A. Notation

Before presenting our proofs, we first include a table summarizing our notation.

Symbol	Meaning
$\mathcal{A}$	set of arms in bandit environment
K	number of arms in the bandit environment $\lvert\mathcal{A}\rvert$
T	Time horizon
A_t	arm pulled at time t by the algorithm $A_t \in \mathcal{A}$
$T_a(t)$	number of times arm a has been pulled by time t
X_{A_t}	reward from choosing arm A_t at time t
θ_a	parameters of likelihood functions such that, $\theta_a \in \mathbb{R}^{d_a}$
d_a	dimension of parameter space for arm a
$p_a(x\lvert\theta_a)$	parametric family of reward distributions for arm a
$\pi_a(\theta_a)$	prior distribution over the parameters for arm a
$\mu_a^{(n)}$	probability measure associated with the posterior over the parameters of arm a after n samples from arm a
$\mu_a^{(n)}[\gamma_a]$	probability measure associated with the (scaled) posterior over the parameters of arm a after n samples from arm a
$\widehat{\mu}_a^{(n)}$	probability measure resulting from running the Langevin MCMC algorithm described in Algorithm 2 which approximates $\mu_a^{(n)}$
$\bar{\mu}_a^{(n)}[\gamma_a]$	probability measure resulting from an approximate sampling method which approximates $\mu_a^{(n)}[\gamma_a]$
θ_a^*	true parameter value for arm a
$\theta_{a,t}$	sampled parameter for arm a at time t of the Thompson Sampling algorithm: $\theta_{a,t} \sim \mu_a^{(n)}$
$\bar{r}_a$	mean of the reward distribution for arm a: $\bar{r}_a = \mathbb{E}[X_a\lvert\theta_a^*]$
α_a^T	vector in $\mathbb{R}^{d_a}$ such that $\bar{r}_a = \alpha_a^T\theta_a^*$
$r_{a,t}(T_a(t))$	estimate of mean of arm a at round t: $r_{a,t}(T_a(t)) = \alpha_a^T\theta_{a,t}$
A_a	norm of α_a
m_a	Strong log-concavity parameter of the family $p_a(x;\theta)$ in θ for all x.
ν_a	Strong log-concavity parameter of the true reward distribution $p_a(x;\theta^*)$ in x.
$F_{n,a}(\theta_a)$	Averaged log likelihood over the data points: $F_{n,a}(\theta_a) = \frac{1}{n}\sum_{i=1}^{n}\log p_a(X_i, \theta_a)$
L_a	Lipschitz constant for the true reward distribution, and likelihood families $p_a(x;\theta^*)$ in x.
κ_a	condition number of the likelihood family $\kappa_a = \max\left(\frac{L_a}{m_a}, \frac{L_a}{\nu_a}\right)$.
B_a	reflects the quality of the prior: $B_a = \frac{\max_\theta \pi_a(\theta)}{\pi_a(\theta^*)}$

We also define a few notations used within the approximate sampling Algorithm 2.

Symbol	Meaning
N	number of steps of the approximate sampling algorithm
$h^{(n)}$	step size of the approximate sampling algorithm after n samples from the arm
$\theta_{ih^{(n)}}$	MCMC sample generated within i-th iteration of Algorithm 2
$\mu_{ih^{(n)}}$	measure of $\theta_{ih^{(n)}}$
k	batch-size of the stochastic gradient Langevin algorithm

B. Posterior Concentration Proof

To begin the proof of Theorem 1, we first prove that under our assumptions, the gradients of the population likelihood function concentrates.

Proposition 2. *If the prior distribution over θ_a satisfies Assumption 3, then we have:*

$$\sup_{\mathbb{R}^{d_a}} \nabla \log \pi_a(\theta_a)^T(\theta_a - \theta_a^*) \le g_a^* - \log \pi_a(\theta_a^*),$$

where $g_a^* = \max_{\theta \in \mathbb{R}^d} \log \pi_a(\theta_a)$.

Proof. Let $\log \pi_a(\theta_a) = g(\theta_a)$. From the concavity of g, we know that

$$\nabla g(\theta_a)^T(\theta_a - \theta_a^*) \le g(\theta_a) - g(\theta_a^*)$$

Since this holds for all $\theta \in \mathbb{R}^{d_a}$, we take the supremum of both sides and get that:

$$\sup_{\mathbb{R}^{d_a}} \nabla g(\theta_a)^T(\theta_a - \theta_a^*) \le g^* - g(\theta_a^*)$$

□

Let $\log B_a := g_a^* - \log \pi_a(\theta_a^*)$. If the prior is centered on the correct value of θ_a^*, then $\log B_a = 0$. Our posterior concentration rates will depend on B_a.

