A NEW APPROACH TO BILEVEL OPTIMIZATION: A VALUE-CENTRIC PERSPECTIVE VIA A DISTRIBUTIONAL FOLLOWER

Anonymous authors
Paper under double-blind review

ABSTRACT

Classical bilevel optimization represents the lower-level response by a selected minimizer $\theta^\star(v)$. This optimizer-centered formulation can be unstable in overparameterized or multimodal settings. Moreover, when multiple lower-level minimizers exist, it requires a response-selection rule that may be difficult to justify in practice. We propose an *optimal-value-centered* formulation that replaces the selected minimizer by a Gibbs distribution over lower-level solutions. Equivalently, this distribution solves an entropy-regularized lower-level problem over probability measures, while the upper level minimizes the expected outer objective. This representation motivates Value-Centric Hypergradient Descent (VCHD) algorithm, a double-loop method that samples from the decision-dependent Gibbs follower using mini-batch SGLD and updates the outer variable with a sample-based hypergradient estimator. We establish nonasymptotic convergence without assuming lower-level convexity and show that VCHD reaches an ε-stationary point with $\mathcal{O}(\varepsilon^{-3})$ sample complexity. We conducted a thorough study of the empirical performance of the VCHD algorithm: a quadratic example examines how temperature and inner-loop length affect its performance, while a double-well example shows that VCHD can approach the correct outer solution even when the lower-level landscape is nonconvex and contains local minima. Results on the hyper-data-cleaning application further suggest that our optimal-value-centered hypergradient can preserve informative update signals when classical bilevel methods suffer from overfitting and vanishing hypergradients.

1 INTRODUCTION

Many learning and decision-making problems naturally exhibit a bilevel structure, including hyperparameter optimization (Franceschi et al., 2018), meta-learning (Rajeswaran et al., 2019), and neural architecture search (Liu et al., 2019). Such problems are commonly formulated as bilevel optimization problems. Specifically, let $\mathcal{S}(v) := \arg\min_{\theta\in\mathbb{R}^d} \ell(v,\theta)$ denote the lower-level solution set. The classical optimistic and pessimistic bilevel formulations are

$$\min_{v\in\mathcal{V}\subseteq\mathbb{R}^p} \Phi_{\mathrm{opt}}(v) := \min_{v\in\mathcal{V}} \min_{\theta\in\mathcal{S}(v)} f(v,\theta), \quad \text{and} \quad \min_{v\in\mathcal{V}\subseteq\mathbb{R}^p} \Phi_{\mathrm{pes}}(v) := \min_{v\in\mathcal{V}} \max_{\theta\in\mathcal{S}(v)} f(v,\theta). \tag{1}$$

When the lower-level problem has multiple minimizers, classical optimistic and pessimistic formulations select the minimizer with the best or worst upper-level objective value, respectively. Both formulations reduce the entire lower-level solution set to a single representative determined by the upper-level objective. We refer to this perspective as **optimizer-centered** modeling.

Rather than selecting a single minimizer, **optimal-value-centered** modeling represents the lower-level response as a distribution that favors smaller lower-level objective values and thus concentrates near the solution set. This optimal-value-centered view is suitable for many practical settings in which optimizer-centered modeling faces the following challenges:

(1) Modeling perspective. In modern machine learning, lower-level problems often involve heavily overparameterized models and may therefore admit a large collection of globally optimal or near-optimal solutions (Vicol et al., 2022; Masiha et al., 2025). In such cases, a single-valued selection $\theta^\star(v) \in \mathcal{S}(v)$ may be unstable: a small perturbation of the upper-level variable can cause the selected

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

optimizer to switch abruptly, leading to non-smoothness or discontinuity of the upper-level objective (Kleniati & Adjiman, 2014; Masiha et al., 2026). By contrast, the lower-level optimal value and near-optimal solution set are often more stable (Royset, 2026), and many applications require only a sufficiently small objective value rather than a particular optimizer (Besançon et al., 2024). These observations motivate an **optimal-value-centered** viewpoint that represents multiple lower-level solutions according to their objective values instead of selecting a single representative. *(2) Free from upper-level response-selection assumption.* When the lower-level minimizer is nonunique, an optimizer-centered formulation must specify which solution the follower returns. Optimistic and pessimistic formulations resolve this ambiguity by selecting the response that is, respectively, most or least favorable to the upper-level objective (Dempe, 2002). These formulations therefore impose a prior favorable or unfavorable selection assumption from the leader's perspective (Besançon et al., 2024). In many applications, such assumptions may be unjustified when the leader neither knows nor controls the follower's selection mechanism. Our optimal-value-centered formulation instead defines a response distribution from lower-level objective values, yielding a well-defined outer objective without imposing a favorable or adversarial selection rule.

We model the inner response by a distribution π_v that favors parameters with smaller lower-level objective values and optimize the expected outer performance under this distribution. Specifically, we consider the following **value-centric** objective based on the Gibbs distribution:

$$J_\tau(v) := \mathbb{E}_{\theta\sim\pi_v}\big[f(v,\theta)\big], \qquad \pi_v(d\theta) := \frac{1}{Z_\tau(v)}\exp\!\left(-\frac{\ell(v,\theta)}{\tau}\right)d\theta,$$

where $Z_\tau(v)$ is the normalizing constant and $\tau > 0$ controls the concentration of π_v around the lower-level solution set. Under mild regularity conditions, J_τ admits the explicit expectation-covariance hypergradient representation

$$\nabla J_\tau(v) = \mathbb{E}_{\pi_v}\big[\nabla_v f(v,\theta)\big] - \frac{1}{\tau}\,\mathrm{Cov}_{\pi_v}\big(f(v,\theta), \nabla_v \ell(v,\theta)\big).$$

This representation naturally motivates our double-loop *Value-Centric Hypergradient Descent* (VCHD) algorithm, which combines mini-batch SGLD sampling with hypergradient estimation. This representation also suggests a Hessian-free approximation to the classical implicit-differentiation hypergradient. Under the classical strongly convex setting, it recovers the latter as a special case: as $\tau \to 0$, the expectation-covariance hypergradient converges to the implicit-differentiation hypergradient, as formalized in Proposition 4.7.

This shift from a single-point optimizer response to a distributional response offers an alternative route for handling nonconvex lower-level problems. Existing approaches often impose a global PL condition (Chen et al., 2024; Huang, 2023) or use a value-function reformulation, as in BOME (Liu et al., 2022). In contrast, our framework accommodates a broad class of lower-level landscapes that arise in many applications. The objective may be arbitrarily nonconvex within a bounded region and need only be strongly convex in the tails, as commonly assumed in analyses of nonconvex Langevin algorithms (Cheng et al., 2018; Li et al., 2024). Moreover, when the lower-level solution sets are nonempty and uniformly bounded, we show that this condition can be enforced through a smooth tail regularizer without changing the solutions of the original bilevel problem.

Our main contributions are: (1) We introduce an **optimal-value-centered** bilevel formulation that represents the lower-level response by a Gibbs distribution rather than by a single optimizer. This formulation is particularly suited to bilevel problems with multimodal and nonconvex lower-level landscapes. (2) We derive an exact expectation-covariance hypergradient representation and develop a computationally efficient double-loop algorithm, termed *Value-Centric Hypergradient Descent* (VCHD). We further establish a nonasymptotic convergence guarantee showing that VCHD reaches an ε-stationary point with $\mathcal{O}(\varepsilon^{-3})$ sample complexity. (3) We evaluate VCHD across three experiments: a controlled quadratic problem characterizes the effects of temperature and inner-loop length, a double-well example demonstrates robustness to lower-level local minima, and hyper-data cleaning shows that the value-centric hypergradient remains informative when classical methods suffer from overfitting and vanishing hypergradients.

2 Related Work

Classical bilevel optimization. Classical gradient-based methods typically assume lower-level strong convexity, yielding a unique response and an implicit-differentiation hypergradient that can be

approximated efficiently (Ghadimi & Wang, 2018; Ji et al., 2021; Arbel & Mairal, 2022; Dagréou et al., 2022; Kong et al., 2026). Recent extensions instead impose a global PL condition (Chen et al., 2024; Huang, 2023) or adopt value-function reformulations (Liu et al., 2021; 2022); the former excludes the cases with suboptimal stationary points, while the latter requires approximating a generally unavailable and potentially nonsmooth value function.

Sampling-based minimizer selection. Gibbs measures and Langevin dynamics are classical tools for global optimization (Raginsky et al., 2017; Dalalyan, 2017). Existing bilevel methods exploiting this principle nevertheless remain optimizer-centered: Masiha et al. (2025) use Gibbs samples for extremal selection over a non-singleton minimizer set, while Masiha et al. (2026) combine best-of-N selection with a pseudoinverse-based hypergradient. Consensus-based particle method Chao et al. (2026) combine lower-level quantile selection with upper-level Gibbs weighting, but analyze a fixed lower-level solution set rather than one that varies with the outer decision.

Gibbs-averaged bilevel optimization. The closest work, Kim & Hospedales (2025), considers essentially the same Gibbs-averaged objective and samples the Gibbs follower using SGLD. HPO-SGLD approximates the hypergradient by differentiating through the SGLD trajectory (Franceschi et al., 2017; 2018), whereas we directly differentiate the stationary Gibbs expectation, obtaining an exact expectation–covariance hypergradient. Moreover, while their analysis assumes an infinitesimal stepsize and exact post-burn-in Gibbs samples, ours accounts for finite-stepsize bias, target variation across outer iterations, and temporal dependence along the SGLD trajectory.

A unifying view. Our formulation is connected to three established lines of work. (A) Viewed over the space of probability measures, our formulation can be considered as a bilevel problem with an entropy-regularized lower level. Regularization has been used to improve the stability of bilevel problems (Lignola & Morgan, 1997), while regularized prediction maps and stochastic perturbations provide smooth alternatives to hard optimizer mappings (Blondel et al., 2020; Berthet et al., 2020). A related distributional variational structure appears in free-energy and marginal-likelihood formulations, which optimize over a distribution through an energy–entropy objective but use the resulting optimal value as the parameter objective (Neal & Hinton, 1998; Kuntz et al., 2023; Akyildiz et al., 2025). (B) On finite action spaces, Gibbs responses reduce to logit choice rules. Such randomized responses underlie quantal-response equilibria, rational-inattention models, security games, and stochastic traffic assignment (McKelvey & Palfrey, 1995; Matějka & McKay, 2015; Fosgerau et al., 2020; Yang et al., 2013; Fisk, 1980). (C) Bayesian local-sensitivity analysis differentiates posterior expectations with respect to model or prior perturbations at a fixed baseline, leading to closely related covariance and linear-response formulas (Gustafson, 1994; Giordano et al., 2018). (D) The expectation-covariance hypergradient identity has classical antecedents in statistical-mechanical linear-response theory and exponential-family analysis (Kubo, 1957; Brown, 1986). Our contribution lies in embedding this identity into a general bilevel optimization framework with a potentially nonconvex and multimodal Gibbs follower. Building on this formulation, we develop a sample-based hypergradient algorithm and establish an nonasymptotic convergence guarantee.

An extended discussion of related work is provided in Appendix C.

3 Value-Centric Bilevel Optimization

This section presents our value-centric bilevel formulation, which models the lower-level response as a Gibbs distribution and minimizes the expected outer objective. Under mild regularity conditions, we derive an exact expectation–covariance hypergradient identity, which motivates Value-Centric Hypergradient Descent (VCHD), a double-loop algorithm that estimates the hypergradient from inner-loop SGLD samples and then uses it to update the outer variable.

3.1 Problem Formulation

Classical bilevel formulations represent the lower-level response by a selected optimizer. In contrast, we adopt an optimal-value-centered distributional response that assigns greater probability to variables with smaller lower-level objective values. Given the lower-level objective $\ell(v,\theta) := \mathbb{E}_{\xi\sim q}\big[l(v,\theta;\xi)\big]$, we define its Gibbs follower at temperature $\tau > 0$ by $\pi_v(\theta) := \frac{1}{Z(v)}\exp(-\ell(v,\theta)/\tau)$, where

$Z(v) := \int_{\mathbb{R}^d} \exp\big(-\frac{1}{\tau}\ell(v,\theta)\big)\mathrm{d}\theta$. Then, the **value-centric** outer objective is

$$J_\tau(v) := \mathbb{E}_{\theta\sim\pi_v}\big[f(v,\theta)\big]. \tag{2}$$

Equivalently, (2) can be viewed as a bilevel problem whose lower-level variable is a probability distribution and whose lower-level objective is entropy regularized. Specifically, for an absolutely continuous probability measure $\rho(d\theta) = r_\rho(\theta)d\theta$, define its differential entropy as $\mathcal{H}(\rho) := -\int_{\mathbb{R}^d} r_\rho(\theta)\log r_\rho(\theta)d\theta$. Under mild conditions, π_v has finite expected energy $E_\rho[\ell(v,\cdot)]$ and differential entropy $\mathcal{H}(\rho)$. The Gibbs variational principle (Dupuis & Ellis, 2011) then implies that π_v is the unique minimizer of $\rho \mapsto \mathbb{E}_\rho[\ell(v,\cdot)] - \tau\mathcal{H}(\rho)$ over all such probability measures for which the objective is well defined and finite.

Under mild regularity conditions, J_τ admits an explicit expectation-covariance hypergradient:

$$\nabla_v J_\tau(v) = \mathbb{E}_{\pi_v}\big[\nabla_v f(v,\theta)\big] - \frac{1}{\tau}\,\mathrm{Cov}_{\pi_v}\big(f(v,\theta), \nabla_v\ell(v,\theta)\big), \tag{3}$$

where the precise conditions and formal statement are provided in Proposition 4.7. Because both terms can be estimated from samples drawn from the Gibbs follower π_v, this identity naturally motivates the Value-Centric Hypergradient Descent algorithm introduced next.

3.2 VALUE-CENTRIC HYPERGRADIENT DESCENT ALGORITHM (VCHD)

We propose a double-loop algorithm that combines mini-batch SGLD sampling with a expectation-covariance hypergradient estimation, see Algorithm 1.

For each outer iteration, the inner loop applies SGLD (Welling & Teh, 2011) to generate approximate samples from the target Gibbs distribution: $\theta_{k,t+1} = \theta_{k,t} - h\nabla_\theta l(v_k,\theta_{k,t};\xi_{k,t}) + \sqrt{2\tau h}\varepsilon_{k,t+1}$, where the mini-batches $\xi_{k,t}$ are drawn i.i.d. from q, the noise vectors satisfy $\varepsilon_{k,t+1}\sim\mathcal{N}(0,I_d)$, and the two sequences are mutually independent.

The outer update requires estimates of $\mathbb{E}_{\pi_v}[\nabla_v f]$ and $\mathrm{Cov}_{\pi_v}(f,\nabla_v\ell)$, which we construct from the final S inner iterates. For any function $\varphi(v,\theta)$ with an unbiased stochastic estimator $\widehat{\varphi}(v,\theta;\xi)$, define the retained-sample average $\widehat{\mathbb{E}}_k[\varphi] := \frac{1}{S}\sum_{t=T-S+1}^{T}\widehat{\varphi}(v_k,\theta_{k,t};\widetilde{\xi}_{k,t})$. The fresh sampled mini-batches $\widetilde{\xi}_{k,t}$ are independent of the mini-batches $\xi_{k,t}$ and Gaussian noises $\varepsilon_{k,t}$ used to generate the SGLD trajectory. For notational convenience, suppress the arguments (v,θ) and define $\varphi_1 := f$, $\varphi_2 := \nabla_v f$, $\varphi_3 := \nabla_v\ell$, and $\varphi_4 := f\nabla_v\ell$. Their unbiased stochastic counterparts are $\widehat{\varphi}_3 := \nabla_v l$ and $\widehat{\varphi}_4 := f\nabla_v l$ (since f and $\nabla_v f$ are evaluated exactly, we set $\widehat{\varphi}_1 := \varphi_1$ and $\widehat{\varphi}_2 := \varphi_2$). We estimate the covariance term by

$$\widehat{g}_k = \widehat{\mathbb{E}}_k[\varphi_2] - \frac{1}{\tau}\widehat{\mathbb{E}}_k[\varphi_4] + \frac{1}{\tau}\widehat{\mathbb{E}}_k[\varphi_1]\widehat{\mathbb{E}}_k[\varphi_3]. \tag{4}$$

For the initialization, $\theta_{k,0} \sim \mu_0$ is resampled at each outer iteration, independently of the algorithmic history and all randomness from current iteration.

4 THEORETICAL ANALYSIS OF VCHD UNDER NON-CONVEX SETTING

4.1 ASSUMPTIONS

We first state the assumptions on the inner-level objective. In addition to standard regularity conditions, we assume strong convexity only outside a compact region, thereby relaxing the requirement of global convexity. We further show that, for nonempty and uniformly bounded lower-level solution sets, exterior strong convexity can be enforced without changing the original problem solution.

Assumption 4.1. The population inner objective $\ell(\cdot,\cdot)$ and the sample loss $l(\cdot,\cdot;\xi)$ are C^2 in (v,θ). Exists constants $R_0,\kappa_0,L_H,G_0,L_\xi > 0$ such that, for every $v,v' \in \mathcal{V}$, $\theta,\theta' \in \mathbb{R}^d$, and ξ, (1) **exterior strong convexity**: $\nabla^2_{\theta\theta}\ell(v,\theta) \succeq \kappa_0 I_d$, $\forall\|\theta\| \geq R_0$; (2) $\|\nabla l(v,\theta;\xi) - \nabla l(v',\theta';\xi)\| \leq L_H\sqrt{\|v-v'\|^2 + \|\theta-\theta'\|^2}$, where $\nabla l = (\nabla_v l, \nabla_\theta l)$; (3) $\sup_{v\in\mathcal{V},\xi}\|\nabla_\theta l(v,0;\xi)\| \leq G_0$; (4) $\mathbb{E}_{\xi\sim q}[\nabla l(v,\theta;\xi)] = \nabla\ell(v,\theta)$; (5) exist a measurable function $\eta : \mathbb{R}^m \to [1,\infty)$ satisfying $\mathbb{E}_q[(1 + \|\xi\|)^4\eta(\xi)^4] < \infty$, such that $\|\nabla_\theta l(v,\theta;\xi) - \nabla_\theta l(v,\theta;\xi')\| \leq L_\xi(\eta(\xi)+\eta(\xi'))(1+\|\theta\|)\|\xi-\xi'\|$.

Algorithm 1 Value-Centric Hypergradient Descent (VCHD)

Input: Outer and inner iteration number K, T, initial outer variable v_0, initialization law μ_0, temperature τ, outer and inner stepsize β, h, and averaging window S
for $k = 0, 1, \ldots, K-1$ **do**
  Independently draw $\theta_{k,0} \sim \mu_0$, initialize $\widehat{\mathbb{E}}_k[\varphi_i] \leftarrow 0$ for $i = 1, 2, 3, 4$ and $s \leftarrow 0$
  for $t = 0, 1, \ldots, T-1$ **do**
    Draw iid $\xi_{k,t} \sim q$ and $\varepsilon_{k,t+1} \sim \mathcal{N}(0, I_d)$
    $\theta_{k,t+1} \leftarrow \theta_{k,t} - h\nabla_\theta l(v_k, \theta_{k,t}; \xi_{k,t}) + \sqrt{2\tau h}\,\varepsilon_{k,t+1}$
    if $t \geq T - S$ **then**
      $s \leftarrow s + 1$ and independently draw $\widetilde{\xi}_{k,t+1}$
      $\widehat{\mathbb{E}}_k[\varphi_i] \leftarrow \frac{s-1}{s}\widehat{\mathbb{E}}_k[\varphi_i] + \frac{1}{s}\widehat{\varphi}_i(v_k, \theta_{k,t+1}; \widetilde{\xi}_{k,t+1})$ for $i = 1, 2, 3, 4$
    end if
  end for
  $\widehat{g}_k \leftarrow \widehat{\mathbb{E}}_k[\varphi_2] - \frac{1}{\tau}\big(\widehat{\mathbb{E}}_k[\varphi_4] - \widehat{\mathbb{E}}_k[\varphi_1]\widehat{\mathbb{E}}_k[\varphi_3]\big)$
  $v_{k+1} \leftarrow \Pi_{\mathcal{V}}(v_k - \beta\widehat{g}_k)$
end for

Remark 4.2. Exterior strong convexity condition, together with joint L_H-smoothness, implies the distant monotonicity condition used for the SGLD contraction result. In particular, with $R := \max\{4R_0(\kappa_0 + L_H)/\kappa_0, 3\}$ and $\kappa := \kappa_0/2$ we have $(\theta_1 - \theta_2)^\top\big(\nabla_\theta\ell(v, \theta_1) - \nabla_\theta\ell(v, \theta_2)\big) \geq \kappa\|\theta_1 - \theta_2\|^2$, for $\|\theta_1 - \theta_2\| > R$.

We next show that exterior strong convexity can be imposed without changing the solution of the original bilevel problem. Consider the bilevel optimization problem in (1).

Proposition 4.3. *Suppose that for every $\xi \in \mathbb{R}^m$, $\ell(\cdot, \cdot)$ and $l(\cdot, \cdot; \xi)$ are C^2 and $S(v) := \arg\min_{\theta \in \mathbb{R}^d} \ell(v, \theta)$ is nonempty and there exists $R_\star > 0$ such that $\bigcup_{v \in \mathcal{V}} S(v) \subset B(0, R_\star)$. Choose $R_\star < R_1 < R_0$ and define*

$$\omega(t) := \begin{cases} 0, & t \leq 0, \\ e^{-1/t}, & t > 0, \end{cases} \qquad \sigma(t) := \frac{\omega(t)}{\omega(t) + \omega(1-t)}, \qquad \chi(s) := \frac{1}{2}\sigma\left(\frac{s - R_1^2}{R_0^2 - R_1^2}\right).$$

Let $\rho_\gamma(\theta) := \gamma\Psi(\|\theta\|^2)$, where $\Psi(s) := \int_0^s \chi(u)\,du$ and constant $\gamma > 0$. Define the regularized loss $\ell_\gamma(v, \theta) := \ell(v, \theta) + \rho_\gamma(\theta)$ and $l_\gamma(v, \theta; \xi) := l(v, \theta; \xi) + \rho_\gamma(\theta)$.

Suppose that all conditions in Assumption 4.1, except exterior strong convexity, hold with constants L_H, G_0 and L_ξ and function η. Let $\gamma = L_H + \kappa_0$. Then, the regularized objective ℓ_γ preserves the lower-level solution set for every $v \in \mathcal{V}$, $\arg\min_{\theta \in \mathbb{R}^d} \ell_\gamma(v, \theta) = S(v)$, and consequently the outer solution set of the original bilevel problem. Moreover, ℓ_γ and l_γ satisfy all conditions in Assumption 4.1 with constant κ_0, $\bar{L}_H$, G_0 and L_ξ and function η, where $\bar{L}_H$ is defined in (7).

Exterior strong convexity yields two key ingredients for our analysis: (1) uniform dissipativity underpins SGLD contraction and moment control; (2) a uniform quadratic lower bound on the inner objective guarantees the existence and regularity of the Gibbs distribution.

Lemma 4.4. *Suppose Assumption 4.1 holds and $\mathcal{V}$ is nonempty and compact. Then, for every $v \in \mathcal{V}$ and $\theta \in \mathbb{R}^d$, $\langle\nabla_\theta\ell(v, \theta), \theta\rangle \geq a\|\theta\|^2 - b$ and $\ell(v, \theta) \geq \underline{a}\|\theta\|^2 - b_\ell$, where constant a, b, $\underline{a}$ and b_ℓ are defined in* (10) *and* (13).

This quadratic lower bound ensures that the Gibbs follower is well defined. Indeed, for every $v \in \mathcal{V}$, $0 < Z(v) = \int_{\mathbb{R}^d} e^{-\ell(v,\theta)/\tau}\,d\theta \leq e^{b_\ell/\tau}(\pi\tau/\underline{a})^{d/2} < \infty$.

We next impose unbiasedness and moment conditions on the stochastic gradient noise arising in the SGLD updates and hypergradient estimation.

Assumption 4.5. Let $\zeta_\theta(v, \theta; \xi) := \nabla_\theta l(v, \theta; \xi) - \nabla_\theta\ell(v, \theta)$ and $\zeta_v(v, \theta; \xi) := \nabla_v l(v, \theta; \xi) - \nabla_v\ell(v, \theta)$ denote the gradient noises. Then, we have $\mathbb{E}[\zeta_\theta(v, \theta; \xi) \mid v, \theta] = 0$ and $\mathbb{E}[\zeta_v(v, \theta; \xi) \mid v, \theta] = 0$. Also, exist constant $\sigma_{\theta,2}, \sigma_{v,2}, \sigma_{\theta,4}, \sigma_{v,4}$ such that $\mathbb{E}[\|\zeta_\theta(v, \theta; \xi)\|^2 \mid v, \theta] \leq \sigma_{\theta,2}^2 d$, $\mathbb{E}[\|\zeta_v(v, \theta; \xi)\|^2 \mid v, \theta] \leq \sigma_{v,2}^2 p$, $\mathbb{E}[\|\zeta_\theta(v, \theta; \xi)\|^4 \mid v, \theta] \leq \sigma_{\theta,4}^4 d^2$, $\mathbb{E}[\|\zeta_v(v, \theta; \xi)\|^4 \mid v, \theta] \leq \sigma_{v,4}^4 p^2$.

Having specified the conditions on the inner objective and Gibbs follower, we complete the problem setup with assumptions on the outer feasible set and objective.

Assumption 4.6. The set $\mathcal{V} \subset \mathbb{R}^p$ is compact and convex. f is continuously differentiable in (v, θ). Exist constants $L_f, L_{f_v}, L_{f\ell_v}, C_f > 0$ such that, for every $v \in \mathcal{V}$ and $\theta, \theta' \in \mathbb{R}^d$, we have $|f(v,\theta) - f(v,\theta')| \le L_f\|\theta - \theta'\|$, $\|\nabla_v f(v,\theta) - \nabla_v f(v,\theta')\| \le L_{f_v}\|\theta - \theta'\|$, $|f(v,\theta)| \le C_f$ and $\|f(v,\theta)\nabla_v \ell(v,\theta) - f(v,\theta')\nabla_v \ell(v,\theta')\| \le L_{f\ell_v}\|\theta - \theta'\|$.

Next, we show that J_τ is finite and continuously differentiable and derive its exact expectation–covariance hypergradient.

