final full-batch refinement phase. A natural question is when to switch between the two. A first observation is that the correctness of the final iterate does not depend on the choice of switching time: as long as the deterministic refinement phase is run to convergence, the output satisfies the empirical EIF estimating-equation (1) exactly, regardless of how many stochastic iterations preceded it. The transition rule therefore only affects efficiency instead of the validity.

Several principled monitoring criteria can be derived directly from quantities already defined in our framework, including the magnitude of the fluctuation parameter $\|\epsilon_k\|$, the periodic full-sample EIF residual $\mathbb{E}[\Delta(q^k)]$, or a fixed stochastic-iteration budget such as a wall-clock time limit or a maximum number of iterations K. Each of these provides a usable trigger for switching to the refinement phase, and all are computable from objects already maintained by the algorithm.

B.2 On the Choice of B and S

The batch size B and coordinate-subset size S are best viewed as resource-accuracy *trade-off parameters*. The batch size B plays the same role as the mini-batch size in classical stochastic optimization: larger B shrinks the noise ball but increases per-iteration cost, with the $1/\sqrt{B}$ scaling in Theorem 3.1 mirroring the standard SGD rate. A moderate value, i.e., a small fraction of n, or on the order of $\sqrt{n}$, is a reasonable starting point, and the hybrid refinement strategy ensures exact convergence regardless of the specific choice. For S, Theorem 4.4 gives a direct guideline: when d is moderate, $S = d$ recovers STMLE without coordinate subsampling; when d is large and memory or per-iteration cost is the bottleneck, S should be chosen so that $\sqrt{1 - S/d}$ is acceptably small, e.g., $S = d/2$ contributes a factor of $\sqrt{1/2} \approx 0.71$ and $S = 3d/4$ reduces it to 0.5.

B.3 Limitations

The stochastic phase of STMLE converges only to a noise ball of radius $\mathcal{O}(1/\sqrt{B})$ rather than to an exact root of (1); we restore exactness via a hybrid refinement step, but a fully stochastic procedure with provable exact convergence remains open. Our convergence and escape results follow standard SGD-style worst-case analyses, and a fully automatic selection rule for B and S is out of reach, consistent with the analogous open question in stochastic optimization. Empirically, our study does not cover very huge datasets ($n \gg 10^6$), longitudinal or survival outcomes, or interactions with highly flexible nuisance estimators.

B.4 Survey-Weighted TMLE

Bertail et al. [28] and Chambaz et al. [29] fit nuisances and perform targeting on a fixed, survey-weighted subsample. Used as a warm start for full-data TMLE, the resulting fit serves as a learned initialization for $\mathbf{Q}_0$ and $\mathbf{G}$, analogous to Super Learner; STMLE instead accelerates targeting from

a given initialization. Extending scalar-EIF sampling designs to vector-valued targets requires an additional coordinate-aggregation rule. Guarantees for the weighted-subsample EIF equation do not imply proximity to the full-data EIF root or fewer refinement steps.

C Proof of Theorem 3.1

Let $(X, \tilde{d})$ be a metric space and let $E \subseteq X$. For $\eta > 0$ and $r > 0$, we denote the closed ball by

$$\mathbb{B}(x, r) \triangleq \{y \in X : \tilde{d}(x, y) \leq r\}, \quad x \in X.$$

Definition C.1 (η-separated Set)**.** A set $\mathcal{N} \subseteq E$ is called η-separated if $\tilde{d}(u, v) > \eta$ for $\forall u, v \in \mathcal{N}$ and $u \neq v$. [2]

Definition C.2 (Covering)**.** A set $\mathcal{N} \subseteq E$ is called an η-net for E if for every $x \in E$ there exists $u \in \mathcal{N}$ such that $\tilde{d}(x, u) \leq \eta$. Equivalently, $E \subseteq \bigcup_{u \in \mathcal{N}} \mathbb{B}(u, \eta)$.

Corollary C.3 (Pairwise Disjointness)**.** *If $\mathcal{N}_\eta \subseteq X$ is η-separated, then for any distinct $u, v \in \mathcal{N}_\eta$,* $\mathbb{B}\Big(u, \frac{\eta}{2}\Big) \cap \mathbb{B}\Big(v, \frac{\eta}{2}\Big) \equiv \varnothing$.

C.1 Case I

Assumption C.4. For each k, the subset $\mathcal{B}^k \subset \{1, \ldots, n\}$ is drawn uniformly at random with replacement.

Under Assumption C.4, we now prove the bound (2).

Proof. For each k there exist random indices $I_1^k, \ldots, I_B^k$ such that conditional on the algorithmic past the variables $I_1^k, \ldots, I_B^k$ are independent and each is uniform on $\{1, \ldots, n\}$. Across different k these collections are independent. Define the history sigma field

$$\mathcal{F}_k := \sigma\big(I_1^0, \ldots, I_B^0,\ I_1^1, \ldots, I_B^1,\ \ldots,\ I_1^{k-1}, \ldots, I_B^{k-1}\big).$$

By construction, $\mathbf{Q}_k$ and the map $\epsilon \mapsto \mathbf{Q}_k(\epsilon \mid \mathbf{G})$ are $\mathcal{F}_k$-measurable for every fixed ϵ. Likewise the maps $\epsilon \mapsto \frac{1}{n}\sum_{i=1}^n \mathcal{L}\left(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\right)(O_i)$ and $\epsilon \mapsto \frac{1}{B}\sum_{j=1}^B \mathcal{L}\left(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\right)(O_{I_j^k})$ are $\mathcal{F}_k$-measurable. Conditional on $\mathcal{F}_k$, the indices $I_1^k, \ldots, I_B^k$ are i.i.d. uniform on $\{1, \ldots, n\}$. For every $k \geq 0$, define the uniform mini-batch deviation

$$\Gamma_k := \sup_{\epsilon \in \mathcal{E}} \left| \frac{1}{B}\sum_{j=1}^B \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{I_j^k}) - \frac{1}{n}\sum_{i=1}^n \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i) \right|. \tag{10}$$

By Assumption 2.5, the following holds for all $\epsilon, \epsilon' \in \mathcal{E}$:

$$\begin{aligned}
&\left| \frac{1}{n}\sum_{i=1}^n \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i) - \frac{1}{n}\sum_{j=1}^n \mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_j) \right| \\
&\leq \frac{1}{n}\sum_{i=1}^n \left| \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i) - \mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_i) \right| \leq \frac{1}{n}\sum_{i=1}^n \Big(L_\epsilon \|\epsilon - \epsilon'\| \Big) \\
&= L_\epsilon \|\epsilon - \epsilon'\|.
\end{aligned} \tag{11}$$

Similarly, for every realization of $I_1^k, \ldots, I_B^k$ we have

$$\begin{aligned}
&\left| \frac{1}{B}\sum_{j=1}^B \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{I_j^k}) - \frac{1}{B}\sum_{i=1}^B \mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_{I_i^k}) \right| \\
&\leq \frac{1}{B}\sum_{i \in \mathcal{B}^k} \left| \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i) - \mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_i) \right| \leq \frac{1}{B}\sum_{i \in \mathcal{B}^k} \Big(L_\epsilon \|\epsilon - \epsilon'\| \Big)
\end{aligned}$$

[2] Some papers use $\tilde{d}(u, v) \geq \eta$. Either convention is standard and only changes strict (weak) inequalities.

$$= L_\epsilon \|\epsilon - \epsilon'\|, \tag{12}$$

with the same Lipschitz constant. Hence $\frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i)$ and $\frac{1}{B}\sum_{j=1}^{B} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{I_j^k})$ are continuous on the compact set $\mathcal{E}$ by Assumption 2.4. It follows that $\arg\min_{\epsilon\in\mathcal{E}}\{\cdot\}$ is nonempty for both objectives. Fix measurable selections

$$\epsilon_n^k \in \arg\min_{\epsilon\in\mathcal{E}} \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i), \quad \epsilon^k \in \arg\min_{\epsilon\in\mathcal{E}} \frac{1}{B}\sum_{j=1}^{B} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{I_j^k}).$$

For all $\epsilon \in \mathcal{E}$, the definition of Γ_k yields

$$-\Gamma_k \le \frac{1}{B}\sum_{j=1}^{B} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{I_j^k}) - \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i) \le \Gamma_k.$$

Using the optimality of ϵ^k for the mini-batch objective and writing the full-data suboptimality as a three-term telescoping decomposition, we obtain

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\left(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\right)(O_i) - \frac{1}{n}\sum_{i'=1}^{n} \mathcal{L}\left(\mathbf{Q}_k(\epsilon_n^k \mid \mathbf{G})\right)(O_{i'}) \\
&= \left(\frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\left(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\right)(O_i) - \frac{1}{B}\sum_{j=1}^{B} \mathcal{L}\left(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\right)(O_{I_j^k})\right) \\
&\quad + \left(\frac{1}{B}\sum_{j=1}^{B} \mathcal{L}\left(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\right)(O_{I_j^k}) - \frac{1}{B}\sum_{j'=1}^{B} \mathcal{L}\left(\mathbf{Q}_k(\epsilon_n^k \mid \mathbf{G})\right)(O_{I_{j'}^k})\right) \\
&\quad + \left(\frac{1}{B}\sum_{j=1}^{B} \mathcal{L}\left(\mathbf{Q}_k(\epsilon_n^k \mid \mathbf{G})\right)(O_{I_j^k}) - \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\left(\mathbf{Q}_k(\epsilon_n^k \mid \mathbf{G})\right)(O_i)\right) \\
&\le \sup_{\epsilon\in\mathcal{E}} \left|\frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i) - \frac{1}{B}\sum_{j=1}^{B} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{I_j^k})\right| + 0 \\
&\quad + \sup_{\epsilon\in\mathcal{E}} \left|\frac{1}{B}\sum_{j=1}^{B} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{I_j^k}) - \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i)\right| \\
&\le \Gamma_k + \Gamma_k = 2\Gamma_k.
\end{aligned}$$

We therefore conclude

$$\frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\left(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\right)(O_i) \le 2\Gamma_k + \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\left(\mathbf{Q}_k(\epsilon_n^k \mid \mathbf{G})\right)(O_i).$$

By the update rule $\mathbf{Q}_{k+1} = \mathbf{Q}_k(\epsilon^k \mid \mathbf{G})$, we have

$$\frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_{k+1}(\mathbf{0} \mid \mathbf{G})\big)(O_i) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\left(\mathbf{Q}_{k+1}\right)(O_i) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\left(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\right)(O_i).$$

Combining this with the identity $\frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\left(\mathbf{Q}_k(\epsilon_n^k \mid \mathbf{G})\right)(O_i) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\mathbf{0} \mid \mathbf{G})\big)(O_i) - \Delta(q^k)$ yields

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_{k+1}(\mathbf{0} \mid \mathbf{G})\big)(O_i) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\big)(O_i) \\
&\le \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon_n^k \mid \mathbf{G})\big)(O_i) + 2\Gamma_k = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\mathbf{0} \mid \mathbf{G})\big)(O_i) - \Delta(q^k) + 2\Gamma_k, \\
&\Delta(q^k) \le \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\mathbf{0} \mid \mathbf{G})\big)(O_i) - \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_{k+1}(\mathbf{0} \mid \mathbf{G})\big)(O_i) + 2\Gamma_k.
\end{aligned} \tag{13}$$

Taking the $\mathcal{F}_k$-conditional expectation followed by the total expectation via the tower property gives

$$
\begin{aligned}
&\mathbb{E}\left[\Delta(q^k)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\Delta(q^k) \mid \mathcal{F}_k\right]\right] \\
&\overset{(13)}{\leq} \mathbb{E}\left[\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\mathbf{0} \mid \mathbf{G})\big)(O_i) - \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_{k+1}(\mathbf{0} \mid \mathbf{G})\big)(O_i) + 2\Gamma_k \,\middle|\, \mathcal{F}_k\right]\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\mathbf{0} \mid \mathbf{G})\big)(O_i)\right] - \mathbb{E}\left[\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_{k+1}(\mathbf{0} \mid \mathbf{G})\big)(O_i) \,\middle|\, \mathcal{F}_k\right]\right] + 2\mathbb{E}\left[\Gamma_k \mid \mathcal{F}_k\right] \\
&\leq \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\mathbf{0} \mid \mathbf{G})\big)(O_i)\right] - \mathbb{E}\left[\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_{k+1}(\mathbf{0} \mid \mathbf{G})\big)(O_i) \,\middle|\, \mathcal{F}_k\right]\right] + 2\mathbb{E}\left[\mathbb{E}\left[\Gamma_k \mid \mathcal{F}_k\right]\right] \\
&\leq 2\mathbb{E}[\Gamma_k] + \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\mathbf{0} \mid \mathbf{G})\big)(O_i)\right] - \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_{k+1}(\mathbf{0} \mid \mathbf{G})\big)(O_i)\right]. \qquad (14)
\end{aligned}
$$

Summing (14) from $k = 0$ to $T-1$ and expanding the telescoping sum yields

$$
\begin{aligned}
&\sum_{k=0}^{T-1}\mathbb{E}[\Delta(q^k)] \\
&\leq 2\sum_{k=0}^{T-1}\mathbb{E}[\Gamma_k] + \sum_{k=0}^{T-1}\left(\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\mathbf{0} \mid \mathbf{G})\big)(O_i)\right] - \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_{k+1}(\mathbf{0} \mid \mathbf{G})\big)(O_i)\right]\right) \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_0(\mathbf{0} \mid \mathbf{G})\big)(O_i)\right] - \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_1(\mathbf{0} \mid \mathbf{G})\big)(O_i)\right] + \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_1(\mathbf{0} \mid \mathbf{G})\big)(O_i)\right] \\
&\quad - \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_2(\mathbf{0} \mid \mathbf{G})\big)(O_i)\right] + \cdots + \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_{T-1}(\mathbf{0} \mid \mathbf{G})\big)(O_i)\right] \\
&\quad - \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_T(\mathbf{0} \mid \mathbf{G})\big)(O_i)\right] + 2\sum_{k=0}^{T-1}\mathbb{E}[\Gamma_k] \\
&= \underbrace{\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_0(\mathbf{0} \mid \mathbf{G})\big)(O_i)\right]}_{\triangleq \xi(\mathbf{0})} - \underbrace{\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_T(\mathbf{0} \mid \mathbf{G})\big)(O_i)\right]}_{\triangleq \xi(T)} + 2\sum_{k=0}^{T-1}\mathbb{E}[\Gamma_k].
\end{aligned}
$$

By Assumption 2.3, we have $\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_T(\mathbf{0} \mid \mathbf{G})\big)(O_i) \geq nL_{\min}$ almost surely and therefore $\xi(T) \geq L_{\min}$. It follows that

$$
\frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}[\Delta(q^k)] \leq \frac{\xi(\mathbf{0}) - L_{\min}}{T} + \frac{2}{T}\sum_{k=0}^{T-1}\mathbb{E}[\Gamma_k]. \qquad (15)
$$

It remains to establish a k-uniform bound on $\mathbb{E}[\Gamma_k]$ and then take $\limsup_{T\to\infty}$. Fix k and work conditional on $\mathcal{F}_k$. Writing $\Gamma_k := \sup_{\epsilon\in\mathcal{E}}\left|\check{Z}_k(\epsilon)\right|$, we obtain the deterministic bound

$$
\begin{aligned}
\left|\check{Z}_k(\epsilon) - \check{Z}_k(\epsilon')\right| &\leq \left|\frac{1}{B}\sum_{j=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{I_j^k}) - \frac{1}{B}\sum_{j=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_{I_j^k})\right| \\
&\quad + \left|\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i) - \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_i)\right| \\
&\overset{(12)}{\leq} 2L_\epsilon\|\epsilon - \epsilon'\|.
\end{aligned}
$$

Fix $\eta > 0$ and let $\mathcal{N}_\eta \subset \mathcal{E}$ be a maximal η-separated set so that $\|u - v\| > \eta$ for distinct $u, v \in \mathcal{N}_\eta$. Maximality implies that this set is also an η-net [45]. Let $N_\eta := \left|\mathcal{N}_\eta\right|$. For every $\epsilon \in \mathcal{E}$ there exists $\pi(\epsilon) \in \mathcal{N}_\eta$ with $\left\|\epsilon - \pi(\epsilon)\right\| \leq \eta$. Hence we have

$$
\Gamma_k = \sup_{\epsilon\in\mathcal{E}}\left|\check{Z}_k(\epsilon)\right| \leq \sup_{\epsilon\in\mathcal{E}}\left[\left|\check{Z}_k(\pi(\epsilon))\right| + \left|\check{Z}_k(\epsilon) - \check{Z}_k(\pi(\epsilon))\right|\right] \leq \sup_{\epsilon\in\mathcal{E}}\left[\left|\check{Z}_k(\pi(\epsilon))\right|\right] + 2L_\epsilon\eta
$$

$$\leq \max_{\epsilon' \in \mathcal{N}_\eta} \left|\check{Z}_k(\epsilon')\right| + 2L_\epsilon \eta. \tag{16}$$

We now bound N_η using the diameter and the isodiametric inequality [46] in its tight diameter form. Let $U := \bigcup_{u \in \mathcal{N}_\eta} \mathbb{B}(u, \eta/2)$. The balls $\{\mathbb{B}(u, \eta/2) : u \in \mathcal{N}_\eta\}$ are disjoint. Then for any $x \in \mathbb{B}(u, \eta/2)$ and $y \in \mathbb{B}(v, \eta/2)$, we have

$$\|x - y\| \leq \|x - u\| + \|u - v\| + \|v - y\| \leq \frac{\eta}{2} + \operatorname{diam}(\mathcal{E}) + \frac{\eta}{2} = R_\mathcal{E} + \eta,$$

which implies $\operatorname{diam}(U) \leq R_\mathcal{E} + \eta$. The isodiametric inequality states that among measurable sets in $\mathbb{R}^d$ with fixed diameter the Euclidean ball has maximal volume. Consequently,

$$\operatorname{Vol}(U) \leq \operatorname{Vol}\left(\mathbb{B}\left(0, \frac{R_\mathcal{E} + \eta}{2}\right)\right). \tag{17}$$

Moreover, disjointness implies

$$\operatorname{Vol}(U) = \sum_{u \in \mathcal{N}_\eta} \operatorname{Vol}\left(\mathbb{B}\left(0, \frac{\eta}{2}\right)\right) = N_\eta \cdot \operatorname{Vol}\left(\mathbb{B}\left(0, \frac{\eta}{2}\right)\right). \tag{18}$$

Using the scaling property of Euclidean ball volume $\operatorname{Vol}(\mathbb{B}(0, r)) = \operatorname{Vol}(\mathbb{B}(0, 1)) \cdot r^d$ together with (17) and (18) gives

$$N_\eta \operatorname{Vol}\left(\mathbb{B}(0,1)\right) \cdot \left(\frac{\eta}{2}\right)^d \leq \operatorname{Vol}\left(\mathbb{B}(0,1)\right) \cdot \left(\frac{R_\mathcal{E} + \eta}{2}\right)^d N_\eta \leq \left(1 + \frac{R_\mathcal{E}}{\eta}\right)^d.$$

Taking logarithms yields

$$\log(2N_\eta) \leq \log 2 + d \log\left(1 + \frac{R_\mathcal{E}}{\eta}\right). \tag{19}$$

Fix $\epsilon' \in \mathcal{N}_\eta$. Conditional on $\mathcal{F}_k$, the random variables $\mathcal{L}\left(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\right)(O_{I_1^k}), \ldots, \mathcal{L}\left(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\right)(O_{I_B^k})$ are i.i.d. and each is supported on $\left[-L_{\max}, L_{\max}\right]$. For any $\lambda > 0$ and any $t > 0$, the Chernoff bound combined with factorization gives [3]

$$\begin{aligned}
&\mathbb{P}\left(\check{Z}_k(\epsilon') \geq t \mid \mathcal{F}_k\right) \\
&= \mathbb{P}\left(\sum_{j=1}^{B} \left(\mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_{I_j^k}) - \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_i)\right) \geq Bt \,\middle|\, \mathcal{F}_k\right) \\
&= \mathbb{P}\left(\exp\left\{\lambda \sum_{j=1}^{B} \left(\mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_{I_j^k}) - \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_i)\right)\right\} \geq \exp\left(\lambda B t\right) \,\middle|\, \mathcal{F}_k\right) \\
&\leq \exp\left(-\lambda B t\right) \mathbb{E}\left[\exp\left(\lambda \sum_{j=1}^{B} \left(\mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_{I_j^k}) - \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_i)\right)\right) \,\middle|\, \mathcal{F}_k\right] \\
&\leq \exp\left(-\lambda B t\right) \prod_{j=1}^{B} \mathbb{E}\left[\exp\left(\lambda \left(\mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_{I_j^k}) - \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_i)\right)\right) \,\middle|\, \mathcal{F}_k\right].
\end{aligned}$$

Since $\mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_{I_j^k}) \in [-L_{\max}, L_{\max}]$, the centered variable $\mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_{I_j^k}) - n^{-1}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_i)$ has range length at most $2L_{\max}$. Therefore, Hoeffding's lemma yields for every λ,

$$\begin{aligned}
&\mathbb{E}\left[\exp\left\{\lambda \left(\mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_{I_j^k}) - \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_i)\right)\right\} \,\middle|\, \mathcal{F}_k\right] \\
&\leq \exp\left(\frac{\lambda^2 (2L_{\max})^2}{8}\right) = \exp\left(\frac{\lambda^2 L_{\max}^2}{2}\right).
\end{aligned}$$

[3] Here λ is a *local* variable limited to proof, not to be confused with Section 4.1.