Before proving the posterior concentration result we first show the empirical likelihood function at θ_a^* is a sub-Gaussian random variable:

Proposition 3. *The random variable* $\|\nabla_\theta F_{a,n}(\theta_a^*)\|$ *is* $L_a\sqrt{\frac{d_a}{n\nu_a}}$*-sub-Gaussian:*

Proof. Recall that the true density $p_a(x|\theta_a^*)$ is ν_a-strongly log-concave in x and that $\nabla_\theta \log p_a(x|\theta_a^*)$ is L_a-Lipschitz in x. Notice that $\nabla_\theta F_a(\theta_a^*) = 0$ since θ_a^* is the point maximizing the population likelihood.

Let's consider the random variable $Z = \nabla_\theta \log p_a(x|\theta_a^*)$. Since $\mathbb{E}[Z] = \nabla_\theta F_a(\theta_a^*)$, the random variable Z is centered.

We start by showing Z is a subgaussian random vector. Let $v \in \mathbb{S}_{d_a}$ be an arbitrary point in the d_a-dimensional sphere and define the function $V : \mathbb{R}^{d_a} \to \mathbb{R}$ as $V(x) = \langle \nabla_\theta \log p_a(x|\theta_a^*), v\rangle$. This function is L_a-Lipschitz. Indeed let $x_1, x_2 \in \mathbb{R}^{d_a}$ be two arbitrary points in $\mathbb{R}^{d_a}$:

$$\begin{aligned}
|V(x_1) - V(x_2)| &= |\langle \nabla_\theta \log p_a(x_1|\theta_a^*) - \nabla_\theta \log p_a(x_2|\theta_a^*), v\rangle| \\
&\le \|\nabla_\theta \log p_a(x_1|\theta_a^*) - \nabla_\theta \log p_a(x_2|\theta_a^*)\|_2 \|v\|_2 \\
&= \|\nabla_\theta \log p_a(x_1|\theta_a^*) - \nabla_\theta \log p_a(x_2|\theta_a^*)\|_2 \\
&\le L_a\|x_1 - x_2\|
\end{aligned}$$

The first inequality follows by Cauchy-Schwartz, the second inequality by the Lipschitz assumption on the gradients. After a simple application of Proposition 2.18 in Ledoux (2001), we conclude that $V(x)$ is subgaussian with parameter $\frac{L_a}{\sqrt{\nu_a}}$.

Since the projection of Z onto an arbitrary direction v of the unit sphere is subgaussian, with a parameter independent of v, we conclude the random vector Z is subgaussian with the same parameter $\frac{L_a}{\sqrt{\nu_a}}$. Consequently, the vector $\nabla_\theta F_{a,n}(\theta_a^*)$, being an average of n i.i.d. subgaussian vectors with parameter $\frac{L_a}{\sqrt{\nu_a}}$ is also subgaussian with parameter $\frac{L_a}{\sqrt{n\nu_a}}$.

Since $\nabla_\theta F_{a,n}(\theta_a^*)$ is a subgaussian vector with parameter $\frac{L_a}{\sqrt{n\nu_a}}$, Lemma 1 of (Jin et al., 2019) implies it is norm subgaussian with parameter $\frac{L_a\sqrt{d_a}}{\sqrt{n\nu_a}}$.

□

Given these results we now prove Theorem 1. For clarity, we restate the theorem below:

Theorem B.1. *1 Suppose that Assumptions 1-3 hold, then given samples* $X_a^{(n)} = X_{a,1}, ..., X_{a,n}$*, the posterior distribution satisfies, for* $\delta \in (0, e^{-1/2})$*:*

$$\mathbb{P}_{\theta \sim \mu_a^{(n)}}[\gamma_a]\left(\|\theta_a - \theta_a^*\|_2 > \sqrt{\frac{2e}{m_a n}\left(\frac{d_a}{\gamma_a} + \log B_a + \left(\frac{32}{\gamma_a} + \frac{8d_a\kappa_a L_a}{\nu_a}\right)\log(1/\delta)\right)}\right) < \delta.$$

Proof. The proof makes use of the techniques used to prove Theorem 1 in Mou et al. (2019): analyzing how a carefully designed potential function evolves along trajectories of the s.d.e. By a careful accounting of terms and constants, however, we are able to keep explicit constants and derive tighter bounds which hold for any finite number of samples. Throughout the proof we drop the dependence on a and condition on the high-probability event, $G_{a,n}(\delta_1)$, defined in Proposition 3, which guarantees that the norm of the likelihood gradients concentrates with probability at least $1 - \delta_1$.

Consider the s.d.e.:

$$d\theta_t = \frac{1}{2}\nabla_\theta F_n(\theta_t)dt + \frac{1}{2n}\nabla_\theta \log \pi(\theta_t))dt + \frac{1}{\sqrt{n\gamma}}dB_t,$$

and the potential function given by:

$$V(\theta) = \frac{1}{2}e^{\alpha t}\|\theta - \theta^*\|_2^2,$$

for a choice of $\alpha > 0$. The idea is that bounds on the p-th moments of $V(\theta_t)$ can be translated into bounds on the p-th moments of $V(\theta)$ where $\theta \sim \mu^{(n)}$, due to the fact that $\lim_{t\to\infty} \theta_t = \theta \sim \mu^{(n)}$. The square-root growth in p of these moments will imply that $\|\theta - \theta^*\|_2$ has subgaussian tails with a rate that we make explicit.