Proposition 4.7. *Suppose Assumptions 4.1 and 4.6 hold, and fix $\tau > 0$. Then, $J_\tau(v)$ is continuous on $\mathcal{V}$ and continuously differentiable on $v \in \operatorname{int}(\mathcal{V})$, with $\nabla_v J_\tau(v) = \mathbb{E}_{\pi_v}[\nabla_v f(v,\theta)] - \frac{1}{\tau}\operatorname{Cov}_{\pi_v}\big(f(v,\theta), \nabla_v \ell(v,\theta)\big)$.*

When the exterior strong convexity condition is replaced by global strongly-convexity condition $\nabla^2_{\theta\theta}\ell(v,\theta) \succeq \kappa_0 I_d$, $\forall v \in \mathcal{V}$, $\theta \in \mathbb{R}^d$, let $\theta^\star(v)$ denote the unique minimizer of $\ell(v,\cdot)$ and define $\Phi(v) := f(v,\theta^\star(v))$, we have that, for every $v \in \operatorname{int}(\mathcal{V})$, $\lim_{\tau\downarrow 0}\nabla_v J_\tau(v) = \nabla\Phi(v) = \nabla_v f(v,\theta^\star(v)) - \nabla^2_{v\theta}\ell(v,\theta^\star(v))\big[\nabla^2_{\theta\theta}\ell(v,\theta^\star(v))\big]^{-1}\nabla_\theta f(v,\theta^\star(v))$.

For the outer-loop analysis, we assume that J_τ is smooth and bounded below, with boundary gradients defined by the continuous extension of the expectation–covariance formula.

Assumption 4.8. The function J_τ is L_J-smooth on $\mathcal{V}$, and $J^\star_\tau := \inf_{v\in\mathcal{V}} J_\tau(v) > -\infty$.

Finally, to control the initial transient and moment growth of the inner SGLD chain, we assume that the initialization distribution μ_0 has finite second and fourth moments.

Assumption 4.9. For the initialization distribution μ_0, there exist constants $D_{2,0}, M_{4,0} < \infty$ such that $\int_{\mathbb{R}^d}\|\theta\|^2\mu_0(d\theta) \le D_{2,0}$, $\int_{\mathbb{R}^d}\|\theta\|^4\mu_0(d\theta) \le M_{4,0}$.

4.2 Main convergence result

This subsection presents the main convergence theorem for VCHD, establishing an $\mathcal{O}(\varepsilon^{-3})$ sample complexity for reaching an ε-stationary point.

Theorem 4.10. *Suppose Assumptions 4.1, 4.6, 4.5, 4.8, and 4.9 hold. Fix $\tau > 0$, let $1 \le S \le T$, choose outer stepsize $0 < \beta \le (2L_J)^{-1}$ and inner stepsize $h > 0$, satisfying* (42)*,* (59)*,* (60) *and $h \le h_4$, where h_4 is defined in* (53)*. Then, for every $K \ge 1$,*

$$\frac{1}{K}\sum_{k=0}^{K-1}\mathbb{E}\|\mathcal{G}_\beta(v_k)\|^2 = O\left(\frac{1}{K} + \frac{1}{S} + \frac{1}{Sh} + h\right)$$

where $\mathcal{G}_\beta(v) := \frac{1}{\beta}\big(v - \Pi_{\mathcal{V}}(v - \beta\nabla J_\tau(v))\big)$. Equivalently, by choosing $h = \Theta(\epsilon)$, $K = \Theta(\epsilon^{-1})$, $S = \Theta(\epsilon^{-2})$, and $T = S + \mathcal{O}(1)$, VCHD achieves an ϵ-stationary point, with overall sample complexity $\Theta(K(T+S)) = \Theta(\epsilon^{-3})$.

The $1/K$ and $1/S$ terms alone would yield an $\mathcal{O}(\epsilon^{-2})$ oracle complexity. The additional factor of ϵ^{-1} arises from temporal dependence within each SGLD trajectory: the variance scales as $\mathcal{O}(1/(Sh))$, while finite-step discretization introduces an $\mathcal{O}(h)$ bias. If independent samples from sufficiently mixed SGLD chains were available at constant cost, the variance would scale as $\mathcal{O}(1/S)$, recovering an $\mathcal{O}(\epsilon^{-2})$ oracle complexity.

For completeness, we also establish convergence using the law-level warm-start initialization $\mu_{k,0} = \mu_{k-1,T}$, see Appendix A.8 for the details.

4.3 Proof sketch

This subsection sketches the proof of Theorem 4.10. Compared with the commonly assumed strongly convex inner problem, the main difficulty is that, in a general nonconvex setting, the inner iterates no longer converge geometrically to the inner solution. We therefore work at the distribution level: although pointwise contraction may fail, the laws of the SGLD iterates contract geometrically in Wasserstein distance. The proof outline proceeds as follows.

Step 1: Relate Convergence to the Hypergradient Error. By the projected descent lemma A.2, we reduce the convergence analysis to controlling the hypergradient error $e_k := \widehat{g}_k - \nabla_v J_\tau(v_k)$. Subtracting the exact hypergradient (3) from its estimator $\widehat{g}_k$ and applying $\|\mathbb{E}_{\pi_{v_k}}[\varphi_3]\| \le C_\ell$ from Remark A.1, and $|\widehat{\mathbb{E}}_k[\varphi_1]| \le C_f$ from Assumption 4.6, we obtain $\mathbb{E}\|e_k\|^2 \le 4\mathbb{E}\|\widehat{\mathbb{E}}_k[\varphi_2] - \mathbb{E}_{\pi_{v_k}}[\varphi_2]\|^2 + \frac{4}{\tau^2}\mathbb{E}\|\widehat{\mathbb{E}}_k[\varphi_4] - \mathbb{E}_{\pi_{v_k}}[\varphi_4]\|^2 + \frac{4C_\ell^2}{\tau^2}\mathbb{E}|\widehat{\mathbb{E}}_k[\varphi_1] - \mathbb{E}_{\pi_{v_k}}[\varphi_1]|^2 + \frac{4C_f^2}{\tau^2}\mathbb{E}\|\widehat{\mathbb{E}}_k[\varphi_3] - \mathbb{E}_{\pi_{v_k}}[\varphi_3]\|^2$. Thus, controlling the hypergradient error reduces to bounding the empirical-average MSE $\mathbb{E}\|\widehat{\mathbb{E}}_k[\varphi_i] - \mathbb{E}_{\pi_{v_k}}[\varphi_i]\|^2$. This MSE can be further decompose into: SGLD distributional bias $\mathbb{E}\big\|\frac{1}{S}\sum_t\big(\mathbb{E}_{\mu_{k,t}}\varphi_i - \mathbb{E}_{\pi_{v_k}}\varphi_i\big)\big\|^2$ and trajectory sampling error $\mathbb{E}\big\|\frac{1}{S}\sum_t\big(\widehat{\varphi}_i(v_k,\theta_{k,t};\widetilde{\xi}_{k,t}) - \mathbb{E}_{\mu_{k,t}}\varphi_i\big)\big\|^2$, where $\mu_{k,t} := \mathrm{Law}(\theta_{k,t})$.

Step 2: Control the SGLD Distributional Bias. By the Lipschitz conditions in Assumptions 4.1 and 4.6, and the Kantorovich–Rubinstein duality, $\mathbb{E}\|\frac{1}{S}\sum_t(\mathbb{E}_{\mu_{k,t}}[\varphi_i] - \mathbb{E}_{\pi_{v_k}}[\varphi_i])\|^2 \le L_{\varphi_i}^2/S\sum_t \mathbb{E}W_1(\mu_{k,t},\pi_{v_k})^2$. To bound $W_1(\mu_{k,t},\pi_{v_k})$, the difficulty is that, for a fixed stepsize h, the invariant law $\bar{\pi}_{v_k,h}$ of SGLD is generally differs from the Gibbs distribution π_{v_k}. Lemma A.10 separates the $W_1(\mu_{k,t},\pi_{v_k})$ into finite-time mixing error $W_1(\mu_{v_k},\bar{\pi}_{v_k,h})$ and invariant-law bias $W_1(\bar{\pi}_{v_k,h},\pi_{v_k})$. The former decays exponentially by contraction of the SGLD kernel, while the latter is $O(\sqrt{h})$, obtained by coupling SGLD initialized from π_{v_k} with the stationary Langevin diffusion over $O(h^{-1})$ steps. Consequently, $W_1(\mu_{k,t},\pi_{v_k}) \lesssim e^{-cht} + \sqrt{h}$, for some constant $c > 0$.

Step 3: Control the Trajectory Sampling Error. We decompose the trajectory sampling error into the centered chain fluctuation $A_{k,t}^{(i)} := \varphi_i(v_k,\theta_{k,t}) - \mathbb{E}_{\mu_{k,t}}[\varphi_i]$ and the mini-batch noise $N_{k,t}^{(i)} := \widehat{\varphi}_i(v_k,\theta_{k,t};\widetilde{\xi}_{k,t}) - \varphi_i(v_k,\theta_{k,t})$. For the mini-batch noise, conditional unbiasedness and bounded variance yield an $\mathcal{O}(\sigma_i^2/S)$ term. The main challenge is the temporal dependence along the SGLD trajectory from the centered SGLD-chain fluctuation term. To control this dependence, we combine the exponential W_1-contraction of the SGLD kernel $Q_{v_k,h}$ with Kantorovich–Rubinstein duality to obtain $\mathrm{Lip}(Q_{v_k,h}^r\varphi_i) \lesssim e^{-chr}L_{\varphi_i}$. Thus, the cross-covariances decay geometrically with the time lag. This further gives $\mathbb{E}\|\frac{1}{S}\sum_t A_{k,t}^{(i)}\|^2 \lesssim \frac{1}{S}(1 + \frac{e^{-ch}}{1-e^{-ch}})$. Together with the mini-batch bound, this yields $\mathbb{E}\|\frac{1}{S}\sum_t(\widehat{\varphi}_i(v_k,\theta_{k,t};\widetilde{\xi}_{k,t}) - \mathbb{E}_{\mu_{k,t}}[\varphi_i])\|^2 \lesssim \frac{1}{S}(1 + \frac{2c_0e^{-ch}}{1-e^{-ch}} + \sigma_i^2)$.

Step 4: Derive the Final Convergence Bound. Combining the above SGLD distributional bias and trajectory-sampling error with the descent lemma completes the proof of Theorem 4.10.

5 Numerical Experiments

In this section, we evaluate VCHD on two toy examples and a hyper-data-cleaning task. The quadratic example uses closed-form ground truth to assess its behavior across temperatures and inner-loop lengths, while the double-well example illustrates its robustness to lower-level local minima. Finally, the hyper-data-cleaning experiment shows that VCHD preserves informative hypergradients in a complex lower-level landscape, whereas the classical bilevel framework may overfit and suffer from vanishing hypergradients. Experimental details, including detailed problem formulation, parameter settings and additional results are provided in Appendix B.

5.1 Quadratic Example with Closed-Form Ground Truth

This controlled quadratic experiment examines how the retained-sample size S, inner-loop length T, and the temperature τ affect VCHD and its hypergradient estimate. We consider the lower-level objective $\ell(v,\theta) := \frac{1}{2}\theta^\top A(v)\theta - b(v)^\top\theta$, where $A(v) \succeq 0$. The outer-level objective is $f(v,\theta) := \frac{1}{2}\|C\theta - y\|^2 + \frac{\gamma}{2}\|v\|^2$, with fixed $C \in \mathbb{R}^{p\times d}$, $y \in \mathbb{R}^p$, and $\gamma > 0$.

The example illustrates two general points. First, as T and S increase, the VCHD trajectory approaches the reference trajectory generated by the exact closed-form hypergradient of $J_\tau(v) = \mathbb{E}_{\theta\sim\pi_v}[f(v,\theta)]$, as shown in Figure 1 (left). This behavior reflects the fact that a longer inner run reduces the mixing error, while a larger averaging window reduces the sample-average estimation error, together improving the accuracy of the estimated hypergradient. Second, Figure 1 (right) illustrates the small-temperature limit: in this strongly convex setting, the exact expectation-covariance hypergradient converges to the implicit-differentiation hypergradient as $\tau \to 0$.

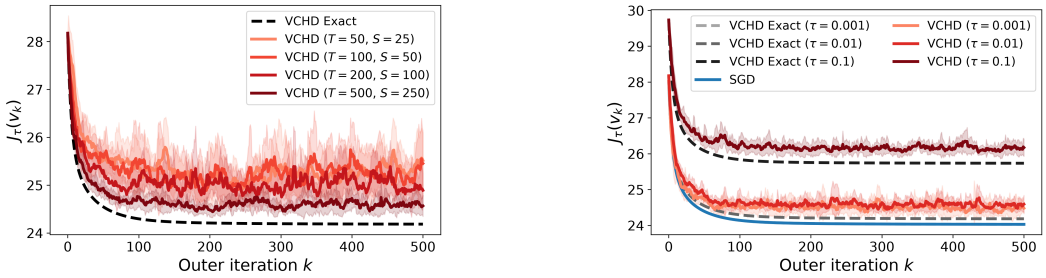

Figure 1: Quadratic Gibbs follower experiment. **(Left)** Outer objective trajectory under different inner steps T. **(Right)** Outer objective trajectory under different τ values.

5.2 Double-well inner loss with a switching global minimum

We use a double-well example to illustrate the value-centric bilevel formulation and demonstrate the advantages of VCHD when the lower-level objective has local minima. The upper- and lower-level losses are $f(v, \theta) = \theta^2 + \frac{\lambda}{2}v^2$ and $\ell(v, \theta) = (\theta^2 - 1)^2 - v\theta$. At $v = 0$, ℓ has two symmetric wells near ± 1; positive v favors the right well, while negative v favors the left.

We compare VCHD with three baselines: *AID-HG* (Ghadimi & Wang, 2018), *HPO-SGLD* (Kim & Hospedales, 2025), and *HG-MS* (Masiha et al., 2026), where the details are in Appendix B.2. As shown in Figure 2 (right), only VCHD and HPO-SGLD approach the optimum $v^\star = 0$, while VCHD has less variation. AID-HG converges to approximately -0.56 and 0.63 for the two seeds. Despite exploring inner problem with N parallel chains, HG-MS converges near $v = 0.58$. The example illustrates some limitations of optimizer-centered methods. In landscapes with disconnected basins, AID-HG and HG-MS can follow a local branch and optimize the wrong hyper-objective, $f(v, \theta_{\text{left}}(v))$ or $f(v, \theta_{\text{right}}(v))$ (Figure 3, left). Sampling does not resolve this issue when used only for point selection: at the low temperatures needed for concentration, HG-MS chains rarely cross the barrier between the wells and Best-of-N may select a nonglobal minimizer. VCHD instead averages both wells with Gibbs weights, producing a smooth J_τ whose covariance hypergradient aggregates both modes. For HPO-SGLD, the results show that the pathwise sensitivity approximation deteriorates as the inner trajectory length increases. In contrast, VCHD directly differentiates the stationary Gibbs expectation, resulting in greater stability.

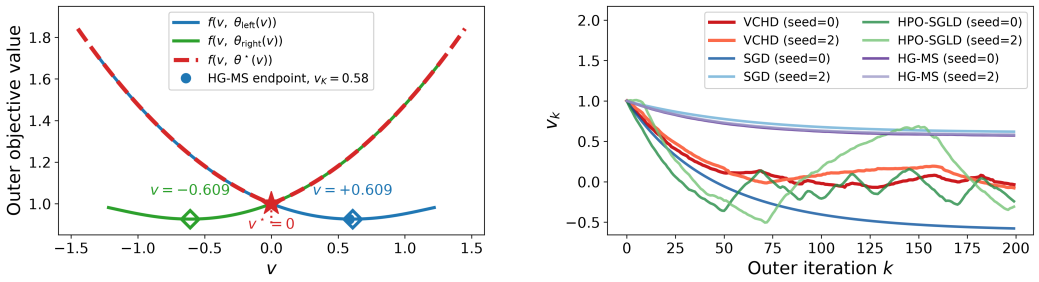

Figure 2: Double-well experiment. **(Left)** Objectives induced by trapping in the left (blue) or right (green) well; the true outer objective is dashed red. **(Right)** Trajectories of the outer iterate v_k.

5.3 Hyper Data Cleaning with a Gibbs Follower

In this subsection, we consider hyper-data cleaning (Franceschi et al., 2017), which uses a clean validation set to learn example-specific weights that downweight mislabeled training samples. Let the classifier's parameter $\theta \in \mathbb{R}^d$ (Neural Network with $d \approx 400$k) be the inner-level variable and $w \in [0, 1]^n$ be the outer variable, where w_i is the weight assigned to the i-th training data. The lower-level objective is the weighted loss evaluated on the corrupted training set: $\ell(w, \theta) := \frac{1}{n}\sum_{i \in I_{\text{train}}} w_i \, \ell_i(\theta) + \frac{\lambda}{2}\|\theta\|_2^2$. The upper-level problem minimizes the empirical loss of the resulting model on the clean validation set: $f(w, \theta) := \frac{1}{m}\sum_{i \in I_{\text{val}}} \ell_i(\theta)$.

We compare VCHD with three baselines: implicit differentiation with a Neumann-series approximation (*Diff+Neumann*) (Lorraine et al., 2020), HPO-SGLD (Kim & Hospedales, 2025), and HG-MS (Masiha et al., 2026). We report two training-set accuracies: *noisy-label accuracy* compares the predictions on the training set against the noisy (corrupted) labels, whereas *clean-label accuracy* compares the same predictions against the clean (uncorrupted) labels. Figure 3 (left) shows that the classical Diff+Neumann method enters an overfitting regime: it attains a high noisy-label accuracy of about 0.9, indicating that it memorizes the corrupted training labels, while its hypergradient vanishes and its clean-label accuracy stalls at around 0.7. In contrast, the Gibbs follower of VCHD keeps its noisy-label accuracy near 0.7, the level consistent with predicting the clean labels rather than memorizing the corrupted ones. And it continues to supply a non-vanishing hypergradient, so that its clean-label accuracy improves steadily (Figure 3, left). The test error of the inner model on the held-out test set follows the same pattern: the best raw test error is $\mathbf{1.55}\%$ for VCHD, compared with 3.42% for HPO-SGLD, 11.09% for Diff+Neumann, and 16.12% for HG-MS.

The preceding results highlight a key difference between the two formulations: the classical hypergradient can vanish under lower-level overfitting, whereas the value-centric hypergradient used by VCHD can remain informative. Under the classical bilevel formulation of the data cleaning problem, the hypergradient takes the form $(\nabla_w F(w))_i = -\frac{1}{n}\langle \nabla_\theta \ell_i^\star, (\nabla^2_{\theta,\theta}\ell^\star)^{-1}\nabla_\theta f^\star\rangle$, which is evaluated at a lower-level minimizer found by the algorithm. As the inner optimization drives the weighted training loss toward zero, the individual gradients may also become negligible, i.e. $\|\nabla_\theta \ell_i(\bar\theta)\| \approx 0$. If the inverse Hessian and the validation-loss gradient remain bounded, the estimated hypergradient will be negligibly small: $\|(\widehat{\nabla} F(w))_i\| \le w_i \big\|\nabla_\theta \ell_i(\bar\theta)\big\|_{\mathrm{op}} \big\|\big(\nabla^2_{\theta\theta}\ell_i(\bar\theta)\big)^{-1}\big\|_{\mathrm{op}} \big\|\nabla_\theta f(w,\bar\theta)\big\| \approx 0$. Consequently, when the lower-level model overfits the training data, the hypergradient carries little useful signal for updating the sample weights or identifying corrupted examples. By contrast, the value-centric formulation gives $(\nabla_w J_\tau(w))_i = -\frac{1}{n\tau}\,\mathrm{Cov}_{\pi_w}(f(w,\theta), \ell_i(\theta))$, this expression depends on loss covariances across a distribution of inner solutions rather than on gradients at a single minimizer. It can therefore remain informative even when some modes memorize the training data and exhibit nearly vanishing gradients. Intuitively, fitting a corrupted example more closely will leads to worse validation performance, and the resulting negative covariance drives its weight downward.

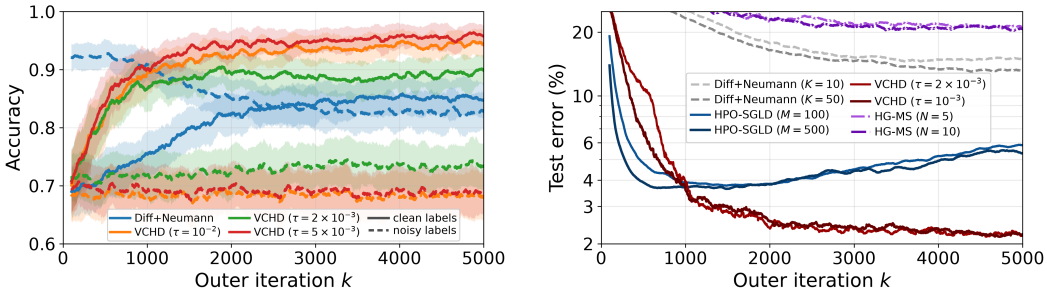

Figure 3: Data-cleaning experiment. **(Left)** Training batch accuracy. **(Right)** Test error.

6 Conclusion

We address a limitation of classical bilevel optimization by shifting from an *optimizer-centered* to an *optimal-value-centered* perspective. Rather than selecting a single lower-level optimizer, we model the lower-level response by the Gibbs distribution $\pi_v(v) \propto \exp(-\ell(v,\theta)/\tau)$ and minimize the expected outer objective $J_\tau(v) = \mathbb{E}_{\theta\sim\pi_v}[f(v,\theta)]$. This formulation naturally accommodates nonconvex and multimodal lower-level landscapes. Under mild regularity conditions, J_τ admits an exact expectation–covariance hypergradient identity. This identity motivates VCHD, a practical double-loop algorithm in which mini-batch SGLD approximates the Gibbs follower and the outer loop uses a plug-in sample-based hypergradient estimator. When the lower-level objective may be nonconvex on a bounded region but is strongly convex in the tails, we establish convergence with $\mathcal{O}(\epsilon^{-3})$ sample complexity. Our experiments show that the Gibbs follower captures multiple modes in nonconvex lower-level landscapes. Especially, in hyper-data cleaning, our formualtion and VCHD retains informative hypergradient updates when deterministic methods suffer from memorization and vanishing hypergradients.

AI USE STATEMENT

Generative AI tools were used to assist with language editing, manuscript organization, LaTeX formatting, and drafting or revising portions of the mathematical exposition, proof arguments, and presentation of experimental results. They were not used to generate the reported datasets or experimental measurements. The authors checked all AI-assisted material against their derivations, source code, experimental records, and cited sources, and take responsibility for the final content of the paper.

REPRODUCIBILITY STATEMENT

The complete assumptions, explicit constants, and proofs are provided in the appendix. The appendix also specifies the experimental problems, data processing, model architectures, optimization schedules, hyperparameters, evaluation protocol, and per-configuration results used to produce the reported comparisons.

REFERENCES

Ömer Deniz Akyildiz, Francesca R. Crucinio, Mark Girolami, Tim Johnston, and Sotirios Sabanis. Interacting particle Langevin algorithm for maximum marginal likelihood estimation. *ESAIM: Probability and Statistics*, 2025. arXiv:2303.13429.

Michael Arbel and Julien Mairal. Amortized implicit differentiation for stochastic bilevel optimization. In *International Conference on Learning Representations*, 2022.

Quentin Berthet, Mathieu Blondel, Olivier Teboul, Marco Cuturi, Jean-Philippe Vert, and Francis Bach. Learning with differentiable perturbed optimizers. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Mathieu Besançon, Miguel F Anjos, and Luce Brotcorne. Robust bilevel optimization for near-optimal lower-level solutions. *Journal of Global Optimization*, 90(4):813–842, 2024.

Mathieu Blondel, André F. T. Martins, and Vlad Niculae. Learning with Fenchel–Young losses. *Journal of Machine Learning Research*, 21(35):1–69, 2020.

Lawrence D. Brown. *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*, volume 9 of *IMS Lecture Notes–Monograph Series*. Institute of Mathematical Statistics, 1986.

Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304, 1995. doi: 10.1214/ss/1177009939.

Yutong Chao, Xudong Sun, Konstantin Riedl, Majid Khadiv, and Jalal Etesami. Convergence of consensus-based particle methods for nonconvex bi-level optimization. *arXiv preprint arXiv:2605.19667*, 2026.

Lesi Chen, Jing Xu, and Jingzhao Zhang. On finding small hyper-gradients in bilevel optimization: Hardness results and improved analysis. In *Proceedings of the Thirty-Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pp. 947–980. PMLR, 2024.

Xiang Cheng, Niladri S. Chatterji, Yasin Abbasi-Yadkori, Peter L. Bartlett, and Michael I. Jordan. Sharp convergence rates for langevin dynamics in the nonconvex setting. *arXiv preprint arXiv:1805.01648*, 2018.

Mathieu Dagréou, Pierre Ablin, Samuel Vaiter, and Thomas Moreau. A framework for bilevel optimization that enables stochastic and global variance reduction algorithms. In *Advances in Neural Information Processing Systems*, volume 35, pp. 26698–26710, 2022.

Arnak Dalalyan. Further and stronger analogy between sampling and optimization: Langevin monte carlo and gradient descent. In *Conference on Learning Theory*, pp. 678–689. PMLR, 2017.

Arnak S Dalalyan and Avetik Karagulyan. User-friendly guarantees for the langevin monte carlo with inaccurate gradient. *Stochastic Processes and their Applications*, 129(12):5278–5311, 2019.

Gary A. Davis. Exact local solution of the continuous network design problem via stochastic user equilibrium assignment. *Transportation Research Part B: Methodological*, 28(1):61–75, 1994.

Stephan Dempe. *Foundations of bilevel programming*. Springer, 2002.

Justin Domke. Generic methods for optimization-based modeling. In *International Conference on Artificial Intelligence and Statistics*, pp. 318–326. PMLR, 2012.

Paul Dupuis and Richard S Ellis. *A weak convergence approach to the theory of large deviations*. John Wiley & Sons, 2011.

Lawrence C Evans. *Measure theory and fine properties of functions*. Chapman and Hall/CRC, 2025.

Caroline Fisk. Some developments in equilibrium traffic assignment. *Transportation Research Part B: Methodological*, 14(3):243–255, 1980.

Gerald B Folland. *Real analysis: modern techniques and their applications*. John Wiley & Sons, 1999.

Mogens Fosgerau, Emerson Melo, André de Palma, and Matthew Shum. Discrete choice and rational inattention: A general equivalence result. *International Economic Review*, 61(4):1569–1589, 2020.

Adam Foster, Martin Jankowiak, Eli Bingham, Paul Horsfall, Yee Whye Teh, Tom Rainforth, and Noah D. Goodman. Variational Bayesian optimal experimental design. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Adam Foster, Desi R Ivanova, Ilyas Malik, and Tom Rainforth. Deep adaptive design: Amortizing sequential bayesian experimental design. In *International Conference on Machine Learning*, pp. 3384–3395. PMLR, 2021.

Luca Franceschi, Michele Donini, Paolo Frasconi, and Massimiliano Pontil. Forward and reverse gradient-based hyperparameter optimization. In *International Conference on Machine Learning*, pp. 1165–1173. PMLR, 2017.

Luca Franceschi, Paolo Frasconi, Saverio Salzo, Riccardo Grazzi, and Massimiliano Pontil. Bilevel programming for hyperparameter optimization and meta-learning. In *International Conference on Machine Learning*, pp. 1568–1577. PMLR, 2018.