This implies

$$\mathbb{P}\big(\check{Z}_k(\epsilon') \geq t \mid \mathcal{F}_k\big) \leq \exp\left(-\lambda B t + B\frac{\lambda^2 L_{\max}^2}{2}\right).$$

Minimizing the exponent over $\lambda > 0$ by solving

$$\frac{d\exp\left(-\lambda B t + B\frac{\lambda^2 L_{\max}^2}{2}\right)}{d\lambda} = 0$$

gives the optimal choice $\lambda^\star = t/L_{\max}^2$ which remains non-negative. Substituting back gives

$$\begin{aligned}\mathbb{P}\left(\check{Z}_k(\epsilon') \geq t \mid \mathcal{F}_k\right) &\leq \exp\left(-\lambda^\star B t + B\frac{(\lambda^\star)^2 L_{\max}^2}{2}\right) \\ &= \exp\left(-\frac{t}{L_{\max}^2}\cdot Bt + B\cdot\frac{t^2}{L_{\max}^4}\cdot\frac{L_{\max}^2}{2}\right) = \exp\left(-\frac{Bt^2}{2L_{\max}^2}\right).\end{aligned}$$

The same bound holds for $-\check{Z}_k(\epsilon')$ by repeating the argument with $-\lambda$. Combining both tails gives

$$\mathbb{P}\left(|\check{Z}_k(\epsilon')| \geq t \mid \mathcal{F}_k\right) \leq 2\exp\left(-\frac{Bt^2}{2L_{\max}^2}\right).$$

Applying a union bound over the η-net yields

$$\mathbb{P}\left(\max_{\epsilon'\in\mathcal{N}_\eta}|\check{Z}_k(\epsilon')| \geq t \;\Big|\; \mathcal{F}_k\right) \leq \sum_{\epsilon'\in\mathcal{N}_\eta}\mathbb{P}\left(|\check{Z}_k(\epsilon')| \geq t \mid \mathcal{F}_k\right) \leq 2N_\eta\exp\left(-\frac{Bt^2}{2L_{\max}^2}\right).$$

Using the tail integral identity for a nonnegative random variable and splitting at the threshold $t_\eta := L_{\max}\sqrt{2\log(2N_\eta)/B}$ where $2N_\eta\exp\left\{-Bt_\eta^2/(2L_{\max}^2)\right\} = 1$, we obtain

$$\begin{aligned}&\mathbb{E}\Big[\max_{\epsilon'\in\mathcal{N}_\eta}\left|\check{Z}_k(\epsilon')\right| \;\Big|\; \mathcal{F}_k\Big] = \int_0^\infty \mathbb{P}\left(\max_{\epsilon'\in\mathcal{N}_\eta}|\check{Z}_k(\epsilon')| \geq t \;\Big|\; \mathcal{F}_k\right)dt \\ &\leq \int_0^{t_\eta} 1\,dt + \int_{t_\eta}^\infty 2N_\eta\exp\left(-\frac{Bt^2}{2L_{\max}^2}\right)dt = t_\eta + 2N_\eta\int_{t_\eta}^\infty\exp\left(-\frac{Bt^2}{2L_{\max}^2}\right)dt. \qquad (20)\end{aligned}$$

We bound the Gaussian tail integral by explicit integration by parts. Let $\alpha := B/(2L_{\max}^2)$. Then for $t \geq t_\eta > 0$, we have

$$\begin{aligned}\frac{d}{dt}\left(\frac{1}{2\alpha t}\exp\{-\alpha t^2\}\right) &= -\exp\{-\alpha t^2\} - \frac{1}{2\alpha t^2}\exp\{-\alpha t^2\}, \\ \exp\{-\alpha t^2\} &= -\frac{d}{dt}\left(\frac{1}{2\alpha t}\exp\{-\alpha t^2\}\right) - \frac{1}{2\alpha t^2}\exp\{-\alpha t^2\}.\end{aligned}$$

Integrating from t_η to ∞ yields

$$\begin{aligned}\int_{t_\eta}^\infty\exp\{-\alpha t^2\}dt &= -\left[\frac{1}{2\alpha t}\exp\{-\alpha t^2\}\right]_{t=t_\eta}^{t=\infty} - \int_{t_\eta}^\infty\frac{1}{2\alpha t^2}\exp\{-\alpha t^2\}dt \\ &\leq \frac{1}{2\alpha t_\eta}\exp\{-\alpha t_\eta^2\}.\end{aligned}$$

Note that $\alpha t_\eta^2 = \frac{B}{2L_{\max}^2}\cdot\frac{2L_{\max}^2\log(2N_\eta)}{B} = \log(2N_\eta)$ so that $\exp\{-\alpha t_\eta^2\} = (2N_\eta)^{-1}$. Therefore,

$$\begin{aligned}&2N_\eta\int_{t_\eta}^\infty\exp\left(-\frac{Bt^2}{2L_{\max}^2}\right)dt \\ &\leq 2N_\eta\cdot\frac{1}{2\alpha t_\eta}\cdot\frac{1}{2N_\eta} \\ &= \frac{1}{2\alpha t_\eta} = \frac{L_{\max}^2}{Bt_\eta} = \frac{L_{\max}^2}{B\cdot L_{\max}\sqrt{2\log(2N_\eta)/B}} = \frac{L_{\max}}{\sqrt{B}}\cdot\frac{1}{\sqrt{2\log(2N_\eta)}}.\end{aligned}$$

In addition, we have

$$t_\eta = L_{\max}\sqrt{\frac{2\log(2N_\eta)}{B}} = \frac{L_{\max}}{\sqrt{B}} \cdot \sqrt{2\log(2N_\eta)}.$$

Substituting these expressions into the expectation bound (20) gives

$$\begin{aligned}\mathbb{E}\Big[\max_{\epsilon' \in \mathcal{N}_\eta} \left|\check{Z}_k(\epsilon')\right| \,\Big|\, \mathcal{F}_k\Big] &\le \frac{L_{\max}}{\sqrt{B}}\left(\sqrt{2\log(2N_\eta)} + \frac{1}{\sqrt{2\log(2N_\eta)}}\right) \\ &\le \frac{L_{\max}}{\sqrt{B}}\left(1 + \sqrt{2\log(2N_\eta)}\right),\end{aligned}$$

where the last inequality uses $\log(2N_\eta) \ge \log 2$. Choosing $\eta = L_{\max}/(L_\epsilon\sqrt{B})$ gives $2L_\epsilon\eta = 2L_{\max}/\sqrt{B}$ and $1 + \frac{R_\mathcal{E}}{\eta} = 1 + \frac{R_\mathcal{E} L_\epsilon \sqrt{B}}{L_{\max}}$. Combining this with (16) yields

$$\begin{aligned}\mathbb{E}\left[\Gamma_k \mid \mathcal{F}_k\right] &\le \frac{L_{\max}}{\sqrt{B}}\left(\sqrt{2\log(2N_\eta)} + 1\right) + 2L_\epsilon\eta \\ &\overset{(19)}{\le} \frac{L_{\max}}{\sqrt{B}}\left(\sqrt{2\left(\log 2 + d\log\left(1 + \frac{R_\mathcal{E}}{\eta}\right)\right)} + 1\right) + 2L_\epsilon\eta \\ &\le \frac{L_{\max}}{\sqrt{B}}\left(\sqrt{2\left(\log 2 + d\log\left(1 + \frac{R_\mathcal{E} L_\epsilon\sqrt{B}}{L_{\max}}\right)\right)} + 1\right) + \frac{2L_{\max}}{\sqrt{B}} \\ &= \frac{L_{\max}}{\sqrt{B}}\left(\sqrt{2\left(\log 2 + d\log\left(1 + \frac{R_\mathcal{E} L_\epsilon\sqrt{B}}{L_{\max}}\right)\right)} + 3\right).\end{aligned}$$

Taking expectation over $\mathcal{F}_k$ yields the same bound for $\mathbb{E}[\Gamma_k]$ uniformly in k. Substituting this uniform bound into (15) and taking $\limsup_{T\to\infty}$ yields

$$\begin{aligned}&\limsup_{T\to\infty} \frac{1}{T}\sum_{k=0}^{T-1} \mathbb{E}\left[\Delta(q^k)\right] \\ &\le \limsup_{T\to\infty}\left[\frac{\xi(\mathbf{0}) - L_{\min}}{T} + \frac{2}{T}\sum_{k=0}^{T-1}\frac{L_{\max}}{\sqrt{B}}\left(3 + \sqrt{2\left(\log 2 + d\log\left(1 + \frac{R_\mathcal{E} L_\epsilon\sqrt{B}}{L_{\max}}\right)\right)}\right)\right] \\ &\le \lim_{T\to\infty}\sup\frac{\xi(\mathbf{0}) - L_{\min}}{T} + \lim_{T\to\infty}\sup 2\frac{L_{\max}}{\sqrt{B}}\left(3 + \sqrt{2\left(\log 2 + d\log\left(1 + \frac{R_\mathcal{E} L_\epsilon\sqrt{B}}{L_{\max}}\right)\right)}\right) \\ &\le 2\frac{L_{\max}}{\sqrt{B}}\left(3 + \sqrt{2\left(\log 2 + d\log\left(1 + \frac{R_\mathcal{E} L_\epsilon\sqrt{B}}{L_{\max}}\right)\right)}\right).\end{aligned}$$

□

C.2 Case II

Assumption C.5. At each k, $\mathcal{B}^k$ is a uniformly random subset of $\{1, \dots, n\}$ of size B sampled without replacement.

Under Assumption C.5, we now prove the bound (3).

Proof. Following the same argument as in Case I, we obtain the bound analogous to (15):

$$\frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}[\Delta(q^k)] \le \frac{\xi(\mathbf{0}) - L_{\min}}{T} + \frac{2}{T}\sum_{k=0}^{T-1}\mathbb{E}[\Gamma_k].$$

Taking $\limsup_{T\to\infty}$ eliminates the first term and yields

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}\big[\Delta(q^k)\big] \le \limsup_{T\to\infty}\left[\frac{\xi(\mathbf{0})-L_{\min}}{T}+\frac{2}{T}\sum_{k=0}^{T-1}\mathbb{E}[\Gamma_k]\right]$$
$$\le \limsup_{T\to\infty}\frac{\xi(\mathbf{0})-L_{\min}}{T}+2\sup_{k\ge 0}\mathbb{E}[\Gamma_k] \ \le 2\sup_{k\ge 0}\mathbb{E}[\Gamma_k]. \tag{21}$$

Fix k and define the same Γ_k on $\mathcal{E}$ as in (10). We write $\Gamma_k = \sup_{\epsilon\in\mathcal{E}}\left|\check{X}_k(\epsilon)\right|$. Then

$$\mathbb{E}\left[\Gamma_k\right] = \mathbb{E}\left[\sup_{\epsilon\in\mathcal{E}}\left|\check{X}_k(\epsilon)\right|\right] \le \mathbb{E}\left[\left|\check{X}_k(\mathbf{0})\right|\right] + \mathbb{E}\left[\sup_{\epsilon\in\mathcal{E}}\left|\check{X}_k(\epsilon)-\check{X}_k(\mathbf{0})\right|\right].$$

We first consider the anchor term. Let $I_i^k = \mathbf{1}_{\{i\in\mathcal{B}^k\}}$. Uniform sampling without replacement of size B implies $\sum_{i=1}^n I_i^k = B$ and $\mathbb{E}\left[I_i^k\right] = \frac{B}{n}$. Then

$$\begin{aligned}
\check{X}_k(\mathbf{0}) &= \frac{1}{B}\sum_{j=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\mathbf{0}\mid\mathbf{G})\big)(O_{I_j^k}) - \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\mathbf{0}\mid\mathbf{G})\big)(O_i)\\
&= \frac{1}{B}\sum_{i=1}^{n} I_i^k\mathcal{L}(\mathbf{Q}_k)(O_i) - \frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}(\mathbf{Q}_k)(O_{i'})\\
&= \frac{1}{B}\sum_{j=1}^{n} I_j^k\mathcal{L}(\mathbf{Q}_k)(O_j) - \frac{1}{Bn}\sum_{i=1}^{n} I_i^k\sum_{i'=1}^{n}\mathcal{L}(\mathbf{Q}_k)(O_{i'})\\
&= \frac{1}{B}\sum_{i=1}^{n} I_i^k\left(\mathcal{L}(\mathbf{Q}_k)(O_i) - \frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}(\mathbf{Q}_k)(O_{i'})\right),
\end{aligned}$$

which gives $\mathbb{E}\left[\check{X}_k(\mathbf{0})\right] = 0$ and

$$\mathbb{E}\left[|\check{X}_k(\mathbf{0})|\right] \le \sqrt{\mathbb{E}\left[\check{X}_k(\mathbf{0})^2\right]} = \sqrt{\mathbb{E}\left[\check{X}_k(\mathbf{0})^2\right] - (\mathbb{E}\left[\check{X}_k(\mathbf{0})\right])^2} = \sqrt{\mathrm{Var}\big[\check{X}_k(\mathbf{0})\big]}.$$

We now compute the variance as

$$\begin{aligned}
&\mathrm{Var}\left[\check{X}_k(\mathbf{0})\right]\\
&= \mathbb{E}\left[\left\{\frac{1}{B}\sum_{i=1}^{n} I_i^k\left(\mathcal{L}(\mathbf{Q}_k)(O_i) - \frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}(\mathbf{Q}_k)(O_{i'})\right)\right\}^2\right]\\
&= \frac{1}{B^2}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} I_i^k I_j^k\left(\mathcal{L}(\mathbf{Q}_k)(O_i) - \frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}(\mathbf{Q}_k)(O_{i'})\right)\left(\mathcal{L}(\mathbf{Q}_k)(O_j) - \frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}(\mathbf{Q}_k)(O_{i'})\right)\right]\\
&= \frac{1}{B^2}\left\{\sum_{i=1}^{n}\mathbb{E}\left[I_i^k\left(\mathcal{L}(\mathbf{Q}_k)(O_i) - \frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}(\mathbf{Q}_k)(O_{i'})\right)^2\right]\right.\\
&\qquad \left. + \sum_{i\ne j}\mathbb{E}\left[I_i^k I_j^k\left(\mathcal{L}(\mathbf{Q}_k)(O_i) - \frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}(\mathbf{Q}_k)(O_{i'})\right)\left(\mathcal{L}(\mathbf{Q}_k)(O_j) - \frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}(\mathbf{Q}_k)(O_{i'})\right)\right]\right\}\\
&= \frac{1}{B^2}\left\{\sum_{i=1}^{n}\frac{B}{n}\left(\mathcal{L}(\mathbf{Q}_k)(O_i) - \frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}(\mathbf{Q}_k)(O_{i'})\right)^2 + \sum_{i\ne j}\frac{B(B-1)}{n(n-1)}\right.\\
&\quad \left.\cdot\left(\mathcal{L}(\mathbf{Q}_k)(O_i) - \frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}(\mathbf{Q}_k)(O_{i'})\right)\left(\mathcal{L}(\mathbf{Q}_k)(O_j) - \frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}(\mathbf{Q}_k)(O_{i'})\right)\right\}\\
&= \frac{1}{B^2}\left\{\frac{B}{n}\sum_{i=1}^{n}\left(\mathcal{L}(\mathbf{Q}_k)(O_i) - \frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}(\mathbf{Q}_k)(O_{i'})\right)^2\right.\\
&\qquad \left. + \frac{B(B-1)}{n(n-1)}\left[\left\{\sum_{i=1}^{n}\left(\mathcal{L}(\mathbf{Q}_k)(O_i) - \frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}(\mathbf{Q}_k)(O_{i'})\right)\right\}^2\right.\right.
\end{aligned}$$

$$
\begin{aligned}
&\quad - \sum_{i=1}^{n} \left(\mathcal{L}(\mathbf{Q}_k)(O_i) - \frac{1}{n} \sum_{i'=1}^{n} \mathcal{L}(\mathbf{Q}_k)(O_{i'}) \right)^2 \Bigg] \Bigg\} \\
&= \frac{1}{B^2} \left\{ \frac{B}{n} \sum_{i=1}^{n} \left(\mathcal{L}(\mathbf{Q}_k)(O_i) - \frac{1}{n} \sum_{i'=1}^{n} \mathcal{L}(\mathbf{Q}_k)(O_{i'}) \right)^2 \right. \\
&\quad \left. - \frac{B(B-1)}{n(n-1)} \sum_{i=1}^{n} \left(\mathcal{L}(\mathbf{Q}_k)(O_i) - \frac{1}{n} \sum_{i'=1}^{n} \mathcal{L}(\mathbf{Q}_k)(O_{i'}) \right)^2 \right\} \\
&= \frac{n-B}{B(n-1)} \cdot \frac{1}{n} \sum_{i=1}^{n} \underbrace{\left(\mathcal{L}(\mathbf{Q}_k)(O_i) - \frac{1}{n} \sum_{i'=1}^{n} \mathcal{L}(\mathbf{Q}_k)(O_{i'}) \right)^2}_{\triangleq \xi(\bar{n})}.
\end{aligned}
$$

By Assumption 2.3, we have $|\mathcal{L}(\mathbf{Q}_k)(O_i)| \le L_{\max}$ and hence $\xi(\bar{n}) \le n L_{\max}^2$. The deterministic inequality $\frac{n-B}{n-1} \le \rho_{B,n}$ holds for the given $\rho_{B,n}$ by direct algebra in the two regimes $B \le n/2$ and $B > n/2$ [47]. Therefore,

$$
\begin{aligned}
\operatorname{Var}\left[\check{X}_k(\mathbf{0})\right] &\le \frac{n-B}{Bn(n-1)} n L_{\max}^2 \le \frac{\rho_{B,n}}{B} L_{\max}^2, \\
\mathbb{E}\left[|\check{X}_k(\mathbf{0})|\right] &\le \sqrt{\operatorname{Var}\left[\check{X}_k(\mathbf{0})\right]} < L_{\max} \sqrt{\frac{2\rho_{B,n}}{B}}.
\end{aligned} \tag{22}
$$

For convenience, we define the increment term $\check{Y}_k(\epsilon)$ as

$$
\check{Y}_k(\epsilon) \triangleq \check{X}_k(\epsilon) - \check{X}_k(\mathbf{0}), \sup_{\epsilon \in \mathcal{E}} \left| \check{X}_k(\epsilon) - \check{X}_k(\mathbf{0}) \right| = \sup_{\epsilon \in \mathcal{E}} \left| \check{Y}_k(\epsilon) \right|.
$$

Fix $\epsilon, \epsilon' \in \mathcal{E}$. Then

$$
\begin{aligned}
\check{Y}_k(\epsilon) - \check{Y}_k(\epsilon') &= \left[\frac{1}{B} \sum_{j \in \mathcal{B}^k} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_j) - \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i) \right] \\
&\quad - \left[\frac{1}{B} \sum_{j \in \mathcal{B}^k} \mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_j) - \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_i) \right] \\
&= \frac{1}{B} \sum_{i \in \mathcal{B}^k} \left[\Big(\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i) - \mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_i) \Big) \right. \\
&\quad \left. - \left(\frac{1}{n} \sum_{i'=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{i'}) - \mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_{i'}) \Big) \right].
\end{aligned} \tag{23}
$$

By Assumption 2.5, we have

$$
\begin{aligned}
\check{\check{Y}}_k(\epsilon) &\triangleq \max_i \left\{ \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i) - \mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_i) \right\} \\
&\quad - \min_i \left\{ \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i) - \mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_i) \right\} \\
&\le L_\epsilon \left\| \epsilon - \epsilon' \right\| + L_\epsilon \left\| \epsilon - \epsilon' \right\| = 2 L_\epsilon \left\| \epsilon - \epsilon' \right\|.
\end{aligned} \tag{24}
$$

The classical Serfling bound for simple random sampling without replacement provides a one-sided tail for (23) with finite-population correction. Using the complement trick when $B > n/2$ yields exactly the piecewise $\rho_{B,n}$ in the statement (4). The two-sided form is

$$
\begin{aligned}
\mathbb{P}\left(\left| \check{Y}_k(\epsilon) - \check{Y}_k(\epsilon') \right| \ge t \right) &\le 2 \exp\left\{ - \frac{2Bt^2}{\rho_{B,n} \left(\check{\check{Y}}_k(\epsilon) \right)^2} \right\} \\
&\overset{(24)}{\le} 2 \exp\left\{ - \frac{2Bt^2}{\rho_{B,n} 4 L_\epsilon^2 \left\| \epsilon - \epsilon' \right\|^2} \right\} = 2 \exp\left\{ - \frac{t^2}{2 L_\epsilon^2 \frac{\rho_{B,n}}{B} \left\| \epsilon - \epsilon' \right\|^2} \right\}.
\end{aligned}
$$

Define the metric on $\mathcal{E}$ by

$$\tilde{d}_{B,n}(\epsilon,\epsilon') = L_\epsilon\sqrt{\frac{\rho_{B,n}}{B}}\,\|\epsilon-\epsilon'\|,$$

so the increment bound becomes

$$\mathbb{P}\left(\left|\check{Y}_k(\epsilon)-Y_k(\epsilon')\right|\geq t\right)\leq 2\exp\left\{-\frac{t^2}{2d_{B,n}(\epsilon,\epsilon')^2}\right\}.$$

Let $N(\mathcal{E},\tilde{d}_{B,n},u)$ denote the minimal covering number at radius u in metric $\tilde{d}_{B,n}$. Fix $m\geq 0$ and let $\mathcal{E}_m\subset\mathcal{E}$ be a $2^{-m}\mathrm{diam}(\mathcal{E},\tilde{d}_{B,n})$-net with cardinality $N\left(\mathcal{E},\tilde{d}_{B,n},2^{-m}\mathrm{diam}(\mathcal{E},\tilde{d}_{B,n})\right)$. For each $\epsilon\in\mathcal{E}$, let $\pi_m(\epsilon)\in\mathcal{E}_m$ satisfy

$$\tilde{d}_{B,n}\left(\epsilon,\pi_m(\epsilon)\right)\leq 2^{-m}\mathrm{diam}(\mathcal{E},\tilde{d}_{B,n}),\quad \pi_0(\epsilon)=\mathbf{0}.$$

For every integer $M\geq 1$ and every $\epsilon\in\mathcal{E}$, we have

$$\begin{aligned}
\sup_{\epsilon\in\mathcal{E}}\left|\check{Y}_k(\epsilon)\right| &= \left|\check{Y}_k(\epsilon)-\check{Y}_k\big(\pi_M(\epsilon)\big)+\sum_{m=1}^{M}\left\{\check{Y}_k\big(\pi_m(\epsilon)\big)-\check{Y}_k\big(\pi_{m-1}(\epsilon)\big)\right\}+\check{Y}_k\big(\pi_0(\epsilon)\big)\right| \\
&= \left|\check{Y}_k(\epsilon)-\check{Y}_k\big(\pi_M(\epsilon)\big)+\sum_{m=1}^{M}\left\{\check{Y}_k\big(\pi_m(\epsilon)\big)-\check{Y}_k\big(\pi_{m-1}(\epsilon)\big)\right\}\right| \\
&\leq \sup_{\epsilon\in\mathcal{E}}\left|\check{Y}_k(\epsilon)-\check{Y}_k\big(\pi_M(\epsilon)\big)\right|+\sum_{m=1}^{M}\sup_{\epsilon\in\mathcal{E}}\left|\check{Y}_k\big(\pi_m(\epsilon)\big)-\check{Y}_k\big(\pi_{m-1}(\epsilon)\big)\right| \\
&\leq \sup_{\epsilon\in\mathcal{E}}\left|\check{Y}_k(\epsilon)-\check{Y}_k\big(\pi_M(\epsilon)\big)\right|+\sum_{m=1}^{M}\max_{a\in\mathcal{E}_m}\max_{b\in\mathcal{E}_{m-1}}\left|\check{Y}_k(a)-\check{Y}_k(b)\right|.
\end{aligned}$$