Saeed Ghadimi and Mengdi Wang. Approximation methods for bilevel programming. *arXiv preprint arXiv:1802.02246*, 2018.

Ryan Giordano, Tamara Broderick, and Michael I. Jordan. Covariances, robustness, and variational Bayes. *Journal of Machine Learning Research*, 19(51):1–49, 2018.

Paul Antony Gustafson. *Local sensitivity of posterior expectations*. Carnegie Mellon University, 1994.

Feihu Huang. On momentum-based gradient methods for bilevel optimization with nonconvex lower-level. *arXiv preprint arXiv:2303.03944*, 2023.

Kaiyi Ji, Junjie Yang, and Yingbin Liang. Bilevel optimization: Convergence analysis and enhanced design. In *International Conference on Machine Learning*, pp. 4882–4892. PMLR, 2021.

Minyoung Kim and Timothy Hospedales. A stochastic approach to bi-level optimization for hyperparameter optimization and meta learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 17913–17920, 2025.

Polyxeni-Margarita Kleniati and Claire S Adjiman. Branch-and-sandwich: a deterministic global optimization algorithm for optimistic bilevel programming problems. part I: Theoretical development. *Journal of Global Optimization*, 60(3):425–458, 2014.

Deyi Kong, Zaiwei Chen, Shuzhong Zhang, and Shancong Mou. Natural hypergradient descent: Algorithm design, convergence analysis, and parallel implementation. In *International Conference on Machine Learning*, 2026.

Ryogo Kubo. Statistical-mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems. *Journal of the Physical Society of Japan*, 12 (6):570–586, 1957.

Juan Kuntz, Jen Ning Lim, and Adam M Johansen. Particle algorithms for maximum likelihood training of latent variable models. In *International Conference on Artificial Intelligence and Statistics*, pp. 5134–5180. PMLR, 2023.

Lei Li, Jian-Guo Liu, and Yuliang Wang. Geometric ergodicity of sgld via reflection coupling. *Stochastics and Dynamics*, 24(05):2450035, 2024.

M. Beatrice Lignola and Jacqueline Morgan. Stability of regularized bilevel programming problems. *Journal of Optimization Theory and Applications*, 93(3):575–596, 1997.

Licong Lin and Tijana Zrnic. Plug-in performative optimization. In *International Conference on Machine Learning*, pp. 30546–30565. PMLR, 2024.

D. V. Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005, 1956. doi: 10.1214/aoms/1177728069.

Bo Liu, Mao Ye, Stephen Wright, Peter Stone, and Qiang Liu. Bome! bilevel optimization made easy: A simple first-order approach. *Advances in neural information processing systems*, 35: 17248–17262, 2022.

Hanxiao Liu, Karen Simonyan, and Yiming Yang. DARTS: Differentiable architecture search. In *International Conference on Learning Representations*, 2019. URL `https://openreview.net/forum?id=S1eYHoC5FX`.

Risheng Liu, Xuan Liu, Xiaoming Yuan, Shangzhi Zeng, and Jin Zhang. A value-function-based interior-point method for non-convex bi-level optimization. In *International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 6882–6892. PMLR, 2021.

Tianyi Liu, Yifan Lin, and Enlu Zhou. Bayesian stochastic gradient descent for stochastic optimization with streaming input data. *SIAM Journal on Optimization*, 34(1):389–418, 2024. doi: 10.1137/22M1478951.

Jonathan Lorraine, Paul Vicol, and David Duvenaud. Optimizing millions of hyperparameters by implicit differentiation. In *International Conference on Artificial Intelligence and Statistics*, pp. 1540–1552. PMLR, 2020.

Dougal Maclaurin, David Duvenaud, and Ryan Adams. Gradient-based hyperparameter optimization through reversible learning. In *International conference on machine learning*, pp. 2113–2122. PMLR, 2015.

Saeed Masiha, Zebang Shen, Negar Kiyavash, and Niao He. Superquantile-gibbs relaxation for minima-selection in bi-level optimization. *arXiv preprint arXiv:2505.05991*, 2025.

Saeed Masiha, Zebang Shen, Negar Kiyavash, and Niao He. Select-then-differentiate: Solving bilevel optimization with manifold lower-level solution sets. *arXiv preprint arXiv:2605.09209*, 2026.

Filip Matějka and Alisdair McKay. Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–298, 2015.

Richard D. McKelvey and Thomas R. Palfrey. Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38, 1995.

Qiang Meng, Hai Yang, and Michael G. H. Bell. Transportation network optimization problems with stochastic user equilibrium constraints. *Transportation Research Record*, 1882:142–151, 2004.

Shakir Mohamed, Mihaela Rosca, Michael Figurnov, and Andriy Mnih. Monte carlo gradient estimation in machine learning. *Journal of Machine Learning Research*, 21(132):1–62, 2020.

Radford M Neal and Geoffrey E Hinton. A view of the em algorithm that justifies incremental, sparse, and other variants. In *Learning in graphical models*, pp. 355–368. Springer, 1998.

Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *International Conference on Machine Learning*, pp. 7599–7609. PMLR, 2020.

Maxim Raginsky, Alexander Rakhlin, and Matus Telgarsky. Non-convex learning via stochastic gradient langevin dynamics: A nonasymptotic analysis. In *Conference on Learning Theory*, volume 65, pp. 1674–1703. PMLR, 2017.

Tom Rainforth, Adam Foster, Desi R Ivanova, and Freddie Bickford Smith. Modern bayesian experimental design. *Statistical Science*, 39(1):100–114, 2024.

Aravind Rajeswaran, Chelsea Finn, Sham M Kakade, and Sergey Levine. Meta-learning with implicit gradients. *Advances in neural information processing systems*, 32, 2019.

Johannes O Royset. On stability in optimistic bilevel optimization. *SIAM Journal on Optimization*, 36(2):987–1016, 2026.

Paul Vicol, Jonathan P. Lorraine, Fabian Pedregosa, David Duvenaud, and Roger B. Grosse. On implicit bias in overparameterized bilevel optimization. In *International Conference on Machine Learning*, volume 162, pp. 22234–22259. PMLR, 2022.

Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.

Max Welling and Yee W Teh. Bayesian learning via stochastic gradient langevin dynamics. In *International Conference on Machine Learning*, pp. 681–688, 2011.

Hai Yang and Michael G. H. Bell. Transport bilevel programming problems: Recent methodological advances and applications. *Transportation Research Part B: Methodological*, 2001.

Rong Yang, Christopher Kiekintveld, Fernando Ordóñez, Milind Tambe, and Richard John. Improving resource allocation strategies against human adversaries in security games: An extended study. *Artificial Intelligence*, 195:440–469, 2013.

Brian D. Ziebart, Andrew L. Maas, J. Andrew Bagnell, and Anind K. Dey. Maximum entropy inverse reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2008.

A Appendix: Proofs

A.1 Proof of Proposition 4.3

Proof. By construction, χ is smooth, $0 \le \chi \le 1/2$, $\chi(s) = 0$ for $s \le R_1^2$, and $\chi(s) = 1/2$ for $s \ge R_0^2$. Hence ρ_γ is smooth and nonnegative, and $\rho_\gamma(\theta) = 0$ whenever $\|\theta\| \le R_1$.

The chain rule gives

$$\begin{aligned} \nabla_\theta \rho_\gamma(\theta) &= 2\gamma\chi(\|\theta\|^2)\theta, \\ \nabla^2_{\theta\theta}\rho_\gamma(\theta) &= 2\gamma\chi(\|\theta\|^2)I_d + 4\gamma\chi'(\|\theta\|^2)\theta\theta^\top. \end{aligned} \tag{5}$$

In particular, for $\|\theta\| \ge R_0$,

$$\nabla^2_{\theta\theta}\rho_\gamma(\theta) = \gamma I_d.$$

Under the additional regularity assumptions, the L_H smoothness of l and gradient unbiasedness from Assumption 4.1 imply

$$\|\nabla_\theta \ell(v,\theta) - \nabla_\theta \ell(v,\theta')\| \le \mathbb{E}_\xi \|\nabla_\theta l(v,\theta;\xi) - \nabla_\theta l(v,\theta';\xi)\| \le L_H \|\theta - \theta'\|.$$

Since ℓ is C^2, it follows that

$$\nabla^2_{\theta\theta}\ell(v,\theta) \succeq -L_H I_d.$$

Consequently, for $\|\theta\| \ge R_0$,

$$\nabla^2_{\theta\theta}\ell_\gamma(v,\theta) = \nabla^2_{\theta\theta}\ell(v,\theta) + \gamma I_d \succeq (\gamma - L_H) I_d = \kappa_0 I_d.$$

We next verify preservation of the lower-level solution sets. Fix $v \in \mathcal{V}$ and write $m(v) := \min_\theta \ell(v,\theta)$. Since $\rho_\gamma \ge 0$,

$$\ell_\gamma(v,\theta) \ge \ell(v,\theta) \ge m(v) \qquad , \forall \theta \in \mathbb{R}^d.$$

For any $\theta^* \in S(v)$, uniform boundedness and $R_\star < R_1$ imply $\rho_\gamma(\theta^*) = 0$. Thus $\ell_\gamma(v,\theta^*) = m(v)$, and therefore

$$\min_\theta \ell_\gamma(v,\theta) = m(v), \qquad S(v) \subseteq \arg\min_\theta \ell_\gamma(v,\theta).$$

Conversely, if $\theta \in \arg\min_z \ell_\gamma(v,z)$, then

$$m(v) = \ell(v,\theta) + \rho_\gamma(\theta) \ge m(v) + \rho_\gamma(\theta).$$

Since $\rho_\gamma(\theta) \ge 0$, this forces $\rho_\gamma(\theta) = 0$ and $\ell(v,\theta) = m(v)$. Hence $\theta \in S(v)$, proving

$$\arg\min_{\theta\in\mathbb{R}^d} \ell_\gamma(v,\theta) = \arg\min_{\theta\in\mathbb{R}^d} \ell(v,\theta) = S(v). \tag{6}$$

Because f is unchanged, the two deterministic bilevel problems have the same feasible graph and outer solutions.

It remains to establish the joint Lipschitz bound and the bound at the origin. Set $\Delta_R := R_0^2 - R_1^2 > 0$. The definition of χ can be written as

$$\chi(s) = \frac{1}{2}\sigma\left(\frac{s - R_1^2}{\Delta_R}\right), \qquad \chi'(s) = \frac{1}{2\Delta_R}\sigma'\left(\frac{s - R_1^2}{\Delta_R}\right).$$

We first show that $0 \le \sigma'(t) \le 2$ for every $t \in \mathbb{R}$. For $t \notin (0,1)$, $\sigma'(t) = 0$. For $0 < t < 1$, using $\cosh^2 x = 1 + \sinh^2 x \ge 1 + x^2$, we obtain

$$\begin{aligned} 0 \le \sigma'(t) &= \frac{t^{-2} + (1-t)^{-2}}{4\cosh^2\left(\frac{1}{2t} - \frac{1}{2(1-t)}\right)} \\ &\le \frac{t^{-2} + (1-t)^{-2}}{4 + \left(t^{-1} - (1-t)^{-1}\right)^2} \\ &= \frac{1}{t^2 + (1-t)^2} \le 2. \end{aligned}$$

Therefore,

$$0 \le \chi'(s) \le \frac{1}{\Delta_R}, \qquad \chi'(s) = 0 \quad \text{for } s \notin (R_1^2, R_0^2).$$

Combining these bounds with (5) gives, for every $\theta \in \mathbb{R}^d$,

$$\begin{aligned}\|\nabla^2_{\theta\theta}\rho_\gamma(\theta)\|_{\mathrm{op}} &\le 2\gamma\chi(\|\theta\|^2) + 4\gamma\chi'(\|\theta\|^2)\|\theta\|^2 \\ &\le \gamma + \frac{4\gamma R_0^2}{\Delta_R} =: L_\rho.\end{aligned}$$

Thus $\nabla_\theta\rho_\gamma$ is globally L_ρ-Lipschitz:

$$\|\nabla_\theta\rho_\gamma(\theta) - \nabla_\theta\rho_\gamma(\theta')\| \le L_\rho\|\theta - \theta'\|.$$

Since ρ_γ is independent of v,

$$\nabla l_\gamma(v, \theta; \xi) = \nabla l(v, \theta; \xi) + \begin{pmatrix} 0_p \\ \nabla_\theta\rho_\gamma(\theta) \end{pmatrix}.$$

Consequently,

$$\begin{aligned}&\|\nabla l_\gamma(v, \theta; \xi) - \nabla l_\gamma(v', \theta'; \xi)\| \\ &\quad\le \|\nabla l(v, \theta; \xi) - \nabla l(v', \theta'; \xi)\| + \|\nabla_\theta\rho_\gamma(\theta) - \nabla_\theta\rho_\gamma(\theta')\| \\ &\quad\le L_H\sqrt{\|v - v'\|^2 + \|\theta - \theta'\|^2} + L_\rho\|\theta - \theta'\| \\ &\quad\le (L_H + L_\rho)\sqrt{\|v - v'\|^2 + \|\theta - \theta'\|^2}.\end{aligned}$$

Substituting $\gamma = L_H + \kappa_0$ yields the constant

$$\bar{L}_H := L_H + (L_H + \kappa_0)\left(1 + \frac{4R_0^2}{R_0^2 - R_1^2}\right). \tag{7}$$

Finally, $\nabla_\theta\rho_\gamma(0) = 0$, so

$$\sup_{v\in\mathcal{V},\xi} \|\nabla_\theta l_\gamma(v, 0; \xi)\| = \sup_{v\in\mathcal{V},\xi} \|\nabla_\theta l(v, 0; \xi)\| \le G_0 = \bar{G}_0.$$

The C^2 regularity of ℓ_γ and l_γ follows from the smoothness of ρ_γ, completing the proof. □

A.2 Proof of Lemma 4.4

Proof. We first prove dissipativity condition. Fix $v \in \mathcal{V}$ and suppose $\|\theta\| \ge R_0$. Write $r := \|\theta\|$, $\nu := \theta/\|\theta\|$, and define, for $s \ge R_0$,

$$q_{v,\nu}(s) := \langle \nabla_\theta\ell(v, s\nu), \nu\rangle.$$

By the chain rule and exterior strong convexity condition in Assumption 4.1,

$$q'_{v,\nu}(s) = \nu^\top \nabla^2_{\theta\theta}\ell(v, s\nu)\nu \ge \kappa_0.$$

Integrating from R_0 to r gives

$$q_{v,\nu}(r) - q_{v,\nu}(R_0) \ge \kappa_0(r - R_0).$$

Multiplying by r and with $b_0 := \max_{\substack{v\in\mathcal{V}\\ \|\nu\|=1}} \{\kappa_0 R_0 - \langle \nabla_\theta\ell(v, R_0\nu), \nu\rangle\}$, we obtain

$$\langle \nabla_\theta\ell(v, \theta), \theta\rangle \ge \kappa_0 r^2 - b_0 r.$$

Young's inequality gives $b_0 r \le (\kappa_0/2)r^2 + b_0^2/(2\kappa_0)$, and therefore

$$\langle \nabla_\theta\ell(v, \theta), \theta\rangle \ge \frac{\kappa_0}{2}\|\theta\|^2 - \frac{b_0^2}{2\kappa_0}, \qquad \|\theta\| \ge R_0. \tag{8}$$

For $\|\theta\| \le R_0$, with $\bar{b}_0 := \max_{\substack{v\in\mathcal{V}\\ \|\theta\|\le R_0}} \left\{\frac{\kappa_0}{2}\|\theta\|^2 - \langle \nabla_\theta\ell(v, \theta), \theta\rangle\right\}$ gives

$$\langle \nabla_\theta\ell(v, \theta), \theta\rangle \ge \frac{\kappa_0}{2}\|\theta\|^2 - \bar{b}_0. \tag{9}$$

Combining (8) and (9) and set

$$a := \frac{\kappa_0}{2}, \qquad b := \max\left\{\bar{b}_0, \frac{b_0^2}{2\kappa_0}\right\}, \tag{10}$$

proves $\langle \nabla_\theta \ell(v,\theta), \theta\rangle \geq a\|\theta\|^2 - b$.

We next prove the quadratic lower bound for ℓ. Suppose first that $\|\theta\| = r \geq R_0$, let $\nu := \theta/\|\theta\|$, set $\bar{\theta} := R_0\nu$, and define

$$\gamma(t) := \bar{\theta} + t(\theta - \bar{\theta}), \qquad t \in [0,1].$$

Then $\|\gamma(t)\| \geq R_0$ for every $t \in [0,1]$. For $g_v(t) := \ell(v, \gamma(t))$, we have

$$\begin{aligned} g_v'(t) &= \left\langle \nabla_\theta \ell(v,\gamma(t)), \theta - \bar{\theta}\right\rangle, \\ g_v''(t) &= (\theta - \bar{\theta})^\top \nabla^2_{\theta\theta} \ell(v,\gamma(t))(\theta - \bar{\theta}) \geq \kappa_0 \|\theta - \bar{\theta}\|^2. \end{aligned}$$

Define $G_\ell := \max_{v\in\mathcal{V},\ \|\theta\|=R_0} \|\nabla_\theta \ell(v,\theta)\|$, the integral form of Taylor's theorem yields

$$\begin{aligned} \ell(v,\theta) &\geq \ell(v,\bar{\theta}) + \left\langle \nabla_\theta \ell(v,\bar{\theta}), \theta - \bar{\theta}\right\rangle + \frac{\kappa_0}{2}\|\theta - \bar{\theta}\|^2 \\ &\geq m_\ell - G_\ell(r - R_0) + \frac{\kappa_0}{2}(r - R_0)^2 \\ &\geq m_\ell - \frac{G_\ell^2}{\kappa_0} + \frac{\kappa_0}{4}(r - R_0)^2, \end{aligned}$$

where the last step uses $G_\ell(r - R_0) \leq (\kappa_0/4)(r - R_0)^2 + G_\ell^2/\kappa_0$. Moreover, since $(r - R_0)^2 \geq \frac{1}{2}r^2 - R_0^2$, let $m_\ell := \min_{v\in\mathcal{V},\ \|\theta\|=R_0} \ell(v,\theta)$, we have

$$\ell(v,\theta) \geq \frac{\kappa_0}{8}\|\theta\|^2 - \left(\frac{G_\ell^2}{\kappa_0} + \frac{\kappa_0 R_0^2}{4} - m_\ell\right), \qquad \|\theta\| \geq R_0. \tag{11}$$

For $\|\theta\| \leq R_0$, define $\bar{m}_\ell := \min_{v\in\mathcal{V},\ \|\theta\|\leq R_0} \ell(v,\theta)$, we obtain

$$\ell(v,\theta) \geq \bar{m}_\ell \geq \frac{\kappa_0}{8}\|\theta\|^2 - \left(\frac{\kappa_0 R_0^2}{8} - \bar{m}_\ell\right). \tag{12}$$

Combining (11) and define

$$\underline{a} := \frac{\kappa_0}{8}, \qquad b_\ell := \max\left\{\frac{\kappa_0 R_0^2}{8} - \bar{m}_\ell, \frac{G_\ell^2}{\kappa_0} + \frac{\kappa_0 R_0^2}{4} - m_\ell\right\}. \tag{13}$$

proves $\ell(v,\theta) \geq \underline{a}\|\theta\|^2 - b_\ell$. □

Remark A.1. The assumptions 4.1 and 4.6 imply the uniform bound on the Gibbs mean of $\nabla_v \ell$ used below. Define $C_{\ell,0} := \max_{v\in\mathcal{V}} \|\nabla_v \ell(v,0)\| < \infty$, $M_{\ell,B} := \max_{\substack{v\in\mathcal{V}\\ \|\theta\|\leq 1}} \ell(v,\theta) < \infty$, and $z_\star := |B(0,1)|e^{-M_{\ell,B}/\tau} > 0$. Then, applying the quadratic bound in Lemma 4.4,

$$\sup_{v\in\mathcal{V}} \mathbb{E}_{\pi_v}\|\theta\| \leq \frac{e^{b_\ell/\tau}}{z_\star}\int_{\mathbb{R}^d} \|\theta\| \exp\Big(-\frac{\underline{a}}{\tau}\|\theta\|^2\Big)\, d\theta =: M_{\pi,1} < \infty.$$

Combining this estimate with the L_H smoothness of l gives

$$\|\mathbb{E}_{\pi_v} \nabla_v \ell(v,\theta)\| \leq C_{\ell,0} + L_H M_{\pi,1} =: C_\ell, \qquad v \in \mathcal{V}. \tag{14}$$

A.3 Proof of Proposition 4.7

Proof. In particular, at every $v \in \mathrm{int}(\mathcal{V})$, By Lemma 4.4, there are constants $\underline{a} > 0$ and $b_\ell \in \mathbb{R}$, independent of v, such that

$$\ell(v,\theta) \geq \underline{a}\|\theta\|^2 - b_\ell, \qquad v \in \mathcal{V}, \quad \theta \in \mathbb{R}^d.$$

Consequently,

$$e^{-\ell(v,\theta)/\tau} \leq e^{b_\ell/\tau} e^{-\underline{a}\|\theta\|^2/\tau}. \tag{15}$$

Compactness of $\mathcal{V}$ and continuity of ℓ also give

$$M_{\ell,B} := \max_{v\in\mathcal{V},\,\|\theta\|\le 1} \ell(v,\theta) < \infty, \qquad Z(v) \ge |B(0,1)|e^{-M_{\ell,B}/\tau} =: z_\star > 0.$$

Together these estimates imply $0 < z_\star \le Z(v) < \infty$ and uniformly bounded Gibbs moments of every polynomial order.

Define the finite constants

$$C_{f,0} := \max_{v\in\mathcal{V}} |f(v,0)|, \quad C_{f_v,0} := \max_{v\in\mathcal{V}} \|\nabla_v f(v,0)\|, \quad C_{\ell_v,0} := \max_{v\in\mathcal{V}} \|\nabla_v \ell(v,0)\|.$$

The Lipschitz assumptions and full-gradient unbiasedness yield

$$|f(v,\theta)| \le C_{f,0} + L_f\|\theta\|, \;\; \|\nabla_v f(v,\theta)\| \le C_{f_v,0} + L_{f_v}\|\theta\|, \;\; \|\nabla_v \ell(v,\theta)\| \le C_{\ell_v,0} + L_H\|\theta\|.$$

Hence there is a finite constant C, independent of v and θ, such that

$$\begin{aligned}(1 + |f(v,\theta)| + \|\nabla_v f(v,\theta)\| + \|\nabla_v \ell(v,\theta)\| + |f(v,\theta)|\,\|\nabla_v \ell(v,\theta)\|)\, e^{-\ell(v,\theta)/\tau} \\ \le C(1+\|\theta\|^2)e^{-\underline{a}\|\theta\|^2/\tau} =: D(\theta), \qquad D \in L^1(\mathbb{R}^d, d\theta).\end{aligned} \tag{16}$$

This proves absolute integrability of every expectation in

$$G_\tau(v) := \mathbb{E}_{\pi_v}[\nabla_v f(v,\theta)] - \frac{1}{\tau}\operatorname{Cov}_{\pi_v}\big(f(v,\theta), \nabla_v \ell(v,\theta)\big). \tag{17}$$

By dominated convergence, the unnormalized integrals defining these expectations are continuous in v. Since $Z(v) \ge z_\star$, both J_τ and G_τ are continuous on $\mathcal{V}$.

Fix $v, w \in \mathcal{V}$, set $u := w - v$ and $v_s := v + su$, and define

$$N(s) := \int f(v_s,\theta)e^{-\ell(v_s,\theta)/\tau}\, d\theta, \qquad Z_s := \int e^{-\ell(v_s,\theta)/\tau}\, d\theta.$$

Convexity implies $v_s \in \mathcal{V}$ for $s \in [0,1]$. For every fixed θ, the chain and product rules give

$$\begin{aligned}\frac{d}{ds}e^{-\ell(v_s,\theta)/\tau} &= -\frac{1}{\tau}\langle \nabla_v \ell(v_s,\theta), u\rangle e^{-\ell(v_s,\theta)/\tau}, \\ \frac{d}{ds}\left(f(v_s,\theta)e^{-\ell(v_s,\theta)/\tau}\right) &= \left\langle \nabla_v f(v_s,\theta) - \frac{f(v_s,\theta)}{\tau}\nabla_v \ell(v_s,\theta), u\right\rangle e^{-\ell(v_s,\theta)/\tau}.\end{aligned}$$

By (15), their absolute values are uniformly bounded in s by constant multiple of the integrable function $\|u\|D(\theta)$. Hence, the dominated convergence theorem yields

$$\begin{aligned}Z_s' &= -\frac{1}{\tau}Z_s\langle \mathbb{E}_{\pi_{v_s}}[\nabla_v \ell(v_s,\theta)], u\rangle, \\ N'(s) &= Z_s\langle \mathbb{E}_{\pi_{v_s}}[\nabla_v f(v_s,\theta)], u\rangle - \frac{1}{\tau}Z_s\langle \mathbb{E}_{\pi_{v_s}}[f(v_s,\theta)\nabla_v \ell(v_s,\theta)], u\rangle\end{aligned}$$

Applying the quotient rule to $J_\tau(v_s) = N(s)/Z_s$ gives

$$\begin{aligned}\frac{d}{ds}J_\tau(v_s) &= \frac{N'(s)}{Z_s} - \frac{N(s)Z_s'}{Z(s)^2} \\ &= \langle \mathbb{E}_{\pi_{v_s}}[\nabla_v f(v_s,\theta)], u\rangle - \frac{1}{\tau}\langle \mathbb{E}_{\pi_{v_s}}[f(v_s,\theta)\nabla_v \ell(v_s,\theta)], u\rangle \\ &\quad + \frac{1}{\tau}\langle \mathbb{E}_{\pi_{v_s}}[f(v_s,\theta)]\mathbb{E}_{\pi_{v_s}}[\nabla_v \ell(v_s,\theta)], u\rangle \\ &= \langle G_\tau(v_s), u\rangle.\end{aligned}$$

Since G_τ is continuous, integration over the segment yields

$$J_\tau(w) - J_\tau(v) = \int_0^1 \langle G_\tau(v + s(w-v)), w - v\rangle\, ds.$$

In particular,

$$\big|J_\tau(w) - J_\tau(v) - \langle G_\tau(v), w - v\rangle\big| \le \|w - v\| \sup_{s\in[0,1]} \|G_\tau(v + s(w-v)) - G_\tau(v)\| = o(\|w - v\|).$$

This proves

$$J_\tau(w) - J_\tau(v) = \langle G_\tau(v), w - v\rangle + o(\|w - v\|), \qquad w \to v, \quad w \in \mathcal{V}. \tag{18}$$

In particular, $\nabla_v J_\tau(v) = G_\tau(v)$ for every $v \in \mathrm{int}(\mathcal{V})$.

Now, we proof the second part of this Proposition. Fix v and abbreviate $\theta^\star := \theta^\star(v)$, $H := \nabla^2_{\theta\theta}\ell(v, \theta^\star)$, $B := \nabla^2_{v\theta}\ell(v, \theta^\star)$ and $a := \nabla_\theta f(v, \theta^\star)$, where $B \in \mathbb{R}^{p\times d}$ and $H \succeq \kappa_0 I_d$. We write π_v^τ for the Gibbs measure to make the temperature dependence explicit.

Let $\Theta_\tau \sim \pi_v^\tau$ and $Y_\tau := (\Theta_\tau - \theta^\star)/\sqrt{\tau}$. The density of Y_τ is

$$q_\tau(y) = \frac{e^{-U_\tau(y)}}{D_\tau}, \qquad U_\tau(y) := \frac{\ell(v, \theta^\star + \sqrt{\tau}y) - \ell(v, \theta^\star)}{\tau}, \qquad D_\tau := \int_{\mathbb{R}^d} e^{-U_\tau(y)}\, dy.$$

Since $\nabla_\theta \ell(v, \theta^\star) = 0$, global strong convexity and L_H-smoothness imply

$$\frac{\kappa_0}{2}\|y\|^2 \le U_\tau(y) \le \frac{L_H}{2}\|y\|^2. \tag{19}$$

Consequently,

$$D_\tau \ge \left(\frac{2\pi}{L_H}\right)^{d/2}, \qquad q_\tau(y) \le \left(\frac{L_H}{2\pi}\right)^{d/2} e^{-\kappa_0\|y\|^2/2}. \tag{20}$$

Moreover, the second-order Taylor expansion of ℓ gives $U_\tau(y) \to \frac{1}{2}y^\top H y$ for every fixed y. The lower bound on U_τ in (19) provides an integrable dominating function; hence, the dominated convergence theorem yields

$$D_\tau \longrightarrow \frac{(2\pi)^{d/2}}{\sqrt{\det H}}, \quad q_\tau(y) \longrightarrow q_0(y) := \frac{\sqrt{\det H}}{(2\pi)^{d/2}} e^{-y^\top H y/2}, \quad \text{as } \tau \to 0.$$

Thus the limiting density is that of $Y \sim \mathcal{N}(0, H^{-1})$. The Gaussian domination (20) also permits dominated convergence for integrands with polynomial growth and, in particular, gives $\sup_{\tau>0} \mathbb{E}\|Y_\tau\|^k < \infty$, for $k = 1, 2$.

Then, define the rescaled differences

$$F_\tau(y) := \frac{f(v, \theta^\star + \sqrt{\tau}y) - f(v, \theta^\star)}{\sqrt{\tau}}, \quad G_\tau(y) := \frac{\nabla_v \ell(v, \theta^\star + \sqrt{\tau}y) - \nabla_v \ell(v, \theta^\star)}{\sqrt{\tau}}.$$

Differentiability gives, for every fixed y,

$$F_\tau(y) \longrightarrow a^\top y, \quad G_\tau(y) \longrightarrow By, \quad \text{as } \tau \to 0.$$

The Lipschitzness from Assumption 4.1 and 4.6 imply $|F_\tau(y)| \le L_f\|y\|$ and $\|G_\tau(y)\| \le L_H\|y\|$. Together with the Gaussian domination of q_τ, the dominated convergence theorem yields

$$\mathbb{E}_{q_\tau}[F_\tau] \longrightarrow \mathbb{E}[a^\top Y] = 0, \quad \mathbb{E}_{q_\tau}[G_\tau] \longrightarrow \mathbb{E}[BY] = 0,$$
$$\mathbb{E}_{q_\tau}[F_\tau G_\tau] \longrightarrow \mathbb{E}[(a^\top Y)BY] = BH^{-1}a.$$

Since covariance is invariant under additive constants, as $\tau \to 0$

$$\frac{1}{\tau}\,\mathrm{Cov}_{\pi_v^\tau}\big(f(v, \theta), \nabla_v \ell(v, \theta)\big) = \mathbb{E}_{q_\tau}[F_\tau G_\tau] - \mathbb{E}_{q_\tau}[F_\tau]\,\mathbb{E}_{q_\tau}[G_\tau] \longrightarrow BH^{-1}a. \tag{21}$$

Then, the Lipschitz continuity of $\nabla_v f(v, \cdot)$ and the boundedness of $\mathbb{E}\|Y_\tau\|$ gives

$$\big\|\mathbb{E}_{\pi_v^\tau}[\nabla_v f(v, \theta)] - \nabla_v f(v, \theta^\star)\big\| \le L_{f_v}\sqrt{\tau}\,\mathbb{E}\|Y_\tau\| \longrightarrow 0.$$

Combining this limit with (21) and the Gibbs hypergradient identity from Proposition 4.7, which applies since global strong convexity implies exterior strong convexity, we obtain

$$\nabla_v J_\tau(v) \longrightarrow \nabla_v f(v, \theta^\star) - BH^{-1}a.$$

Finally, the implicit function theorem applied to $\nabla_\theta \ell(v, \theta^\star(v)) = 0$ yields $D_v\theta^\star(v) = -H^{-1}B^\top$. The chain rule therefore gives

$$\nabla\Phi(v) = \nabla_v f(v, \theta^\star) + [D_v\theta^\star(v)]^\top \nabla_\theta f(v, \theta^\star) = \nabla_v f(v, \theta^\star) - BH^{-1}a,$$

which completes the proof. □

A.4 Statement and Proof of Lemma A.2

Lemma A.2. *Suppose Assumption 4.8 holds,* $0 < \beta \le (2L_J)^{-1}$ *and set* $\mathcal{V}$ *is nonempty closed convex. For the update* $v_{k+1} = \Pi_{\mathcal{V}}(v_k - \beta(\nabla J_\tau(v_k) + e_k))$*,*

$$\mathbb{E}J_\tau(v_{k+1}) \le \mathbb{E}J_\tau(v_k) - \frac{\beta}{8}\mathbb{E}\|\mathcal{G}_\beta(v_k)\|^2 + \frac{3\beta}{4}\mathbb{E}\|e_k\|^2, \tag{22}$$

where $\mathcal{G}_\beta(v) = \frac{1}{\beta}(v - \Pi_{\mathcal{V}}(v - \beta\nabla J_\tau(v)))$*.*

Proof. Set $d_k := v_{k+1} - v_k$ and $g_k := \nabla J_\tau(v_k) + e_k$. Since $v_{k+1} = \Pi_{\mathcal{V}}(v_k - \beta g_k)$, the variational characterization of Euclidean projection for nonempty closed convex set gives, for every $v \in \mathcal{V}$,

$$\langle v_{k+1} - (v_k - \beta g_k), v - v_{k+1}\rangle \ge 0.$$

Taking $v = v_k$ gives $0 \le \langle d_k + \beta g_k, -d_k\rangle = -\|d_k\|^2 - \beta\langle g_k, d_k\rangle$, so

$$\langle \nabla J_\tau(v_k) + e_k, d_k\rangle \le -\frac{1}{\beta}\|d_k\|^2. \tag{23}$$

By L_J-smoothness from Assumption 4.8,

$$\begin{aligned} J_\tau(v_{k+1}) &\le J_\tau(v_k) + \langle \nabla J_\tau(v_k), d_k\rangle + \frac{L_J}{2}\|d_k\|^2 \\ &= J_\tau(v_k) + \langle \nabla J_\tau(v_k) + e_k, d_k\rangle - \langle e_k, d_k\rangle + \frac{L_J}{2}\|d_k\|^2 \\ &\le J_\tau(v_k) - \left(\frac{1}{\beta} - \frac{L_J}{2}\right)\|d_k\|^2 - \langle e_k, d_k\rangle. \end{aligned}$$

Then, apply Young's inequality $-\langle e_k, d_k\rangle \le \frac{\beta}{2}\|e_k\|^2 + \frac{1}{2\beta}\|d_k\|^2$, we obtain

$$\begin{aligned} J_\tau(v_{k+1}) &\le J_\tau(v_k) - \left(\frac{1}{2\beta} - \frac{L_J}{2}\right)\|d_k\|^2 + \frac{\beta}{2}\|e_k\|^2 \\ &\le J_\tau(v_k) - \frac{1}{4\beta}\|d_k\|^2 + \frac{\beta}{2}\|e_k\|^2, \end{aligned} \tag{24}$$

where the last inequality uses $\beta L_J \le 1/2$.

Define the exact projected step $v_k^+ := \Pi_{\mathcal{V}}(v_k - \beta\nabla J_\tau(v_k))$. Nonexpansiveness of projection gives

$$\|v_{k+1} - v_k^+\| \le \|(v_k - \beta(\nabla J_\tau(v_k) + e_k)) - (v_k - \beta\nabla J_\tau(v_k))\| = \beta\|e_k\|.$$

Moreover,

$$\beta\|\mathcal{G}_\beta(v_k)\| = \|v_k - v_k^+\| \le \|v_k - v_{k+1}\| + \|v_{k+1} - v_k^+\| \le \|d_k\| + \beta\|e_k\|.$$

Squaring and rearranging gives, $\|d_k\|^2 \ge \frac{\beta^2}{2}\|\mathcal{G}_\beta(v_k)\|^2 - \beta^2\|e_k\|^2$. Substituting this into (24) gives

$$\begin{aligned} J_\tau(v_{k+1}) &\le J_\tau(v_k) - \frac{\beta}{8}\|\mathcal{G}_\beta(v_k)\|^2 + \frac{\beta}{4}\|e_k\|^2 + \frac{\beta}{2}\|e_k\|^2 \\ &= J_\tau(v_k) - \frac{\beta}{8}\|\mathcal{G}_\beta(v_k)\|^2 + \frac{3\beta}{4}\|e_k\|^2. \end{aligned}$$

□

A.5 Statement and Proof for Proposition A.3

Proposition A.3. *Suppose Assumptions 4.1, 4.6, 4.5, and 4.9 hold, and Algorithm 1 is run with independent restart. Let* $1 \le S \le T$*. Suppose* (42) *holds,* $h \le h_4$ *with* h_4 *defined in* (53)*, and* (59) *and* (60) *hold. Then, for every* $k \ge 0$ *and* $i \in \{1, 2, 3, 4\}$*,*

$$\begin{aligned} \mathbb{E}\|\widehat{\mathbb{E}}_k[\varphi_i] - \mathbb{E}_{\pi_{v_k}}[\varphi_i]\|^2 \le \frac{4}{S}\left[L_{\varphi_i}^2 D_2\left(1 + \frac{2c_0 e^{-ch}}{1 - e^{-ch}}\right) + \sigma_i^2\right] \\ + \frac{2L_{\varphi_i}^2 C_{\mathrm{erg}}^2 (M_{4,0} + 1)^2}{chS} e^{-2ch(T-S)} + 4L_{\varphi_i}^2 C_{\mathrm{b}}^2 h, \end{aligned} \tag{25}$$

where D_2 *is defined in* (44)*,* c, c_0 *in* (64)*,* C_{b} *in* (67)*, and* C_{erg} *in* (70)*. Moreover,* $\sigma_1^2 = \sigma_2^2 = 0$*,* $\sigma_3^2 = \sigma_{v,2}^2 p$*,* $\sigma_4^2 = C_f^2 \sigma_{v,2}^2 p$*,* $L_{\varphi_1} := L_f$*,* $L_{\varphi_2} := L_{f_v}$*,* $L_{\varphi_3} := L_H$*, and* $L_{\varphi_4} := L_{f\ell_v}$*.*

Proof. Fix $i \in \{1,2,3,4\}$ and for each $t \in I_k := \{T-S+1,\dots,T\}$, we have

$$\widehat{\mathbb{E}}_k[\varphi_i] - \mathbb{E}_{\pi_{v_k}}[\varphi_i] = \frac{1}{S}\sum_{t\in I_k}\left(\mathbb{E}_{\mu_{k,t}}[\varphi_i] - \mathbb{E}_{\pi_{v_k}}[\varphi_i]\right) + \frac{1}{S}\sum_{t\in I_k}\left(\widehat{\varphi}_i(v_k,\theta_{k,t};\widetilde{\xi}_{k,t}) - \mathbb{E}_{\mu_{k,t}}[\varphi_i]\right).$$

Thus, taking expectation of all the randomness from both inner and outer iteration and using $\|x+y\|^2 \le 2\|x\|^2 + 2\|y\|^2$,

$$\begin{aligned}\mathbb{E}\left\|\widehat{\mathbb{E}}_k[\varphi_i] - \mathbb{E}_{\pi_{v_k}}[\varphi_i]\right\|^2 \le{}& 2\mathbb{E}\left\|\frac{1}{S}\sum_{t\in I_k}\left(\mathbb{E}_{\mu_{k,t}}[\varphi_i] - \mathbb{E}_{\pi_{v_k}}[\varphi_i]\right)\right\|^2 \\ &+ 2\mathbb{E}\left\|\frac{1}{S}\sum_{t\in I_k}\left(\widehat{\varphi}_i(v_k,\theta_{k,t};\widetilde{\xi}_{k,t}) - \mathbb{E}_{\mu_{k,t}}[\varphi_i]\right)\right\|^2.\end{aligned}$$

For the first term, by the L_{φ_i}-Lipschitzness from Assumptions 4.1 and 4.6, and apply duality

$$\left\|\mathbb{E}_{\mu_{k,t}}[\varphi_i] - \mathbb{E}_{\pi_{v_k}}[\varphi_i]\right\| = \sup_{\|u\|=1}\left|\mathbb{E}_{\mu_{k,t}}\langle u,\varphi_i\rangle - \mathbb{E}_{\pi_{v_k}}\langle u,\varphi_i\rangle\right| \le L_{\varphi_i} W_1(\mu_{k,t},\pi_{v_k}).$$

Thus, for the finite average,

$$\begin{aligned}2\mathbb{E}\left\|\frac{1}{S}\sum_{t\in I_k}\left(\mathbb{E}_{\mu_{k,t}}[\varphi_i] - \mathbb{E}_{\pi_{v_k}}[\varphi_i]\right)\right\|^2 &\le \frac{2}{S}\sum_{t\in I_k}\mathbb{E}\left\|\mathbb{E}_{\mu_{k,t}}[\varphi_i] - \mathbb{E}_{\pi_{v_k}}[\varphi_i]\right\|^2 \\ &\le \frac{2L_{\varphi_i}^2}{S}\sum_{t\in I_k}\mathbb{E}W_1(\mu_{k,t},\pi_{v_k})^2. \qquad (26)\end{aligned}$$

Under independent restart, $\mu_{k,0} = \mu_0$ conditionally on $\mathcal{G}_k$. Lemma A.10, applied pathwise at the random but $\mathcal{G}_k$-measurable parameter v_k, yields

$$W_1(\mu_{k,t},\pi_{v_k}) \le C_{\mathrm{erg}}(M_{4,0}+1)e^{-cht} + C_{\mathrm{b}}\sqrt{h} \qquad \mathbb{P}\text{-almost surely.}$$

Consequently, the bound (26) becomes

$$\begin{aligned}2\mathbb{E}\left\|\frac{1}{S}\sum_{t\in I_k}\left(\mathbb{E}_{\mu_{k,t}}[\varphi_i] - \mathbb{E}_{\pi_{v_k}}[\varphi_i]\right)\right\|^2 &\le \frac{2L_{\varphi_i}^2}{S}\sum_{t\in I_k}\mathbb{E}W_1(\mu_{k,t},\pi_{v_k})^2 \\ &\le \frac{4L_{\varphi_i}^2 C_{\mathrm{erg}}^2(M_{4,0}+1)^2}{S}\frac{e^{-2ch(T-S+1)}(1-e^{-2chS})}{1-e^{-2ch}} + 4L_{\varphi_i}^2 C_{\mathrm{b}}^2 h \\ &\le \frac{2L_{\varphi_i}^2 C_{\mathrm{erg}}^2(M_{4,0}+1)^2}{chS}e^{-2ch(T-S)} + 4L_{\varphi_i}^2 C_{\mathrm{b}}^2 h. \qquad (27)\end{aligned}$$

For the second term, define $A_{k,t}^{(i)} := \varphi_i(v_k,\theta_{k,t}) - \mathbb{E}_{\mu_{k,t}}[\varphi_i]$, $N_{k,t}^{(i)} := \widehat{\varphi}_i(v_k,\theta_{k,t};\widetilde{\xi}_{k,t}) - \varphi_i(v_k,\theta_{k,t})$. Then,

$$\mathbb{E}\left\|\frac{1}{S}\sum_{t\in I_k}\left(\widehat{\varphi}_i(v_k,\theta_{k,t};\widetilde{\xi}_{k,t}) - \mathbb{E}_{\mu_{k,t}}[\varphi_i]\right)\right\|^2 \le 2\mathbb{E}\left\|\frac{1}{S}\sum_{t\in I_k}A_{k,t}^{(i)}\right\|^2 + 2\mathbb{E}\left\|\frac{1}{S}\sum_{t\in I_k}N_{k,t}^{(i)}\right\|^2. \qquad (28)$$

By Lemma A.5,

$$\sup_{k\ge 0}\sup_{0\le t\le T}\mathbb{E}\|\theta_{k,t}\|^2 \le D_2,$$

where D_2 is defined in (44). Moreover, (59)–(60) allow us to apply the contraction estimate in Lemma A.8. Therefore, Lemma A.11, gives

$$\mathbb{E}\left\|\frac{1}{S}\sum_{t\in I_k}A_{k,t}^{(i)}\right\|^2 = \mathbb{E}\left[\mathbb{E}\left[\left\|\frac{1}{S}\sum_{t\in I_k}(\varphi_i(v_k,\theta_{k,t}) - \mathbb{E}_{\mu_{k,t}}[\varphi_i])\right\|^2\middle|\mathcal{G}_k\right]\right] \le \frac{L_{\varphi_i}^2 D_2}{S}\left(1 + \frac{2c_0e^{-ch}}{1-e^{-ch}}\right), \qquad (29)$$

where $\mathcal{G}_k$ contain all randomness generated before the initialization at outer iteration k. For the observation-noise part, expand the squared norm rather than suppressing the cross terms:

$$\mathbb{E}\left\|\frac{1}{S}\sum_{t\in I_k} N_{k,t}^{(i)}\right\|^2 = \frac{1}{S^2}\sum_{t\in I_k}\mathbb{E}\|N_{k,t}^{(i)}\|^2 + \frac{2}{S^2}\sum_{\substack{s,t\in I_k\\ s<t}}\mathbb{E}\left\langle N_{k,s}^{(i)}, N_{k,t}^{(i)}\right\rangle.$$

For $s<t$, the conditional expectation has that $\mathbb{E}\Big[N_{k,t}^{(i)} \mid \mathcal{F}_{k,t}\Big] = 0$. Therefore,

$$\mathbb{E}\left\langle N_{k,s}^{(i)}, N_{k,t}^{(i)}\right\rangle = \mathbb{E}\left[\left\langle N_{k,s}^{(i)}, \mathbb{E}\Big[N_{k,t}^{(i)} \mid \mathcal{F}_{k,t}\Big]\right\rangle\right] = 0,$$

where

$$\mathcal{F}_{k,t} := \sigma\Big(\mathcal{G}_k,\, \theta_{k,0},\, \{\xi_{k,r} : 0\le r<t\},\, \{\varepsilon_{k,r} : 1\le r\le t\},\, \{\widetilde{\xi}_{k,r} : 1\le r<t\}\Big), \quad 1\le t\le T. \tag{30}$$

Thus, we have

$$\mathbb{E}\left\|\frac{1}{S}\sum_{t\in I_k} N_{k,t}^{(i)}\right\|^2 = \frac{1}{S^2}\sum_{t\in I_k}\mathbb{E}\|N_{k,t}^{(i)}\|^2 \le \frac{1}{S^2}\sum_{t\in I_k}\sigma_i^2 = \frac{\sigma_i^2}{S}, \tag{31}$$

where by Assumption 4.5, we have

$$\sigma_1^2 = \sigma_2^2 = 0, \quad \sigma_3^2 = \sigma_{v,2}^2 p, \quad \sigma_4^2 = C_f^2\sigma_{v,2}^2 p.$$

Combining (28), (29), and (31),

$$2\mathbb{E}\left\|\frac{1}{S}\sum_{t\in I_k}\Big(\widehat{\varphi}_i(v_k,\theta_{k,t};\widetilde{\xi}_{k,t}) - \mathbb{E}_{\mu_{k,t}}[\varphi_i]\Big)\right\|^2 \le \frac{4}{S}\left(L_{\varphi_i}^2 D_2\left(1+\frac{2c_0e^{-ch}}{1-e^{-ch}}\right)+\sigma_i^2\right). \tag{32}$$

Finally, (27) and (32) prove (25). □

A.6 Proof for Theorem 4.10

Proof. The descent Lemma A.2 shows that the key remaining task is to control the hypergradient error $\mathbb{E}\|e_k\|^2$, where $e_k := \widehat{g}_k - \nabla_v J_\tau(v_k)$. The exact Gibbs hypergradient is $\nabla_v J_\tau(v_k) = \mathbb{E}_{\pi_{v_k}}[\varphi_2] - \frac{1}{\tau}\mathbb{E}_{\pi_{v_k}}[\varphi_4] + \frac{1}{\tau}\mathbb{E}_{\pi_{v_k}}[\varphi_1]\mathbb{E}_{\pi_{v_k}}[\varphi_3]$, whereas the estimator is $\widehat{g}_k = \widehat{\mathbb{E}}_k[\varphi_2] - \frac{1}{\tau}\widehat{\mathbb{E}}_k[\varphi_4] + \frac{1}{\tau}\widehat{\mathbb{E}}_k[\varphi_1]\widehat{\mathbb{E}}_k[\varphi_3]$. Using $a_1b_1 - a_2b_2 = b_2(a_1-a_2) + a_1(b_1-b_2)$ gives

$$e_k = \widehat{\mathbb{E}}_k[\varphi_2] - \mathbb{E}_{\pi_{v_k}}[\varphi_2] - \frac{1}{\tau}\big(\widehat{\mathbb{E}}_k[\varphi_4] - \mathbb{E}_{\pi_{v_k}}[\varphi_4]\big) + \frac{1}{\tau}\mathbb{E}_{\pi_{v_k}}[\varphi_3]\big(\widehat{\mathbb{E}}_k[\varphi_1] - \mathbb{E}_{\pi_{v_k}}[\varphi_1]\big) + \frac{1}{\tau}\widehat{\mathbb{E}}_k[\varphi_1]\big(\widehat{\mathbb{E}}_k[\varphi_3] - \mathbb{E}_{\pi_{v_k}}[\varphi_3]\big).$$

By Remark A.1, $\|\mathbb{E}_{\pi_{v_k}}[\varphi_3]\| \le C_\ell$, and Assumption 4.6 gives $|\widehat{\mathbb{E}}_k[\varphi_1]| \le C_f$. Therefore,

$$\begin{aligned}\mathbb{E}\|e_k\|^2 \le{}& 4\mathbb{E}\|\widehat{\mathbb{E}}_k[\varphi_2] - \mathbb{E}_{\pi_{v_k}}[\varphi_2]\|^2 + \frac{4}{\tau^2}\mathbb{E}\|\widehat{\mathbb{E}}_k[\varphi_4] - \mathbb{E}_{\pi_{v_k}}[\varphi_4]\|^2\\ &+ \frac{4C_\ell^2}{\tau^2}\mathbb{E}|\widehat{\mathbb{E}}_k[\varphi_1] - \mathbb{E}_{\pi_{v_k}}[\varphi_1]|^2 + \frac{4C_f^2}{\tau^2}\mathbb{E}\|\widehat{\mathbb{E}}_k[\varphi_3] - \mathbb{E}_{\pi_{v_k}}[\varphi_3]\|^2.\end{aligned} \tag{33}$$

Thus, controlling the hypergradient error reduces to controlling $\mathbb{E}\|\widehat{\mathbb{E}}_k[\varphi_i] - \mathbb{E}_{\pi_{v_k}}[\varphi_i]\|^2$.