For the residual term, note that for each $\epsilon\in\mathcal{E}$ we have $\tilde{d}_{B,n}\big(\epsilon,\pi_M(\epsilon)\big)\leq 2^{-M}\mathrm{diam}(\mathcal{E},d_{B,n})$. The increment tail bound then yields

$$\begin{aligned}
\mathbb{E}\left|\check{Y}_k(\epsilon)-\check{Y}_k\big(\pi_M(\epsilon)\big)\right| &= \int_0^\infty \mathbb{P}\left(\left|\check{Y}_k(\epsilon)-\check{Y}_k\big(\pi_M(\epsilon)\big)\right|\geq t\right)dt \\
&\leq \int_0^\infty 2\exp\left\{-\frac{t^2}{2\cdot 2^{-2M}\mathrm{diam}(\mathcal{E},\tilde{d}_{B,n})^2}\right\}dt \\
&= 2^{-M}\mathrm{diam}(\mathcal{E},\tilde{d}_{B,n})\int_0^\infty 2\exp\left\{-\frac{s^2}{2}\right\}ds = 2^{-M}\mathrm{diam}(\mathcal{E},\tilde{d}_{B,n})\sqrt{2\pi},
\end{aligned}$$

so taking $\sup_{\epsilon\in\mathcal{E}}$ and then letting $M\to\infty$ yields vanishing contribution in expectation. For the chaining term, note that for any $a\in\mathcal{E}_m$ and $b\in\mathcal{E}_{m-1}$ that arise as $\pi_m(\epsilon)$ and $\pi_{m-1}(\epsilon)$ we have

$$\begin{aligned}
\tilde{d}_{B,n}(a,b) &\leq \tilde{d}_{B,n}(a,\epsilon)+\tilde{d}_{B,n}(\epsilon,b) \\
&\leq 2^{-m}\mathrm{diam}(\mathcal{E},\tilde{d}_{B,n})+2^{-(m-1)}\mathrm{diam}(\mathcal{E},\tilde{d}_{B,n}) = 3\cdot 2^{-m}\mathrm{diam}(\mathcal{E},\tilde{d}_{B,n}),
\end{aligned}$$

and hence the sub-Gaussian tail gives for every $t>0$

$$\mathbb{P}\left(\left|\check{Y}_k(a)-\check{Y}_k(b)\right|\geq t\right)\leq 2\exp\left\{-\frac{t^2}{2\cdot 9\cdot 2^{-2m}\mathrm{diam}(\mathcal{E},\tilde{d}_{B,n})^2}\right\}.$$

Let $M_m=|\mathcal{E}_m||\mathcal{E}_{m-1}|$. By a union bound and tail integration,

$$\begin{aligned}
\mathbb{E}\left[\max_{a\in\mathcal{E}_m}\max_{b\in\mathcal{E}_{m-1}}\left|\check{Y}_k(a)-\check{Y}_k(b)\right|\right] &= \int_0^\infty \mathbb{P}\left(\max_{a\in\mathcal{E}_m}\max_{b\in\mathcal{E}_{m-1}}\left|\check{Y}_k(a)-\check{Y}_k(b)\right|\geq t\right)dt \\
&\leq \int_0^\infty \min\left\{1,\,2M_m\exp\left[-\frac{t^2}{18\cdot 2^{-2m}\mathrm{diam}(\mathcal{E},\tilde{d}_{B,n})^2}\right]\right\}dt \\
&\leq 3\cdot 2^{-m}\mathrm{diam}(\mathcal{E},\tilde{d}_{B,n})\left\{\sqrt{2\log 2M_m}+\frac{\sqrt{\pi}}{2}\right\},
\end{aligned}$$

and since $M_m \le |\mathcal{E}_m|^2$ and $|\mathcal{E}_m| = N\left(\mathcal{E}, \tilde{d}_{B,n}, 2^{-m}\mathrm{diam}(\mathcal{E}, \tilde{d}_{B,n})\right)$ we have

$$\sqrt{\log 2M_m} \le \sqrt{\log 2 + 2\log|\mathcal{E}_m|} \le \sqrt{\log 2} + \sqrt{2\log|\mathcal{E}_m|}.$$

Summing over the dyadic scales and using monotonicity of $u \mapsto \sqrt{\log N(\mathcal{E}, \tilde{d}_{B,n}, u)}$ gives the entropy integral bound

$$\mathbb{E}\left[\sup_{\epsilon\in\mathcal{E}} \check{Y}_k(\epsilon)\right] \le 12 \int_0^{\mathrm{diam}(\mathcal{E},\tilde{d}_{B,n})} \sqrt{\log N(\mathcal{E}, \tilde{d}_{B,n}, u)}\, du,$$

which is justified in Lemma C.6. Hence we have

$$\begin{aligned}\mathbb{E}\left[\sup_{\epsilon\in\mathcal{E}} \left|\check{Y}_k(\epsilon)\right|\right] &\le \mathbb{E}\left[\sup_{\epsilon\in\mathcal{E}} \check{Y}_k(\epsilon)\right] + \mathbb{E}\left[\sup_{\epsilon\in\mathcal{E}} -\check{Y}_k(\epsilon)\right] \\ &\le 24 \int_0^{\mathrm{diam}(\mathcal{E},\tilde{d}_{B,n})} \sqrt{\log N(\mathcal{E}, \tilde{d}_{B,n}, u)}\, du. \end{aligned} \tag{25}$$

Using Assumption 2.4 and the volumetric bound for Euclidean covering numbers under the scaled metric,

$$\begin{aligned}\mathrm{diam}(\mathcal{E}, \tilde{d}_{B,n}) &= \sup_{\epsilon,\epsilon'\in\mathcal{E}} L_\epsilon \sqrt{\frac{\rho_{B,n}}{B}}\, \|\epsilon - \epsilon'\| = L_\epsilon R_\mathcal{E} \sqrt{\frac{\rho_{B,n}}{B}}, \\ N(\mathcal{E}, \tilde{d}_{B,n}, u) &= N\left(\mathcal{E}, \|\epsilon - \epsilon'\|, \frac{u}{L_\epsilon\sqrt{\rho_{B,n}/B}}\right) \le \left(1 + \frac{L_\epsilon R_\mathcal{E}\sqrt{\rho_{B,n}/B}}{u}\right)^d,\end{aligned}$$

and hence

$$\begin{aligned}\mathbb{E}\left[\sup_{\epsilon\in\mathcal{E}} \left|\check{Y}_k(\epsilon)\right|\right] &\overset{(25)}{\le} 24 \int_0^{L_\epsilon R_\mathcal{E}\sqrt{\rho_{B,n}/B}} \sqrt{d\log\left(1 + \frac{L_\epsilon R_\mathcal{E}\sqrt{\rho_{B,n}/B}}{u}\right)}\, du \\ &= 24 L_\epsilon R_\mathcal{E} \sqrt{\frac{\rho_{B,n} d}{B}} \int_0^1 \sqrt{\log\left(1 + \frac{1}{t}\right)}\, dt. \end{aligned} \tag{26}$$

For $t \in (0, 1]$, we have $1 + 1/t \le 2/t$. For $\varepsilon \in (0, 1)$, consider the improper integral in (26) as

$$\begin{aligned}&\int_0^1 \sqrt{\log\left(1 + \frac{1}{t}\right)}\, dt = \lim_{\varepsilon\to 0^+} \int_\varepsilon^1 \sqrt{\log\left(1 + \frac{1}{t}\right)}\, dt \le \lim_{\varepsilon\to 0^+} \int_\varepsilon^1 \sqrt{\log\left(\frac{2}{t}\right)} \\ &\overset{(t=e^{-s})}{=} \lim_{\varepsilon\to 0^+} \int_0^{-\log\varepsilon} \sqrt{\log 2 + s}\, e^{-s}\, ds \le \lim_{\varepsilon\to 0^+} \int_0^{-\log\varepsilon} \left[\sqrt{\log 2} + \sqrt{s}\right] e^{-s}\, ds.\end{aligned}$$

Using monotone convergence and letting $\varepsilon \to 0^+$, we obtain

$$\int_0^1 \sqrt{\log\left(1 + \frac{1}{t}\right)}\, dt \le \sqrt{\log 2} \int_0^\infty e^{-s}\, ds + \int_0^\infty s^{1/2} e^{-s}\, ds = \sqrt{\log 2} + \hat{\Gamma}\left(\frac{3}{2}\right),$$

where the Gamma function satisfies $\hat{\Gamma}\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2}$. Combining the anchor bound (22) and the increment bound (26) yields

$$\begin{aligned}\mathbb{E}\left[\Gamma_k\right] &\le \mathbb{E}\left[\left|\check{X}_k(\mathbf{0})\right|\right] + \mathbb{E}\left[\sup_{\epsilon\in\mathcal{E}} \left|\check{Y}_k(\epsilon)\right|\right] \\ &\le L_{\max}\sqrt{\frac{2\rho_{B,n}}{B}} + 24\left(\sqrt{\log 2} + \frac{\sqrt{\pi}}{2}\right) L_\epsilon R_\mathcal{E} \sqrt{\frac{\rho_{B,n} d}{B}},\end{aligned}$$

uniformly in k. Hence $\sup_{k\ge 0} \mathbb{E}[\Gamma_k]$ satisfies the same bound. Inserting this into (21) yields

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{E}\left[\Delta\left(q^k\right)\right] \le 2\left(24\left(\sqrt{\log 2} + \frac{\sqrt{\pi}}{2}\right) L_\epsilon R_\mathcal{E} \sqrt{\frac{\rho_{B,n} d}{B}} + L_{\max}\sqrt{\frac{2\rho_{B,n}}{B}}\right).$$

□

Lemma C.6. *We have*

$$\mathbb{E}\left[\sup_{\epsilon\in\mathcal{E}} \check{Y}_k(\epsilon)\,\middle|\,\mathscr{F}_k\right] \le 12\int_0^{\operatorname{diam}(\mathcal{E},\tilde{d}_{B,n})} \sqrt{\log N(u)}\,du,$$

where $d_{B,n}$ is the metric defined in the proof of Case II, $N(u)$ is the minimal cardinality of a $\tilde{d}_{B,n}$-cover of $\mathcal{E}$ at radius u, and $\check{Y}_k(\epsilon)$ is the centered empirical-process term in the above STMLE analysis.

We now prove Lemma C.6, which was used in the proof above.

Proof. Fix k and condition on $\mathscr{F}_k$ throughout the proof. All expectations and probabilities below are conditional on $\mathscr{F}_k$, so $\mathbf{Q}_k$ and $\{O_i\}_{i=1}^n$ are deterministic and the only randomness is the uniform size-B subset $\mathcal{B}^k$. Recall that $\tilde{d}_{B,n}(\epsilon,\epsilon') := L_\epsilon\sqrt{\rho_{B,n}/B}\,\|\epsilon-\epsilon'\|$, that $\operatorname{diam}(\mathcal{E},\tilde{d}_{B,n})<\infty$ by Assumption 2.4, and that $N(u)$ denotes the minimal cardinality of a $\tilde{d}_{B,n}$-cover of $\mathcal{E}$ at radius u.

If $B=n$, then $\mathcal{B}^k=\{1,\dots,n\}$ deterministically so $\check{Y}_k(\epsilon)\equiv 0$ for all ϵ and the desired inequality holds trivially since $\operatorname{diam}(\mathcal{E},\tilde{d}_{B,n})=0$. We therefore assume $B<n$, which implies $\rho_{B,n}>0$ and $\operatorname{diam}(\mathcal{E},\tilde{d}_{B,n})>0$. By Assumption 2.5, we have $\left|\mathcal{L}\big(\mathbf{Q}_k(\epsilon\mid\mathbf{G})\big)(o)-\mathcal{L}\big(\mathbf{Q}_k(\epsilon'\mid\mathbf{G})\big)(o)\right|\le L_\epsilon\|\epsilon-\epsilon'\|$ for every $o=O_i$, and hence

$$-L_\epsilon\,\|\epsilon-\epsilon'\| \le \mathcal{L}\big(\mathbf{Q}_k(\epsilon\mid\mathbf{G})\big)(O_i)-\mathcal{L}\big(\mathbf{Q}_k(\epsilon'\mid\mathbf{G})\big)(O_i) \le L_\epsilon\,\|\epsilon-\epsilon'\|\,.$$

Using the standard exponential-moment inequality underlying the Hoeffding-Serfling bound [47], we have for every $\lambda\in\mathbb{R}$

$$\begin{aligned}
&\log\mathbb{E}\left[\exp\left(\lambda\left(\frac{1}{B}\sum_{i\in\mathcal{B}^k}\big(\mathcal{L}\big(\mathbf{Q}_k(\epsilon\mid\mathbf{G})\big)(O_i)-\mathcal{L}\big(\mathbf{Q}_k(\epsilon'\mid\mathbf{G})\big)(O_i)\big)\right.\right.\right.\\
&\left.\left.\left.-\frac{1}{n}\sum_{i=1}^n\Big(\mathcal{L}\big(\mathbf{Q}_k(\epsilon\mid\mathbf{G})\big)-\mathcal{L}\big(\mathbf{Q}_k(\epsilon'\mid\mathbf{G})\big)\Big)(O_i)\right)\right)\right]\\
&\le\frac{\lambda^2}{2}\cdot\frac{\rho_{B,n}}{B}\left(L_\epsilon\,\|\epsilon-\epsilon'\|\right)^2,
\end{aligned}$$

where $n^{-1}\sum_{i=1}^n(\cdot)(O_i)$ is the population mean and $\rho_{B,n}$ is exactly the piecewise finite-population factor used in (4). Therefore, for all $\lambda\in\mathbb{R}$ and all $\epsilon,\epsilon'\in\mathcal{E}$, the definition of $\tilde{d}_{B,n}$ yields the conditional increment bound

$$\log\mathbb{E}\left[\exp\Big(\lambda\left(\check{Y}_k(\epsilon)-\check{Y}_k(\epsilon')\right)\Big)\,\middle|\,\mathscr{F}_k\right]\overset{(23)}{\le}\frac{\lambda^2}{2}\,L_\epsilon^2\|\epsilon-\epsilon'\|^2\,\frac{\rho_{B,n}}{B}=\frac{\lambda^2}{2}\tilde{d}_{B,n}(\epsilon,\epsilon')^2. \tag{27}$$

Define dyadic radii $u_j:=2^{-j}\operatorname{diam}(\mathcal{E},\tilde{d}_{B,n})$ for $j\ge 0$. For each $j\ge 0$, choose a minimal u_j-net $\mathcal{E}_j\subset\mathcal{E}$ so that

$$|\mathcal{E}_j|=N(u_j),\quad\forall\epsilon\in\mathcal{E},\ \exists\pi_j(\epsilon)\in\mathcal{E}_j\text{ s.t. }\tilde{d}_{B,n}(\epsilon,\pi_j(\epsilon))\le u_j.$$

We take $\mathcal{E}_0=\{\mathbf{0}\}$ and $\pi_0(\epsilon)\equiv\mathbf{0}$, which is valid since $u_0=\operatorname{diam}(\mathcal{E},\tilde{d}_{B,n})$ covers $\mathcal{E}$ by one ball and also ensures $\check{Y}_k(\pi_0(\epsilon))=\check{Y}_k(\mathbf{0})=0$. For any integer $m\ge 1$ and any $\epsilon\in\mathcal{E}$, the telescoping identity gives

$$\begin{aligned}
\sup_{\epsilon\in\mathcal{E}}\check{Y}_k(\epsilon)&=\sup_{\epsilon\in\mathcal{E}}\left[\check{Y}_k(\pi_m(\epsilon))+\big(\check{Y}_k(\epsilon)-\check{Y}_k(\pi_m(\epsilon))\big)\right]\\
&=\sup_{\epsilon\in\mathcal{E}}\left[\sum_{j=1}^m\big(\check{Y}_k(\pi_j(\epsilon))-\check{Y}_k(\pi_{j-1}(\epsilon))\big)+\big(\check{Y}_k(\epsilon)-\check{Y}_k(\pi_m(\epsilon))\big)\right]\\
&\le\sum_{j=1}^m\sup_{\epsilon\in\mathcal{E}}\big(\check{Y}_k(\pi_j(\epsilon))-\check{Y}_k(\pi_{j-1}(\epsilon))\big)+\sup_{\epsilon\in\mathcal{E}}\left|\check{Y}_k(\epsilon)-\check{Y}_k(\pi_m(\epsilon))\right|.
\end{aligned}\tag{28}$$

The remainder term is deterministically controlled using Assumption 2.5. For any ϵ, ϵ', we have

$$\begin{aligned}
\left|\check{Y}_k(\epsilon) - \check{Y}_k(\epsilon')\right| &\le \left|\frac{1}{B}\sum_{j\in\mathcal{B}^k} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_j) - \frac{1}{B}\sum_{j\in\mathcal{B}^k} \mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_j)\right| \\
&\quad + \left|\frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i) - \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon' \mid \mathbf{G})\big)(O_i)\right| \\
&\overset{(24)}{\le} 2L_\epsilon\|\epsilon - \epsilon'\| = 2\sqrt{\frac{B}{\rho_{B,n}}}\,\tilde{d}_{B,n}(\epsilon, \epsilon'),
\end{aligned}$$

so with $\epsilon' = \pi_m(\epsilon)$ and $\tilde{d}_{B,n}(\epsilon, \pi_m(\epsilon)) \le u_m$ we obtain

$$\sup_{\epsilon\in\mathcal{E}} \left|\check{Y}_k(\epsilon) - \check{Y}_k(\pi_m(\epsilon))\right| \le 2\sqrt{\frac{B}{\rho_{B,n}}}\,u_m = 2\sqrt{\frac{B}{\rho_{B,n}}}\,2^{-m}\operatorname{diam}(\mathcal{E}, \tilde{d}_{B,n}).$$

Letting $m \to \infty$ yields

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\epsilon\in\mathcal{E}} \check{Y}_k(\epsilon) \,\middle|\, \mathscr{F}_k\right] \\
&\overset{(28)}{\le} \lim_{m\to\infty} \mathbb{E}\left[\sum_{j=1}^{m} \sup_{\epsilon\in\mathcal{E}} \left(\check{Y}_k(\pi_j(\epsilon)) - \check{Y}_k(\pi_{j-1}(\epsilon))\right) + \sup_{\epsilon\in\mathcal{E}} \left|\check{Y}_k(\epsilon) - \check{Y}_k(\pi_m(\epsilon))\right| \,\middle|\, \mathscr{F}_k\right] \\
&\le \sum_{j=1}^{\infty} \mathbb{E}\left[\sup_{\epsilon\in\mathcal{E}} \left(\check{Y}_k(\pi_j(\epsilon)) - \check{Y}_k(\pi_{j-1}(\epsilon))\right) \,\middle|\, \mathscr{F}_k\right] + \lim_{m\to\infty} 2\sqrt{\frac{B}{\rho_{B,n}}}\,2^{-m}\operatorname{diam}(\mathcal{E}, \tilde{d}_{B,n}) \\
&= \sum_{j=1}^{\infty} \mathbb{E}\left[\sup_{\epsilon\in\mathcal{E}} \left(\check{Y}_k(\pi_j(\epsilon)) - \check{Y}_k(\pi_{j-1}(\epsilon))\right) \,\middle|\, \mathscr{F}_k\right]. \qquad (29)
\end{aligned}$$

Fix $j \ge 1$ and define the finite edge set of pairs

$$\mathcal{A}_j := \left\{\left(\pi_j(\epsilon), \pi_{j-1}(\epsilon)\right) : \epsilon \in \mathcal{E}\right\} \subset \mathcal{E}_j \times \mathcal{E}_{j-1}.$$

Then

$$|\mathcal{A}_j| \le |\mathcal{E}_j| \cdot |\mathcal{E}_{j-1}| = N(u_j)N(u_{j-1}) \le N(u_j)^2$$

since $u_{j-1} = 2u_j$ and $N(\cdot)$ is nonincreasing. Moreover, for any $(a, b) \in \mathcal{A}_j$ there exists ϵ with $a = \pi_j(\epsilon)$ and $b = \pi_{j-1}(\epsilon)$, and by the triangle inequality we have

$$\tilde{d}_{B,n}(a, b) \le \tilde{d}_{B,n}(a, \epsilon) + \tilde{d}_{B,n}(\epsilon, b) \le u_j + u_{j-1} = u_j + 2u_j = 3u_j. \qquad (30)$$

Therefore,

$$\sup_{\epsilon\in\mathcal{E}} \left(\check{Y}_k(\pi_j(\epsilon)) - \check{Y}_k(\pi_{j-1}(\epsilon))\right) = \max_{(a,b)\in\mathcal{A}_j} \left(\check{Y}_k(a) - \check{Y}_k(b)\right). \qquad (31)$$

To bound the conditional expectation of this maximum, fix $\lambda > 0$ and apply Jensen's inequality:

$$\begin{aligned}
&\mathbb{E}\left[\max_{(a,b)\in\mathcal{A}_j} \left(\check{Y}_k(a) - \check{Y}_k(b)\right) \,\middle|\, \mathscr{F}_k\right] = \frac{1}{\lambda}\,\mathbb{E}\left[\log\exp\left(\lambda \max_{(a,b)\in\mathcal{A}_j} \left(\check{Y}_k(a) - \check{Y}_k(b)\right)\right) \,\middle|\, \mathscr{F}_k\right] \\
&\le \frac{1}{\lambda}\,\log\mathbb{E}\left[\exp\left(\lambda \max_{(a,b)\in\mathcal{A}_j} \left(\check{Y}_k(a) - \check{Y}_k(b)\right)\right) \,\middle|\, \mathscr{F}_k\right] \\
&\le \frac{1}{\lambda}\,\log\mathbb{E}\left[\sum_{(a,b)\in\mathcal{A}_j} \exp\left(\lambda\left(\check{Y}_k(a) - \check{Y}_k(b)\right)\right) \,\middle|\, \mathscr{F}_k\right] \\
&= \frac{1}{\lambda}\,\log \sum_{(a,b)\in\mathcal{A}_j} \mathbb{E}\left[\exp\left(\lambda(\check{Y}_k(a) - \check{Y}_k(b))\right) \,\middle|\, \mathscr{F}_k\right] \overset{(27)}{\le} \frac{1}{\lambda}\,\log \sum_{(a,b)\in\mathcal{A}_j} \exp\left(\frac{\lambda^2}{2}\tilde{d}_{B,n}(a, b)^2\right) \\
&\overset{(30)}{\le} \frac{1}{\lambda}\,\log\left(|\mathcal{A}_j|\exp\left(\frac{\lambda^2}{2}(3u_j)^2\right)\right) = \frac{\log|\mathcal{A}_j|}{\lambda} + \frac{\lambda}{2}(3u_j)^2. \qquad (32)
\end{aligned}$$

The last expression in (32) is minimized over $\lambda > 0$ by taking the derivative and setting it to zero, which gives $-\log|\mathcal{A}_j|/\lambda^2 + (3u_j)^2/2 = 0$ and hence

$$\lambda^\star = \frac{\sqrt{2\log|\mathcal{A}_j|}}{3u_j}.$$

Substituting $\lambda^\star$ into (32) yields

$$\begin{aligned}\mathbb{E}\left[\max_{(a,b)\in\mathcal{A}_j}\left(\check{Y}_k(a)-\check{Y}_k(b)\right)\middle|\mathscr{F}_k\right] &\le 3u_j\sqrt{2\log|\mathcal{A}_j|}\\ &\le 3u_j\sqrt{2\log\left(N(u_j)^2\right)} = 6u_j\sqrt{\log N(u_j)}. \end{aligned} \tag{33}$$