Define

$$A_\delta := 4\left(L_{f_v}^2 + \frac{L_{f\ell_v}^2}{\tau^2} + \frac{C_\ell^2 L_f^2}{\tau^2} + \frac{C_f^2 L_H^2}{\tau^2}\right). \tag{34}$$

Substituting Proposition A.3 into (33) yields

$$\mathbb{E}\|e_k\|^2 \le \frac{2A_\delta C_{\mathrm{erg}}^2(M_{4,0}+1)^2}{chS}e^{-2ch(T-S)} + 4A_\delta C_{\mathrm{b}}^2 h + \frac{4A_\delta D_2}{S}\left(1+\frac{2c_0e^{-ch}}{1-e^{-ch}}\right) + \frac{32C_f^2\sigma_{v,2}^2 p}{S\tau^2}. \tag{35}$$

Summing (22) over $k = 0, \dots, K-1$, using $J_\tau(v_K) \geq J_\tau^\star$, and dividing by $\beta K/8$ gives

$$\frac{1}{K}\sum_{k=0}^{K-1}\mathbb{E}\|\mathcal{G}_\beta(v_k)\|^2 \leq \frac{8(J_\tau(v_0)-J_\tau^\star)}{\beta K} + \frac{6}{K}\sum_{k=0}^{K-1}\mathbb{E}\|e_k\|^2. \tag{36}$$

Combining (35) and (36) proves

$$\frac{1}{K}\sum_{k=0}^{K-1}\mathbb{E}\|\mathcal{G}_\beta(v_k)\|^2 \leq \frac{8(J_\tau(v_0)-J_\tau^\star)}{\beta K} + \frac{12A_\delta C_{\mathrm{erg}}^2(M_{4,0}+1)^2}{chS}e^{-2ch(T-S)} + 24A_\delta C_{\mathrm{b}}^2 h + \frac{24A_\delta D_2}{S}\left(1+\frac{2c_0e^{-ch}}{1-e^{-ch}}\right) + \frac{192C_f^2\sigma_{v,2}^2 p}{S\tau^2}, \tag{37}$$

where A_δ is defined in (34), D_2 in (44), c, c_0 in (64), C_{b} in (67), and C_{erg} in (70). If $T = S + B$ for a fixed $B \geq 0$, then $1 - e^{-chS} \leq 1$ and $(1-e^{-xh})^{-1} = O(h^{-1})$ for fixed $x > 0$. This proves

$$\frac{1}{K}\sum_{k=0}^{K-1}\mathbb{E}\|\mathcal{G}_\beta(v_k)\|^2 = O\left(\frac{1}{K}+\frac{1}{S}+\frac{1}{Sh}+h\right)$$

and completes the proof of Theorem 4.10. □

A.7 Proofs Needed for Theorem 4.10

Lemma A.4. *Under Assumption 4.1, for $q = 2, 3, 4$ we obatined*

$$\|H_v(\theta,\xi)\|^q \leq K_q(1+\|\theta\|^q), \quad \textit{where} \quad K_q := 2^{q-1}\max\{L_H^q, G_0^q\}. \tag{38}$$

Moreover, for every fixed $v \in \mathcal{V}$, when stepsize $h \leq \min\{1, a/K_2\}$,

$$(Q_{v,h}V_2)(\theta) \leq (1-ah)V_2(\theta) + B_2 h, \tag{39}$$

where

$$V_2(\theta) := 1 + \|\theta\|^2, \quad \textit{and} \quad B_2 := a + 2b + K_2 + 2\tau d. \tag{40}$$

Proof. By L_H-smoothness of l and $\sup_{v\in\mathcal{V},\xi}\|\nabla_\theta l(v,0;\xi)\| \leq G_0$ of Assumption 4.1,

$$\|H_v(\theta,\xi)\| \leq \|H_v(\theta,\xi) - H_v(0,\xi)\| + \|H_v(0,\xi)\| \leq L_H\|\theta\| + G_0.$$

For every $q \geq 1$, $(x+y)^q \leq 2^{q-1}(x^q + y^q)$ for $x, y \geq 0$. Hence

$$\|H_v(\theta,\xi)\|^q \leq 2^{q-1}L_H^q\|\theta\|^q + 2^{q-1}G_0^q \leq K_q(1+\|\theta\|^q),$$

where one may take

$$K_q := 2^{q-1}\max\{L_H^q, G_0^q\}, \qquad q \in \{2,3,4\}. \tag{41}$$

Now condition on the present state θ. Expanding one SGLD step $\theta^+ = \theta - hH_v(\theta,\xi) + \sqrt{2\tau h}\varepsilon$ and using $\mathbb{E}\varepsilon = 0$, $\mathbb{E}\|\varepsilon\|^2 = d$, and independence of ε and ξ, we obtain

$$\begin{aligned}\mathbb{E}[\|\theta^+\|^2 \mid \theta] &= \mathbb{E}\left[\|\theta - hH_v(\theta,\xi)\|^2 \mid \theta\right] + 2\tau hd \\ &= \|\theta\|^2 - 2h\langle\theta, \mathbb{E}[H_v(\theta,\xi)\mid\theta]\rangle + h^2\mathbb{E}[\|H_v(\theta,\xi)\|^2\mid\theta] + 2\tau hd \\ &= \|\theta\|^2 - 2h\langle\theta, h_v(\theta)\rangle + h^2\mathbb{E}\|H_v(\theta,\xi)\|^2 + 2\tau hd.\end{aligned}$$

By dissipativity condition (see Lemma 4.4) implied by Assumption 4.1 and the gradient growth bound (38) derived above,

$$\begin{aligned}\mathbb{E}[\|\theta^+\|^2 \mid \theta] &\leq \|\theta\|^2 - 2h(a\|\theta\|^2 - b) + K_2h^2(1+\|\theta\|^2) + 2\tau hd \\ &= (1-2ah+K_2h^2)\|\theta\|^2 + \big(2b + K_2h + 2\tau d\big)h.\end{aligned}$$

Set $h \leq \min\{1, a/K_2\}$, we have $1 - 2ah + K_2h^2 \leq 1 - ah$ and

$$\begin{aligned}(Q_{v,h}V_2)(\theta) &= 1 + \mathbb{E}[\|\theta^+\|^2 \mid \theta] \\ &\leq 1 + (1-ah)\|\theta\|^2 + \big(2b + K_2 + 2\tau d\big)h \\ &= (1-ah)(1+\|\theta\|^2) + \big(a + 2b + K_2 + 2\tau d\big)h.\end{aligned}$$

This completes the proof. □

Lemma A.5. *Suppose Assumptions 4.1 and 4.9 hold. Consider either the independently restarted algorithm, for which $\mu_{k,0} = \mu_0$ for every k, or the law-level warm-start variant satisfying $\mu_{k,0} = \mu_{k-1,T}$ and start with $\mu_{0,0} = \mu_0$. If*

$$0 < h \le \min\{1, a/K_2, 1/a\}, \tag{42}$$

then

$$\sup_{k\ge 0}\sup_{0\le t\le T}\int_{\mathbb{R}^d}\|\theta\|^2\mu_{k,t}(\omega, d\theta) \le D_2 \qquad \mathbb{P}\text{-almost surely}, \tag{43}$$

where

$$D_2 := \max\left\{D_{2,0}, \frac{2b + K_2 + 2\tau d}{a}\right\}, \quad K_2 = 2\max\{L_H^2, G_0^2\}. \tag{44}$$

Consequently, $\sup_{k\ge 0}\sup_{0\le t\le T}\mathbb{E}\|\theta_{k,t}\|^2 \le D_2$.

Proof. Set $\rho := 1 - ah$. Since $0 < h \le 1/a$, we have $\rho \in [0, 1)$. Under Assumption 4.1 and $h \le \min\{1, a/K_2, 1/a\}$, we have the drift inequality (39) from Lemma A.4, uniformly over $v \in \mathcal{V}$,

$$\begin{aligned}(Q_{v,h}\|\cdot\|^2)(\theta) &= (Q_{v,h}V_2)(\theta) - 1\\ &\le (1 - ah)(1 + \|\theta\|^2) + B_2h - 1\\ &= \rho\|\theta\|^2 + (1-\rho)\frac{B_2 - a}{a}.\end{aligned} \tag{45}$$

When the law-level warm-start identity $\mu_{k,0} = \mu_{k-1,T}$ are used, fix $\omega \in \Omega_0$ and define

$$M_{k,t}(\omega) := \int_{\mathbb{R}^d}\|\theta\|^2\,\mu_{k,t}(\omega, d\theta).$$

For $t = 0, \dots, T-1$, the conditional kernel evolution and (45) give

$$M_{k,t+1}(\omega) = \int_{\mathbb{R}^d}\big(Q_{v_k(\omega),h}\|\cdot\|^2\big)(\theta)\,\mu_{k,t}(\omega, d\theta) \le \rho M_{k,t}(\omega) + (1-\rho)\frac{B_2 - a}{a}.$$

For the independent restart, $M_{k,0} \le D_{2,0}$ for every k by Assumption 4.9. So, with D_2 defined in 44,

$$M_{k,t}(w) \le \rho^t D_{2,0} + (1 - \rho^t)\frac{B_2 - a}{a} \le D_2.$$

For the law-level warm start case, at the boundary between two consecutive outer iterations, the law-level warm-start identity $\mu_{k,0} = \mu_{k-1,T}$ gives $M_{k,0}(\omega) = M_{k-1,T}(\omega)$, for $k \ge 1$. Therefore, iterating the preceding recursion across both inner and outer iterations yields

$$M_{k,t}(\omega) \le \rho^{kT+t}M_{0,0}(\omega) + \big(1 - \rho^{kT+t}\big)\frac{B_2 - a}{a}. \tag{46}$$

By Assumption 4.9, $M_{0,0}(\omega) \le D_{2,0}$. Hence,

$$M_{k,t}(\omega) \le \max\{D_{2,0}, \frac{B_2 - a}{a}\} = D_2$$

for every $k \ge 0$ and $0 \le t \le T$. This proves (43).

Finally, by the definition of the conditional law and the tower property,

$$\mathbb{E}\|\theta_{k,t}\|^2 = \mathbb{E}\left[\int_{\mathbb{R}^d}\|\theta\|^2\,\mu_{k,t}(d\theta)\right] \le D_2.$$

Taking the supremum over k and t proves $\sup_{k\ge 0}\sup_{0\le t\le T}\mathbb{E}\|\theta_{k,t}\|^2 \le D_2$. □

Lemma A.6. *Suppose Assumption 4.1 holds. For any fixed $v \in \mathcal{V}$, let $Y := \theta - hH_v(\theta, \xi)$. Then, for every stepsize $h \le h_Y$ satisfying* (50)*,*

$$\mathbb{E}[\|Y\|^4 \mid \theta] \le (1 - ah)\|\theta\|^4 + B_Y h, \tag{47}$$

where

$$B_Y := \frac{(4b + 6K_2)^2}{a} + 3K_3\left(\frac{4K_3}{a}\right)^{1/3} + K_4. \tag{48}$$

Proof. For brevity, let $r = \|\theta\|$, $A := \langle \theta, H_v(\theta, \xi)\rangle$ and $B := \|H_v(\theta, \xi)\|^2$. Then,

$$\begin{aligned}\|Y\|^4 &= r^4 - 4hr^2A + 2h^2r^2B + 4h^2A^2 - 4h^3AB + h^4B^2 \\ &\le r^4 - 4hr^2A + 6h^2r^2B + 4h^3rB^{3/2} + h^4B^2,\end{aligned}$$

where the last inequality follows from Cauchy–Schwarz $A^2 \le r^2B$ and $|AB| \le rB^{3/2}$.

Taking conditional expectation given θ, using $\mathbb{E}[A \mid \theta] = \langle \theta, h_v(\theta)\rangle$, and then using dissipativity condition (see Lemma 4.4) implied by Assumption 4.1, we obtain

$$\begin{aligned}\mathbb{E}[\|Y\|^4 \mid \theta] &\le r^4 - 4hr^2\langle \theta, h_v(\theta)\rangle + 6h^2r^2\mathbb{E}[B \mid \theta] + 4h^3r\mathbb{E}[B^{3/2} \mid \theta] + h^4\mathbb{E}[B^2 \mid \theta] \\ &\le r^4 - 4ahr^4 + 4bhr^2 \\ &\quad + 6h^2r^2\mathbb{E}\|H_v(\theta, \xi)\|^2 \\ &\quad + 4h^3r\mathbb{E}\|H_v(\theta, \xi)\|^3 + h^4\mathbb{E}\|H_v(\theta, \xi)\|^4.\end{aligned}$$

Applying the gradient growth bound (38) with $q = 2, 3, 4$ gives

$$\begin{aligned}\mathbb{E}[\|Y\|^4 \mid \theta] \le \big(1 - 4ah + 6K_2h^2 + 4K_3h^3 + K_4h^4\big)r^4 \\ + \big(4bh + 6K_2h^2\big)r^2 + 4K_3h^3r + K_4h^4.\end{aligned} \tag{49}$$

Choose $h_Y \le 1$ sufficiently small that

$$6K_2h^2 + 4K_3h^3 + K_4h^4 \le ah, \qquad 0 < h \le h_Y. \tag{50}$$

Then the coefficient of r^4 in (49) is at most $1 - 3ah$.

Since $h \le 1$, $(4bh + 6K_2h^2)r^2 \le (4b + 6K_2)hr^2$ and $4K_3h^3r \le 4K_3hr$. By Young's inequality, for every $\epsilon_2 > 0$ and $r \ge 0$, we have $r^2 \le \epsilon_2 r^4 + 1/4\epsilon_2$. Taking $\epsilon_2 = a/(4(4b + 6K_2))$ gives

$$(4b + 6K_2)hr^2 \le \frac{ah}{4}r^4 + \frac{(4b + 6K_2)^2}{a}h. \tag{51}$$

Similarly, for every $\epsilon_1 > 0$, $r \le \epsilon_1 r^4 + \frac{3}{4}(4\epsilon_1)^{-1/3}$. Taking $\epsilon_1 = a/(16K_3)$ gives

$$4K_3hr \le \frac{ah}{4}r^4 + 3K_3\left(\frac{4K_3}{a}\right)^{1/3}h. \tag{52}$$

Substituting (51) and (52) into (49), and using $K_4h^4 \le K_4h$, yields

$$\begin{aligned}\mathbb{E}[\|Y\|^4 \mid \theta] &\le \left(1 - 3ah + \frac{ah}{4} + \frac{ah}{4}\right)r^4 + \left(\frac{(4b + 6K_2)^2}{a} + 3K_3\left(\frac{4K_3}{a}\right)^{1/3} + K_4\right)h \\ &\le (1 - ah)r^4 + \left(\frac{(4b + 6K_2)^2}{a} + 3K_3\left(\frac{4K_3}{a}\right)^{1/3} + K_4\right)h.\end{aligned}$$

This completes the proof. □

Lemma A.7. *Suppose Assumption 4.1 holds, and let*

$$h_4 := \min\left\{h_Y, 1, \frac{a}{K_2}\right\}, \tag{53}$$

where h_Y is defined in (50)*. Then, for every fixed $v \in \mathcal{V}$ and every stepsize $0 < h \le h_4$,*

$$(Q_{v,h}V_4)(\theta) \le (1 - \alpha_4 h)V_4(\theta) + B_4h + B_4'h^2, \qquad V_4(\theta) := 1 + \|\theta\|^4, \tag{54}$$

where constants

$$\alpha_4 := \frac{a}{2}, \quad B_4 := \frac{a}{2} + B_Y + \frac{8\tau^2(d + 2)^2}{a}, \quad B_4' := 4\tau(2b + K_2)(d + 2) + 4\tau^2d(d + 2), \tag{55}$$

and B_Y is defined in (48)*.*

Proof. Let $Z := \sqrt{2\tau h}\,\varepsilon$. For any random vector $Y \in \mathbb{R}^d$ that is independent of Z,

$$\|Y+Z\|^4 = \|Y\|^4 + 4\langle Y,Z\rangle^2 + \|Z\|^4 + 4\|Y\|^2\langle Y,Z\rangle + 2\|Y\|^2\|Z\|^2 + 4\langle Y,Z\rangle\|Z\|^2.$$

The two odd terms have expectation zero. Moreover, conditioning on Y ad taking expectation over Z,

$$\mathbb{E}_\varepsilon\|Z\|^2 = 2\tau h d, \quad \mathbb{E}_\varepsilon\langle Y,Z\rangle^2 = 2\tau h\|Y\|^2, \quad \mathbb{E}_\varepsilon\|Z\|^4 = (2\tau h)^2\mathbb{E}\|\varepsilon\|^4 = 4\tau^2h^2d(d+2).$$

Consequently,

$$\begin{aligned}\mathbb{E}_\varepsilon\|Y+Z\|^4 &= \|Y\|^4 + 4(2\tau h)\|Y\|^2 + 2\|Y\|^2(2\tau h d) + 4\tau^2h^2d(d+2)\\ &= \|Y\|^4 + 4\tau h(d+2)\|Y\|^2 + 4\tau^2h^2d(d+2). \end{aligned} \tag{56}$$

Apply (56) conditionally with $Y = \theta - hH_v(\theta,\xi)$ and then average over ξ:

$$\mathbb{E}[\|\theta^+\|^4 \mid \theta] = \mathbb{E}[\|Y\|^4 \mid \theta] + 4\tau h(d+2)\mathbb{E}[\|Y\|^2 \mid \theta] + 4\tau^2h^2d(d+2). \tag{57}$$

The same second-moment expansion as in Lemma A.4, but without the Gaussian term, gives

$$\begin{aligned}\mathbb{E}[\|Y\|^2 \mid \theta] &= \|\theta\|^2 - 2h\langle h_v(\theta),\theta\rangle + h^2\mathbb{E}\|H_v(\theta,\xi)\|^2\\ &\le (1-2ah+K_2h^2)\|\theta\|^2 + (2b+K_2h)h.\end{aligned}$$

Take $h \le \min\{1, a/K_2\}$, this implies $1-2ah+K_2h^2 \le 1$, we have

$$\mathbb{E}[\|Y\|^2 \mid \theta] \le \|\theta\|^2 + (2b+K_2)h. \tag{58}$$

Combining (47) from Lemma A.6, (57), and (58) yields

$$\begin{aligned}\mathbb{E}[\|\theta^+\|^4 \mid \theta] \le{}& (1-ah)\|\theta\|^4 + B_Yh + 4\tau h(d+2)\|\theta\|^2\\ &+ 4\tau(2b+K_2)(d+2)h^2 + 4\tau^2d(d+2)h^2.\end{aligned}$$

Applying $4\tau(d+2)\|\theta\|^2 \le \frac{a}{2}\|\theta\|^4 + \frac{8\tau^2(d+2)^2}{a}$ by Young's inequality, we have

$$\begin{aligned}\mathbb{E}[\|\theta^+\|^4 \mid \theta] \le{}& \left(1-\frac{a}{2}h\right)\|\theta\|^4\\ &+ \left(B_Y + \frac{8\tau^2(d+2)^2}{a}\right)h\\ &+ \big(4\tau(2b+K_2)(d+2) + 4\tau^2d(d+2)\big)\,h^2.\end{aligned}$$

Adding 1 to both sides gives (54) and completes the proof. □

Lemma A.8. *Suppose that Assumption 4.1 holds and that the step size h satisfies*

$$\begin{aligned}h^{1/2}|\log h|^{1/2} &\le \min\left(\frac{1}{6c'}e^{-2c_fR}\kappa, \sqrt{\bar c}\right),\\ h &\le \bar c^{-1/2}h^{1/2}|\log h|^{1/2}(L_HR)^{-1},\\ h &\le \min\left(\frac{1}{2L_H}, \frac{R^2}{9}, 1\right).\end{aligned} \tag{59}$$

and

$$h \le \min\left\{\frac{\bar cR^2}{128\log 5\,\tau}, \frac{\bar cR^2}{128\big(2c_fR+\log(18/\kappa)\big)\tau}, \frac{\bar cR^2/32}{\tau\log\Big(\frac{45}{2}\left(1+\sqrt{2\tau^{-1}}\,c_f^{-1}e^{2c_fR}L_HR\right)\Big)}\right\}, \tag{60}$$

where $c_f > 0$ is a constant satisfying

$$\frac{\sqrt{2\tau}}{2}\,c_fR^{-1} - L_H \ge 0, \tag{61}$$

and

$$\bar{c} := \frac{1}{2e(1+C)}, \qquad c' := \bar{c}^{-1/2}\left(\frac{2L_H}{R/2-1} + L_H c_f e^{-c_f(R/2-1)}\right) + \frac{4\bar{c}^{-1/2}}{R}, \tag{62}$$

with C depending only on the underlying Hilbert space. For every $v \in \mathcal{V}$ and every pair of initial distributions $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$, the parameterized Markov kernel $Q_{v,h}$ satisfies

$$W_1(\mu Q_{v,h}^n, \nu Q_{v,h}^n) \le c_0 e^{-chn} W_1(\mu, \nu), \qquad n \ge 0, \tag{63}$$

where

$$c = \frac{1}{3} e^{-2c_f R} \min\left(\sqrt{2\tau}\,\frac{c_f R}{2}, \kappa\right) \quad \textit{and} \quad c_0 = e^{2c_f R}. \tag{64}$$

Proof. Assumption 4.1, together with Remark 4.2, ensures that Assumption 2.1 of Li et al. (2024) is satisfied. Moreover, (59) and (60) ensure that the step-size conditions required by Theorem 2.1 of Li et al. (2024) hold. The desired contraction bound therefore follows directly from that theorem. □

Lemma A.9. *Suppose Assumption 4.1 holds and the constant stepsize h satisfies $0 < h \le h_4$, with h_4 defined in* (53)*, as well as the contraction step-size conditions* (59) *and* (60)*. Then, for every deterministic $v \in \mathcal{V}$, the SGLD kernel $Q_{v,h}$ admits a unique invariant probability measure $\bar{\pi}_{v,h}$. Moreover,*

$$\sup_{v\in\mathcal{V}} \int V_4(\theta)\,\bar{\pi}_{v,h}(d\theta) \le \frac{B_4 + B_4' h}{\alpha_4}, \tag{65}$$

where α_4, B_4, B_4' are defined in (55)*.*

Proof. Under Assumption 4.1 and the stepsize restriction in Lemma A.8, the conditions of Corollary 2.1 in Li et al. (2024) are satisfied for every deterministic $v \in \mathcal{V}$. Hence, the SGLD kernel $Q_{v,h}$ admits a unique invariant probability measure $\bar{\pi}_{v,h}$.

Fix $v \in \mathcal{V}$ and $r > 0$, and define the bounded truncation $V_{4,r}(\cdot) = \min\{V_4(\cdot), r\}$. On the set $\{V_4(\theta) \le r\}$, we have $V_{4,r}(\theta) = V_4(\theta)$ and $V_{4,r} \le V_4$. Therefore,

$$Q_{v,h}V_{4,r}(\theta) - V_{4,r}(\theta) \le Q_{v,h}V_4(\theta) - V_4(\theta) \le -\alpha_4 h V_4(\theta) + (B_4 + B_4' h)h,$$

where the last inequality follows from Lemma A.7.

On the other hand, on the set $\{V_4(\theta) > r\}$, we have $V_{4,r}(\theta) = r$ and $V_{4,r} \le r$. Hence,

$$Q_{v,h}V_{4,r}(\theta) - V_{4,r}(\theta) \le Q_{v,h}r - r = 0.$$

Combining the two cases yields

$$Q_{v,h}V_{4,r}(\theta) - V_{4,r}(\theta) \le 1_{\{V_4(\theta)\le r\}}\left[-\alpha_4 h V_4(\theta) + (B_4 + B_4' h)h\right].$$

Integrate against the invariant distribution $\bar{\pi}_{v,h}$, we have

$$0 = \int Q_{v,h}V_{4,r} - V_{4,r}\,d\bar{\pi}_{v,h} \le -\alpha_4 h \int_{\{V_4\le r\}} V_4\,d\bar{\pi}_{v,h} + (B_4 h + B_4' h^2)\bar{\pi}_{v,h}(V_4 \le r).$$

Since $\bar{\pi}_{v,h}(V_4 \le r) \le 1$, it follows that

$$\int_{\{V_4\le r\}} V_4\,d\bar{\pi}_{v,h} \le \frac{B_4 + B_4' h}{\alpha_4}.$$

Letting $r \to \infty$ and applying the monotone convergence theorem(Folland, 1999, Theorem 2.14), we obtain

$$\int V_4(\theta)\,\bar{\pi}_{v,h}(d\theta) \le \frac{B_4 + B_4' h}{\alpha_4}.$$

Since the right-hand side is independent of v, taking the supremum over $v \in \mathcal{V}$ proves (65), which completes the proof.

□

Lemma A.10. *Suppose Assumptions 4.1 and 4.5 hold. Suppose* $0 < h \le h_4$*, with* h_4 *defined in* (53)*, and that* (59) *and* (60) *hold. Then, we have*

$$\sup_{v\in\mathcal{V}} W_1(\bar{\pi}_{v,h}, \pi_v) \le C_{\mathrm{b}}\sqrt{h}, \tag{66}$$

where

$$C_{\mathrm{b}} := 2\sqrt{\frac{C_2}{C_1}\left(e^{C_1(\max\{1,c^{-1}\log(2c_0)\}+h_4)} - 1\right)}, \tag{67}$$

$$C_1 := 1 + 4L_H + 2L_H^2, \tag{68}$$

$$C_2 := 2(\sigma_{\theta,2}^2 d + 2L_H^2(L_H^2 M_{\pi,2} + G_0^2 + \tau d)), \tag{69}$$

and $M_{\pi,2}$ *is defined in* (72)*. Moreover, define*

$$M_{\bar{\pi},1} := \left(\frac{B_4 + B_4' h_4}{\alpha_4}\right)^{1/4}, \qquad C_{\mathrm{erg}} := c_0(1 + M_{\bar{\pi},1}), \tag{70}$$

where α_4*,* B_4*,* B_4' *are defined in 55 and c,* c_0 *are defined in* (64)*. Then, for every probability measure* μ *with finite fourth moment and every* $n \ge 0$*, the SGLD kernel* $Q_{v,h}$ *has the following contraction relation*

$$W_1(\mu Q_{v,h}^n, \pi_v) \le C_{\mathrm{erg}} e^{-chn}\left(\int_{\mathbb{R}^d} \|\theta\|^4 \mu(d\theta) + 1\right) + C_{\mathrm{b}}\sqrt{h}, \tag{71}$$

uniformly in $v \in \mathcal{V}$*.*

Proof. Fix a $v \in \mathcal{V}$. Recall that $H_v(x,\xi) := \nabla_\theta l(v,x;\xi)$ and $h_v(x) := \mathbb{E}_\xi[H_v(x,\xi)] = \nabla_\theta \ell(v,x)$. Also, Assumption 4.1 implies that, uniformly in v, we have $\|h_v(x) - h_v(y)\| \le L_H\|x-y\|$ and $\|h_v(x)\| \le L_H\|x\| + G_0$.

We first verify a uniform second-moment bound for the Gibbs measures. By Lemma 4.4, there are constants $\alpha_\ell > 0$ and b_ℓ, independent of v, such that $\ell(v,x) \ge \alpha_\ell\|x\|^2 - b_\ell$. Moreover, compactness of $\mathcal{V}$ and continuity of ℓ imply $M_{\ell,B} := \sup_{\substack{v\in\mathcal{V}\\ \|x\|\le 1}} \ell(v,x) < \infty$. Consequently, the normalizing constant satisfies

$$Z(v) = \int_{\mathbb{R}^d} e^{-\ell(v,x)/\tau}\, dx \ge |B(0,1)| e^{-M_{\ell,B}/\tau} =: z_\star > 0.$$

It follows that

$$\sup_{v\in\mathcal{V}} \int_{\mathbb{R}^d} \|x\|^2\, \pi_v(dx) \le \frac{e^{b_\ell/\tau}}{z_\star} \int_{\mathbb{R}^d} \|x\|^2 e^{-\alpha_\ell\|x\|^2/\tau}\, dx =: M_{\pi,2} < \infty. \tag{72}$$

The same argument gives finite moments of every polynomial order.

Consider the Langevin diffusion

$$dL_t = -h_v(L_t)\, dt + \sqrt{2\tau}\, dB_t, \qquad L_0 \sim \pi_v,$$

where $(B_t)_{t\ge 0}$ be a standard d-dimensional Brownian motion and L_0 is independent of B. The global Lipschitz continuity of h_v ensures that this SDE has a unique strong solution. Since $h_v = \nabla_\theta \ell(v,\cdot)$, the Gibbs measure π_v is invariant. Thus, $\mathcal{L}(L_t) = \pi_v$, $t \ge 0$. In particular,

$$\sup_{v\in\mathcal{V}} \sup_{t\ge 0} \mathbb{E}\|h_v(L_t)\|^2 \le 2L_H^2 M_{\pi,2} + 2G_0^2 =: M_H.$$

On the same probability space, let $(\xi_j)_{j\ge 0}$ be independent mini-batches with the prescribed data law, independent of (L_0, B). Recall the SGLD

$$X_0 = L_0, \qquad X_{j+1} = X_j - hH_v(X_j,\xi_j) + \sqrt{2\tau}\big(B_{(j+1)h} - B_{jh}\big).$$

This chain has transition kernel $Q_{v,h}$, and $\mathcal{L}(X_j) = \pi_v Q_{v,h}^j$.

Define the filtration

$$\mathcal{F}_j := \sigma\big(L_0,\ B_s : 0 \le s \le jh,\ \xi_0, \dots, \xi_{j-1}\big).$$

Let $e_j := X_j - L_{jh}$. Integrating the Langevin SDE over $[jh, (j+1)h]$ and subtracting it from the SGLD update gives the exact identity

$$e_{j+1} = e_j - h\big(h_v(X_j) - h_v(L_{jh})\big) - h\underbrace{\big(H_v(X_j, \xi_j) - h_v(X_j)\big)}_{\zeta_{\theta,j}} + \underbrace{\int_{jh}^{(j+1)h} \big(h_v(L_s) - h_v(L_{jh})\big)\, ds}_{r_j}.$$

We next bound the remainder r_j. For $s \in [jh, (j+1)h]$, write $u = s - jh$. The integrated SDE, the inequality $\|x+y\|^2 \le 2\|x\|^2 + 2\|y\|^2$, and Cauchy–Schwarz give

$$\begin{aligned}
\mathbb{E}\|L_s - L_{jh}\|^2 &\le 2\mathbb{E}\left\|\int_{jh}^{s} h_v(L_r)\, dr\right\|^2 + 4\tau\mathbb{E}\|B_s - B_{jh}\|^2 \\
&\le 2u\int_{jh}^{s} \mathbb{E}\|h_v(L_r)\|^2\, dr + 4\tau du \\
&\le 2M_H u^2 + 4\tau du \\
&\le (2M_H + 4\tau d)(s - jh),
\end{aligned}$$

where the last inequality is due to $u \le h \le h_4 \le 1$.

Applying Cauchy–Schwarz once more, together with the L_H-Lipschitz continuity of h_v, yields

$$\begin{aligned}
\mathbb{E}\|r_j\|^2 &\le h\int_{jh}^{(j+1)h} \mathbb{E}\|h_v(L_s) - h_v(L_{jh})\|^2\, ds \\
&\le hL_H^2 \int_{jh}^{(j+1)h} \mathbb{E}\|L_s - L_{jh}\|^2\, ds \\
&\le hL_H^2(2M_H + 4\tau d)\int_0^h u\, du = \frac{L_H^2(2M_H + 4\tau d)}{2} h^3.
\end{aligned}$$

To derive the error recursion, define $A_j := e_j - h\big(h_v(X_j) - h_v(L_{jh})\big)$. Then A_j is $\mathcal{F}_j$-measurable, the L_H-Lipschitz continuity of h_v and the conditional centering of $\zeta_{\theta,j}$ by Assumption 4.5 implies

$$\|A_j\| \le (1 + L_H h)\|e_j\| \quad \text{and} \quad \mathbb{E}\langle A_j, \zeta_{\theta,j}\rangle = \mathbb{E}\left[\langle A_j, \mathbb{E}[\zeta_{\theta,j} \mid \mathcal{F}_j]\rangle\right] = 0.$$

Hence

$$\mathbb{E}\|A_j - h\zeta_{\theta,j}\|^2 = \mathbb{E}\|A_j\|^2 + h^2\mathbb{E}\|\zeta_{\theta,j}\|^2 \le (1 + L_H h)^2\mathbb{E}\|e_j\|^2 + \sigma_{\theta,2}^2 dh^2.$$

Using Young's inequality in the form $\|x+y\|^2 \le (1+h)\|x\|^2 + (1+h^{-1})\|y\|^2$, we conclude that

$$\begin{aligned}
\mathbb{E}\|e_{j+1}\|^2 &\le (1+h)\mathbb{E}\|A_j - h\zeta_{\theta,j}\|^2 + (1+h^{-1})\mathbb{E}\|r_j\|^2 \\
&\le (1+h)(1 + L_H h)^2\mathbb{E}\|e_j\|^2 + (1+h)d\sigma_{\theta,2}^2 h^2 + (1+h^{-1})\frac{L_H^2(2M_H + 4\tau d)}{2} h^3 \\
&\le (1 + C_1 h)\mathbb{E}\|e_j\|^2 + C_2 h^2,
\end{aligned}$$

where, since $h \le 1$, one may take $C_1 := 1 + 4L_H + 2L_H^2$ and $C_2 := 2(\sigma_{\theta,2}^2 d + \frac{L_H^2(2M_H + 4\tau d)}{2})$.

Since $e_0 = 0$, iterating this recursion gives, for every $j \ge 1$,

$$\mathbb{E}\|e_j\|^2 \le C_2 h^2 \sum_{q=0}^{j-1} (1 + C_1 h)^q = \frac{C_2}{C_1} h\big((1 + C_1 h)^j - 1\big) \le \frac{C_2}{C_1} h\big(e^{C_1 jh} - 1\big). \tag{73}$$

According to Lemma A.8, with stepsize satisfying conditions (59) and (60), we have

$$W_1(\mu Q_{v,h}^n, \nu Q_{v,h}^n) \le c_0 e^{-chn} W_1(\mu, \nu),$$

where c and c_0 are defined in (64). Choose $t_\circ := \max\{1, c^{-1}\log(2c_0)\}$ and $n_h := \lceil \frac{t_\circ}{h} \rceil$. As stepsize h also satisfies $h < h_4$, we have

$$t_\circ \le n_h h < t_\circ + h \le t_\circ + h_4.$$

Then, by (73),

$$\mathbb{E}\|X_{n_h} - L_{n_h h}\|^2 \le \frac{C_2}{C_1}\left(e^{C_1 n_h h} - 1\right) h \le \frac{C_2}{C_1}\left(e^{C_1(t_\circ + h_4)} - 1\right) h,$$

Since $\mathcal{L}(L_{n_h h}) = \pi_v$, the definition of W_1 and Cauchy–Schwarz imply

$$W_1(\pi_v Q_{v,h}^{n_h}, \pi_v) \le \mathbb{E}\|X_{n_h} - L_{n_h h}\| \le \left(\mathbb{E}\|X_{n_h} - L_{n_h h}\|^2\right)^{1/2} \le \sqrt{\frac{C_2}{C_1}\left(e^{C_1(t_\circ + h_4)} - 1\right) h}.$$

By Lemma A.9, $Q_{v,h}$ has an invariant probability measure $\bar{\pi}_{v,h}$ with a finite fourth moment. Thus $W_1(\bar{\pi}_{v,h}, \pi_v) < \infty$. Using invariance, the triangle inequality, and the contraction estimate gives

$$\begin{aligned} W_1(\bar{\pi}_{v,h}, \pi_v) &\le W_1\big(\bar{\pi}_{v,h} Q_{v,h}^{n_h}, \pi_v Q_{v,h}^{n_h}\big) + W_1(\pi_v Q_{v,h}^{n_h}, \pi_v) \\ &\le c_0 e^{-chn_h} W_1(\bar{\pi}_{v,h}, \pi_v) + \sqrt{\frac{C_2}{C_1}\left(e^{C_1(t_\circ + h_4)} - 1\right) h} \\ &\le \frac{1}{2} W_1(\bar{\pi}_{v,h}, \pi_v) + \sqrt{\frac{C_2}{C_1}\left(e^{C_1(t_\circ + h_4)} - 1\right) h}, \end{aligned}$$

where the last inequality follows from $c_0 e^{-chn_h} \le c_0 e^{-ct_\circ} \le 1/2$. Rearranging and then taking the supremum over v yields

$$\sup_{v \in \mathcal{V}} W_1(\bar{\pi}_{v,h}, \pi_v) \le 2\sqrt{\frac{C_2}{C_1}\left(e^{C_1(t_\circ + h_4)} - 1\right)}\sqrt{h}.$$

Finally, let μ be a probability measure with finite fourth moment. By invariance of $\bar{\pi}_{v,h}$, contraction, and the bias estimate just proved,

$$\begin{aligned} W_1(\mu Q_{v,h}^n, \pi_v) &\le W_1(\mu Q_{v,h}^n, \bar{\pi}_{v,h}) + W_1(\bar{\pi}_{v,h}, \pi_v) \\ &\le c_0 e^{-chn} W_1(\mu, \bar{\pi}_{v,h}) + 2\sqrt{\frac{C_2}{C_1}\left(e^{C_1(t_\circ + h_4)} - 1\right)}\sqrt{h}. \end{aligned}$$

Lemma A.9 and Hölder's inequality give

$$\int_{\mathbb{R}^d} \|x\|\, \bar{\pi}_{v,h}(dx) \le \left(\int_{\mathbb{R}^d} \|x\|^4\, \bar{\pi}_{v,h}(dx)\right)^{1/4} \le \left(\frac{B_4 + B_4' h_4}{\alpha_4}\right)^{1/4} = M_{\bar{\pi},1}.$$

Consequently,

$$\begin{aligned} W_1(\mu, \bar{\pi}_{v,h}) &\le \int_{\mathbb{R}^d} \|x\|\, \mu(dx) + \int_{\mathbb{R}^d} \|x\|\, \bar{\pi}_{v,h}(dx) \\ &\le \left(\int_{\mathbb{R}^d} \|x\|^4\, \mu(dx)\right)^{1/4} + M_{\bar{\pi},1} \\ &\le (1 + M_{\bar{\pi},1}) \left(1 + \int_{\mathbb{R}^d} \|x\|^4\, \mu(dx)\right). \end{aligned}$$

With $C_{\mathrm{erg}} := c_0(1 + M_{\bar{\pi},1})$, we obtain

$$W_1(\mu Q_{v,h}^n, \pi_v) \le C_{\mathrm{erg}} e^{-chn} \left(1 + \int_{\mathbb{R}^d} \|x\|^4\, \mu(dx)\right) + 2\sqrt{\frac{C_2}{C_1}\left(e^{C_1(t_\circ + h_4)} - 1\right)}\sqrt{h}, \quad n \ge 0.$$

This completes the proof. □

Lemma A.11. *For fixed parameter v, let $(X_t)_{t \ge 0}$ be a Markov chain with kernel $Q_{v,h}$, let $\mu_t = \mathcal{L}(X_t)$, and suppose the following contraction relation holds with constant $c_0, c > 0$*

$$W_1(\mu Q_{v,h}^t, \nu Q_{v,h}^t) \le c_0 e^{-cht} W_1(\mu, \nu), \qquad t \ge 0, \tag{74}$$

and $\sup_t \mathbb{E}\|X_t\|^2 \le D$ holds. If $\phi : \mathbb{R}^d \to \mathbb{R}^m$ is L_ϕ-Lipschitz, then, for every consecutive block I of S indices,

$$\mathbb{E}\left\| \frac{1}{S} \sum_{t \in I} \big(\phi(X_t) - \mathbb{E}_{\mu_t}\phi\big) \right\|^2 \le \frac{\Gamma_\phi(h)}{S}, \tag{75}$$

where

$$\Gamma_\phi(h) := L_\phi^2 D \left(1 + \frac{2c_0 e^{-ch}}{1 - e^{-ch}}\right). \tag{76}$$

Proof. Let $i < j$ and write $r := j - i$. By the Markov property, $\mathbb{E}\left[\phi(X_j) - \mathbb{E}_{\mu_j}\phi \mid X_i\right] = Q^r_{v,h}\phi(X_i) - \mu_i Q^r_{v,h}\phi$, because $\mu_j = \mu_i Q^r_{v,h}$. Hence

$$\mathbb{E}\left\langle \phi(X_i) - \mu_i\phi, \phi(X_j) - \mu_j\phi\right\rangle = \mathbb{E}_{\mu_i}\left\langle \phi - \mu_i\phi, Q^r_{v,h}\phi - \mu_i Q^r_{v,h}\phi\right\rangle. \tag{77}$$

For $x, y \in \mathbb{R}^d$, applying Kantorovich–Rubinstein Theorem (Villani, 2021, Theorem 1.14) and the contraction relation (74), we have

$$\|Q^r_{v,h}\phi(x) - Q^r_{v,h}\phi(y)\| \le L_\phi W_1(Q^r_{v,h}(x,\cdot), Q^r_{v,h}(y,\cdot)) \le c_0 e^{-chr} L_\phi \|x - y\|.$$

Thus, the Lipschitz constant

$$\mathrm{Lip}(Q^r_{v,h}\phi) \le c_0 e^{-chr} L_\phi. \tag{78}$$

For any Hilbert-valued L-Lipschitz map F, and independent $X, X' \sim \mu$, we have

$$\begin{aligned}
\mathbb{E}_\mu\|F - \mu F\|^2 &= \frac{1}{2}\mathbb{E}\|F(X) - F(X')\|^2 \\
&\le \frac{L^2}{2}\mathbb{E}\|X - X'\|^2 \\
&= L^2\left(\mathbb{E}\|X\|^2 - \|\mathbb{E}X\|^2\right) \\
&\le L^2\mathbb{E}\|X\|^2.
\end{aligned} \tag{79}$$

Applying (79) first to $F = \phi$ and then to $F = Q^r_{v,h}\phi$, and using (78), gives

$$\mathbb{E}_{\mu_i}\|\phi - \mu_i\phi\|^2 \le L^2_\phi D, \qquad \mathbb{E}_{\mu_i}\|Q^r_{v,h}\phi - \mu_i Q^r_{v,h}\phi\|^2 \le c_0^2 e^{-2chr} L^2_\phi D.$$

Applying Cauchy–Schwarz for (77) yields

$$|\mathbb{E}\left\langle \phi(X_i) - \mu_i\phi, \phi(X_j) - \mu_j\phi\right\rangle| \le c_0 e^{-chr} L^2_\phi D. \tag{80}$$

The diagonal terms satisfy $\mathbb{E}\|\phi(X_i) - \mu_i\phi\|^2 \le L^2_\phi D$. Expanding the squared norm of the average,

$$\begin{aligned}
&\mathbb{E}\left\|\frac{1}{S}\sum_{t\in I}(\phi(X_t) - \mu_t\phi)\right\|^2 \\
&= \frac{1}{S^2}\sum_{i\in I}\mathbb{E}\|\phi(X_i) - \mu_i\phi\|^2 + \frac{2}{S^2}\sum_{\substack{i,j\in I\\ i<j}}\mathbb{E}\left\langle \phi(X_i) - \mu_i\phi, \phi(X_j) - \mu_j\phi\right\rangle \\
&\le \frac{L^2_\phi D}{S^2}\left[S + 2c_0\sum_{r=1}^{S-1}(S-r)e^{-chr}\right] \\
&\le \frac{L^2_\phi D}{S}\left[1 + 2c_0\sum_{r=1}^{\infty}e^{-chr}\right] \\
&= \frac{L^2_\phi D}{S}\left(1 + \frac{2c_0 e^{-ch}}{1 - e^{-ch}}\right),
\end{aligned}$$

which proves (75). □

A.8 Statement and Proof for Result using warm-start

Let $\mathcal{G}_k$ denote the information available before initializing the inner chain at outer iteration k. Define $\mu_{k,t} := \mathcal{L}(\theta_{k,t} \mid \mathcal{G}_k)$. Conditional on $\mathcal{G}_k$, the outer variable v_k is fixed, and the inner laws satisfy $\mu_{k,t} = \mu_{k,0}Q^t_{v_k,h}$, almost surely, where the $Q^t_{v_k,h}$ is the transition kernel of inner SGLD. We use the same notation for these conditional probability measures and their realizations. We now state the convergence result of VCHD when applying warm-start initialization.

For the warm-start setting, to control fluctuations of this Gibbs family uniformly in v, we further impose the following Poincaré inequality assumption.

Assumption A.12. There exists a constant $C_P > 0$ such that, for every $v \in \mathcal{V}$ and $q \in H^1(\pi_v)$, $\mathrm{Var}_{\pi_v}(q) \le C_P\mathbb{E}_{\pi_v}\|\nabla_\theta q\|^2$.

Proposition A.13. *Suppose Assumptions 4.1, 4.6, A.12, 4.5, 4.8, and 4.9 hold. Suppose the law-level warm-start condition $\mu_{k,0} = \mu_{k-1,T}$ holds, with $\mu_{0,0} = \mu_0$. Let $1 \le S \le T$, fix $\tau > 0$, and choose $0 < \beta \le (2L_J)^{-1}$. Choose an inner stepsize $h > 0$, constant throughout the algorithm, such that* (42) *holds, $h \le h_4$ with h_4 defined in* (53)*, and* (59) *and* (60) *hold. If*

$$T - S \ge \max\left\{0,\ \left\lceil \frac{\log(6c_0^2(1 + 14A_\delta(C_pL_H/\tau)^2\beta^2))}{2ch} \right\rceil \right\}$$

then, for every $K \ge 1$,

$$\frac{1}{K}\sum_{k=0}^{K-1} \mathbb{E}\|\mathcal{G}_\beta(v_k)\|^2 = O\left(\frac{1}{K} + \frac{1}{S} + \frac{1}{Sh} + h\right).$$

Proof. We first define the retained-window distributional tracking bias

$$\delta_k := \max_{t=T-S+1,\dots,T} \mathbb{E}W_1(\mu_{k,t}, \pi_{v_k})^2. \tag{81}$$

Proceeding as in the proof of Theorem 4.10, the projected descent lemma A.2, gives

$$\frac{1}{K}\sum_{k=0}^{K-1} \mathbb{E}\|\mathcal{G}_\beta(v_k)\|^2 \le \frac{8}{\beta}\frac{J_\tau(v_0) - J_\tau^\star}{K} + \frac{6}{K}\sum_{k=0}^{K-1} \mathbb{E}\|e_k\|^2. \tag{82}$$

Thus, to establish the convergence rate, it remains to bound the mean-square hypergradient $\mathbb{E}\|e_k\|^2$. Lemma A.15 controls this error by the distributional tracking bias δ_k and the nonstationary sampling error $A_\delta(h)/S$:

$$\mathbb{E}\|e_k\|^2 \le A_\delta\delta_k + \frac{A_\delta(h)}{S}, \tag{83}$$

where $A_\delta(h) := A_\delta D_2\left(1 + \frac{2c_0e^{-ch}}{1-e^{-ch}}\right) + \frac{8C_f^2\sigma_{v,2}^2 p}{\tau^2}$ and D_2 defined in (44).

Then, according to Proposition A.14, the distributional tracking error δ_k has a recursion relation due to the warm-start initialization and yields the bound for δ_k:

$$\delta_k \le a_{T,S}^k\delta_0 + b_{T,S}\sum_{i=1}^{k} a_{T,S}^{k-i}\mathbb{E}\|v_i - v_{i-1}\|^2 + r_h\sum_{i=1}^{k} a_{T,S}^{k-i},$$

where $a_{T,S} := 3q_{T,S}^2$, $b_{T,S} := 3q_{T,S}^2(C_PL_H/\tau)^2$, $r_h := 3(1 + q_{T,S})^2C_{\mathrm{b}}^2h$ and $q_{T,S} := c_0e^{-ch(T-S)}$.

Combing with (83), we obtain that

$$\begin{aligned}
\mathbb{E}\|e_k\|^2 &\le A_\delta\left(a_{T,S}^k\delta_0 + b_{T,S}\sum_{i=1}^{k} a_{T,S}^{k-i}\mathbb{E}\|v_i - v_{i-1}\|^2\right) + A_\delta r_h\sum_{i=1}^{k} a_{T,S}^{k-i} + \frac{A_\delta(h)}{S} \\
&\le A_\delta\left(a_{T,S}^k\delta_0 + r_h\sum_{i=1}^{k} a_{T,S}^{k-i}\right) + 2A_\delta b_{T,S}\beta^2\sum_{i=1}^{k} a_{T,S}^{k-i}\mathbb{E}\|\mathcal{G}_\beta(v_{i-1})\|^2 \\
&\quad + 2A_\delta b_{T,S}\beta^2\sum_{i=1}^{k} a_{T,S}^{k-i}\mathbb{E}\|e_{i-1}\|^2 + \frac{A_\delta(h)}{S}.
\end{aligned}$$

Provided that

$$D_\beta := 1 - a_{T,S} - 2A_\delta b_{T,S}\beta^2 > 0,$$

summing both sides and rearranging, we obtain

$$\begin{aligned}
\frac{1}{K}\sum_{k=0}^{K-1} \mathbb{E}\|e_k\|^2 &\le \frac{A_\delta\delta_0(1 - a_{T,S}^K)}{KD_\beta} + \frac{A_\delta r_h}{D_\beta} + \frac{(1 - a_{T,S})A_\delta(h)}{SD_\beta} \\
&\quad + \frac{2A_\delta b_{T,S}\beta^2}{D_\beta}\frac{1}{K}\sum_{k=0}^{K-1} \mathbb{E}\|\mathcal{G}_\beta(v_k)\|^2.
\end{aligned} \tag{84}$$

Combining the above estimate with (82) and rearranging, provided that $D_\beta > 12A_\delta b_{T,S}\beta^2$, we obtain

$$\frac{1}{K}\sum_{k=0}^{K-1}\mathbb{E}\|\mathcal{G}_\beta(v_k)\|^2 \le \frac{8D_\beta\big(J_\tau(v_0)-J_\tau^*\big)}{\beta K\big(D_\beta - 12A_\delta b_{T,S}\beta^2\big)} + \frac{6}{D_\beta - 12A_\delta b_{T,S}\beta^2}\left[\frac{A_\delta\delta_0(1-a_{T,S}^K)}{K} + A_\delta r_h + \frac{(1-a_{T,S})A_\delta(h)}{S}\right]. \tag{85}$$

Let $\Gamma_\beta := 1 + 14A_\delta(C_P L_H/\tau)^2\beta^2$, by the definitions of $a_{T,S}$ and $b_{T,S}$,

$$D_\beta - 12A_\delta b_{T,S}\beta^2 = 1 - a_{T,S} - 14A_\delta b_{T,S}\beta^2 = 1 - 3c_0^2\Gamma_\beta e^{-2ch(T-S)}.$$

Therefore, the condition $D_\beta > 12A_\delta b_{T,S}\beta^2$ is equivalent to

$$T - S > \frac{\log(3c_0^2\Gamma_\beta)}{2ch}.$$

To keep the denominator bounded away from zero uniformly in the algorithmic parameters, it suffices to impose the stronger condition

$$T - S \ge \max\left\{0,\ \left\lceil\frac{\log(6c_0^2\Gamma_\beta)}{2ch}\right\rceil\right\}. \tag{86}$$

Under this condition, $D_\beta - 12A_\delta b_{T,S}\beta^2 \ge \frac{1}{2}$, $0 < D_\beta \le 1$, $0 \le a_{T,S} < 1$ and $q_{T,S} \le \frac{1}{\sqrt{6\Gamma_\beta}} \le \frac{1}{\sqrt{6}}$. Consequently, substituting $r_h = 3(1+q_{T,S})^2 C_{\mathrm{b}}^2 h$ and (94) into the preceding bound gives

$$\frac{1}{K}\sum_{k=0}^{K-1}\mathbb{E}\|\mathcal{G}_\beta(v_k)\|^2 \le \frac{16\big(J_\tau(v_0)-J_\tau^*\big)}{\beta K} + \frac{12A_\delta\delta_0}{K} + 36A_\delta C_{\mathrm{b}}^2\left(1+\frac{1}{\sqrt{6}}\right)^2 h + \frac{12A_\delta D_2}{S}\left(1+\frac{2c_0}{e^{ch}-1}\right) + \frac{96C_f^2\sigma_{v,2}^2 p}{S\tau^2}. \tag{87}$$

Since $e^{ch} - 1 \ge ch$, we have

$$\frac{1}{S}\left(1+\frac{2c_0}{e^{ch}-1}\right) \le \frac{1}{S} + \frac{2c_0}{cSh}.$$

Thus, for fixed $\tau > 0$, a fixed outer stepsize $\beta > 0$, and a uniform bound $\delta_0 = O(1)$, we obtain

$$\frac{1}{K}\sum_{k=0}^{K-1}\mathbb{E}\|\mathcal{G}_\beta(v_k)\|^2 = O\left(\frac{1}{K} + \frac{1}{S} + \frac{1}{Sh} + h\right),$$

provided that (86) holds. Here the implicit constant is independent of K, S, T, h.

To achieve ϵ-stationarity in the sense that $\frac{1}{K}\sum_{k=0}^{K-1}\mathbb{E}\|\mathcal{G}_\beta(v_k)\|^2 \le \epsilon$, choose, with suitable constants, $h = \Theta(\epsilon)$, $K = \Theta(\epsilon^{-1})$ and $S = \Theta(\epsilon^{-2})$, where h also satisfies the stated stepsize restrictions (42),(53), (59) and (60). Then, choose T and S to ensure (86) holds, this implies $T - S = \Theta(\epsilon^{-1})$ and $T = \Theta(\epsilon^{-2})$.