Combining (29), (31), and (33) gives

$$\mathbb{E}\left[\sup_{\epsilon\in\mathcal{E}}\check{Y}_k(\epsilon)\middle|\mathscr{F}_k\right] \le \sum_{j=1}^{\infty} 6u_j\sqrt{\log N(u_j)}. \tag{34}$$

It remains to convert the dyadic sum (34) into the entropy integral. Since $u \mapsto N(u)$ is nonincreasing, for each $j \ge 1$ and each $u \in [u_{j+1}, u_j]$ we have $N(u) \ge N(u_j)$ and hence $\sqrt{\log N(u)} \ge \sqrt{\log N(u_j)}$. Therefore,

$$\int_{u_{j+1}}^{u_j}\sqrt{\log N(u)}\,du \ge (u_j-u_{j+1})\sqrt{\log N(u_j)} = \left(u_j-\frac{u_j}{2}\right)\sqrt{\log N(u_j)} = \frac{u_j}{2}\sqrt{\log N(u_j)},$$

which implies

$$6u_j\sqrt{\log N(u_j)} \le 12\int_{u_{j+1}}^{u_j}\sqrt{\log N(u)}\,du.$$

Summing over $j \ge 1$ and using that the intervals $[u_{j+1}, u_j]$ partition $(0, u_1]$ with $u_1 = \operatorname{diam}(\mathcal{E}, \tilde{d}_{B,n})/2 \le \operatorname{diam}(\mathcal{E}, \tilde{d}_{B,n})$, we obtain

$$\begin{aligned}\sum_{j=1}^{\infty} 6u_j\sqrt{\log N(u_j)} &\le 12\sum_{j=1}^{\infty}\int_{u_{j+1}}^{u_j}\sqrt{\log N(u)}\,du = 12\int_0^{u_1}\sqrt{\log N(u)}\,du\\ &\le 12\int_0^{\operatorname{diam}(\mathcal{E},\tilde{d}_{B,n})}\sqrt{\log N(u)}\,du.\end{aligned}$$

This yields the desired inequality

$$\mathbb{E}\left[\sup_{\epsilon\in\mathcal{E}}\check{Y}_k(\epsilon)\middle|\mathscr{F}_k\right] \overset{(34)}{\le} 12\int_0^{\operatorname{diam}(\mathcal{E},\tilde{d}_{B,n})}\sqrt{\log N(u)}\,du,$$

which completes the proof. □

D Proof of Lemma 3.2

Proof. We first note that $\mathbf{J}(\mathbf{Q}) \in \mathbb{R}$ is finite and in particular $\mathbf{J}(\mathbf{Q})$ is integrable. Using the law of total expectation and expanding over the finite support of $(\mathcal{B}^0, \ldots, \mathcal{B}^{K-1})$ with conditional uniformity, we obtain

$$\begin{aligned}&\mathbb{E}\Big[\min\left\{\mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}),\,\mathbf{J}(\mathbf{Q}_{\mathrm{ST}})\right\}\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\min\left\{\mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}),\,\mathbf{J}(\mathbf{Q}_{\mathrm{ST}})\right\}\Big|\,\mathcal{B}^0,\mathcal{B}^1,\ldots,\mathcal{B}^{K-1}\Big]\Big]\\
&= \sum_{\substack{\mathcal{B}^0\subset\{1,\ldots,n\}\\|\mathcal{B}^0|=B}}\sum_{\substack{\mathcal{B}^1\subset\{1,\ldots,n\}\\|\mathcal{B}^1|=B}}\cdots\sum_{\substack{\mathcal{B}^{K-1}\subset\{1,\ldots,n\}\\|\mathcal{B}^{K-1}|=B}}\min\left\{\mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}),\mathbf{J}(\mathbf{Q}_{\mathrm{ST}}(\mathcal{B}^0,\ldots,\mathcal{B}^{K-1}))\right\}\mathbb{P}\left(\mathcal{B}^0,\ldots,\mathcal{B}^{K-1}\right)\\
&= \sum_{\substack{\mathcal{B}^0\subset\{1,\ldots,n\}\\|\mathcal{B}^0|=B}}\cdots\sum_{\substack{\mathcal{B}^{K-1}\subset\{1,\ldots,n\}\\|\mathcal{B}^{K-1}|=B}}\min\left\{\mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}),\,\mathbf{J}(\mathbf{Q}_{\mathrm{ST}}(\mathcal{B}^0,\ldots,\mathcal{B}^{K-1}))\right\}\prod_{k=0}^{K-1}\mathbb{P}\left(\mathcal{B}^k\,\middle|\,\mathcal{B}^0,\ldots,\mathcal{B}^{k-1}\right)\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\substack{\mathcal{B}^0\subset\{1,\dots,n\}\\|\mathcal{B}^0|=B}} \cdots \sum_{\substack{\mathcal{B}^{K-1}\subset\{1,\dots,n\}\\|\mathcal{B}^{K-1}|=B}} \min\Big\{\mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}})\,,\ \mathbf{J}\big(\mathbf{Q}_{\mathrm{ST}}(\mathcal{B}^0,\dots,\mathcal{B}^{K-1})\big)\Big\} \prod_{k=0}^{K-1}\binom{n}{B}^{-1} \\
&= \binom{n}{B}^{-K} \sum_{\substack{\mathcal{B}^0\subset\{1,\dots,n\}\\|\mathcal{B}^0|=B}} \cdots \sum_{\substack{\mathcal{B}^{K-1}\subset\{1,\dots,n\}\\|\mathcal{B}^{K-1}|=B}} \min\Big\{\mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}})\,,\ \mathbf{J}\big(\mathbf{Q}_{\mathrm{ST}}(\mathcal{B}^0,\dots,\mathcal{B}^{K-1})\big)\Big\} \\
&= \binom{n}{B}^{-K} \sum_{\substack{\mathcal{B}^0\subset\{1,\dots,n\}\\|\mathcal{B}^0|=B}} \cdots \sum_{\substack{\mathcal{B}^{K-1}\subset\{1,\dots,n\}\\|\mathcal{B}^{K-1}|=B}} \Bigg(\mathbf{1}\Big\{\big(\mathcal{B}^0,\dots,\mathcal{B}^{K-1}\big)=\big(\mathcal{B}^{0,\star},\dots,\mathcal{B}^{K-1,\star}\big)\Big\} \\
&\quad + \mathbf{1}\Big\{\big(\mathcal{B}^0,\dots,\mathcal{B}^{K-1}\big)\neq\big(\mathcal{B}^{0,\star},\dots,\mathcal{B}^{K-1,\star}\big)\Big\}\Bigg) \min\Big\{\mathbf{J}\big(\mathbf{Q}_{\mathrm{ST}}(\mathcal{B}^0,\dots,\mathcal{B}^{K-1})\big)\,,\ \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}})\Big\} \\
&= \binom{n}{B}^{-K}\Bigg[\min\Big\{\mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}})\,,\ \mathbf{J}\big(\mathbf{Q}_{\mathrm{ST}}(\mathcal{B}^{0,\star},\dots,\mathcal{B}^{K-1,\star})\big)\Big\} \\
&\quad + \sum_{\substack{\mathcal{B}^0,\dots,\mathcal{B}^{K-1}\\|\mathcal{B}^0|=\cdots=|\mathcal{B}^{K-1}|=B\\(\mathcal{B}^0,\dots,\mathcal{B}^{K-1})\neq(\mathcal{B}^{0,\star},\dots,\mathcal{B}^{K-1,\star})}} \min\Big\{\mathbf{J}\big(\mathbf{Q}_{\mathrm{ST}}(\mathcal{B}^0,\dots,\mathcal{B}^{K-1})\big)\,,\ \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}})\Big\}\Bigg] \\
&\le \binom{n}{B}^{-K}\Bigg[\mathbf{J}\big(\mathbf{Q}_{\mathrm{ST}}(\mathcal{B}^{0,\star},\dots,\mathcal{B}^{K-1,\star})\big) \sum_{\substack{\mathcal{B}^0,\dots,\mathcal{B}^{K-1}\\|\mathcal{B}^0|=\cdots=|\mathcal{B}^{K-1}|=B}} \mathbf{1}\Big\{\big(\mathcal{B}^0,\dots,\mathcal{B}^{K-1}\big)=\big(\mathcal{B}^{0,\star},\dots,\mathcal{B}^{K-1,\star}\big)\Big\} \\
&\quad + \sum_{\substack{\mathcal{B}^0,\dots,\mathcal{B}^{K-1}\\|\mathcal{B}^0|=\cdots=|\mathcal{B}^{K-1}|=B\\(\mathcal{B}^0,\dots,\mathcal{B}^{K-1})\neq(\mathcal{B}^{0,\star},\dots,\mathcal{B}^{K-1,\star})}} \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}})\Bigg] \\
&\overset{(5)}{\le} \binom{n}{B}^{-K}\Bigg[\mathbf{J}_K^\star\cdot 1 + \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) \sum_{\substack{\mathcal{B}^0,\dots,\mathcal{B}^{K-1}\\|\mathcal{B}^0|=\cdots=|\mathcal{B}^{K-1}|=B\\(\mathcal{B}^0,\dots,\mathcal{B}^{K-1})\neq(\mathcal{B}^{0,\star},\dots,\mathcal{B}^{K-1,\star})}} 1\Bigg] \\
&= \binom{n}{B}^{-K}\Bigg[\mathbf{J}_K^\star + \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) \sum_{\substack{\mathcal{B}^0,\dots,\mathcal{B}^{K-1}\\|\mathcal{B}^0|=\cdots=|\mathcal{B}^{K-1}|=B}} \mathbf{1}\Big\{\big(\mathcal{B}^0,\dots,\mathcal{B}^{K-1}\big)\neq\big(\mathcal{B}^{0,\star},\dots,\mathcal{B}^{K-1,\star}\big)\Big\}\Bigg] \\
&= \binom{n}{B}^{-K}\Bigg[\mathbf{J}_K^\star + \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}})\Bigg(\sum_{\substack{\mathcal{B}^0\subset\{1,\dots,n\}\\|\mathcal{B}^0|=B}} \cdots \sum_{\substack{\mathcal{B}^{K-1}\subset\{1,\dots,n\}\\|\mathcal{B}^{K-1}|=B}} 1 - \underbrace{1}_{\text{removing }\left(\mathcal{B}^{0,\star},\dots,\mathcal{B}^{K-1,\star}\right)}\Bigg)\Bigg] \\
&= \binom{n}{B}^{-K}\Big[\mathbf{J}_K^\star + \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}})\Big(\binom{n}{B}^K - 1\Big)\Big] \\
&= \binom{n}{B}^{-K}\mathbf{J}_K^\star + \left(\binom{n}{B}^{-K}\binom{n}{B}^K - \binom{n}{B}^{-K}\right)\mathbf{J}\left(\mathbf{Q}_{\mathrm{TMLE}}\right) \\
&= \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \Big(\mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbf{J}_K^\star\Big)\binom{n}{B}^{-K},
\end{aligned}
$$

where the key inequality is $\min\{x,y\}\le y$. Hence we obtain the desired claim. □

E Proof of Theorem 3.5

Proof. Throughout this proof, $\mathbb{P}(\cdot)$ and $\mathbb{E}[\cdot]$ are conditional on $\{O_i\}_{i=1}^n$, and the only randomness is the mini-batch $\mathcal{B}$. By Assumption 3.4, we can define the random index

$$
\widehat{j} \in \arg\min_{1\le j\le m} \frac{1}{B}\sum_{i\in\mathcal{B}} \mathcal{L}\big(\mathbf{Q}_{(j)}\big)(O_i), \tag{35}
$$

and we have assumed that the hybrid output satisfies the algorithm-to-root identification

$$
\mathbf{Q}_{\mathrm{ST}} = \mathbf{Q}_{(\widehat{j})}.
$$

We first reduce the target inequality to a bound on $\mathbb{P}\big(\widehat{j} \neq j^\star\big)$. Starting from the definition (35) and expanding $\mathbf{J}\big(\mathbf{Q}_{\mathrm{ST}}\big)$ by conditioning on the discrete outcome $\widehat{j} \in \{1, \ldots, m\}$, we obtain

$$
\begin{aligned}
&\mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbb{E}[\mathbf{J}(\mathbf{Q}_{\mathrm{ST}})] \\
&= \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbb{E}\Big[\mathbf{J}\big(\mathbf{Q}_{(\widehat{j})}\big)\Big] = \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbb{E}\left[\sum_{r=1}^{m} \mathbf{J}\big(\mathbf{Q}_{(r)}\big)\mathbf{1}\{\widehat{j} = r\}\right] \\
&= \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \sum_{r=1}^{m} \mathbf{J}\big(\mathbf{Q}_{(r)}\big)\mathbb{E}\Big[\mathbf{1}\{\widehat{j} = r\}\Big] = \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \sum_{r=1}^{m} \mathbf{J}\big(\mathbf{Q}_{(r)}\big)\mathbb{P}\big(\widehat{j} = r\big) \\
&= \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big)\mathbb{P}\big(\widehat{j} = j^\star\big) - \sum_{r \neq j^\star} \mathbf{J}\big(\mathbf{Q}_{(r)}\big)\mathbb{P}\big(\widehat{j} = r\big) \\
&= \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big)\Big[\mathbb{P}\big(\widehat{j} = j^\star\big) + \sum_{r \neq j^\star} \mathbb{P}\big(\widehat{j} = r\big)\Big] - \sum_{r \neq j^\star} \Big[\mathbf{J}\big(\mathbf{Q}_{(r)}\big) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big)\Big]\mathbb{P}\big(\widehat{j} = r\big) \\
&= \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big) \cdot 1 - \sum_{r \neq j^\star} \Big[\mathbf{J}\big(\mathbf{Q}_{(r)}\big) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big)\Big]\mathbb{P}\big(\widehat{j} = r\big) \\
&\geq \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big) - \sum_{r \neq j^\star} \Big|\mathbf{J}\big(\mathbf{Q}_{(r)}\big) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big)\Big|\,\mathbb{P}\big(\widehat{j} = r\big) \\
&= \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big) - \sum_{r \neq j^\star} \left|\sum_{i=1}^{n} \Big[\mathcal{L}\big(\mathbf{Q}_{(r)}\big)(O_i) - \mathcal{L}\big(\mathbf{Q}_{(j^\star)}\big)(O_i)\Big]\right| \mathbb{P}\big(\widehat{j} = r\big) \\
&\geq \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big) - \sum_{r \neq j^\star} \sum_{i=1}^{n} \Big|\mathcal{L}\big(\mathbf{Q}_{(r)}\big)(O_i) - \mathcal{L}\big(\mathbf{Q}_{(j^\star)}\big)(O_i)\Big|\,\mathbb{P}\big(\widehat{j} = r\big) \\
&\geq \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big) - \sum_{r \neq j^\star} \sum_{i=1}^{n} \Big[\big|\mathcal{L}\big(\mathbf{Q}_{(r)}\big)(O_i)\big| + \big|\mathcal{L}\big(\mathbf{Q}_{(j^\star)}\big)(O_i)\big|\Big]\mathbb{P}\big(\widehat{j} = r\big) \\
&\geq \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big) - \sum_{r \neq j^\star} \sum_{i=1}^{n} \Big[L_{\max} + L_{\max}\Big]\mathbb{P}\big(\widehat{j} = r\big) \\
&= \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big) - 2nL_{\max} \sum_{r \neq j^\star} \mathbb{P}\big(\widehat{j} = r\big) \\
&= \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big) - 2nL_{\max}\,\mathbb{P}\big(\widehat{j} \neq j^\star\big),
\end{aligned}
$$

where we used Assumption 2.3 and the fact that $\mathbb{P}\left(\widehat{j} = j^\star\right) + \sum_{j \neq j^\star} \mathbb{P}\left(\widehat{j} = j\right) = 1$. Consequently, it remains to bound $\mathbb{P}\left(\widehat{j} \neq j^\star\right)$. We begin from the definition of $\widehat{j}$. The event $\left\{\widehat{j} \neq j^\star\right\}$ implies that there exists some index $j \neq j^\star$ whose mini-batch empirical loss is no larger than that of $j^\star$:

$$
\left\{\widehat{j} \neq j^\star\right\} \subseteq \bigcup_{j \neq j^\star} \left\{\frac{1}{B}\sum_{i \in \mathcal{B}} \mathcal{L}\big(\mathbf{Q}_{(j)}\big)(O_i) \leq \frac{1}{B}\sum_{i \in \mathcal{B}} \mathcal{L}\big(\mathbf{Q}_{(j^\star)}\big)(O_i)\right\}. \tag{36}
$$

Hence, by the union bound,

$$
\begin{aligned}
\mathbb{P}\left(\widehat{j} \neq j^\star\right) &\overset{(36)}{\leq} \mathbb{P}\left(\exists j \neq j^\star : \frac{1}{B}\sum_{i \in \mathcal{B}} \mathcal{L}\big(\mathbf{Q}_{(j)}\big)(O_i) \leq \frac{1}{B}\sum_{i \in \mathcal{B}} \mathcal{L}\big(\mathbf{Q}_{(j^\star)}\big)(O_i)\right) \\
&= \mathbb{P}\left(\bigcup_{j \neq j^\star} \left\{\frac{1}{B}\sum_{i \in \mathcal{B}} \Big[\mathcal{L}\big(\mathbf{Q}_{(j)}\big)(O_i) - \mathcal{L}\big(\mathbf{Q}_{(j^\star)}\big)(O_i)\Big] \leq 0\right\}\right) \\
&\leq \sum_{j \neq j^\star} \mathbb{P}\left(\frac{1}{B}\sum_{i \in \mathcal{B}} \Big[\mathcal{L}\big(\mathbf{Q}_{(j)}\big)(O_i) - \mathcal{L}\big(\mathbf{Q}_{(j^\star)}\big)(O_i)\Big] \leq 0\right).
\end{aligned} \tag{37}
$$

Fix any $j \neq j^\star$. The event in (37) can be rewritten as a downward deviation relative to its full-data mean:

$$
\begin{aligned}
&\left\{\frac{1}{B}\sum_{i \in \mathcal{B}} \Big[\mathcal{L}\big(\mathbf{Q}_{(j)}\big)(O_i) - \mathcal{L}\big(\mathbf{Q}_{(j^\star)}\big)(O_i)\Big] \leq 0\right\} \\
&= \left\{\sum_{i \in \mathcal{B}} \Big[\mathcal{L}\big(\mathbf{Q}_{(j)}\big)(O_i) - \mathcal{L}\big(\mathbf{Q}_{(j^\star)}\big)(O_i)\Big] \leq 0\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= \left\{ \sum_{i\in\mathcal{B}} \left[\mathcal{L}(\mathbf{Q}_{(j)})(O_i) - \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_i)\right] - \frac{B}{n}\left[\mathbf{J}(\mathbf{Q}_{(j)}) - \mathbf{J}(\mathbf{Q}_{(j^\star)})\right] \le -\frac{B}{n}\left[\mathbf{J}(\mathbf{Q}_{(j)}) - \mathbf{J}(\mathbf{Q}_{(j^\star)})\right] \right\} \\
&= \left\{ \sum_{i\in\mathcal{B}} \left[\left[\mathcal{L}(\mathbf{Q}_{(j)})(O_i) - \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_i)\right] - \frac{1}{n}\left[\mathbf{J}(\mathbf{Q}_{(j)}) - \mathbf{J}(\mathbf{Q}_{(j^\star)})\right] \right] \right. \\
&\quad \left. \le -\frac{B}{n}\left[\mathbf{J}(\mathbf{Q}_{(j)}) - \mathbf{J}(\mathbf{Q}_{(j^\star)})\right] \right\} \\
&\subseteq \left\{ \sum_{i\in\mathcal{B}} \left[\left[\mathcal{L}(\mathbf{Q}_{(j)})(O_i) - \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_i)\right] - \frac{1}{n}\left[\mathbf{J}(\mathbf{Q}_{(j)}) - \mathbf{J}(\mathbf{Q}_{(j^\star)})\right] \right] \le -B\Omega \right\} \\
&= \left\{ -\sum_{i\in\mathcal{B}} \left[\left[\mathcal{L}(\mathbf{Q}_{(j)})(O_i) - \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_i)\right] - \frac{1}{n}\left[\mathbf{J}(\mathbf{Q}_{(j)}) - \mathbf{J}(\mathbf{Q}_{(j^\star)})\right] \right] \ge B\Omega \right\}.
\end{aligned}
$$

At this point, we apply an exponential Chernoff bound. Let $\lambda > 0$. Then

$$
\begin{aligned}
&\mathbb{P}\left(-\sum_{i\in\mathcal{B}} \left[\left[\mathcal{L}(\mathbf{Q}_{(j)})(O_i) - \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_i)\right] - \frac{1}{n}\left[\mathbf{J}(\mathbf{Q}_{(j)}) - \mathbf{J}(\mathbf{Q}_{(j^\star)})\right] \right] \ge B\Omega \right) \\
&= \mathbb{P}\left(\exp\left(\lambda \left(-\sum_{i\in\mathcal{B}} \left[\left[\mathcal{L}(\mathbf{Q}_{(j)})(O_i) - \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_i)\right] - \frac{1}{n}\left[\mathbf{J}(\mathbf{Q}_{(j)}) - \mathbf{J}(\mathbf{Q}_{(j^\star)})\right] \right] \right) \right) \\
&\quad \ge \exp\left(\lambda B\Omega\right)) \\
&\le \exp\left(-\lambda B\Omega\right) \\
&\quad \mathbb{E}\left[\exp\left(\lambda \Big(-\sum_{i\in\mathcal{B}} \left[\left[\mathcal{L}(\mathbf{Q}_{(j)})(O_i) - \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_i)\right] - \frac{1}{n}\left[\mathbf{J}(\mathbf{Q}_{(j)}) - \mathbf{J}(\mathbf{Q}_{(j^\star)})\right] \right] \Big) \right) \right] \\
&= \exp\left(-\lambda B\Omega\right) \mathbb{E}\left[\exp\left(\sum_{i\in\mathcal{B}} \lambda \left[\frac{1}{n}\left[\mathbf{J}(\mathbf{Q}_{(j)}) - \mathbf{J}(\mathbf{Q}_{(j^\star)})\right] - \left[\mathcal{L}(\mathbf{Q}_{(j)})(O_i) - \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_i)\right] \right] \right) \right].
\end{aligned}
$$

We now use the standard Hoeffding convex order comparison [48] for sampling without replacement, applied to the convex function $u \mapsto \exp u$, which yields the domination

$$
\begin{aligned}
&\mathbb{E}\left[\exp\left(\sum_{i\in\mathcal{B}} \lambda \left[\frac{1}{n}\left[\mathbf{J}(\mathbf{Q}_{(j)}) - \mathbf{J}(\mathbf{Q}_{(j^\star)})\right] - \left[\mathcal{L}(\mathbf{Q}_{(j)})(O_i) - \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_i)\right] \right] \right) \right] \\
&\quad \le \left[\frac{1}{n}\sum_{s=1}^{n} \exp\left(\lambda \left[\frac{1}{n}\left[\mathbf{J}(\mathbf{Q}_{(j)}) - \mathbf{J}(\mathbf{Q}_{(j^\star)})\right] - \left[\mathcal{L}(\mathbf{Q}_{(j)})(O_s) - \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_s)\right] \right] \right) \right]^B.
\end{aligned}
$$