For fixed mini-batch sizes, under a per-iteration cost of $\Theta(T+S)$ gradient evaluations, the total gradient complexity is

$$\Theta\big(K(T+S)\big) = \Theta\big(K(S+h^{-1})\big) = \Theta(\epsilon^{-3}).$$

□

A.8.1 Auxiliary lemmas for Proposition A.13

Proposition A.14. *Suppose Assumptions 4.1, 4.6, A.12, 4.5, and 4.9 hold, and suppose the law-level warm-start condition $\mu_{k,0} = \mu_{k-1,T}$ holds. Suppose* (42) *holds, $h \le h_4$ with h_4 defined in* (53)*, and* (59) *and* (60) *hold. Then, for every $k \ge 1$,*

$$\delta_k \le a_{T,S}\delta_{k-1} + b_{T,S}\mathbb{E}\|v_k - v_{k-1}\|^2 + r_h, \tag{88}$$

where

$$q_{T,S} := c_0 e^{-ch(T-S)}, \quad a_{T,S} := 3q_{T,S}^2, \tag{89}$$

$$b_{T,S} := 3q_{T,S}^2 (C_P L_H/\tau)^2, \quad r_h := 3(1+q_{T,S})^2 C_{\mathrm{b}}^2 h, \tag{90}$$

and c, c_0 are given in (64)*, L_π is the constant in* (100)*, and C_{b} is defined in* (67)*. In particular, the recursion is contractive whenever $a_{T,S} = 3c_0^2 e^{-2ch(T-S)} < 1$. Iterating the recursion in* (88) *yields*

$$\delta_k \le a_{T,S}^k \delta_0 + b_{T,S} \sum_{i=1}^{k} a_{T,S}^{k-i} \mathbb{E}\|v_i - v_{i-1}\|^2 + r_h \sum_{i=1}^{k} a_{T,S}^{k-i}.$$

Proof. By the contraction estimate (63) from Lemma A.8 and the invariant-measure bias bound (66) from Lemma A.10, for any $t \in \{T-S+1, \dots, T\}$, we have

$$\begin{aligned} W_1(\mu_{k,t}, \pi_{v_k}) &\le W_1(\mu_{k,t}, \bar{\pi}_{v_k,h}) + W_1(\bar{\pi}_{v_k,h}, \pi_{v_k}) \\ &= W_1(\mu_{k,0} Q_{v_k,h}^t, \bar{\pi}_{v_k,h} Q_{v_k,h}^t) + W_1(\bar{\pi}_{v_k,h}, \pi_{v_k}) \\ &\le c_0 e^{-cht} W_1(\mu_{k,0}, \bar{\pi}_{v_k,h}) + C_{\mathrm{b}}\sqrt{h} \\ &\le q_{T,S} W_1(\mu_{k,0}, \bar{\pi}_{v_k,h}) + C_{\mathrm{b}}\sqrt{h}. \end{aligned} \tag{91}$$

By the law-level warm-start condition, $\mu_{k,0} = \mu_{k-1,T}$. Applying the triangle inequality with the intermediate Gibbs measures $\pi_{v_{k-1}}$ and π_{v_k} and the Lemma A.18 yields

$$\begin{aligned} W_1(\mu_{k,0}, \bar{\pi}_{v_k,h}) &= W_1(\mu_{k-1,T}, \bar{\pi}_{v_k,h}) \\ &\le W_1(\mu_{k-1,T}, \pi_{v_{k-1}}) + W_1(\pi_{v_{k-1}}, \pi_{v_k}) + W_1(\pi_{v_k}, \bar{\pi}_{v_k,h}) \\ &\le W_1(\mu_{k-1,T}, \pi_{v_{k-1}}) + L_\pi \|v_k - v_{k-1}\| + C_{\mathrm{b}}\sqrt{h}. \end{aligned} \tag{92}$$

Substituting (92) into (91),

$$W_1(\mu_{k,t}, \pi_{v_k}) \le q_{T,S} W_1(\mu_{k-1,T}, \pi_{v_{k-1}}) + q_{T,S} L_\pi \|v_k - v_{k-1}\| + (1+q_{T,S}) C_{\mathrm{b}}\sqrt{h}. \tag{93}$$

Using $(x+y+z)^2 \le 3x^2 + 3y^2 + 3z^2$ in (93) and taking expectation yields

$$\mathbb{E}W_1(\mu_{k,t}, \pi_{v_k})^2 \le 3q_{T,S}^2 \mathbb{E}W_1(\mu_{k-1,T}, \pi_{v_{k-1}})^2 + 3q_{T,S}^2 L_\pi^2 \mathbb{E}\|v_k - v_{k-1}\|^2 + 3(1+q_{T,S})^2 C_{\mathrm{b}}^2 h.$$

Because T belongs to the retained window at iteration $k-1$, by the definition of δ_{k-1},

$$\mathbb{E}W_1(\mu_{k-1,T}, \pi_{v_{k-1}})^2 \le \delta_{k-1}.$$

Maximizing over retained t proves (88).

$$\begin{aligned} \delta_k &= \max_{t=T-S+1,\dots,T} \mathbb{E}W_1(\mu_{k,t}, \pi_{v_k})^2 \\ &\le 3q_{T,S}^2 \delta_{k-1} + 3q_{T,S}^2 L_\pi^2 \mathbb{E}\|v_k - v_{k-1}\|^2 + 3(1+q_{T,S})^2 C_{\mathrm{b}}^2 h. \end{aligned}$$

□

Lemma A.15. *Suppose Assumptions 4.1, 4.6, 4.5, and 4.9 hold, and suppose the law-level warm-start condition $\mu_{k,0} = \mu_{k-1,T}$ holds. Let $1 \le S \le T$. Suppose* (42) *holds, $h \le h_4$ with h_4 defined in* (53)*, and* (59) *and* (60) *hold. Then,*

$$\mathbb{E}\|e_k\|^2 \le A_\delta \delta_k + \frac{A_\delta(h)}{S},$$

where A_δ is defined in (34) *and*

$$A_\delta(h) := A_\delta D_2 \left(1 + \frac{2c_0 e^{-ch}}{1-e^{-ch}}\right) + \frac{8C_f^2 \sigma_{v,2}^2 p}{\tau^2}, \quad D_2 \text{ defined in } (44). \tag{94}$$

Proof. Use the notations in the proof of Proposition A.3. For $i \in \{1,2,3,4\}$ and $I_k := \{T-S+1, \dots, T\}$, set

$$B_k^{(i)} := \frac{1}{S} \sum_{t \in I_k} \left(\mathbb{E}_{\mu_{k,t}}[\varphi_i] - \mathbb{E}_{\pi_{v_k}}[\varphi_i]\right), \quad A_k^{(i)} := \frac{1}{S} \sum_{t \in I_k} A_{k,t}^{(i)}, \quad N_k^{(i)} := \frac{1}{S} \sum_{t \in I_k} N_{k,t}^{(i)}.$$

Then

$$\widehat{\mathbb{E}}_k[\varphi_i] - \mathbb{E}_{\pi_{v_k}}[\varphi_i] = B_k^{(i)} + A_k^{(i)} + N_k^{(i)}.$$

Here $B_k^{(i)}$ is $\mathcal{G}_k$-measurable and $\mathbb{E}[A_k^{(i)} \mid \mathcal{G}_k] = 0$. Moreover, with $\mathcal{H}_k := \sigma(\mathcal{G}_k, \theta_{k,0}, \dots, \theta_{k,T})$, both $B_k^{(i)}$ and $A_k^{(i)}$ are $\mathcal{H}_k$-measurable, while independence and centering of the observation mini-batches imply $\mathbb{E}[N_k^{(i)} \mid \mathcal{H}_k] = 0$. Thus all three cross terms vanish.

Using the same component estimates as in the proof of Proposition A.3, with the bias bounded by $L_{\varphi_i}^2 \delta_k$, gives

$$\begin{aligned}\mathbb{E}\left\|\widehat{\mathbb{E}}_k[\varphi_i] - \mathbb{E}_{\pi_{v_k}}[\varphi_i]\right\|^2 &= \mathbb{E}\|B_k^{(i)}\|^2 + \mathbb{E}\|A_k^{(i)}\|^2 + \mathbb{E}\|N_k^{(i)}\|^2 \\ &\le L_{\varphi_i}^2 \delta_k + \frac{L_{\varphi_i}^2 D_2}{S}\left(1 + \frac{2c_0 e^{-ch}}{1 - e^{-ch}}\right) + \frac{\sigma_i^2}{S},\end{aligned}$$

where $\sigma_1^2 = \sigma_2^2 = 0$, $\sigma_3^2 = \sigma_{v,2}^2 p$, and $\sigma_4^2 = C_f^2 \sigma_{v,2}^2 p$. With A_δ defined in (34), we obtain

$$\mathbb{E}\|e_k\|^2 \le A_\delta \delta_k + \frac{A_\delta D_2}{S}\left(1 + \frac{2c_0 e^{-ch}}{1 - e^{-ch}}\right) + \frac{8C_f^2 \sigma_{v,2}^2 p}{S\tau^2} = A_\delta \delta_k + \frac{A_\delta(h)}{S},$$

which proves the claim. □

Lemma A.16. *Let $v_s = v_0 + s(v_1 - v_0)$ for $s \in [0,1]$, set $u := v_1 - v_0$. Assume that $s \mapsto \ell(v_s, \theta)$ is continuously differentiable with derivative $H_s(\theta) := \langle \nabla_v \ell(v_s, \theta), u\rangle$ for every θ. Let $g : \mathbb{R}^d \to \mathbb{R}$ be measurable. Suppose there exists $D \in L^1(\mathbb{R}^d, d\theta)$ such that, for every $s \in [0,1]$ and almost every θ,*

$$\big(|g(\theta)| + 1\big)|H_s(\theta)| e^{-\ell(v_s,\theta)/\tau} + |g(\theta)| e^{-\ell(v_s,\theta)/\tau} \le D(\theta), \tag{95}$$

and $0 < \int_{\mathbb{R}^d} e^{-\ell(v_s,\theta)/\tau} < \infty$ for all $s \in [0,1]$. Then $s \mapsto \mathbb{E}_{\pi_{v_s}}[g]$ is differentiable on $(0,1)$ and

$$\frac{d}{ds}\mathbb{E}_{\pi_{v_s}}[g] = -\frac{1}{\tau}\operatorname{Cov}_{\pi_{v_s}}(g, H_s). \tag{96}$$

The corresponding one-sided derivatives exist at $s = 0$ and $s = 1$.

Proof. Define

$$N_g(s) := \int_{\mathbb{R}^d} g(\theta) e^{-\ell(v_s,\theta)/\tau}\, d\theta, \qquad Z(s) := \int_{\mathbb{R}^d} e^{-\ell(v_s,\theta)/\tau}\, d\theta.$$

For fixed θ, the chain rule gives

$$\frac{d}{ds} e^{-\ell(v_s,\theta)/\tau} = -\frac{1}{\tau} H_s(\theta) e^{-\ell(v_s,\theta)/\tau}.$$

The domination condition (95) permits differentiation under both integrals by the dominated convergence theorem. Consequently,

$$N_g'(s) = -\frac{1}{\tau}\int g(\theta) H_s(\theta) e^{-\ell(v_s,\theta)/\tau}\, d\theta, \qquad Z'(s) = -\frac{1}{\tau}\int H_s(\theta) e^{-\ell(v_s,\theta)/\tau}\, d\theta.$$

Since $Z(s) > 0$ and $\mathbb{E}_{\pi_{v_s}}[g] = N_g(s)/Z(s)$, the quotient rule yields

$$\begin{aligned}\frac{d}{ds}\mathbb{E}_{\pi_{v_s}}[g] &= -\frac{1}{\tau}\mathbb{E}_{\pi_{v_s}}[gH_s] + \frac{1}{\tau}\mathbb{E}_{\pi_{v_s}}[g]\,\mathbb{E}_{\pi_{v_s}}[H_s] \\ &= -\frac{1}{\tau}\operatorname{Cov}_{\pi_{v_s}}(g, H_s),\end{aligned}$$

which proves (96). □

Lemma A.17. *Suppose Assumptions 4.1 and 4.6 hold. Let $v_0, v_1 \in \mathcal{V}$ be such that*

$$\{v_0 + s(v_1 - v_0) : s \in [0,1]\} \subseteq \mathcal{V}.$$

Define $v_s := v_0 + s(v_1 - v_0)$, $u := v_1 - v_0$. Then, for every 1-Lipschitz function $g : \mathbb{R}^d \to \mathbb{R}$, the map $s \longmapsto \mathbb{E}_{\pi_{v_s}}[g]$ is differentiable on $(0,1)$, and

$$\frac{d}{ds}\mathbb{E}_{\pi_{v_s}}[g] = -\frac{1}{\tau}\operatorname{Cov}_{\pi_{v_s}}\left(g, \langle \nabla_v \ell(v_s, \theta), u\rangle\right). \tag{97}$$

The same identity holds at $s = 0$ and $s = 1$ with the corresponding one-sided derivatives.

Proof. For $s \in [0,1]$, define

$$N_g(s) := \int_{\mathbb{R}^d} g(\theta) e^{-\ell(v_s,\theta)/\tau}\, d\theta, \qquad Z(s) := \int_{\mathbb{R}^d} e^{-\ell(v_s,\theta)/\tau}\, d\theta,$$

and set $H_s(\theta) := \langle \nabla_v \ell(v_s, \theta), u\rangle$.

We first verify that differentiation under the integral sign is justified. Since g is 1-Lipschitz,

$$|g(\theta)| \le |g(0)| + \|\theta\|. \tag{98}$$

Moreover, by Assumption 4.1,

$$\begin{aligned} |H_s(\theta)| &\le |H_s(0)| + |H_s(\theta) - H_s(0)| \\ &= |H_s(0)| + |\langle \nabla_v \ell(v_s,\theta) - \nabla_v \ell(v_s,0), u\rangle| \\ &\le |H_s(0)| + L_{\varphi_3}\|u\|\,\|\theta\|. \end{aligned}$$

The map $s \mapsto \nabla_v \ell(v_s, 0)$ is continuous, and the interval $[0,1]$ is compact. Therefore,

$$M_0 := \sup_{s\in[0,1]} \|\nabla_v \ell(v_s,0)\| < \infty,$$

so that, uniformly in $s \in [0,1]$, $|H_s(\theta)| \le \|u\|\big(M_0 + L_{\varphi_3}\|\theta\|\big)$. Together with (98) and the quadratic lower bound from Lemma 4.4, we obtain a constant $C < \infty$, independent of s and θ, such that

$$\big(1 + |g(\theta)|\big)\big(1 + |H_s(\theta)|\big) e^{-\ell(v_s,\theta)/\tau} \le D(\theta), \tag{99}$$

where

$$D(\theta) := C(1 + \|\theta\|^2) \exp\left(-\frac{a}{\tau}\|\theta\|^2\right).$$

Since a polynomial multiplied by a Gaussian density is integrable, $D \in L^1(\mathbb{R}^d, d\theta)$. In particular, $N_g(s)$ and $Z(s)$ are finite, and $Z(s) > 0$.

Furthermore, (99) gives an integrable bound, uniform in s, for the derivatives of both integrands in $N_g(s)$ and $Z(s)$. Thus, the dominated convergence theorem applied to the corresponding difference quotients justifies differentiation under the integral sign. We obtain

$$N_g'(s) = -\frac{1}{\tau}\int_{\mathbb{R}^d} g(\theta) H_s(\theta) e^{-\ell(v_s,\theta)/\tau}\, d\theta = -\frac{Z(s)}{\tau}\mathbb{E}_{\pi_{v_s}}[gH_s],$$

and similarly,

$$Z'(s) = -\frac{1}{\tau}\int_{\mathbb{R}^d} H_s(\theta) e^{-\ell(v_s,\theta)/\tau}\, d\theta = -\frac{Z(s)}{\tau}\mathbb{E}_{\pi_{v_s}}[H_s].$$

Since $\mathbb{E}_{\pi_{v_s}}[g] = \frac{N_g(s)}{Z(s)}$, and $Z(s) > 0$, the quotient rule yields

$$\begin{aligned} \frac{d}{ds}\mathbb{E}_{\pi_{v_s}}[g] &= \frac{N_g'(s)}{Z(s)} - \frac{N_g(s)Z'(s)}{Z(s)^2} \\ &= -\frac{1}{\tau}\mathbb{E}_{\pi_{v_s}}[gH_s] + \frac{1}{\tau}\mathbb{E}_{\pi_{v_s}}[g]\,\mathbb{E}_{\pi_{v_s}}[H_s] \\ &= -\frac{1}{\tau}\operatorname{Cov}_{\pi_{v_s}}(g, H_s). \end{aligned}$$

Substituting the definition of H_s gives (97). □

Lemma A.18. *Under Assumptions 4.1 and A.12,*

$$W_1(\pi_v, \pi_{v'}) \le L_\pi \|v - v'\|, \qquad L_\pi := \frac{C_{\mathrm{P}} L_H}{\tau}. \tag{100}$$

Proof. Fix a scalar 1-Lipschitz function g, set $u = v' - v$, and let $v_s = v + su$. By Rademacher's theorem (Evans, 2025, Theorem 3.2), g is differentiable almost everywhere and $\|\nabla_\theta g\| \le 1$ almost everywhere.

Let $H_s(\theta) := \langle \nabla_v \ell(v_s, \theta), u \rangle$, then $\nabla_\theta H_s(\theta) = \nabla^2_{\theta v} \ell(v_s, \theta) u$. Therefore, applying Assumptions 4.1 and A.12,

$$\begin{aligned} \mathrm{Var}_{\pi_{v_s}}(H_s) &\le C_{\mathrm{P}} \mathbb{E}_{\pi_{v_s}} \|\nabla_\theta H_s(\theta)\|^2 \\ &= C_{\mathrm{P}} \mathbb{E}_{\pi_{v_s}} \|\nabla^2_{\theta v} \ell(v_s, \theta) u\|^2 \\ &\le C_{\mathrm{P}} \|u\|^2 \mathbb{E}_{\pi_{v_s}} \|\nabla^2_{\theta v} \ell(v_s, \theta)\|^2_{\mathrm{op}} \\ &\le C_{\mathrm{P}} L_H^2 \|u\|^2. \end{aligned} \tag{101}$$

Using Lemma A.17, followed by covariance Cauchy–Schwarz,

$$\begin{aligned} \left| \frac{d}{ds} \mathbb{E}_{\pi_{v_s}} g \right| &= \frac{1}{\tau} \left| \mathrm{Cov}_{\pi_{v_s}}(g, H_s) \right| \\ &\le \frac{1}{\tau} \sqrt{\mathrm{Var}_{\pi_{v_s}}(g)} \sqrt{\mathrm{Var}_{\pi_{v_s}}(H_s)} \\ &\le \frac{C_{\mathrm{P}} L_H}{\tau} \|u\|. \end{aligned}$$

Integrating from 0 to 1 gives

$$\left| \mathbb{E}_{\pi_{v'}} g - \mathbb{E}_{\pi_v} g \right| \le \int_0^1 \left| \frac{d}{ds} \mathbb{E}_{\pi_{v_s}} g \right| ds \le \frac{C_{\mathrm{P}} L_H}{\tau} \|v' - v\|.$$

Taking the supremum over all scalar 1-Lipschitz g and using the Kantorovich–Rubinstein dual representation (Villani, 2021, Theorem 1.14) of W_1 proves (100). □

B Experimental Details and Additional Results

All experiments are run on NVIDIA V100.

B.1 Quadratic Gibbs follower experiment (Section 5.1)

Formulation details. To ensure $A(v)$ remains uniformly positive definite, we parameterize it as $A(v) := Q^\top \mathrm{diag}(\exp(u + Rv)) Q + \lambda I_d$, $\lambda > 0$, where $Q \in \mathbb{R}^{d \times d}$ is a fixed orthogonal matrix, $u \in \mathbb{R}^d$ and $R \in \mathbb{R}^{d \times d_v}$ are fixed parameters. We further set $b(v) := b_0 + Bv$, for fixed $b_0 \in \mathbb{R}^d$ and $B \in \mathbb{R}^{d \times d_v}$.

Because the quadratic lower-level objective induces a Gaussian distribution, the value-centric objective admits the closed form $J_\tau(v) = \frac{1}{2}\Big(\|CA(v)^{-1} b(v) - y\|^2 + \mathrm{tr}\big(\tau C A(v)^{-1} C^\top\big) \Big) + \frac{\gamma}{2}\|v\|^2$. Differentiating this expression yields the exact hypergradient, which serves as the ground truth for evaluation.

Parameter setting. We use dimensions $d = 50$ (inner), $p = 10$ (outer), and $p = 50$, with $\lambda = 0.5$ for $A(v) := Q^\top \mathrm{diag}\big(\exp(u + Rv)\big) Q + \lambda I_d$, $\gamma = 10^{-2}$, and temperature $\tau = 10^{-2}$. The outer loop uses step size $\beta = 0.05$ for $K = 500$ iterations, starting from $v_0 \sim \mathcal{N}(0, 0.2^2 I_{d_v})$. The inner sampler is persistent-chain SGLD with step size $h = 0.05$, T inner steps per outer iteration, additive gradient noise $\mathcal{N}(0, \sigma^2 I)$ and $\sigma = 0.1$, and tail averaging over the last $S = \lfloor 0.5T \rfloor$ inner states. We average curves over 10 random seeds (each seed initializes (v_0, θ_0) and, by default, the random problem instance), and use $T \in \{50, 100, 200, 500\}$ in the trajectory plot and $T = 500$ for the τ-sweep plots.

B.2 Double-well experiment (Section 5.2)

Baselines. *(i) AID-HG*, the standard deterministic bilevel algorithm (Ghadimi & Wang, 2018; Lorraine et al., 2020). At the k-th iteration, it performs T gradient steps on $\ell(v_k, \cdot)$ to obtain an approximate lower-level solution $\theta_{k,T}$, and then takes a projected gradient step using the implicit-differentiation hypergradient

$$\nabla \Phi(v) \;=\; \nabla_v f \;-\; \nabla^2_{v\theta} \ell \left[\nabla^2_{\theta\theta} \ell \right]^{-1} \nabla_\theta f, \tag{102}$$

evaluated at $(v_k, \theta_{k,T})$. When computing the inverse Hessian directly is impractical, a truncated Neumann-series approximation or the conjugate-gradient method (Lorraine et al., 2020) are used.

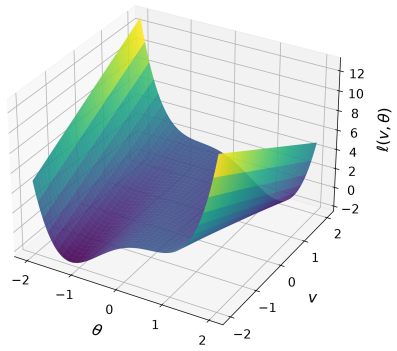

Figure 4: Surface plot of the double-well inner loss $\ell(v, \theta)$ as a function of v and θ. The landscape exhibits two prominent wells when $v \approx 0$, illustrating the multi-modal nature of the inner objective.

(ii) HPO-SGLD (Kim & Hospedales, 2025) replaces the lower-level solution $\theta^\star(v)$ with a Gibbs distribution and draws samples approximately from the Gibbs distribution using SGLD. Unlike VCHD, however, it computes the hypergradient by differentiating through the SGLD recursion with respect to v, rather than using a expectation-covariance hypergradient.

(iii) HG-MS (Masiha et al., 2026) is a "select-then-differentiate" scheme for optimistic bilevel problems whose lower-level solution set $S(v) := \arg\min_\theta \ell(v, \theta)$ is a manifold. It runs N parallel Unadjusted Langevin Algorithm (ULA) chains so that the candidates concentrate near $S(v)$. It then selects based on the upper objective value $\hat{\theta} \in \arg\min_{i\in[N]} f(v, X_i)$ and computes a ridge-regularized version of (102) at $\widehat{\theta}$.

For our VCHD method, the hypergradient is $\nabla_v J_\tau(v) = \lambda v + \frac{1}{\tau}\mathrm{Cov}_{\pi_v}\left(\theta^2, \theta\right)$. In the classical bilevel setting, the corresponding implicit-function hypergradient is $\nabla\Phi(v) = \lambda v + 2\theta^\star(v) \cdot \frac{1}{12(\theta^\star(v))^2-4}$, which is used for AID-HG and HG-MS. For the HPO-SGLD, it use the hypegradient's recursion formula in equaltion (22) of Kim & Hospedales (2025).

Parameter Setup. **Common setting.** All methods solve the same problem over the outer feasible set $\mathcal{V} = [-4, 4]$ and take the projected step $v_{k+1} = \Pi_{\mathcal{V}}(v_k - \beta\, \hat{g}_k)$ with $\beta = 5 \times 10^{-2}$ for $K = 200$ iterations from $v_0 = 1.0$, with $\lambda = 0.5$ in the outer loss and inner stepsize $h = 10^{-3}$. The Gibbs temperature is $\tau = 0.5$. For each seed we draw a single $\theta_0 \sim \mathrm{Unif}[-4, 4]$ and hand the same θ_0 to every method, so that any difference between methods is attributable to the hypergradient and not to the initialization. We report seeds $\{0, 2\}$, for which $\theta_0 = +1.096$ and $\theta_0 = -1.907$ fall in different basins of $\ell(v_0, \cdot)$.

VCHD (ours). Persistent-chain SGLD targeting π_{v_k}, warm-started from the inner state of the previous outer iteration: $\theta \leftarrow \theta - h\, \partial_\theta \ell(v, \theta) + \sqrt{2\tau h}\, \xi$ with $\xi \sim \mathcal{N}(0, 1)$, run for $T = 400$ steps per outer iteration of which the first $B = 100$ are discarded, leaving $S = T - B = 300$ retained samples for expectation-covariance estimatio.

SGD. $T = 400$ warm-started deterministic gradient steps $\theta \leftarrow \theta - h\, \partial_\theta \ell(v, \theta)$ at the same $h = 10^{-3}$, followed by the implicit-differentiation hypergradient evaluated at the resulting θ_k, which here is $\hat{g} = \lambda v + \frac{2\theta_k}{12\theta_k^2-4}$. In one dimension the inverse Hessian is a scalar and is computed exactly, we guard it by clipping $|\partial^2_{\theta\theta}\ell|$ below at 10^{-8}.

HPO-SGLD (Kim & Hospedales, 2025). The inner SGLD chain is identical to VCHD's, their update $\theta \leftarrow \theta + \frac{\epsilon}{2}\nabla_\theta \log p + \sqrt{\epsilon}\, z$ coincides with ours under $\epsilon = 2\tau h$. Their Algorithm 1 re-initialises the pathwise accumulator at every outer iteration and treats $\theta^{(0)}$ as independent of v, so $g_0 = 0$ and $\mathrm{d}\theta^{(0)}/\mathrm{d}v = 0$, and their first-order recursion

$$g_m \;=\; g_{m-1} \;+\; \tfrac{\epsilon}{2}\left(\partial^2_{v\theta} \log p\right)^\top \nabla_\theta f\left(v, \theta^{(m)}\right) \tag{103}$$

specialises to $g_m = g_{m-1} + 2h\,\theta^{(m)}$. The reported hypergradient is $\hat{g} = \frac{1}{S}\sum_{m=B+1}^{B+S}(\lambda v + g_m)$ with burn-in $B = 100$ and $S = T - B = 300$ retained samples, matching VCHD's budget exactly. Under this choice $\mathbb{E}[\hat{g}] = \lambda v + G\,\mathbb{E}_{\pi_v}[\theta]$ with effective gain $G = 2h\big(B + \frac{S+1}{2}\big) = 0.501$, a constant fixed by the chain length rather than by the local curvature. We additionally implement the exact forward-mode variant $d_m = \big(1 - h\,\partial^2_{\theta\theta}\ell\big)\,d_{m-1} + h$, $g_m = 2\theta^{(m)}d_m$, which restores the contraction; it is used only in the ablation of Figure 6, not in the main comparison.

HG-MS (Masiha et al., 2026). $N = 10$ parallel ULA chains, each run for $T = 400$ steps at Gibbs temperature $\lambda_{\mathrm{G}} = 10^{-3}$ with the same stepsize $h = 10^{-3}$: $X \leftarrow X - h\,\partial_\theta \ell(v, X) + \sqrt{2\lambda_{\mathrm{G}} h}\,\xi$. Their Algorithm 1 takes the chain starts as an input, so we draw $X^{(0)} \sim \mathrm{Unif}[-4, 4]^N$ once per seed and restart all N chains from that same $X^{(0)}$ at every outer iteration. Best-of-N selection is $\hat{\theta} = X_{i^*}$ with $i^* \in \arg\min_{i\in[N]} f(v, X_i)$, followed by the ridge-regularised solve $\big(\partial^2_{\theta\theta}\ell(v, \hat{\theta}) + \gamma\big)\nu = \partial_\theta f(v, \hat{\theta})$ with $\gamma = 10^{-2}$, giving $\hat{g} = \lambda v + \nu$.

Additional Results. **Ground truth.** $J_\tau(v) = \mathbb{E}_{\pi_v}[f(v, \theta)]$ is evaluated by numerical quadrature on a uniform grid over $\theta \in [-4, 4]$ with 4001 points. Minimising over a 20,001-point grid on $v \in [-2, 2]$ gives $v^\star = 0.0000$ and $J_\tau(v^\star) = 0.85214$.

Table 1: Double-well experiment. Statistics are computed over the final 50 of $K = 200$ outer iterations. "$\bar{v}$ per seed" is the per-seed time average of v_k; "sd" is the within-run temporal standard deviation of v_k, averaged over seeds.