Define

$$
Z_s^{(j)} \triangleq \frac{1}{n}\left[\mathbf{J}(\mathbf{Q}_{(j)}) - \mathbf{J}(\mathbf{Q}_{(j^\star)})\right] - \left[\mathcal{L}(\mathbf{Q}_{(j)})(O_s) - \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_s)\right].
$$

Then

$$
\frac{1}{n}\sum_{s=1}^{n} Z_s^{(j)} = \frac{1}{n}\sum_{r=1}^{n} \left[\mathcal{L}(\mathbf{Q}_{(j)})(O_r) - \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_r)\right] - \frac{1}{n}\sum_{s=1}^{n} \left[\mathcal{L}(\mathbf{Q}_{(j)})(O_s) - \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_s)\right] = 0. \tag{38}
$$

Also, $\left|\mathcal{L}(\mathbf{Q}_{(j)})(O_s) - \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_s)\right| \le 2L_{\max}$ implies

$$
\begin{aligned}
&-2L_{\max} \le \mathcal{L}(\mathbf{Q}_{(j)})(O_s) - \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_s) \le 2L_{\max} \\
&\Leftrightarrow \quad \frac{1}{n}\sum_{r=1}^{n} \left[\mathcal{L}(\mathbf{Q}_{(j)})(O_r) - \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_r)\right] - 2L_{\max} \le Z_s^{(j)} \\
&\le \frac{1}{n}\sum_{r=1}^{n} \left[\mathcal{L}(\mathbf{Q}_{(j)})(O_r) - \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_r)\right] + 2L_{\max} \Leftrightarrow \quad \max_s Z_s^{(j)} - \min_s Z_s^{(j)} \le 4L_{\max}.
\end{aligned}
$$

We now bound the discrete uniform by an explicit second derivative argument that yields the Hoeffding lemma constant. Consider

$$
h(\lambda) := \log\left[\frac{1}{n}\sum_{s=1}^{n} \exp\left(\lambda Z_s^{(j)}\right) \right].
$$

Computing the derivatives exactly, we obtain

$$h'(\lambda) = \frac{\frac{1}{n}\sum_{s=1}^{n} Z_s^{(j)} \exp\left(\lambda Z_s^{(j)}\right)}{\frac{1}{n}\sum_{s=1}^{n} \exp\left(\lambda Z_s^{(j)}\right)} = \frac{\sum_{s=1}^{n} Z_s^{(j)} \exp\left(\lambda Z_s^{(j)}\right)}{\sum_{s=1}^{n} \exp\left(\lambda Z_s^{(j)}\right)} =: \mathbb{E}_\lambda[Z],$$

$$h''(\lambda) = \frac{\left[\sum_{s=1}^{n} \left[Z_s^{(j)}\right]^2 \exp\left(\lambda Z_s^{(j)}\right)\right]\left[\sum_{s=1}^{n} \exp\left(\lambda Z_s^{(j)}\right)\right] - \left[\sum_{s=1}^{n} Z_s^{(j)} \exp\left(\lambda Z_s^{(j)}\right)\right]^2}{\left[\sum_{s=1}^{n} \exp\left(\lambda Z_s^{(j)}\right)\right]^2}.$$

The right-hand side is the variance of $Z_s^{(j)}$ under the exponentially tilted discrete distribution with weights proportional to $e^{-\lambda Z_s}$, and hence $h''(\lambda) = \mathrm{Var}_\lambda(Z)$ for some distribution supported on $[\alpha, \beta]$. For any distribution supported on an interval $[\alpha, \beta]$, the variance is bounded by [49] as

$$\begin{aligned}\mathrm{Var}_\lambda(Z) = \mathbb{E}_\lambda\left[(Z - \mathbb{E}_\lambda[Z])^2\right] &\le \mathbb{E}_\lambda\left[\left(Z - \frac{\alpha+\beta}{2}\right)^2\right] \\ &\le \sup_{z\in[\alpha,\beta]}\left(z - \frac{\alpha+\beta}{2}\right)^2 = \frac{(\beta-\alpha)^2}{4}.\end{aligned}$$

Therefore,

$$h''(\lambda) \le \frac{\left[\max_s Z_s^{(j)} - \min_s Z_s^{(j)}\right]^2}{4} \le \frac{(4L_{\max})^2}{4} = 4L_{\max}^2.$$

Using Taylor's theorem with integral remainder and substituting $h(0) = \log 1 = 0$, we obtain

$$\begin{aligned}h(\lambda) &= h(0) + \lambda h'(0) + \int_0^\lambda \int_0^u h''(t)\, dt\, du \overset{(38)}{=} 0 + 0 + \int_0^\lambda (\lambda - t) h''(t)\, dt \\ &\le \int_0^\lambda (\lambda - t)\, 4L_{\max}^2\, dt \\ &= 4L_{\max}^2 \int_0^\lambda (\lambda - t)\, dt = 4L_{\max}^2 \left[\lambda t - \frac{t^2}{2}\right]_{t=0}^{t=\lambda} = 2\lambda^2 L_{\max}^2.\end{aligned}$$

Exponentiating and raising to power B yields

$$\left[\frac{1}{n}\sum_{s=1}^{n} \exp\left(\lambda Z_s^{(j)}\right)\right]^B = \exp\left(Bh(\lambda)\right) \le \exp\left(2\lambda^2 L_{\max}^2 B\right).$$

Substituting back into the Chernoff chain and completing the square, we obtain

$$\begin{aligned}&\mathbb{P}\left(\frac{1}{B}\sum_{i\in\mathcal{B}}\left[\mathcal{L}(\mathbf{Q}_{(j)})(O_i) - \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_i)\right] \le 0\right) \\ &\le \mathbb{P}\left(-\sum_{i\in\mathcal{B}}\left[\mathcal{L}(\mathbf{Q}_{(j)})(O_i) - \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_i) - \frac{1}{n}\left[\mathbf{J}(\mathbf{Q}_{(j)}) - \mathbf{J}(\mathbf{Q}_{(j^\star)})\right]\right] \ge B\Omega\right) \\ &\le \exp\left(-\lambda B\Omega\right)\mathbb{E}\left[\exp\left[\sum_{i\in\mathcal{B}}\lambda\left[\frac{1}{n}\left[\mathbf{J}(\mathbf{Q}_{(j)}) - \mathbf{J}(\mathbf{Q}_{(j^\star)})\right] - \mathcal{L}(\mathbf{Q}_{(j)})(O_i) + \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_i)\right]\right]\right] \\ &\le \exp\left(-\lambda B\Omega\right)\left[\frac{1}{n}\sum_{s=1}^{n}\exp\left[\lambda\left[\frac{1}{n}\left[\mathbf{J}(\mathbf{Q}_{(j)}) - \mathbf{J}(\mathbf{Q}_{(j^\star)})\right] \;-\mathcal{L}(\mathbf{Q}_{(j)})(O_s) + \mathcal{L}(\mathbf{Q}_{(j^\star)})(O_s)\right]\right]\right]^B \\ &= \exp\left(-\lambda B\Omega\right)\left[\frac{1}{n}\sum_{s=1}^{n}\exp\left(\lambda Z_s^{(j)}\right)\right]^B \le \exp\left(-\lambda B\Omega\right)\cdot\exp\left(2\lambda^2 L_{\max}^2 B\right) \\ &= \exp\left(B\left(2L_{\max}^2\left(\lambda^2 - \frac{\Omega}{2L_{\max}^2}\lambda\right)\right)\right) = \exp\left(B\left(2L_{\max}^2\left(\lambda - \frac{\Omega}{4L_{\max}^2}\right)^2 - \frac{\Omega^2}{8L_{\max}^2}\right)\right),\end{aligned}$$

where the last inequality holds for every $\lambda > 0$. Plugging this bound into the earlier union bound gives

$$\begin{aligned}\mathbb{P}\left(\widehat{j} \neq j^\star\right) &\leq \min\left\{1, \sum_{j\neq j^\star} \mathbb{P}\left(\frac{1}{B}\sum_{i\in\mathcal{B}}\left[\mathcal{L}\big(\mathbf{Q}_{(j)}\big)(O_i) - \mathcal{L}\big(\mathbf{Q}_{(j^\star)}\big)(O_i)\right] \leq 0\right)\right\} \\ &\leq \min\left\{1, \inf_{\lambda>0}\sum_{j\neq j^\star}\exp\left(B\left(2L_{\max}^2\left(\lambda - \frac{\Omega}{4L_{\max}^2}\right)^2 - \frac{\Omega^2}{8L_{\max}^2}\right)\right)\right\} \\ &= \min\left\{1, \ (m-1)\exp\left(-\frac{B\Omega^2}{8L_{\max}^2}\right)\right\}.\end{aligned}$$

Substituting this probability bound into the earlier lower bound on $\mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbb{E}[\mathbf{J}(\mathbf{Q}_{\mathrm{ST}})]$, we obtain

$$\begin{aligned}&\mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbb{E}[\mathbf{J}(\mathbf{Q}_{\mathrm{ST}})] \\ &\geq \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big) - 2nL_{\max}\,\mathbb{P}\big(\widehat{j} \neq j^\star\big) \\ &\geq \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big) - 2nL_{\max}\min\left\{1, (m-1)\exp\left(-\frac{B\Omega^2}{8L_{\max}^2}\right)\right\},\end{aligned}$$

which is the claimed inequality. □

Remark E.1. In particular, whenever

$$2nL_{\max}\min\left\{1, \ (m-1)\exp\left(-\frac{B\Omega^2}{8L_{\max}^2}\right)\right\} < \mathbf{J}(\mathbf{Q}_{\mathrm{TMLE}}) - \mathbf{J}\big(\mathbf{Q}_{(j^\star)}\big),$$

we obtain the result of Remark 3.6.

F Proof of Lemma 4.2

Proof. Fix k and λ. The function $f(\mathbf{v}, \lambda)$ is real-valued on $\{\mathbf{v} \in \mathbb{R}^n : \mathbf{0} \preceq \mathbf{v} \preceq \mathbf{1}\}$. Using $\mathbf{0} \in \mathcal{E}$ and $\mathbf{Q}_k = \mathbf{Q}_k(\mathbf{0} \mid \mathbf{G})$, we have for all $i \in \{1, \ldots, n\}$

$$\left|\mathcal{L}(\mathbf{Q}_k)(O_i)\right| = \left|\mathcal{L}\big(\mathbf{Q}_k(\mathbf{0} \mid \mathbf{G})\big)(O_i)\right| \leq L_{\max}. \tag{39}$$

Define the feasible set

$$\mathcal{V} \triangleq \left\{\mathbf{v} \in \mathbb{R}^n : \mathbf{0} \preceq \mathbf{v} \preceq \mathbf{1}\right\} = \left\{\mathbf{v} = (v_1, \ldots, v_n)^\top \in \mathbb{R}^n : \forall i, \ 0 \leq v_i \leq 1\right\},$$

and define the objective

$$\vec{\mathcal{J}}(\mathbf{v}) = \sum_{i=1}^n v_i\,\mathcal{L}(\mathbf{Q}_k)(O_i) + f(\mathbf{v}, \lambda), \qquad \mathbf{v} \in \mathcal{V}. \tag{40}$$

For each coordinate i, define the continuous coordinate projection $\pi_i(\mathbf{v}) = v_i$. Then

$$\mathcal{V} = \bigcap_{i=1}^n \pi_i^{-1}\big([0, \infty)\big) \cap \bigcap_{i=1}^n \pi_i^{-1}\big((-\infty, 1]\big).$$

Since $[0, \infty)$ and $(-\infty, 1]$ are closed subsets of $\mathbb{R}$ and π_i is continuous, each preimage $\pi_i^{-1}([0, \infty))$ and $\pi_i^{-1}((-\infty, 1])$ is closed in $\mathbb{R}^n$ [50]. A finite intersection of closed sets is closed, and hence $\mathcal{V}$ is closed.

Take any sequence $\{\mathbf{v}^m\}_{m\geq 1} \subset \mathcal{V}$ and write $\mathbf{v}^m = (v_1^m, \ldots, v_n^m)^\top$. For each fixed coordinate i, the scalar sequence $\{v_i^m\}_{m\geq 1}$ lies in $[0, 1]$ and hence is bounded and admits a convergent subsequence. We can apply this coordinatewise with a diagonal selection [51]. Concretely, pick indices $m_1(j)$ such that $v_1^{m_1(j)} \rightsquigarrow v_1^\star$ for some $v_1^\star \in [0, 1]$. From the subsequence $\{\mathbf{v}^{m_1(j)}\}$, pick indices $m_2(j)$ such

that $v_2^{m_2(j)} \rightsquigarrow v_2^\star \in [0,1]$ while still $v_1^{m_2(j)} \rightsquigarrow v_1^\star$. Continuing this construction through coordinate n yields a subsequence $\{\mathbf{v}^{m_n(j)}\}_{j\geq 1}$ such that for each $i \in \{1,\dots,n\}$,

$$v_i^{m_n(j)} \rightsquigarrow v_i^\star \quad \text{with } v_i^\star \in [0,1].$$

Define $\mathbf{v}^\star = (v_1^\star,\dots,v_n^\star)^\top$. Then $\mathbf{v}^{m_n(j)} \rightsquigarrow \mathbf{v}^\star$ in $\mathbb{R}^n$ because

$$\left\|\mathbf{v}^{m_n(j)} - \mathbf{v}^\star\right\|^2 = \sum_{i=1}^n \left(v_i^{m_n(j)} - v_i^\star\right)^2 \rightsquigarrow \sum_{i=1}^n 0 = 0,$$

and $\mathbf{v}^\star \in \mathcal{V}$ because $0 \leq v_i^\star \leq 1$ for all i. Thus every sequence in $\mathcal{V}$ has a convergent subsequence whose limit remains in $\mathcal{V}$, so $\mathcal{V}$ is sequentially compact and hence compact in $\mathbb{R}^n$. Let $\mathbf{v}, \mathbf{u} \in \mathcal{V}$ with $\mathbf{u} = (u_1,\dots,u_n)^\top$. Using the triangle inequality, we obtain

$$\begin{aligned}
&\left|\vec{\mathcal{J}}(\mathbf{v}) - \vec{\mathcal{J}}(\mathbf{u})\right| = \left|\sum_{i=1}^n (v_i - u_i)\,\mathcal{L}(\mathbf{Q}_k)(O_i) + f(\mathbf{v},\lambda) - f(\mathbf{u},\lambda)\right| \\
&\leq \sum_{i=1}^n |v_i - u_i|\,\left|\mathcal{L}(\mathbf{Q}_k)(O_i)\right| + \left|f(\mathbf{v},\lambda) - f(\mathbf{u},\lambda)\right| \\
&\overset{(39)}{\leq} L_{\max} \sum_{i=1}^n |v_i - u_i| + \left|f(\mathbf{v},\lambda) - f(\mathbf{u},\lambda)\right| = L_{\max}\left\|\mathbf{v} - \mathbf{u}\right\|_1 + \left|f(\mathbf{v},\lambda) - f(\mathbf{u},\lambda)\right| \\
&\leq L_{\max} \cdot \left(\sum_{i=1}^n 1^2\right)^{1/2} \cdot \left(\sum_{i=1}^n (v_i - u_i)^2\right)^{1/2} + \left|f(\mathbf{v},\lambda) - f(\mathbf{u},\lambda)\right| \\
&= L_{\max}\sqrt{n}\left\|\mathbf{v} - \mathbf{u}\right\| + \left|f(\mathbf{v},\lambda) - f(\mathbf{u},\lambda)\right|. \qquad (41)
\end{aligned}$$

Since $f(\cdot,\lambda)$ is continuous, $\mathbf{v}^m \rightsquigarrow \mathbf{v}$ implies $f(\mathbf{v}^m,\lambda) \rightsquigarrow f(\mathbf{v},\lambda)$, and the inequality (41) implies $\vec{\mathcal{J}}(\mathbf{v}^m) \rightsquigarrow \vec{\mathcal{J}}(\mathbf{v})$. Hence the objective $\vec{\mathcal{J}}$ is continuous on $\mathcal{V}$. We next show that $\vec{\mathcal{J}}$ is bounded below on $\mathcal{V}$ in a way that yields a finite infimum and supports a minimizing-sequence argument. For any $\mathbf{v} \in \mathcal{V}$,

$$\begin{aligned}
&\sum_{i=1}^n v_i\,\mathcal{L}(\mathbf{Q}_k)(O_i) \geq -\left|\sum_{i=1}^n v_i \cdot \mathcal{L}(\mathbf{Q}_k)(O_i)\right| \geq -\sum_{i=1}^n v_i \cdot \left|\mathcal{L}(\mathbf{Q}_k)(O_i)\right| \\
&\overset{(39)}{\geq} -\sum_{i=1}^n v_i\,L_{\max} \geq -L_{\max}\sum_{i=1}^n 1 = -nL_{\max}. \qquad (42)
\end{aligned}$$

We now consider the continuity of $f(\cdot,\lambda)$ on $\mathcal{V}$. Suppose for contradiction that $f(\cdot,\lambda)$ is not bounded below on $\mathcal{V}$. Then for each integer $m \geq 1$ there exists $\mathbf{v}^m \in \mathcal{V}$ such that $f(\mathbf{v}^m,\lambda) \leq -m$. By compactness of $\mathcal{V}$, we can extract a convergent subsequence $\mathbf{v}^{m_j} \rightsquigarrow \mathbf{v}^\star \in \mathcal{V}$. By continuity of $f(\cdot,\lambda)$,

$$f(\mathbf{v}^{m_j},\lambda) \rightsquigarrow f(\mathbf{v}^\star,\lambda) \in \mathbb{R},$$

while simultaneously

$$f(\mathbf{v}^{m_j},\lambda) \leq -m_j \rightsquigarrow -\infty,$$

so

$$f(\mathbf{v}^\star,\lambda) = \lim_{j\to\infty} f(\mathbf{v}^{m_j},\lambda) \leq \lim_{j\to\infty}(-m_j) = -\infty,$$

contradicting that f is real-valued on $\mathcal{V}$. Therefore, $\inf_{\mathbf{u}\in\mathcal{V}} f(\mathbf{u},\lambda)$ is a finite real number. Combining this with the bound for the linear term (42) gives, for all $\mathbf{v} \in \mathcal{V}$,

$$\begin{aligned}
\vec{\mathcal{J}}(\mathbf{v}) &= \sum_{i=1}^n v_i\,\mathcal{L}(\mathbf{Q}_k)(O_i) + f(\mathbf{v},\lambda) \overset{(42)}{\geq} -nL_{\max} + f(\mathbf{v},\lambda) \\
&\geq -nL_{\max} + \inf_{\mathbf{u}\in\mathcal{V}} f(\mathbf{u},\lambda) > -\infty,
\end{aligned}$$

and hence the infimum $J_\star = \inf_{\mathbf{v}\in\mathcal{V}} \vec{\mathcal{J}}(\mathbf{v})$ is a finite real number. By the definition of infimum, for each integer $m \geq 1$ there exists $\mathbf{v}^m \in \mathcal{V}$ such that

$$\vec{\mathcal{J}}(\mathbf{v}^m) \leq J_\star + \frac{1}{m}.$$

The sequence $\{\mathbf{v}^m\} \subset \mathcal{V}$ admits a convergent subsequence $\mathbf{v}^{m_j} \rightsquigarrow \mathbf{v}^\star \in \mathcal{V}$ by compactness. Using continuity of $\vec{\mathcal{J}}$ and the continuity estimate already established, we obtain

$$\left|\vec{\mathcal{J}}(\mathbf{v}^{m_j}) - \vec{\mathcal{J}}(\mathbf{v}^\star)\right| \leq L_{\max}\sqrt{n}\left\|\mathbf{v}^{m_j} - \mathbf{v}^\star\right\| + \left|f(\mathbf{v}^{m_j}, \lambda) - f(\mathbf{v}^\star, \lambda)\right| \rightsquigarrow 0,$$

so $\vec{\mathcal{J}}(\mathbf{v}^{m_j}) \rightsquigarrow \vec{\mathcal{J}}(\mathbf{v}^\star)$. On the other hand,

$$J_\star \leq \vec{\mathcal{J}}(\mathbf{v}^{m_j}) \leq J_\star + \frac{1}{m_j} \rightsquigarrow J_\star,$$

and hence $\vec{\mathcal{J}}(\mathbf{v}^{m_j}) \rightsquigarrow J_\star$. Identifying the limits yields

$$\vec{\mathcal{J}}(\mathbf{v}^\star) = J_\star, \quad \text{so that } \mathbf{v}^\star \in \arg\min_{\mathbf{v}\in\mathcal{V}} \vec{\mathcal{J}}(\mathbf{v}),$$

or equivalently,

$$\mathbf{v}^\star \in \arg\min_{\mathbf{0}\preceq\mathbf{v}\preceq\mathbf{1}} \left\{\sum_{i=1}^n v_i\,\mathcal{L}(\mathbf{Q}_k)(O_i) + f(\mathbf{v}, \lambda)\right\}.$$

This proves that the argmin set is nonempty and thus the optimization defining $\mathbf{v}^k$ has at least one solution. □

G Proof of Theorem 4.3

Proof. Similar to (40), for each regularizer we define the weight objective at iteration k as

$$\vec{\mathcal{J}}_k(\mathbf{v}) := \sum_{i=1}^n \Big(v_i\,\mathcal{L}(\mathbf{Q}_k)(O_i) + f(v_i, \lambda_k)\Big), \qquad \mathbf{v} \in \mathcal{V}. \tag{43}$$

We prove that there exists K_0 such that for all $k \geq K_0$ the strict gap property holds. Formally, for all $\mathbf{v} \in \mathcal{V}\backslash\{\mathbf{1}\}$,

$$\vec{\mathcal{J}}_k(\mathbf{v}) - \vec{\mathcal{J}}_k(\mathbf{1}) > 0, \tag{44}$$

which forces $\mathbf{v}^k = \mathbf{1}$.