Method	$\bar{v}$ (seeds 0/2)	$\mathrm{sd}(v_k)$	$J_\tau(v_k)$
VCHD (ours)	$+0.030$ / $+0.060$	0.065	**0.8575**
HPO-SGLD	-0.114 / $+0.126$	0.225	0.9038
AID-HG	-0.558 / $+0.626$	0.009	1.0693
HG-MS	$+0.579$ / $+0.588$	0.005	1.0638

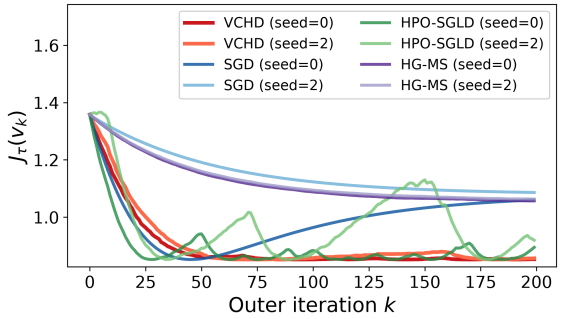

Figure 5: Double-well experiment. Hyper-objective $J_\tau(v_k)$, evaluated by numerical quadrature.

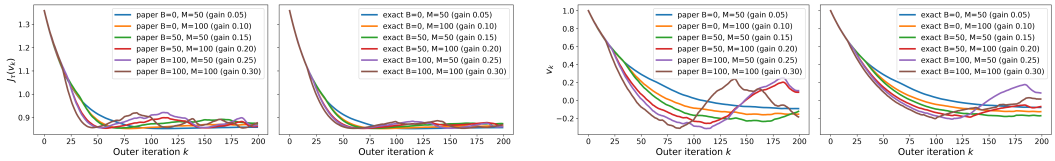

Figure 6: HPO-SGLD ablation over burn-in B and retained-sample count M at $h = 10^{-3}$ (left panel of each pair: the published recursion (103); right panel: the exact forward-mode variant that keeps the contraction factor). Legends report the effective gain $G = 2h(B + \frac{M+1}{2})$. Under the published recursion the trajectory depends systematically on B and M through G, which is the signature of the dropped contraction: the implied sensitivity grows linearly in the chain length instead of saturating at $[\partial^2_{\theta\theta}\ell]^{-1}$. The exact variant removes that dependence.

B.3 HYPER DATA CLEANING EXPERIMENT (SECTION 5.3)

Dataset and corruption. We use MNIST with a targeted *pairflip* corruption at rate 50%: half of the training examples of class 1 are relabeled to 7 and half of 7 to 1, and similarly for the pairs $(4, 6)$ and $(3, 8)$. From the 60,000 original training images we draw a fixed random subset of 50,000, hold out 5,000 of them as a *clean* validation set, and corrupt only the remaining $n = 45{,}000$ training points. Because six of the ten classes are affected and half of each is flipped, the overall corruption rate is $0.5 \times 0.6 \approx 30\%$. The validation split and the 10,000-image test split keep their ground-truth labels. Inputs are normalized with the usual MNIST statistics $(0.1307, 0.3081)$, and all inner updates use minibatches of size 128.

Model. The inner model is a small CNN with two convolutional layers and two fully connected layers with $d = 421{,}642$ parameters in total. The outer variable is the per-sample weight vector $w \in [0, 1]^n$ with $n = 45{,}000$, initialized at $w_i = 0.5$ and projected back to $[0, 1]$ after every outer update. The inner objective is the weighted risk $\ell(w, \theta) := \frac{1}{n} \sum_{i=1}^{n} w_i \, \ell_i(\theta) + \frac{\lambda}{2} \|\theta\|_2^2$ with $\lambda = 5 \times 10^{-4}$.

Baselines. Common setting. All four methods share the dataset, split, architecture, batch size, initialization $w_i = 0.5$, outer optimizer (Adam on w with constant step size $\beta = 10^{-2}$), and total outer iteration number $K = 5000$. They differ only in how the inner problem is solved and how the hypergradient is formed.

VCHD (ours). The target follower is the Gibbs law $\pi_v(\theta) \propto \exp(-\ell(v, \theta)/\tau)$. In this large-scale experiment it is approximated by a warm-started Adam-drift Langevin heuristic: an Adam drift step (learning rate h, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$) followed by the injection $\theta \leftarrow \theta + \sqrt{2\tau h}\, \xi$, $\xi \sim \mathcal{N}(0, I)$. Because its state-dependent drift is paired with isotropic noise, we do not claim that this practical chain preserves the exact Gibbs law; the results in this section are therefore empirical evidence for a Gibbs-inspired approximation rather than a direct validation of the exact sampler analyzed in the theory. The outer update uses the expectation–covariance hypergradient (3). We evaluate 12 combinations of $\tau \in \{10^{-2}, 5 \times 10^{-3}, 2 \times 10^{-3}, 10^{-3}\}$ and $h \in \{10^{-3}, 5 \times 10^{-3}, 10^{-2}\}$.

Diff+Neumann. The inner problem is minimized with Adam at a constant step size $h = 10^{-3}$. The outer hypergradient uses implicit-differentiation hypergradient , with $H^{-1} g$ approximated by a truncated Neumann series of $K_{\mathrm{N}} \in \{10, 50, 500\}$ terms and step size $\alpha = 5 \times 10^{-3}$, implemented with Hessian-vector products.

HPO-SGLD (Kim & Hospedales, 2025). The inner chain is SGLD with drift step $\varepsilon/(2\tau) = 10^{-3}$ and noise $\sqrt{\varepsilon}\, \kappa$ with $\kappa = 10^{-2}$, and the hypergradient is accumulated by the forward-mode recursion over the last $M \in \{5, 10, 20, 50, 100, 500\}$ inner iterations.

HG-MS (Masiha et al., 2026). Select-then-differentiate: N parallel ULA chains at Gibbs temperature λ_{G} with step size $h = 10^{-3}$, a Best-of-N selection $\hat{\theta} \in \arg\min_{i \in [N]} f(v, \theta_i)$ against the current validation batch, and a ridge-regularized way $(\nabla^2_{\theta\theta} \ell + \gamma I)\nu = \nabla_\theta f$ solved by conjugate gradients (with 50 iterations). We evaluate all 24 combinations of $N \in \{5, 10, 50, 100\}$, $\lambda_{\mathrm{G}} \in \{10^{-3}, 5 \times 10^{-3}, 10^{-2}\}$, and $\gamma \in \{10^{-2}, 10\}$. The chains are warm-started across outer iterations rather than restarted from fixed initial states , and the drift uses minibatch gradients rather than full gradients, as for the other stochastic baselines.

Inner/outer schedules and averaging. At outer iteration k we run T_k inner steps and estimate the required expectations by an *online tail average* over the last $S_k \le T_k$ iterates,

$$\widehat{\mathbb{E}}_k[\varphi] = \frac{1}{S_k} \sum_{s=1}^{S_k} \widehat{\varphi}(v_k, \theta_{k, T_k - S_k + s}, \widetilde{\xi}_{k,t}), \qquad S_k = \min\{S, T_k\}.$$

Every method uses a long burn-in at the first outer iteration, $T_0 = 10^5$, and short tracking updates thereafter, $T_k = 100$ for $k \ge 1$; HG-MS runs T_k steps in *each* of its N chains, so its inner cost per outer iteration is NT_k. We set $S = 5 \times 10^4$ for the first outer iteration $k = 0$ and $S = 500$ for the rest outer iterations.

Evaluation protocol. After each outer update we record: (a) clean validation and test accuracy of the current inner model, evaluated on at most 60 minibatches (4,992 validation and 7,680 test images, in a fixed order); (b) the outer objective value; (c) training-batch accuracy scored against both the corrupted labels used for training and the original labels, on the same batch and the same predictions; and (d) the ability of the learned weights to separate clean from corrupted points, measured as the accuracy of the rule $w_i > 0.5$ against the known corruption mask over 60 training minibatches. Trajectory plots show a moving average over 100 outer iterations. All numbers use seed 1; each configuration is a single run on one NVIDIA V100.

Additional Results. As shown in Figure 7, the deterministic Diff+Neumann method attains the lowest inner-objective value, consistent with our explanation that it overfits the lower-level training objective. For VCHD, the temperature τ controls the trade-off between exploration and bias. If τ is too large, the sampling error of hypergradient will increase. As shown in Table 2, $\tau = 2 \times 10^{-3}$ provides the best balance and achieves the lowest test error (about 1.55%).

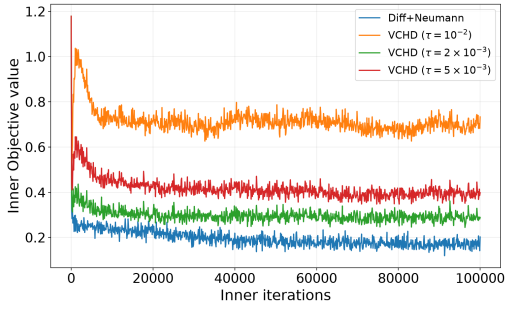

Figure 7: Data-cleaning experiment (inner dynamics). Inner training loss trajectory.

Three secondary observations are worth recording. First, increasing the Neumann truncation is not uniformly helpful: $K_{\mathrm{N}} = 500$ is markedly worse (21.99%) than $K_{\mathrm{N}} = 50$ (11.09%), because the truncated series diverges once the inner Hessian is indefinite, relative residual $\|Hp - g\|/\|g\|$ of the linear solve has median 2.93 at $K_{\mathrm{N}} = 500$ versus 0.92 at $K_{\mathrm{N}} = 50$. Second, HG-MS is extremely sensitive to its Gibbs temperature: at $\lambda_{\mathrm{G}} = 10^{-3}$ it reaches 16.1%, but at $\lambda_{\mathrm{G}} \geq 5 \times 10^{-3}$ the chains no longer concentrate near the solution manifold and Best-of-N selects among uninformative candidates, leaving the method near chance (82–89% error). Third, the ridge γ matters far less than the sampler: $\gamma = 10^{-2}$ and $\gamma = 10$ differ by about one point (17.73% vs. 16.68% at $N = 5$) even though the CG solve is accurate only for the latter (median relative residual 3.7×10^{-2} versus 0.62).

We also measure how well the learned weights identify the corrupted examples, by thresholding each weight at $w_i > 0.5$ and comparing the resulting clean/corrupted decision with the known corruption mask. Since 30.65% of the training labels are flipped, declaring every point clean already scores 0.694, and w starts at 0.5 everywhere. VCHD produces the cleanest weights, at 0.925, against 0.893 for HPO-SGLD, 0.783 for HG-MS and 0.760 for Diff+Neumann.

We also provide a offline evaluation of $\Phi(v_k)$, shown in Figure 10. The training log records the surrogate $J_\tau(v_k)$, not the original objective $\Phi(v) = f(v, \theta^\star(v))$. To evaluate the latter we save w_k every 50 outer iterations of the $\tau = 5 \times 10^{-3}$ run (101 checkpoints) and, for each, re-solve the inner problem offline: starting from a *fixed* initialization (seed 42) we run 2000 plain Adam steps at learning rate 10^{-3} on $\ell(w_k, \cdot)$ with no injected noise, so that $\theta^\star(w_k)$ is a deterministic function of w_k alone, and then evaluate the clean cross-entropy on the full validation set. Along the trajectory $\Phi(v_k)$ decreases monotonically from 0.477 at $k = 0$ to 0.255, with the test accuracy of $\theta^\star(v_k)$ rising from 67.7% to 98.0%.

C Detailed Related Works

Classical bilevel optimization. In the nonconvex-strongly-convex setting, lower-level strongly convexity ensures a unique response map and yields a closed-form hypergradient involving an inverse lower-level Hessian (Ghadimi & Wang, 2018; Ji et al., 2021). A substantial literature has

Table 2: Hyper data cleaning on MNIST with 50% pairflip corruption. Best raw test error over $K = 5000$ outer iterations, and wall-clock time of the full run.

Method	Configuration	Test err.	Wall
Diff+Neumann	$K_{\mathrm{N}} = 10$	12.66%	2.29h
	$K_{\mathrm{N}} = 50$	11.09%	3.08h
	$K_{\mathrm{N}} = 500$	21.99%	5.79h
HPO-SGLD	$M = 5, \tau = 10^{-2}$	6.63%	2.66h
	$M = 10, \tau = 10^{-2}$	5.72%	2.35h
	$M = 20, \tau = 10^{-2}$	4.92%	2.54h
	$M = 50, \tau = 10^{-3}$	4.17%	3.29h
	$M = 100, \tau = 10^{-3}$	3.57%	3.40h
	$M = 500, \tau = 5 \times 10^{-3}$	3.42%	7.74h
HG-MS	$N = 5, \lambda_{\mathrm{G}} = 10^{-3}, \gamma = 10^{-2}$	17.73%	5.45h
	$N = 5, \lambda_{\mathrm{G}} = 10^{-3}, \gamma = 10$	16.68%	5.40h
	$N = 10, \lambda_{\mathrm{G}} = 10^{-3}, \gamma = 10$	16.12%	9.01h
	$N = 10, \lambda_{\mathrm{G}} = 5 \times 10^{-3}, \gamma = 10$	81.65%	8.85h
	$N = 5, \lambda_{\mathrm{G}} = 10^{-2}, \gamma = 10$	88.66%	5.42h
VCHD (ours)	$\tau = 10^{-2}, h = 10^{-3}$	3.45%	3.27h
	$\tau = 5 \times 10^{-3}, h = 10^{-3}$	2.83%	3.30h
	$\tau = 2 \times 10^{-3}, h = 10^{-3}$	6.82%	3.51h
	$\tau = 5 \times 10^{-3}, h = 10^{-2}$	2.50%	3.49h
	$\tau = 2 \times 10^{-3}, h = 10^{-2}$	2.80%	3.33h
	$T_k = 500, \tau = 5 \times 10^{-3}, h = 10^{-3}$	1.99%	6.28h
	$T_k = 500, \tau = 10^{-3}, h = 5 \times 10^{-3}$	1.58%	6.18h
	$T_k = 500, \tau = 2 \times 10^{-3}, h = 5 \times 10^{-3}$	**1.55**%	6.21h

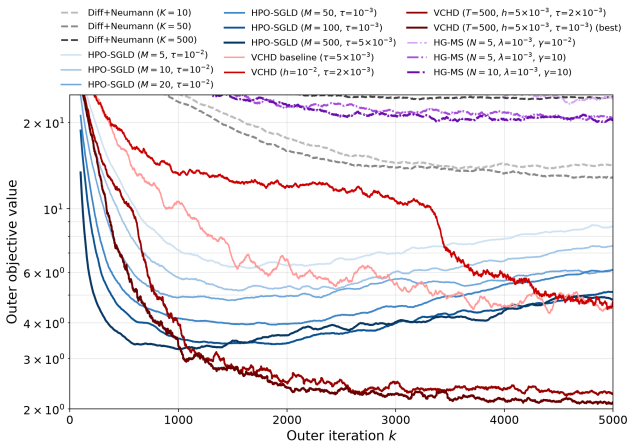

Figure 8: Data-cleaning experiment (outer dynamics). Outer validation loss trajectory.

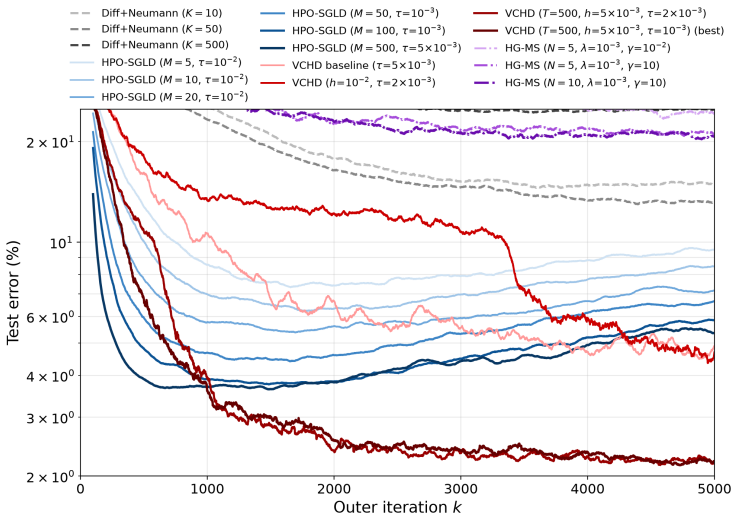

Figure 9: Data-cleaning experiment (outer dynamics). Test error.

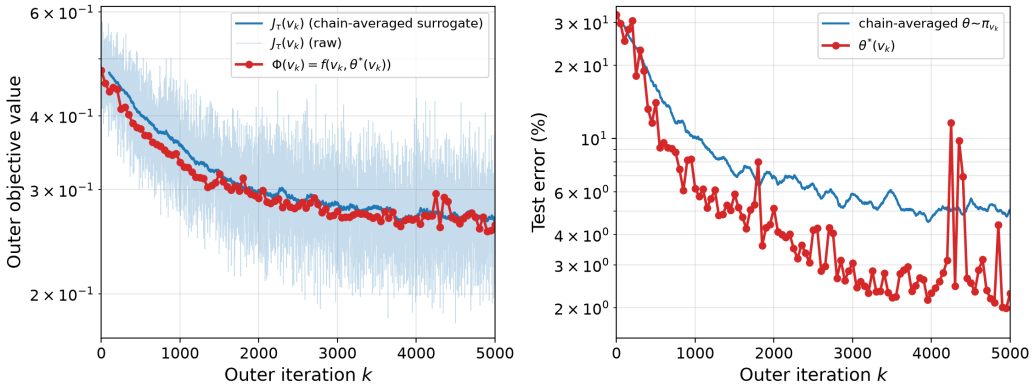

Figure 10: Tracking the original objective Φ along the VCHD trajectory ($\tau = 5 \times 10^{-3}$, per-sample weights w_k checkpointed every 50 outer iterations). **(Left)** Outer-objective value on the validation split: the surrogate $J_\tau(v_k)$ actually optimized (raw, and smoothed with a 100-iteration moving average) against $\Phi(v_k) = f(v_k, \theta^*(v_k))$, obtained by re-solving the inner problem offline from a fixed initialization at each checkpoint. The two curves stay close and decrease together, with Φ falling from 0.477 at $k = 0$ to 0.255; optimizing the Gibbs surrogate therefore also drives down the original bilevel objective. **(Right)** Test error of a chain sample $\theta \sim \pi_{v_k}$ against that of the deterministic re-solve $\theta^*(v_k)$: the re-solved model is consistently better, reaching 98.0% test accuracy, so the learned weights are good beyond the particular chain state used to compute the hypergradient.

developed efficient approximations of this hypergradient (Arbel & Mairal, 2022; Dagréou et al., 2022; Kong et al., 2026). However, lower-level strongly convexity rules out nonunique responses and can be restrictive for overparameterized or multimodal learning problems. One line of work relaxes strong convexity to a global Polyak–Łojasiewicz (PL) condition (Chen et al., 2024; Huang, 2023). Although global PL allows certain nonconvex objectives, it implies that every stationary point is globally optimal and therefore excludes landscapes with suboptimal local minima. Value-function methods (Liu et al., 2021) encode lower-level optimality through $\ell(v,\theta) - \min_{\vartheta} \ell(v,\vartheta) \leq 0$. While this reformulation is exact, the value function $\ell^\star$ is generally unavailable and may be nonsmooth, evaluating or approximating it still requires solving a nonconvex problem.

Sampling-based minima selection. Gibbs measures and Langevin dynamics have close connection with global optimization: as the temperature decreases, a Gibbs measure concentrates near the global minimizers of its potential (Raginsky et al., 2017; Dalalyan, 2017). This principle has also been exploited in bilevel optimization. In most existing approaches, however, sampling serves only as an algorithmic device for locating or selecting a lower-level optimizer, so the final formulation remains optimizer-centered. Masiha et al. (2025) introduced a Superquantile–Gibbs relaxation that uses Gibbs samples to approximate extremal selection over a non-singleton lower-level minimizer set. Under the same geometric framework, Masiha et al. (2026) studied the optimistic formulation and proposed a select-then-differentiate method combining best-of-N selection from Gibbs samples with a pseudoinverse-based hypergradient. In both methods, the Gibbs law facilitates the selection of an extremal optimizer rather than serving as the follower's response. Moreover, their PL°-based framework yields connected, compact, manifold-structured minimizer sets. Our assumptions instead allow arbitrary nonconvexity within a bounded region, including separated local and global modes, while requiring strong convexity only outside that region, as used in analyses of nonconvex Langevin algorithms (Cheng et al., 2018; Li et al., 2024). Consensus-based particle methods provide another sampling-based approach to hierarchical selection. In particular, Chao et al. (2026) combine smooth lower-level quantile selection with Gibbs-type Laplace weighting based on the upper-level objective, driving the particle system toward a selected bilevel solution. Their main convergence theory, however, concerns the fixed-set problem $\min_{\theta \in \arg\min_{\vartheta} L(\vartheta)} G(\theta)$, rather than a parameterized lower-level problem whose solution set varies with the outer decision.

Gibbs-averaged bilevel optimization. The work most closely related to ours is Kim & Hospedales (2025), which considers essentially the same Gibbs-averaged objective and uses SGLD to sample from the lower-level Gibbs distribution. However, It estimates the hypergradient by differentiating pathwise through the SGLD recursion and applying a first-order approximation to the resulting forward-sensitivity recursion, in the spirit of iterative differentiation through an inner trajectory (Domke, 2012; Maclaurin et al., 2015; Franceschi et al., 2017; 2018). We instead differentiate the stationary Gibbs expectation directly, obtaining an exact expectation–covariance identity involving only zeroth- and first-order quantities, which mirrors the classical hypergradient descent algorithm Ghadimi & Wang (2018); Ji et al. (2021); Lorraine et al. (2020) in a distributional follower setting. In contrast, the proposed VCHD algorithm therefore requires neither differentiation through the sampling trajectory nor Hessian–vector products. We further provide a nonasymptotic convergence analysis that explicitly accounts for finite-stepsize SGLD discretization bias, variation in the target Gibbs distribution across outer iterations, and temporal dependence among samples from the same trajectory. By contrast, the analysis of Kim & Hospedales (2025) uses an infinitesimal-stepsize approximation and treats the post-burn-in chain as sampling from the exact Gibbs law.

Performative prediction and decision-dependent optimization. Beyond bilevel optimization, our formulation relates to performative prediction and decision-dependent optimization, where decisions affect the distribution under which the objective is evaluated. Perdomo et al. (2020) formalized performative risk and analyzed repeated retraining toward stable solutions. Subsequent work estimates distributional responses through batch procedures (Lin & Zrnic, 2024) or learns their parameters online from decision-dependent streaming data (Liu et al., 2024). In these settings, the distributional response is typically learned from data generated under deployed decisions. In our setting, it arises from a structured bilevel mechanism: the outer variable parameterizes a lower-level objective, which in turn induces a Gibbs follower. This structure yields an analytic expectation-covariance hypergradient and leads to the problem of tracking a moving Gibbs law across outer iterations.

Logit and Gibbs response models. Probabilistic responses of Gibbs or logit form have a long history in models of bounded rationality and information-processing costs. In game theory, the logit specification of quantal response equilibrium replaces exact best responses with stochastic choice probabilities proportional to exponentiated expected payoffs (McKelvey & Palfrey, 1995). In economics, rational-inattention models derive generalized multinomial-logit choice from Shannon information costs (Matějka & McKay, 2015), while broader equivalences between generalized information costs and random-utility discrete-choice models are established by Fosgerau et al. (2020). These works primarily concern discrete choices, with the game-theoretic literature additionally imposing strategic equilibrium conditions. By contrast, our follower is a high-dimensional Gibbs law over continuous model parameters, and our focus is on exact hypergradient estimation and persistent-chain tracking as this law changes with the outer variable.

Bilevel optimization under stochastic equilibrium constraints (transportation/OR). Bilevel formulations with probabilistic lower-level responses are also classical in transportation network design and pricing, where the lower level is often a *logit-based stochastic user equilibrium* (SUE) and the upper level optimizes network objectives (Davis, 1994; Yang & Bell, 2001; Meng et al., 2004). This line of work provides a clear cross-domain analogue of a "distributional follower" in bilevel optimization, but it differs technically from our setting: their lower-level distributions typically range over discrete route choices and are characterized via equilibrium constraints, whereas we study a general bilevel formulation with a continuous inner optimization problem, derive an explicit Gibbs hypergradient, and analyze SGLD-based tracking and estimation.

Bayesian optimal experimental design (BOED) and relation to distribution-dependent objectives. BOED (also called optimal experimental design / active learning for parameter inference) chooses a *design* or *query* variable x to maximize an *expected utility* $\mathcal{U}(x) = \mathbb{E}_{y\sim p(y|x)}[u(x, y, p(\theta \mid y, x))]$, where the utility is defined through the posterior update and common choices include *information-gain* criteria (e.g., expected reduction in posterior uncertainty, mutual information between θ and y) (Chaloner & Verdinelli, 1995; Lindley, 1956; Rainforth et al., 2024). A central computational challenge in BOED is that $\mathcal{U}(x)$ typically involves *nested* expectations (over outcomes y and over the posterior), so practical methods rely on nested Monte Carlo, variational bounds, and amortized/neural design strategies (Rainforth et al., 2024; Foster et al., 2019; 2021).

In BOED, uncertainty is primarily *epistemic*—it reflects posterior uncertainty about an unknown parameter after observing data from a chosen experiment. In contrast, our setting is *optimization-induced* (operational/algorithmic): the outer decision v changes the *law of the inner response* via a Gibbs follower $\pi_v(\theta) \propto \exp(-\ell(v, \theta)/\tau)$, and we optimize $J_\tau(v) = \mathbb{E}_{\theta\sim\pi_v}[f(v, \theta)]$. This yields an *exact* expectation-covariance hypergradient for a decision-dependent Gibbs law and supports a *persistent-chain* mini-batch SGLD estimator that tracks a *moving target* $\{\pi_{v_k}\}$. Thus, while both BOED and our work involve objectives evaluated under distributions that depend on the decision variable, the mechanism (posterior updating vs. Gibbs follower from an inner loss) and the technical core (nested expected-utility estimation vs. exact Gibbs hypergradient + moving-target tracking analysis) are fundamentally different.

Maximum-entropy modeling and Monte Carlo gradient identities. Maximum-entropy principles lead to Gibbs/Boltzmann distributions in control and inverse reinforcement learning, e.g., maximum-entropy IRL models behavior via globally normalized exponential-family trajectory distributions (Ziebart et al., 2008). Methodologically, our covariance term is a score-function style sensitivity of an expectation w.r.t. distribution parameters, closely related to standard Monte Carlo gradient estimators surveyed in Mohamed et al. (2020).

Langevin sampling and SGLD. SGLD (Welling & Teh, 2011) and its variants enable scalable sampling using stochastic gradients. Nonasymptotic guarantees for Langevin Monte Carlo with inaccurate/noisy gradients are available (e.g., Dalalyan & Karagulyan, 2019), which we leverage to obtain clean Wasserstein tracking recursions with explicit discretization and mini-batch noise floors.