Hard SP-regularizer For the hard regularizer, $f(v, \lambda_k) = -\lambda_k v$. We compute the exact gap for any $\mathbf{v} \in [0,1]^n$:

$$\begin{aligned}
&\vec{\mathcal{J}}_k(\mathbf{v}) - \vec{\mathcal{J}}_k(\mathbf{1}) = \sum_{i=1}^n \Big(v_i\,\mathcal{L}(\mathbf{Q}_k)(O_i) - \lambda_k\,v_i\Big) - \sum_{i=1}^n \Big(1\cdot\mathcal{L}(\mathbf{Q}_k)(O_i) - \lambda_k\cdot 1\Big) \\
&= \sum_{i=1}^n (1-v_i)\Big(\lambda_k - \mathcal{L}(\mathbf{Q}_k)(O_i)\Big) \geq \sum_{i=1}^n (1-v_i)\Big(\lambda_k - \sup_{k'\geq 0}\max_{1\leq j\leq n}\mathcal{L}(\mathbf{Q}_{k'})(O_j)\Big) \\
&\geq (\lambda_k - L_{\max})\sum_{i=1}^n (1-v_i)^2 = (\lambda_k - L_{\max})\,\|\mathbf{1}-\mathbf{v}\|_2^2 \\
&\geq (\lambda_k - L_{\max})\cdot\frac{1}{n}\Big(\sum_{i=1}^n (1-v_i)\Big)^2 \geq (\lambda_k - L_{\max})\cdot\frac{(n-\|\mathbf{v}\|_1)^3}{n^2} \geq 0.
\end{aligned} \tag{45}$$

Since $\{\lambda_k\}$ is monotone, K_0 admits the equivalent exact identities

$$K_0 = 1 + \max\{k \geq 0 : \lambda_k \leq L_{\max}\} = \sum_{k=0}^\infty \mathbf{1}\{\lambda_k \leq L_{\max}\}.$$

For $k \geq K_0$ in the hard case, we have $\lambda_k > L_{\max}$ and hence $\lambda_k - L_{\max} > 0$. Thus every inequality in (45) is strict whenever $\mathbf{v} \neq \mathbf{1}$, or equivalently $n - \|\mathbf{v}\|_1 > 0$, which gives (44). Therefore, $\mathbf{v}^k = \mathbf{1}$ for all $k \geq K_0$ in the hard case.

LOG SP-regularizer Fix $k \geq K_0$ and define $\vec{\mathcal{J}}_k$ by (43) with f^L. For any $\mathbf{v} \in (0,1]^n$, noting that the log term diverges as $v_i \downarrow 0$ so minimizers lie in $(0,1]^n$, we compute the exact gap:

$$\begin{aligned}
\vec{\mathcal{J}}_k(\mathbf{v}) - \vec{\mathcal{J}}_k(\mathbf{1}) &= \sum_{i=1}^{n} \left[v_i \mathcal{L}(\mathbf{Q}_k)(O_i) + \frac{1}{\alpha}\Big((1+\alpha\lambda_k)\log\frac{1+\alpha\lambda_k}{v_i} - (1+\alpha\lambda_k) + v_i\Big)\right] \\
&\quad - \sum_{i=1}^{n} \left[1 \cdot \mathcal{L}(\mathbf{Q}_k)(O_i) + \frac{1}{\alpha}\Big((1+\alpha\lambda_k)\log(1+\alpha\lambda_k) - (1+\alpha\lambda_k) + 1\Big)\right] \\
&= \sum_{i=1}^{n}(v_i - 1)\mathcal{L}(\mathbf{Q}_k)(O_i) + \frac{1}{\alpha}\sum_{i=1}^{n}\Big(-(1+\alpha\lambda_k)\log v_i + (v_i - 1)\Big) \\
&= \sum_{i=1}^{n}\Big((1-v_i)\lambda_k - (1-v_i)\mathcal{L}(\mathbf{Q}_k)(O_i) + \frac{1+\alpha\lambda_k}{\alpha}\big(-\log v_i - (1-v_i)\big)\Big) \\
&= \sum_{i=1}^{n}(1-v_i)\big(\lambda_k - \mathcal{L}(\mathbf{Q}_k)(O_i)\big) + \frac{1+\alpha\lambda_k}{\alpha}\sum_{i=1}^{n}\Big(-\log v_i - (1-v_i)\Big).
\end{aligned}$$

We now rewrite the LOG remainder term through integral and geometric-series representations. For $v_i \in (0,1]$,

$$\begin{aligned}
-\log v_i - (1-v_i) = \int_{v_i}^{1}\Big(\frac{1}{t} - 1\Big)\,dt &= \int_{0}^{1-v_i}\frac{x}{1-x}\,dx \\
&= \sum_{m=0}^{\infty}\int_{0}^{1-v_i} x^{m+1}\,dx = \sum_{m=2}^{\infty}\frac{(1-v_i)^m}{m}. \qquad (46)
\end{aligned}$$

We then apply the Jensen-on-powers lower bound to collapse the double sum into a one-dimensional series in $n - \|\mathbf{v}\|_1$:

$$\begin{aligned}
&\vec{\mathcal{J}}_k(\mathbf{v}) - \vec{\mathcal{J}}_k(\mathbf{1}) \\
&\overset{(46)}{=} \sum_{i=1}^{n}(1-v_i)\big(\lambda_k - \mathcal{L}(\mathbf{Q}_k)(O_i)\big) + \frac{1+\alpha\lambda_k}{\alpha}\sum_{i=1}^{n}\sum_{m=2}^{\infty}\frac{(1-v_i)^m}{m} \\
&\geq \sum_{i=1}^{n}(1-v_i)\big(\lambda_k - L_{\max}\big) + \frac{1+\alpha\lambda_k}{\alpha}\sum_{m=2}^{\infty}\frac{1}{m}\sum_{i=1}^{n}(1-v_i)^m \\
&= (\lambda_k - L_{\max})\sum_{i=1}^{n}(1-v_i) + \frac{1+\alpha\lambda_k}{\alpha}\sum_{m=2}^{\infty}\frac{1}{m}\sum_{i=1}^{n}(1-v_i)^m \\
&\geq (\lambda_k - L_{\max})\big(n - \|\mathbf{v}\|_1\big) + \frac{1+\alpha\lambda_k}{\alpha}\sum_{m=2}^{\infty}\frac{1}{m}\cdot n\left(\frac{1}{n}\sum_{i=1}^{n}(1-v_i)\right)^m \\
&= (\lambda_k - L_{\max})\big(n - \|\mathbf{v}\|_1\big) + \frac{1+\alpha\lambda_k}{\alpha}\cdot n\left[\sum_{m=1}^{\infty}\frac{1}{m}\left(\frac{n-\|\mathbf{v}\|_1}{n}\right)^m - \left(\frac{n-\|\mathbf{v}\|_1}{n}\right)\right]. \qquad (47)
\end{aligned}$$

We now turn (47) into a polynomial coercivity chain by expanding the log term and truncating:

$$-\log(1-x) - x = \sum_{m=2}^{\infty}\frac{x^m}{m} \;\geq\; \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \ldots, \qquad \forall x \in [0,1),$$

with $\dfrac{n - \|\mathbf{v}\|_1}{n} \in [0,1]$. Substituting this yields the lower bound

$$\begin{aligned}
&\vec{\mathcal{J}}_k(\mathbf{v}) - \vec{\mathcal{J}}_k(\mathbf{1}) \\
&\geq (\lambda_k - L_{\max})\big(n - \|\mathbf{v}\|_1\big) \\
&\quad + \frac{1+\alpha\lambda_k}{\alpha}\cdot n\left[\frac{1}{2}\left(\frac{n-\|\mathbf{v}\|_1}{n}\right)^2 + \frac{1}{3}\left(\frac{n-\|\mathbf{v}\|_1}{n}\right)^3 + \frac{1}{4}\left(\frac{n-\|\mathbf{v}\|_1}{n}\right)^4 + \ldots\right] \\
&= (\lambda_k - L_{\max})\big(n - \|\mathbf{v}\|_1\big) + \frac{1+\alpha\lambda_k}{\alpha}\left[\frac{(n-\|\mathbf{v}\|_1)^2}{2n} + \frac{(n-\|\mathbf{v}\|_1)^3}{3n^2} + \frac{(n-\|\mathbf{v}\|_1)^4}{4n^3} + \ldots\right] \\
&\geq \frac{1+\alpha\lambda_k}{\alpha}\cdot\frac{(n-\|\mathbf{v}\|_1)^2}{2n}, \qquad (48)
\end{aligned}$$

since all coefficients are nonnegative. Since $\{\lambda_k\}$ is monotone, K_0 admits the equivalent exact identities

$$K_0 = 1 + \max\left\{k \geq 0 : \lambda_k < L_{\max}\right\} = \sum_{k=0}^{\infty} \mathbf{1}\left\{\lambda_k < L_{\max}\right\}.$$

Finally, (48) implies the strict gap property

$$\mathbf{v} \neq \mathbf{1} \quad \Leftrightarrow \quad \vec{\mathcal{J}}_k(\mathbf{v}) - \vec{\mathcal{J}}_k(\mathbf{1}) \geq \frac{(1+\alpha\lambda_k)\,(n - \|\mathbf{v}\|_1)^2}{2\alpha n} > 0,$$

which is exactly (44). Therefore, $\mathbf{v}^k = \mathbf{1}$ for all $k \geq K_0$ in the LOG regularizer case.

Mixture SP-regularizer Recall that the activation threshold of the Mixture SP-regularizer is given by

$$\tau(\lambda) := \left(\frac{\lambda\gamma}{\lambda+\gamma}\right)^2.$$

We can find a deterministic constant Λ_M such that for all $\lambda > \Lambda_M$ we have $\tau(\lambda) > L_{\max}$. To this end, we enforce the scalar inequality

$$\begin{aligned} L_{\max} \leq \left(\frac{\lambda_k\gamma}{\lambda_k+\gamma}\right)^2 \quad &\Leftrightarrow \quad \gamma^2\lambda_k^2 \geq L_{\max}\left(\lambda_k^2 + 2\lambda_k\gamma + \gamma^2\right) \\ &\Leftrightarrow \quad \left(\gamma^2 - L_{\max}\right)\lambda_k^2 - 2L_{\max}\gamma\lambda_k - L_{\max}\gamma^2 \geq 0. \end{aligned} \tag{49}$$

We compute the discriminant of (49) as

$$\begin{aligned} \Delta &:= \left(2L_{\max}\gamma\right)^2 - 4\cdot\left(\gamma^2 - L_{\max}\right)\cdot\left(-L_{\max}\gamma^2\right) \\ &= 4L_{\max}^2\gamma^2 + 4L_{\max}\gamma^2\left(\gamma^2 - L_{\max}\right) = 4L_{\max}\gamma^4, \end{aligned}$$

which gives $\sqrt{\Delta} = 2\gamma^2\sqrt{L_{\max}}$. The two real roots are

$$\lambda_{k,\pm} = \frac{-\left(-2L_{\max}\right)\gamma \pm 2\gamma^2\sqrt{L_{\max}}}{2\cdot\left(\gamma^2 - L_{\max}\right)} = \frac{L_{\max}\gamma \pm \gamma^2\sqrt{L_{\max}}}{\gamma^2 - L_{\max}}.$$

Since $\gamma^2 - L_{\max} > 0$ by construction in Equation (9), the inequality (49) holds if and only if $\lambda > \lambda_{k,+}$ or $\lambda < \lambda_{k,-}$. Since $\lambda \geq 0$ and $\lambda_{k,-} < 0$, the relevant condition is $\lambda > \lambda_{k,+}$. Thus we may take the positive root as

$$\Lambda_M := \lambda_{k,+} = \frac{L_{\max}\gamma + \gamma^2\sqrt{L_{\max}}}{\gamma^2 - L_{\max}} = \frac{\gamma\sqrt{L_{\max}}\left(\sqrt{L_{\max}} + \gamma\right)}{\left(\gamma - \sqrt{L_{\max}}\right)\left(\gamma + \sqrt{L_{\max}}\right)} = \frac{\gamma\sqrt{L_{\max}}}{\gamma - \sqrt{L_{\max}}}.$$

Since $\{\lambda_k\}$ is monotone, K_0 admits the equivalent exact identities

$$K_0 = 1 + \max\left\{k \geq 0 : \lambda_k < \frac{\gamma\sqrt{L_{\max}}}{\gamma - \sqrt{L_{\max}}}\right\} = \sum_{k=0}^{\infty} \mathbf{1}\left\{\lambda_k < \frac{\gamma\sqrt{L_{\max}}}{\gamma - \sqrt{L_{\max}}}\right\}.$$

Thus, for $k \geq K_0$ in the mixture case, we have $v_i^k = 1$ for all i, that is, $\mathbf{v}^k = \mathbf{1}$. □

H Proof of Theorem 4.4

Assumption H.1. At each iteration k, the mini-batch indices $\mathcal{B}^k$ are sampled uniformly at random with replacement, and the coordinate subset $\mathcal{S}^k$ is sampled uniformly at random with replacement. The draws are independent across k, and $\mathcal{B}^k$ is independent of $\mathcal{S}^k$.

Under Assumption H.1, we prove the bound (7).

Proof. Fix $T \geq 1$. Under uniform mini-batch sampling, we represent $\mathcal{B}^k$ as indices $I_{k,1}, \ldots, I_{k,B}$. For each k, denote

$$\mathcal{E}_{\mathcal{S}^k} = \{\epsilon \in \mathcal{E} : \epsilon_j = 0 \text{ for all } j \notin \mathcal{S}^k\},$$

$$\epsilon^k \in \arg\min_{\epsilon\in\mathcal{E}_{\mathcal{S}^k}} \frac{1}{B}\sum_{b=1}^{B} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{I_{k,b}}),\ \ \mathbf{Q}_{k+1} \leftarrow \mathbf{Q}_k(\epsilon^k \mid \mathbf{G}).$$

Starting from the identity for $\Delta(q^k)$, averaging and taking expectation gives

$$\begin{aligned}
&\frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}[\Delta(q^k)] \\
&= \frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\mathbf{0} \mid \mathbf{G})\big)(O_i) - \frac{1}{n}\sum_{j'=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\big)(O_{j'})\right. \\
&\quad \left. + \frac{1}{n}\sum_{j=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\big)(O_j) - \inf_{\epsilon\in\mathcal{E}}\frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{i'})\right] \\
&= \frac{1}{T}\,\mathbb{E}\underbrace{\sum_{k=0}^{T-1}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}(\mathbf{Q}_k)(O_i) - \frac{1}{n}\sum_{j=1}^{n}\mathcal{L}(\mathbf{Q}_{k+1})(O_j)\right]}_{\triangleq\ \xi(\hat{n})} \\
&\quad + \frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}\left[\frac{1}{n}\sum_{j=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\big)(O_j) - \inf_{\epsilon\in\mathcal{E}}\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i)\right].
\end{aligned}$$

By Assumption 2.3,

$$\begin{aligned}
&\xi(\hat{n}) \\
&= \left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}(\mathbf{Q}_0)(O_i) - \frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}(\mathbf{Q}_1)(O_{i'})\right] + \left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}(\mathbf{Q}_1)(O_i) - \frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}(\mathbf{Q}_2)(O_{i'})\right] + \cdots \\
&\quad + \left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}(\mathbf{Q}_{T-1})(O_i) - \frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}(\mathbf{Q}_T)(O_{i'})\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}(\mathbf{Q}_0)(O_i) - \frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}(\mathbf{Q}_T)(O_{i'}) \\
&\le \left|\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}(\mathbf{Q}_0)(O_i)\right| + \left|\frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}(\mathbf{Q}_T)(O_{i'})\right| \le 2L_{\max}.
\end{aligned}$$

Using $\mathcal{E}_{\mathcal{S}^k} \subset \mathcal{E}$ for each k and taking expectation, we obtain

$$\begin{aligned}
&\frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}[\Delta(q^k)] \\
&= \frac{\mathbb{E}\,[\xi(\hat{n})]}{T} + \frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\big)(O_i) - \inf_{\epsilon\in\mathcal{E}}\frac{1}{n}\sum_{j=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_j)\right] \\
&\le \frac{2L_{\max}}{T} + \frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\big)(O_i) - \inf_{\epsilon\in\mathcal{E}}\frac{1}{n}\sum_{j=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_j)\right] \\
&\le \frac{2L_{\max}}{T} + \frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}\left[\inf_{\epsilon\in\mathcal{E}_{\mathcal{S}^k}}\frac{1}{n}\sum_{j'=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{j'}) - \inf_{\epsilon\in\mathcal{E}}\frac{1}{n}\sum_{j=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_j)\right. \\
&\quad \left. + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\big)(O_i) - \inf_{\epsilon\in\mathcal{E}_{\mathcal{S}^k}}\frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{i'})}_{\triangleq\ \bigstar(\mathrm{i})}\right].
\end{aligned}$$

We now bound the expectation of each term on the right. By Assumption 2.4, $\mathcal{E}_{\mathcal{S}^k}$ is compact and nonempty, and hence the infimum $\inf_{\epsilon\in\mathcal{E}_{\mathcal{S}^k}} n^{-1}\sum_i \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i)$ is attained by some $\bar{\epsilon}^k \in \mathcal{E}_{\mathcal{S}^k}$. The next inequality converts restricted empirical minimization into a uniform deviation term using optimality and the triangle inequality:

$$\begin{aligned}
&\bigstar(\mathrm{i}) \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\left(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\right)(O_i) - \frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}\left(\mathbf{Q}_k(\bar{\epsilon}^k \mid \mathbf{G})\right)(O_{i'})
\end{aligned}$$

$$
\begin{aligned}
&= \left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\left(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\right)(O_i) - \frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\left(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\right)(O_{I_{k,b}})\right] \\
&\quad + \left[\frac{1}{B}\sum_{b'=1}^{B}\mathcal{L}\left(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\right)(O_{I_{k,b'}}) - \frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\left(\mathbf{Q}_k(\bar{\epsilon}^k \mid \mathbf{G})\right)(O_{I_{k,b}})\right] \\
&\quad + \left[\frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\left(\mathbf{Q}_k(\bar{\epsilon}^k \mid \mathbf{G})\right)(O_{I_{k,b}}) - \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\left(\mathbf{Q}_k(\bar{\epsilon}^k \mid \mathbf{G})\right)(O_i)\right] \\
&= \left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\left(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\right)(O_i) - \frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\left(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\right)(O_{I_{k,b}})\right] \\
&\quad + \left[\frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\left(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\right)(O_{I_{k,b}}) - \inf_{\epsilon\in\mathcal{E}_{\mathcal{S}^k}}\frac{1}{B}\sum_{b'=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{I_{k,b'}})\right] \\
&\quad + \left[\inf_{\epsilon\in\mathcal{E}_{\mathcal{S}^k}}\frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{I_{k,b}}) - \frac{1}{B}\sum_{b'=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\bar{\epsilon}^k \mid \mathbf{G})\big)(O_{I_{k,b'}})\right] \\
&\quad + \left[\frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\left(\mathbf{Q}_k(\bar{\epsilon}^k \mid \mathbf{G})\right)(O_{I_{k,b}}) - \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\left(\mathbf{Q}_k(\bar{\epsilon}^k \mid \mathbf{G})\right)(O_i)\right] \\
&\le \left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\left(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\right)(O_i) - \frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\left(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\right)(O_{I_{k,b}})\right] + 0 + 0 \\
&\quad + \left[\frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\left(\mathbf{Q}_k(\bar{\epsilon}^k \mid \mathbf{G})\right)(O_{I_{k,b}}) - \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\left(\mathbf{Q}_k(\bar{\epsilon}^k \mid \mathbf{G})\right)(O_i)\right] \\
&\le \left|\frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\big)(O_{I_{k,b}}) - \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon^k \mid \mathbf{G})\big)(O_i)\right| \\
&\quad + \left|\frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\bar{\epsilon}^k \mid \mathbf{G})\big)(O_{I_{k,b}}) - \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\bar{\epsilon}^k \mid \mathbf{G})\big)(O_i)\right| \\
&\le \sup_{\epsilon\in\mathcal{E}_{\mathcal{S}^k}}\left|\frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{I_{k,b}}) - \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i)\right| \\
&\quad + \sup_{\epsilon\in\mathcal{E}_{\mathcal{S}^k}}\left|\frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{I_{k,b}}) - \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i)\right| \\
&= 2\sup_{\epsilon\in\mathcal{E}_{\mathcal{S}^k}}\left|\frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{I_{k,b}}) - \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i)\right| \\
&\le 2\sup_{\epsilon\in\mathcal{E}}\left|\frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{I_{k,b}}) - \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i)\right|.
\end{aligned}
$$

Taking expectation and using this bound inside the earlier inequality yields

$$
\begin{aligned}
&\mathbb{E}\left[\bigstar(\mathrm{i})\right] \\
&\le 2\,\mathbb{E}\left[\sup_{\epsilon\in\mathcal{E}}\left|\frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{I_{k,b}}) - \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i)\right|\right] \triangleq 2\,\mathbb{E}\left[\sup_{\epsilon\in\mathcal{E}}|\bigstar(\mathrm{ii})|\right], \qquad (50)
\end{aligned}
$$

so it remains to bound (50) explicitly. The following discretization argument exposes all dependencies on $d, B, L_\epsilon, R_{\mathcal{E}}$. Since $\mathbf{0} \in \mathcal{E}$ and $\sup_{\epsilon,\epsilon'\in\mathcal{E}}\|\epsilon - \epsilon'\| = R_{\mathcal{E}}$, for every $\epsilon \in \mathcal{E}$,

$$
\|\epsilon\| = \|\epsilon - \mathbf{0}\| \le \sup_{\epsilon'\in\mathcal{E}}\|\epsilon - \epsilon'\| \le \sup_{\epsilon',\epsilon''\in\mathcal{E}}\|\epsilon' - \epsilon''\| = R_{\mathcal{E}}, \qquad (51)
$$

so $\mathcal{E} \subset \{\epsilon : \|\epsilon\| \le R_{\mathcal{E}}\}$. Fix any $u > 0$ and take $\mathcal{N}_u \subset \mathcal{E}$ maximal with $\|\epsilon - \epsilon'\| > u$ for all distinct $\epsilon, \epsilon' \in \mathcal{N}_u$. Maximality implies that $\mathcal{N}_u$ is a u-net of $\mathcal{E}$, because if some $\tilde{\epsilon} \in \mathcal{E}$ satisfied $\|\tilde{\epsilon} - \epsilon\| > u$ for all $\epsilon \in \mathcal{N}_u$, then $\mathcal{N}_u \cup \{\tilde{\epsilon}\}$ would contradict maximality. For $\epsilon \neq \epsilon'$ in $\mathcal{N}_u$, we have $\mathbb{B}(\epsilon, u/2) \cap \mathbb{B}(\epsilon', u/2) = \varnothing$. For $\epsilon \in \mathcal{N}_u$ with $\|\epsilon\| \le R_{\mathcal{E}}$, the packing bound is obtained by disjoint Euclidean balls and exact volume formulas:

$$
\mathbb{B}(\epsilon, u/2) \subset \mathbb{B}(\mathbf{0}, R_{\mathcal{E}} + u/2),
$$

and hence

$$\sum_{\epsilon\in\mathcal{N}_u}\mathrm{Vol}\big(\mathbb{B}(\epsilon,u/2)\big)\le\mathrm{Vol}\big(\mathbb{B}(\mathbf{0},R_{\mathcal{E}}+u/2)\big).$$

Therefore, similar to the derivation of (19), we have

$$|\mathcal{N}_u|\cdot\frac{\pi^{d/2}}{\hat{\Gamma}(d/2+1)}\left(\frac{u}{2}\right)^d\le\frac{\pi^{d/2}}{\hat{\Gamma}(d/2+1)}\left(R_{\mathcal{E}}+\frac{u}{2}\right)^d,$$
$$\log\big(2|\mathcal{N}_u|\big)\le\log 2+d\log\left(1+\frac{2R_{\mathcal{E}}}{u}\right).\tag{52}$$

We now reduce the supremum over $\mathcal{E}$ to a supremum over $\mathcal{N}_u$ with an explicit Lipschitz error. For any $\epsilon\in\mathcal{E}$, pick $\pi(\epsilon)\in\mathcal{N}_u$ with $\big\|\epsilon-\pi(\epsilon)\big\|\le u$. Applying the established Lipschitz bounds, we obtain

$$\begin{aligned}
&|\bigstar(\text{ii})|\\
&=\Bigg|\frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\epsilon\mid\mathbf{G})\big)(O_{I_{k,b}})-\frac{1}{B}\sum_{b'=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\pi(\epsilon)\mid\mathbf{G})\big)(O_{I_{k,b'}})\\
&\quad+\frac{1}{B}\sum_{b''=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\pi(\epsilon)\mid\mathbf{G})\big)(O_{I_{k,b''}})-\frac{1}{n}\sum_{i''=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\pi(\epsilon)\mid\mathbf{G})\big)(O_{i''})\\
&\quad+\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\pi(\epsilon)\mid\mathbf{G})\big)(O_i)-\frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon\mid\mathbf{G})\big)(O_{i'})\Bigg|\\
&\le\Bigg|\frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\epsilon\mid\mathbf{G})\big)(O_{I_{k,b}})-\frac{1}{B}\sum_{b'=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\pi(\epsilon)\mid\mathbf{G})\big)(O_{I_{k,b'}})\Bigg|\\
&\quad+\Bigg|\frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\pi(\epsilon)\mid\mathbf{G})\big)(O_{I_{k,b}})-\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\pi(\epsilon)\mid\mathbf{G})\big)(O_i)\Bigg|\\
&\quad+\Bigg|\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\pi(\epsilon)\mid\mathbf{G})\big)(O_i)-\frac{1}{n}\sum_{i'=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon\mid\mathbf{G})\big)(O_{i'})\Bigg|\\
&\overset{(12)}{\le}L_\epsilon\big\|\epsilon-\pi(\epsilon)\big\|+\Bigg|\frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\pi(\epsilon)\mid\mathbf{G})\big)(O_{I_{k,b}})-\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\pi(\epsilon)\mid\mathbf{G})\big)(O_i)\Bigg|+L_\epsilon\big\|\epsilon-\pi(\epsilon)\big\|\\
&\le 2L_\epsilon u+\Bigg|\frac{1}{B}\sum_{b=1}^{B}\mathcal{L}\big(\mathbf{Q}_k(\pi(\epsilon)\mid\mathbf{G})\big)(O_{I_{k,b}})-\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\pi(\epsilon)\mid\mathbf{G})\big)(O_i)\Bigg|.
\end{aligned}$$

Taking the supremum over $\epsilon\in\mathcal{E}$ and then taking expectation yields

$$\mathbb{E}\left[\sup_{\epsilon\in\mathcal{E}}|\bigstar(\text{ii})|\right]\le 2L_\epsilon u+\mathbb{E}\left[\sup_{\epsilon\in\mathcal{N}_u}|\bigstar(\text{ii})|\right].\tag{53}$$

We now bound the finite supremum by an explicit tail integration argument. Fix $\epsilon\in\mathcal{N}_u$. For any $\lambda>0$, using Markov's inequality and conditional independence, we have

$$\begin{aligned}
&\mathbb{P}\left(\bigstar(\text{ii})\ge t\,\middle|\,\mathbf{Q}_k,\{O_i\}_{i=1}^n\right)\\
&=\mathbb{P}\left(\frac{1}{B}\sum_{b=1}^{B}\left(\mathcal{L}\big(\mathbf{Q}_k(\epsilon\mid\mathbf{G})\big)(O_{I_{k,b}})-\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon\mid\mathbf{G})\big)(O_i)\right)\ge t\,\middle|\,\mathbf{Q}_k,\{O_i\}_{i=1}^n\right)\\
&=\mathbb{P}\left(\exp\left\{\lambda\sum_{b=1}^{B}\left(\mathcal{L}\big(\mathbf{Q}_k(\epsilon\mid\mathbf{G})\big)(O_{I_{k,b}})-\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon\mid\mathbf{G})\big)(O_i)\right)\right\}\ge e^{\lambda Bt}\,\middle|\,\mathbf{Q}_k,\{O_i\}_{i=1}^n\right)\\
&\le e^{-\lambda Bt}\,\mathbb{E}\left[\exp\left\{\lambda\sum_{b=1}^{B}\left(\mathcal{L}\big(\mathbf{Q}_k(\epsilon\mid\mathbf{G})\big)(O_{I_{k,b}})-\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon\mid\mathbf{G})\big)(O_i)\right)\right\}\,\middle|\,\mathbf{Q}_k,\{O_i\}_{i=1}^n\right]\\
&\le e^{-\lambda Bt}\prod_{b=1}^{B}\mathbb{E}\left[\exp\left\{\lambda\left(\mathcal{L}\big(\mathbf{Q}_k(\epsilon\mid\mathbf{G})\big)(O_{I_{k,b}})-\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon\mid\mathbf{G})\big)(O_i)\right)\right\}\,\middle|\,\mathbf{Q}_k,\{O_i\}_{i=1}^n\right].
\end{aligned}\tag{54}$$

Conditional on $\mathbf{Q}_k$ and $\{O_i\}$,

$$\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{I_{k,b}}) - \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i)$$
$$\in \left[-L_{\max} - \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i),\ L_{\max} - \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i)\right].$$

This interval has length

$$\left(L_{\max} - \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i)\right) - \left(-L_{\max} - \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i)\right) = 2L_{\max},$$

so Hoeffding's lemma applies with range length $2L_{\max}$ and yields

$$\mathbb{E}\left[\exp\left\{\lambda\left(\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{I_{k,b}}) - \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i)\right)\right\} \,\middle|\, \mathbf{Q}_k, \{O_i\}_{i=1}^n\right]$$
$$\leq \exp\left(\frac{\lambda^2(2L_{\max})^2}{8}\right) = \exp\left(\frac{\lambda^2 L_{\max}^2}{2}\right).$$

Substituting into the product (54) gives

$$\mathbb{P}\left(\bigstar(\text{ii}) \geq t \,\middle|\, \mathbf{Q}_k, \{O_i\}_{i=1}^n\right) \leq e^{-\lambda B t}\prod_{b=1}^{B}\exp\left(\frac{\lambda^2 L_{\max}^2}{2}\right) = \exp\left(-\lambda B t + \frac{B\lambda^2 L_{\max}^2}{2}\right).$$

Optimizing the exponent over $\lambda > 0$, the minimizer solves

$$\frac{d}{d\lambda}\left(-\lambda B t + \frac{B\lambda^2 L_{\max}^2}{2}\right) = -Bt + B\lambda L_{\max}^2 = 0,$$

and substituting $\lambda^\star = \frac{t}{L_{\max}^2}$ back yields

$$\mathbb{P}\left(\bigstar(\text{ii}) \geq t \,\middle|\, \mathbf{Q}_k, \{O_i\}_{i=1}^n\right) \leq \exp\left(-\lambda^\star B t + \frac{B(\lambda^\star)^2 L_{\max}^2}{2}\right)$$
$$= \exp\left(-\frac{t}{L_{\max}^2}Bt + \frac{B}{2}\frac{t^2}{L_{\max}^4}L_{\max}^2\right) = \exp\left(-\frac{Bt^2}{2L_{\max}^2}\right).$$

Applying the same derivation to $-\bigstar(\text{ii})$ gives the symmetric bound for the lower tail. By a union bound over the two events, we obtain the standard bound

$$\mathbb{P}\left(\,|\bigstar(\text{ii})| \geq t \,\middle|\, \mathbf{Q}_k, \{O_i\}_{i=1}^n\right) \leq 2\exp\left(-\frac{Bt^2}{2L_{\max}^2}\right).$$

Applying a union bound over the finite net $\mathcal{N}_u$, we have

$$\begin{aligned}\mathbb{P}\left(\sup_{\epsilon\in\mathcal{N}_u}|\bigstar(\text{ii})| \geq t \,\middle|\, \mathbf{Q}_k, \{O_i\}_{i=1}^n\right) &= \mathbb{P}\left(\bigcup_{\epsilon\in\mathcal{N}_u}\{|\bigstar(\text{ii})| \geq t\} \,\middle|\, \mathbf{Q}_k, \{O_i\}_{i=1}^n\right)\\ &\leq \sum_{\epsilon\in\mathcal{N}_u}\mathbb{P}\left(\,|\bigstar(\text{ii})| \geq t \,\middle|\, \mathbf{Q}_k, \{O_i\}_{i=1}^n\right)\\ &\leq \sum_{\epsilon\in\mathcal{N}_u} 2\exp\left(-\frac{Bt^2}{2L_{\max}^2}\right) = 2|\mathcal{N}_u|\exp\left(-\frac{Bt^2}{2L_{\max}^2}\right).\end{aligned} \tag{55}$$

We convert this into an expectation bound using tail integration and an explicit split point. For the nonnegative random variable, the tail bound yields

$$\mathbb{E}\left[\sup_{\epsilon\in\mathcal{N}_u}|\bigstar(\text{ii})| \,\middle|\, \mathbf{Q}_k, \{O_i\}_{i=1}^n\right] = \int_0^\infty \mathbb{P}\left(\sup_{\epsilon\in\mathcal{N}_u}|\bigstar(\text{ii})| \geq t \,\middle|\, \mathbf{Q}_k, \{O_i\}_{i=1}^n\right)dt$$

$$\overset{(55)}{\leq} \int_0^\infty \min\left\{1,\ 2\,|\mathcal{N}_u| \exp\left(-\frac{Bt^2}{2L_{\max}^2}\right)\right\} dt.$$

Define the split point t_0 by the equation

$$2\,|\mathcal{N}_u| \exp\left\{-\frac{Bt^2}{2L_{\max}^2}\right\}\Big|_{t=t_0} = 1, t_0 = L_{\max}\sqrt{\frac{2\log\big(2|\mathcal{N}_u|\big)}{B}}. \tag{56}$$

Then

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\epsilon\in\mathcal{N}_u} |\bigstar(\text{ii})| \,\middle|\, \mathbf{Q}_k, \{O_i\}_{i=1}^n\right] \\
&= \int_0^{t_0} \mathbb{P}\left(\sup_{\epsilon\in\mathcal{N}_u} |\bigstar(\text{ii})| \geq t \,\middle|\, \mathbf{Q}_k, \{O_i\}_{i=1}^n\right) dt + \int_{t_0}^\infty \mathbb{P}\left(\sup_{\epsilon\in\mathcal{N}_u} |\bigstar(\text{ii})| \geq t \,\middle|\, \mathbf{Q}_k, \{O_i\}_{i=1}^n\right) dt \\
&\leq \int_0^{t_0} 1\, dt + \int_{t_0}^\infty 2|\mathcal{N}_u| \exp\left\{-\frac{Bt^2}{2L_{\max}^2}\right\} dt = t_0 + 2|\mathcal{N}_u| \int_{t_0}^\infty \exp\left\{-\frac{Bt^2}{2L_{\max}^2}\right\} dt.
\end{aligned}$$

We evaluate the Gaussian tail integral explicitly by substitution $s = \sqrt{B}\ t/(\sqrt{2}L_{\max})$, so $t = \sqrt{2}L_{\max}s/\sqrt{B}$ and $dt = \sqrt{2}L_{\max}\ ds/\sqrt{B}$, giving

$$\begin{aligned}
\int_{t_0}^\infty \exp\left\{-\frac{Bt^2}{2L_{\max}^2}\right\} dt &= \int_{\frac{\sqrt{B}t_0}{\sqrt{2}L_{\max}}}^\infty e^{-s^2}\ \frac{\sqrt{2}L_{\max}}{\sqrt{B}}\ ds \\
&= \frac{\sqrt{2}L_{\max}}{\sqrt{B}} \int_{s_0}^\infty e^{-s^2} ds, \quad s_0 = \frac{\sqrt{B}t_0}{\sqrt{2}L_{\max}} = \sqrt{\log\big(2|\mathcal{N}_u|\big)}.
\end{aligned}$$

Applying the inequality

$$\int_{s_0}^\infty e^{-s^2} ds = \int_{s_0}^\infty \frac{s}{s} e^{-s^2} ds \leq \frac{1}{s_0} \int_{s_0}^\infty s e^{-s^2} ds = \frac{1}{s_0}\left[-\frac{1}{2}e^{-s^2}\right]_{s=s_0}^{s=\infty} = \frac{e^{-s_0^2}}{2s_0},$$

we obtain

$$\begin{aligned}
&2|\mathcal{N}_u| \int_{t_0}^\infty \exp\left\{-\frac{Bt^2}{2L_{\max}^2}\right\} dt = 2|\mathcal{N}_u| \cdot \frac{\sqrt{2}L_{\max}}{\sqrt{B}} \int_{s_0}^\infty e^{-s^2} ds \leq 2|\mathcal{N}_u| \cdot \frac{\sqrt{2}L_{\max}}{\sqrt{B}} \cdot \frac{e^{-s_0^2}}{2s_0} \\
&= \frac{\sqrt{2}L_{\max}}{\sqrt{B}} \cdot \frac{|\mathcal{N}_u| e^{-s_0^2}}{s_0}.
\end{aligned}$$

Since $s_0^2 = \log(2|\mathcal{N}_u|)$, we have

$$e^{-s_0^2} = \frac{1}{2|\mathcal{N}_u|}, \quad |\mathcal{N}_u| e^{-s_0^2} = \frac{1}{2},$$

and thus

$$2|\mathcal{N}_u| \int_{t_0}^\infty \exp\left\{-\frac{Bt^2}{2L_{\max}^2}\right\} dt \leq \frac{\sqrt{2}L_{\max}}{\sqrt{B}} \cdot \frac{1/2}{s_0} = \frac{\sqrt{2}L_{\max}}{\sqrt{B}} \cdot \frac{1}{2\sqrt{\log(2\,|\mathcal{N}_u|)}}.$$

Using $|\mathcal{N}_u| \geq 1$ so that $\log(2\,|\mathcal{N}_u|) \geq \log 2 > 1/2$, we have $1/\sqrt{\log(2\,|\mathcal{N}_u|)} \leq 2\sqrt{\log(2\,|\mathcal{N}_u|)}$, which implies

$$\begin{aligned}
\frac{\sqrt{2}L_{\max}}{\sqrt{B}} \cdot \frac{1}{2\sqrt{\log(2\,|\mathcal{N}_u|)}} &\leq \frac{\sqrt{2}L_{\max}}{\sqrt{B}} \cdot \frac{1}{2} \cdot 2\sqrt{\log(2\,|\mathcal{N}_u|)} \\
&= L_{\max}\sqrt{\frac{2\log(2\,|\mathcal{N}_u|)}{B}} \overset{(56)}{=} t_0.
\end{aligned} \tag{57}$$

Taking the unconditional expectation does not change the right-hand side of (57). Inserting this into the net reduction bound, we have

$$\mathbb{E}\left[\sup_{\epsilon\in\mathcal{E}} |\bigstar(\text{ii})|\right] \overset{(53)}{\leq} 2L_\epsilon u + t_0 + 2|\mathcal{N}_u| \int_{t_0}^\infty \exp\left\{-\frac{Bt^2}{2L_{\max}^2}\right\} dt \leq 2L_\epsilon u + t_0 + t_0$$

$$= 2L_\epsilon u + 2L_{\max}\sqrt{\frac{2\log(2\,|\mathcal{N}_u|)}{B}} \overset{(52)}{\leq} 2L_\epsilon u + 2L_{\max}\sqrt{\frac{2\left[\log 2 + d\log\left(1+\frac{2R_\mathcal{E}}{u}\right)\right]}{B}}.$$

We choose $u = R_\mathcal{E}/\sqrt{B}$. Then

$$\mathbb{E}\left[\sup_{\epsilon\in\mathcal{E}} |\bigstar(\text{ii})|\right] \leq 2L_\epsilon \cdot \frac{R_\mathcal{E}}{\sqrt{B}} + 2L_{\max}\sqrt{\frac{2\left[\log 2 + d\log\left(1+\frac{2R_\mathcal{E}}{R_\mathcal{E}/\sqrt{B}}\right)\right]}{B}}$$
$$= \frac{2L_\epsilon R_\mathcal{E}}{\sqrt{B}} + 2L_{\max}\sqrt{\frac{2\left[\log 2 + d\log(1+2\sqrt{B})\right]}{B}}.$$

Multiplying by the factor 2 from the basic inequality, we obtain

$$\mathbb{E}\left[\bigstar(\text{i})\right] \leq \frac{4L_\epsilon R_\mathcal{E}}{\sqrt{B}} + 4L_{\max}\sqrt{\frac{2\left[\log 2 + d\log(1+2\sqrt{B})\right]}{B}}. \tag{58}$$

We now bound the coordinate restriction term. Let $\epsilon_k^\star \in \arg\min_{\epsilon\in\mathcal{E}} n^{-1}\sum_{i=1}^n \mathcal{L}\left(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\right)(O_i)$, which exists since $\mathcal{L}$ is continuous and $\mathcal{E}$ is compact. Let $\epsilon_{k,\mathcal{S}^k}^\star$ be obtained from $\epsilon_k^\star$ by setting coordinates outside $\mathcal{S}^k$ to zero. By Assumption 2.4, $\epsilon_{k,\mathcal{S}^k}^\star \in \mathcal{E}_{\mathcal{S}^k}$, so

$$\begin{aligned}
&\inf_{\epsilon\in\mathcal{E}_{\mathcal{S}^k}} \frac{1}{n}\sum_{i=1}^n \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i) \\
&\leq \frac{1}{n}\sum_{i=1}^n \mathcal{L}\big(\mathbf{Q}_k(\epsilon_{k,\mathcal{S}^k}^\star \mid \mathbf{G})\big)(O_i) \\
&= \frac{1}{n}\sum_{i'=1}^n \mathcal{L}\big(\mathbf{Q}_k(\epsilon_k^\star \mid \mathbf{G})\big)(O_{i'}) + \frac{1}{n}\sum_{i=1}^n \mathcal{L}\big(\mathbf{Q}_k(\epsilon_{k,\mathcal{S}^k}^\star \mid \mathbf{G})\big)(O_i) - \frac{1}{n}\sum_{i''=1}^n \mathcal{L}\big(\mathbf{Q}_k(\epsilon_k^\star \mid \mathbf{G})\big)(O_{i''}),
\end{aligned}$$

and hence by Lipschitzness,

$$\begin{aligned}
&\inf_{\epsilon\in\mathcal{E}_{\mathcal{S}^k}} \frac{1}{n}\sum_{i'=1}^n \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{i'}) - \inf_{\epsilon\in\mathcal{E}} \frac{1}{n}\sum_{i=1}^n \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_i) \\
&\leq \frac{1}{n}\sum_{i=1}^n \mathcal{L}\big(\mathbf{Q}_k(\epsilon_{k,\mathcal{S}^k}^\star \mid \mathbf{G})\big)(O_i) - \frac{1}{n}\sum_{i'=1}^n \mathcal{L}\big(\mathbf{Q}_k(\epsilon_k^\star \mid \mathbf{G})\big)(O_{i'}) \overset{(11)}{\leq} L_\epsilon \left\|\epsilon_{k,\mathcal{S}^k}^\star - \epsilon_k^\star\right\|.
\end{aligned}$$

Write $\epsilon_k^\star = (\epsilon_{k,1}^\star, \ldots, \epsilon_{k,d}^\star)$. Conditioning on $\epsilon_k^\star$ and taking expectation over $\mathcal{S}^k$, we have

$$\begin{aligned}
\mathbb{E}\left[\left\|\epsilon_{k,\mathcal{S}^k}^\star - \epsilon_k^\star\right\|^2 \,\middle|\, \epsilon_k^\star\right] &= \mathbb{E}\left[\sum_{j=1}^d \left(\epsilon_{k,j}^\star\right)^2 \mathbf{1}\{j \notin \mathcal{S}^k\} \,\middle|\, \epsilon_k^\star\right] \\
&= \sum_{j=1}^d \left(\epsilon_{k,j}^\star\right)^2 \mathbb{E}\left[\mathbf{1}\{j \notin \mathcal{S}^k\} \mid \epsilon_k^\star\right] = \sum_{j=1}^d \left(\epsilon_{k,j}^\star\right)^2 \mathbb{P}(j \notin \mathcal{S}^k).
\end{aligned}$$

We compute $\mathbb{P}(j \notin \mathcal{S}^k)$ under uniform sampling of size-S subsets:

$$\mathbb{P}(j \notin \mathcal{S}^k) = \frac{\#\{\mathcal{S} \subset \{1,\ldots,d\} :\, |\mathcal{S}| = S,\ j \notin \mathcal{S}\}}{\#\{\mathcal{S} \subset \{1,\ldots,d\} :\, |\mathcal{S}| = S\}} = \frac{(d-1)!}{S!(d-1-S)!} \cdot \frac{S!(d-S)!}{d!} = \frac{d-S}{d}.$$

Applying Jensen's inequality, we obtain

$$\mathbb{E}\left[\left\|\epsilon_{k,\mathcal{S}^k}^\star - \epsilon_k^\star\right\| \,\middle|\, \epsilon_k^\star\right] \leq \sqrt{\mathbb{E}\left[\left\|\epsilon_{k,\mathcal{S}^k}^\star - \epsilon_k^\star\right\|^2 \,\middle|\, \epsilon_k^\star\right]} = \sqrt{\sum_{j=1}^d (\epsilon_{k,j}^\star)^2 \left(\frac{d-S}{d}\right)}$$

$$= \sqrt{\frac{d-S}{d} \cdot \|\epsilon_k^\star\|^2} = \|\epsilon_k^\star\| \sqrt{1 - \frac{S}{d}}. \tag{59}$$

Combining the two expectation bounds for the mini-batch term and the coordinate term, we obtain the time average:

$$\begin{aligned}
&\frac{1}{T}\sum_{k=0}^{T-1} \mathbb{E}[\Delta(q^k)] \\
&\leq \frac{2L_{\max}}{T} + \frac{1}{T}\sum_{k=0}^{T-1} \mathbb{E}\left[\inf_{\epsilon \in \mathcal{E}_{\mathcal{S}^k}} \frac{1}{n}\sum_{j'=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{j'}) - \inf_{\epsilon \in \mathcal{E}} \frac{1}{n}\sum_{j=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_j) + \bigstar(\mathrm{i})\right] \\
&\leq \frac{2L_{\max}}{T} + \frac{1}{T}\sum_{k=0}^{T-1} \Big\{\mathbb{E}\left[\bigstar(\mathrm{i})\right] \\
&\quad + \mathbb{E}\left[\inf_{\epsilon \in \mathcal{E}_{\mathcal{S}^k}} \frac{1}{n}\sum_{j'=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_{j'}) - \inf_{\epsilon \in \mathcal{E}} \frac{1}{n}\sum_{j=1}^{n} \mathcal{L}\big(\mathbf{Q}_k(\epsilon \mid \mathbf{G})\big)(O_j)\right]\Big\} \\
&\overset{(58)}{\leq} \frac{2L_{\max}}{T} + \frac{1}{T}\sum_{k=0}^{T-1}\left\{\frac{4L_\epsilon R_{\mathcal{E}}}{\sqrt{B}}\right\} + \frac{1}{T}\sum_{k=0}^{T-1}\left\{4L_{\max}\sqrt{\frac{2\left[\log 2 + d\log(1+2\sqrt{B})\right]}{B}}\right\} \\
&\quad + \mathbb{E}\left[\frac{1}{T}\sum_{k=0}^{T-1}\left\{L_\epsilon \mathbb{E}\left[\left\|\epsilon_{k,\mathcal{S}^k}^\star - \epsilon_k^\star\right\| \mid \epsilon_k^\star\right]\right\}\right] \\
&\overset{(59)}{\leq} \frac{2L_{\max}}{T} + \frac{4L_\epsilon R_{\mathcal{E}}}{\sqrt{B}} + 4L_{\max}\sqrt{\frac{2\left[\log 2 + d\log(1+2\sqrt{B})\right]}{B}} + \mathbb{E}\left[\frac{1}{T}\sum_{k=0}^{T-1}\left\{L_\epsilon \|\epsilon_k^\star\|\sqrt{1-\frac{S}{d}}\right\}\right] \\
&\overset{(51)}{\leq} \frac{2L_{\max}}{T} + \frac{4L_\epsilon R_{\mathcal{E}}}{\sqrt{B}} + 4L_{\max}\sqrt{\frac{2\left[\log 2 + d\log(1+2\sqrt{B})\right]}{B}} + \frac{1}{T}\sum_{k=0}^{T-1}\left\{L_\epsilon R_{\mathcal{E}}\sqrt{1-\frac{S}{d}}\right\}.
\end{aligned}$$

Taking $\limsup_{T\to\infty}$ yields

$$\begin{aligned}
&\limsup_{T\to\infty} \frac{1}{T}\sum_{k=0}^{T-1} \mathbb{E}[\Delta(q^k)] \\
&\leq \limsup_{T\to\infty} \frac{2L_{\max}}{T} + \limsup_{T\to\infty}\left\{\frac{4L_\epsilon R_{\mathcal{E}}}{\sqrt{B}} + 4L_{\max}\sqrt{\frac{2\left[\log 2 + d\log(1+2\sqrt{B})\right]}{B}} + L_\epsilon R_{\mathcal{E}}\sqrt{1-\frac{S}{d}}\right\} \\
&= \frac{4L_\epsilon R_{\mathcal{E}}}{\sqrt{B}} + 4L_{\max}\sqrt{\frac{2\left[\log 2 + d\log(1+2\sqrt{B})\right]}{B}} + L_\epsilon R_{\mathcal{E}}\sqrt{1-\frac{S}{d}},
\end{aligned}$$

which is the claimed bound. □

I Implementation Details

We use MATLAB R2021a on a workstation running Windows 10 with 32GB RAM and an Intel Core i7 CPU. No GPU acceleration is employed, ensuring that the reported timing comparisons reflect pure algorithmic efficiency rather than hardware-specific optimizations. For $\Psi_{(1)}$ and $\Psi_{(2)}$, we obtain the initial fits $\mathbf{Q}_0$ using standard scikit-learn estimators. For $\Psi_{(3)}$ and $\Psi_{(4)}$, we follow the Super Learner procedure of van der Laan et al. [52]. Due to the heavy-tailed nature of the convergence trajectories, we recenter the horizontal axis for improved visual clarity.

I.1 Target Parameters

We consider four target parameters of varying structure and dimension.

The first target $\Psi_{(1)}$ is a distributional functional based on a scalar observation $O \in \mathbb{R}$. Fix a deterministic grid $x_j := -3 + 6(j-1)/29$ for $j \in \{1, \ldots, 30\}$, and define the CDF values $q_j(\mathbf{P}) := \mathbf{P}(O \leq x_j)$. The target parameter is the $d = 30$ vector $\Psi_{(1)}(\mathbf{P}) := (q_1(\mathbf{P})^2, \ldots, q_{30}(\mathbf{P})^2)$. We use the separable Bernoulli log-likelihood loss with $Y_j^{\mathrm{cdf}}(O) := \mathbb{I}(O \leq x_j)$, and the componentwise logistic tilting submodel $\mathrm{logit}(q_{k,j}(\epsilon)) = \mathrm{logit}(q_{k,j}) + \epsilon_j$ for $\epsilon \in \mathbb{R}^{30}$. This yields $\mathcal{E} \subset \mathbb{R}^{30}$, supporting indefinite outer-loop iteration without feasibility constraints. We solve the subproblem using mini-batch SGD on ϵ, where the gradient admits the closed form $\nabla_{\epsilon_j} \mathcal{L}(\mathbf{Q}_k(\epsilon))(O_i) = q_{k,j}(\epsilon) - Y_j^{\mathrm{cdf}}(O_i)$.

The second target $\Psi_{(2)}$ is a 30-dimensional median regression parameter. Here $O = (X, Y)$ with $X \in \mathbb{R}^{29}$ and $Y \in \mathbb{R}$. Define $X^{\sim} := (1, X^\top)^\top \in \mathbb{R}^{30}$ and the target $\Psi_{(2)}(\mathbf{P}) := \arg\min_{\beta \in \mathbb{R}^{30}} \mathbb{E}_{\mathbf{P}}|Y - \beta^\top X^{\sim}|$. We use the absolute deviation loss $\mathcal{L}(\beta)(X, Y) := |Y - \beta^\top X^{\sim}|$ and the additive submodel $\beta_k(\epsilon) := \beta_k + \epsilon$ with $\mathcal{E} = \mathbb{R}^{30}$. Since the loss is convex but non-smooth, we solve the subproblem using mini-batch stochastic subgradient descent with Polyak–Ruppert averaging to stabilize the optimization.

The third target $\Psi_{(3)}$ is the treatment-specific mean under a causal structure. Here $O = (Y, A, W)$ with $A \in \{0, 1\}$, $W \in \mathbb{R}^{28}$, and $Y \in \mathbb{R}$. Define the propensity score $\mathbf{G}(\mathbf{P})(1 \mid W) := \mathbf{P}(A = 1 \mid W)$ and the outcome regression $\mathbf{Q}(\mathbf{P})(A, W) := \mathbb{E}_{\mathbf{P}}[Y \mid A, W]$. The scalar target is $\Psi_{(3)}(\mathbf{P}) := \mathbb{E}_{\mathbf{P}}[\mathbf{Q}(\mathbf{P})(1, W)]$, evaluated via the empirical distribution of $\{W_i\}_{i=1}^n$. We use the squared-error loss and the least-squares linear tilting submodel

$$\mathbf{Q}_k(\epsilon \mid \hat{\mathbf{G}})(A, W) = \mathbf{Q}_k(A, W) + \epsilon\, H_{(3)}(A, W; \hat{\mathbf{G}}), \quad H_{(3)}(A, W; \mathbf{G}) := \frac{\mathbb{I}(A = 1)}{\mathbf{G}(1 \mid W)},$$

with $\epsilon \in \mathbb{R}$. We solve the subproblem using mini-batch SGD on the scalar ϵ.

The fourth target $\Psi_{(4)}$ is the variance of the conditional treatment effect (VTE), retaining the same data structure and nuisance components as $\Psi_{(3)}$. Define the blip function $b_{\mathbf{Q}}(W) := \mathbf{Q}(1, W) - \mathbf{Q}(0, W)$ and the scalar target $\Psi_{(4)}(\mathbf{P}) := \mathrm{Var}_{\mathbf{P}}(b_{\mathbf{Q}(\mathbf{P})}(W))$. We use the squared-error loss and the VTE linear tilting submodel

$$\mathbf{Q}_k(\epsilon \mid \hat{\mathbf{G}})(A, W) = \mathbf{Q}_k(A, W) + \epsilon\, H_{(4),k}(A, W), \quad H_{(4),k}(A, W) := 2(b_k(W) - \bar{b}_k)\, \frac{2A - 1}{\hat{\mathbf{G}}(A \mid W)},$$

where $b_k(W) := b_{\mathbf{Q}_k}(W)$ and $\bar{b}_k := n^{-1} \sum_{i=1}^n b_k(W_i)$. Since $H_{(4),k}$ depends on $\mathbf{Q}_k$ through b_k, the outer-loop updates remain nontrivial and iterate until convergence.

I.2 Synthetic Data-Generating Distributions

For each target $\Psi_{(m)}$, we consider two synthetic data-generating distributions. In all scenarios, we generate $n = 10^6$ i.i.d. observations from the specified $\mathbf{P}_0$. We write $\mathrm{expit}(t) := \{1 + \exp(-t)\}^{-1}$ and define the clipping operator $\mathrm{clip}(u; \varepsilon) := \min\{1 - \varepsilon, \max\{\varepsilon, u\}\}$.

For $\Psi_{(1)}$, the observation is scalar $O \in \mathbb{R}$. In DGP-1A, we generate $O \sim \mathcal{N}(0, 1)$. In DGP-1B, we generate a two-component Gaussian mixture: sample $U \sim \mathrm{Bernoulli}(0.7)$ and draw $O \sim \mathcal{N}(-1, 0.5^2)$ if $U = 1$, otherwise $O \sim \mathcal{N}(2, 1.2^2)$. The ground-truth $\Psi_{(1)}(\mathbf{P}_0)$ is computable in closed form using the corresponding CDF.

For $\Psi_{(2)}$, the observation is $O = (X, Y)$ with $X \in \mathbb{R}^{29}$ and $Y \in \mathbb{R}$. Let $\Sigma_X \in \mathbb{R}^{29 \times 29}$ denote the AR(1) covariance matrix $(\Sigma_X)_{jk} := 0.3^{|j-k|}$. In DGP-2A, we generate $X \sim \mathcal{N}(0, \Sigma_X)$ and $Y = \beta^{\star\top} X + \xi$ with $\xi \sim \mathrm{Laplace}(0, 1)$ and $\beta_j^\star = 0.5(-1)^{j+1}/j$ for $j \geq 1$, $\beta_0^\star = 0$. In DGP-2B, we generate X from a Gaussian mixture and define a nonlinear conditional median function $m(X)$ involving interactions and trigonometric terms; we then generate $Y = m(X) + \xi$ with heteroscedastic noise. The ground-truth $\beta(\mathbf{P}_0)$ for DGP-2B is approximated by Monte Carlo.

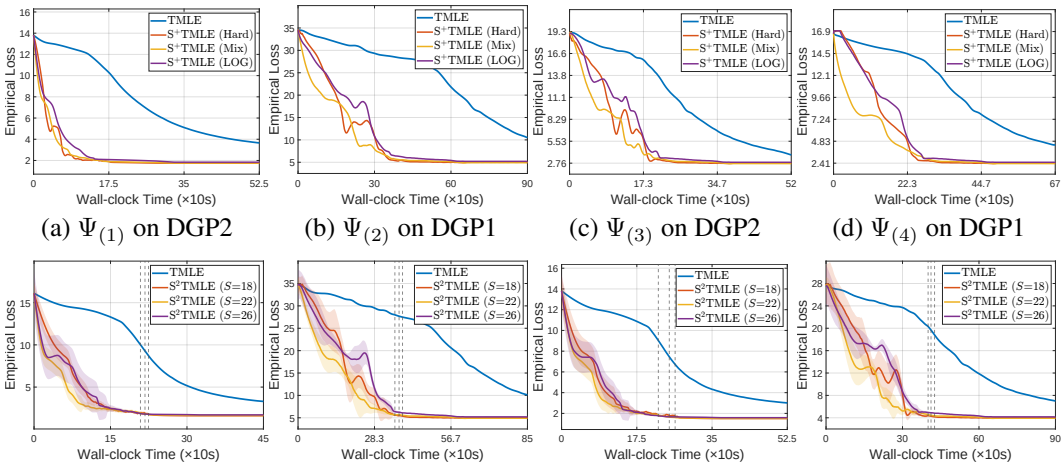
For $\Psi_{(3)}$, the observation is $O = (Y, A, W)$ with $W \in \mathbb{R}^{28}$, $A \in \{0, 1\}$, and $Y \in \mathbb{R}$. Let Σ_W denote the AR(1) covariance matrix $(\Sigma_W)_{jk} := 0.3^{|j-k|}$. In DGP-3A, we generate $W \sim \mathcal{N}(0, \Sigma_W)$, $A \mid W \sim \mathrm{Bernoulli}(\mathrm{clip}(\mathrm{expit}(\ell(W)); 0.05))$ with a linear $\ell(W)$, and $Y = \mu(W) + A\tau + \varepsilon$ with constant treatment effect $\tau = 1.0$ and $\varepsilon \sim \mathcal{N}(0, 1)$. In DGP-3B, we generate heavy-tailed covariates via a scale mixture of normals, use a nonlinear propensity score with interactions and trigonometric

terms, and generate $Y = \mu(W) + A\tau(W) + \varepsilon$ with heterogeneous treatment effect $\tau(W)$ and t_3-distributed errors. The ground-truth $\Psi_{(3)}(\mathbf{P}_0)$ equals 1.2 for DGP-3A and is approximated by Monte Carlo for DGP-3B.

For $\Psi_{(4)}$, the observation structure matches $\Psi_{(3)}$. In DGP-4A, we generate $W \sim \mathcal{N}(0, \Sigma_W)$ and $Y = \mu(W) + A\tau(W) + \varepsilon$ with linear $\tau(W) = 0.5 + 0.25W_1 + 0.15W_2 - 0.10W_3 + 0.05W_4$ and $\varepsilon \sim \mathcal{N}(0, 1)$. The ground-truth VTE admits the closed form $\Psi_{(4)}(\mathbf{P}_0) = c^\top \Sigma_{W,1:4} c$ where $c := (0.25, 0.15, -0.10, 0.05)^\top$. In DGP-4B, we generate W from a two-component Gaussian mixture, use a nonlinear propensity score, and generate Y with indicator-based heterogeneous treatment effect and t_3-distributed errors. The ground-truth $\Psi_{(4)}(\mathbf{P}_0)$ is approximated by Monte Carlo.

J Additional Results

Following a similar experimental setup in the main body, we provide additional results along with analysis in Figure 4.



(a) $\Psi_{(1)}$ on DGP2 (b) $\Psi_{(2)}$ on DGP1 (c) $\Psi_{(3)}$ on DGP2 (d) $\Psi_{(4)}$ on DGP1

(e) $\Psi_{(1)}$ on DGP1 (f) $\Psi_{(2)}$ on DGP1 (g) $\Psi_{(1)}$ on DGP2 (h) $\Psi_{(2)}$ on DGP2

Figure 4: Empirical loss vs. wall-clock time for S$^+$TMLE (top row) and S^{2}TMLE (bottom row). Top: different choices of SP-regularizer lead to different adaptive smoothing mechanisms. The transition toward full-batch targeting is adaptive, so no explicit vertical switching line is shown. Bottom: S is varied while $B = 0.5n$ is fixed. When sampling simultaneously along both the coordinate of Ψ and sample dimensions, the subproblem size is significantly smaller than the full problem, resulting in low per-iteration cost. However, reducing S does not always lead to speedup; overly small values of S may cause stability issues during the optimization process.

Ablation on Hyperparameters Figure 5 reports a sweep over (B, S) values. When either is taken extremely small, the mini-batch subproblem no longer approximates the perturbation step and updates degenerate into near-random moves, consistent with the $1/\sqrt{B}$ and $\sqrt{1 - S/d}$ scalings predicted by our theory.

Optimization Diagnostics Figure 6 compares optimization progress and target error per outer iteration. Figure 7 shows the evolution of nominal confidence-interval coverage for the two scalar targets. Figure 8 compares full-data EIF residuals and empirical-loss gaps under matched computational work, separating stochastic targeting from stochastic optimization of the full-sample fluctuation problem.

Ablation on Inner Solver To further isolate the effect of stochastic targeting from any randomness introduced by the inner optimization routine, we conduct an additional ablation in which all methods share a common deterministic inner solver (gradient descent, with no solver-level SGD). Under this controlled setting, the only source of stochasticity is the targeting step itself. As shown in Figure 9, the advantage of stochastic targeting becomes even more pronounced: the gains reported in the main text are therefore not driven by the randomness of the inner solver. On the contrary, allowing solver-level randomness in the baselines partially narrows the gap, which means our main results are in fact *conservative* estimates of the benefit of stochastic targeting.

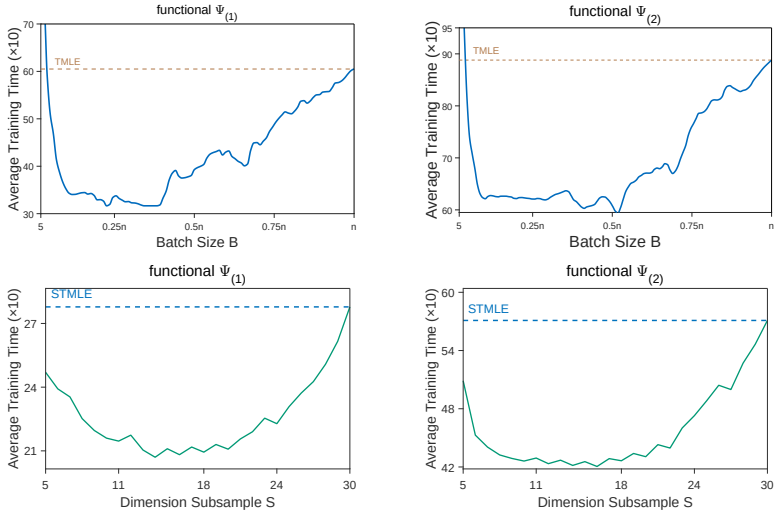
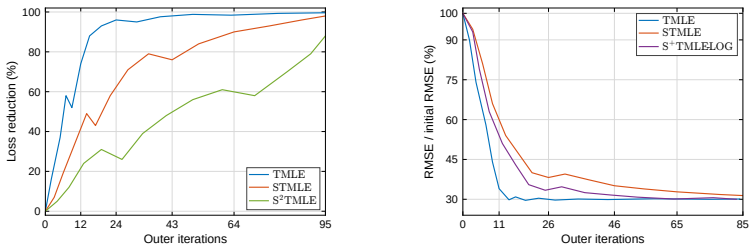
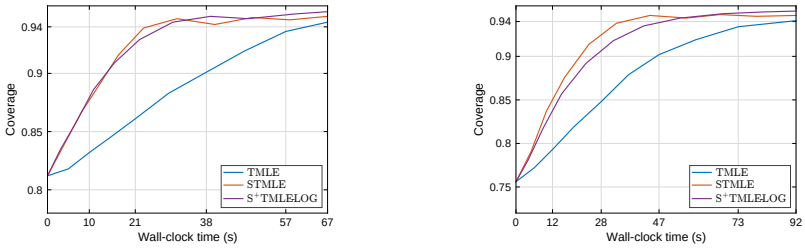


Figure 5: Analysis of convergence time for STMLE and S^{2}TMLE with respect to the algorithmic hyperparameters B and S. The dashed line at the top of each panel marks the degenerate regime attained when B (or S) is set to its maximum value. In the bottom two panels, the batch size is fixed at $B = 0.5n$ to isolate the effect of the number of splits S.



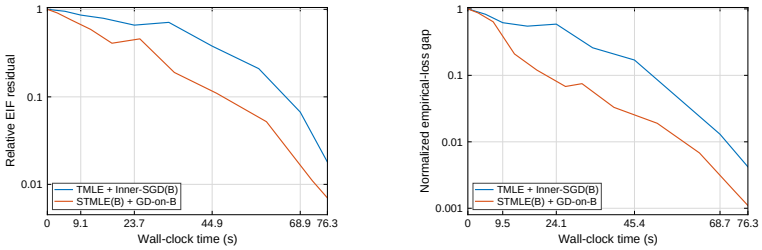
(a) Full-sample loss reduction (b) Relative target RMSE

Figure 6: Additional iteration-based diagnostics under the original simulation setup. Panel (a) shows the reduction in full-sample loss for $\Psi_{(2)}$ on a small set of DGP1, comparing TMLE, STMLE, and S^{2}TMLE. Panel (b) shows target RMSE for $\Psi_{(3)}$, comparing TMLE, STMLE, and S$^+$TMLE-LOG.



(a) $\Psi_{(3)}$ on DGP1: empirical coverage (b) $\Psi_{(4)}$ on DGP1: empirical coverage

Figure 7: Coverage trajectories for nominal 95% Wald confidence intervals for the two scalar targets. Panel (a) reports empirical coverage for $\Psi_{(3)}$ on a small set of DGP1 and panel (b) reports empirical coverage for $\Psi_{(4)}$, each as a function of wall-clock time.



(a) Relative EIF residual (b) Normalized empirical-loss gap

Figure 8: Compute-matched comparison for $\Psi_{(2)}$ on DGP1 with $B = 0.5n$. We compare TMLE + Inner-SGD(B) and STMLE(B) + GD-on-B under a common wall-clock budget. Panel (a) shows the relative full-data EIF residual and panel (b) shows the normalized empirical-loss gap. Vertical axes are logarithmic.

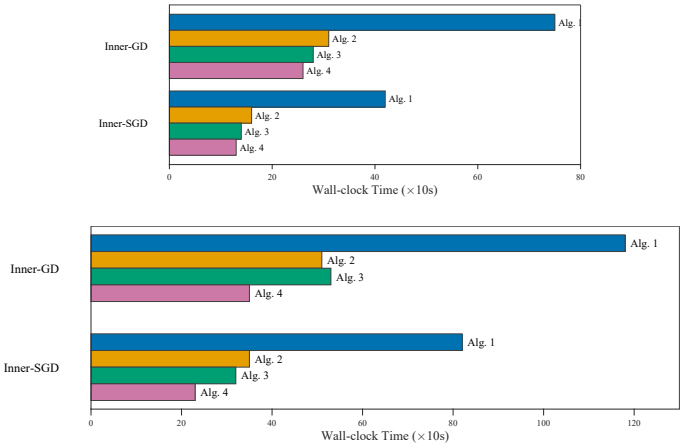


Figure 9: Ablation study separating the contributions of stochastic targeting from stochastic inner solvers. The top and bottom panels correspond to $\Psi_{(1)}$ and $\Psi_{(2)}$, respectively. For methods where applicable, we set $B = 0.25n$ and $S = 24$. GD refers to (full-batch) gradient descent, i.e., the inner solver evaluates gradients over the entire sample.