We begin by using Ito's Lemma on $V(\theta_t)$:

$$V(\theta_t) = T1 + T2 + T3 + T4,$$

where:

$$\begin{aligned}
T1 &= -\frac{1}{2}\int_0^t e^{\alpha s}\langle \theta^* - \theta_s, \nabla_\theta F_n(\theta_s)\rangle ds + \frac{\alpha}{2}\int_0^t e^{\alpha s}\|\theta_s - \theta^*\|_2^2 ds \\
T2 &= \frac{1}{2n}\int_0^t e^{\alpha s}\langle \theta_s - \theta^*, \nabla_\theta \log \pi(\theta_s)\rangle ds \\
T3 &= \frac{d}{2n\gamma}\int_0^t e^{\alpha s} ds \\
T4 &= \frac{1}{\sqrt{n\gamma}}\int_0^t e^{\alpha s}\langle \theta_s - \theta^*, dB_s\rangle
\end{aligned}$$

Let us first upper bound $T1$:

$$\begin{aligned}
T1 &= -\frac{1}{2}\int_0^t e^{\alpha s}\langle \theta^* - \theta_s, \nabla_\theta F_n(\theta_s)\rangle ds + \frac{\alpha}{2}\int_0^t e^{\alpha s}\|\theta_s - \theta^*\|_2^2 ds \\
&= -\frac{1}{2}\int_0^t e^{\alpha s}\langle \theta^* - \theta_s, \nabla_\theta F_n(\theta_s) - \nabla_\theta F_n(\theta^*)\rangle ds + \frac{\alpha}{2}\int_0^t e^{\alpha s}\|\theta_s - \theta^*\|_2^2 ds \\
&\quad - \frac{1}{2}\int_0^t e^{\alpha s}\langle \theta^* - \theta_s, \nabla_\theta F_n(\theta^*)\rangle ds \\
&\overset{(i)}{\leq} \frac{\alpha - m}{2}\int_0^t e^{\alpha s}\|\theta_s - \theta^*\|_2^2 ds - \frac{1}{2}\int_0^t e^{\alpha s}\langle \theta^* - \theta_s, \nabla_\theta F_n(\theta^*)\rangle ds \\
&\overset{(ii)}{\leq} \frac{\alpha - m}{2}\int_0^t e^{\alpha s}\|\theta_s - \theta^*\|_2^2 ds + \frac{1}{2}\int_0^t e^{\alpha s}\|\theta^* - \theta_s\| \underbrace{\|\nabla_\theta F_n(\theta^*)\|}_{:=\epsilon(n)} ds
\end{aligned}$$

where in (i) we use the strong-concavity property from Assumption 1-Local, and in (ii) we use Cauchy-Shwartz.

Using Young's inequality for products, where the constant is m, gives:

$$T1 \leq \frac{2\alpha - m}{4}\int_0^t e^{\alpha s}\|\theta_s - \theta^*\|_2^2 ds + \frac{\epsilon(n)^2}{4m}\int_0^t e^{\alpha s} ds$$

Finally, choosing $\alpha = m/2$ the first term on the RHS vanishes. Evaluating the integral in the second term on the RHS gives:

$$T1 \le \frac{\epsilon(n)^2}{2m^2}\left(e^{\alpha t} - 1\right) \le \frac{\epsilon(n)^2}{m^2} e^{\alpha t}.$$

Given our assumption on the prior, our choice of $\alpha = \frac{m}{2}$ and simple algebra, we can upper bound $T2$ and $T3$ as:

$$T2 = \frac{1}{2n}\int_0^t e^{\alpha s}\langle \theta_s - \theta^*, \nabla_\theta \log \pi(\theta_s)\rangle ds \le \frac{\log B}{2\alpha n}(e^{\alpha t} - 1) \le \frac{\log B}{nm} e^{\alpha t}$$

$$T3 = \frac{d}{2n\gamma}\int_0^t e^{\alpha s} ds \le \frac{d}{\gamma n m} e^{\alpha t}.$$

We proceed to bound $T4$. Let's start by defining:

$$M_t = \int_0^t e^{\alpha s}\langle \theta_s - \theta^*, dB_s\rangle,$$

so that:

$$T4 = \frac{M_t}{\sqrt{n\gamma}}.$$

Combining all the upper bounds of $T1, T2, T3$, and $T4$ we have that:

$$V(\theta_t) \le \left(\frac{\epsilon(n)^2}{m^2} + \frac{d}{\gamma n m} + \frac{\log B}{nm}\right) e^{\alpha t} + \frac{M_t}{\sqrt{\gamma n}}.$$

To find a bound for the $p-$th moments of V, we upper bound the p-th moments of the supremum of M_t where $p \ge 1$:

$$\begin{aligned}
\mathbb{E}\left[\sup_{0\le t\le T} |M_t|^p\right] &\overset{(i)}{\le} (8p)^{\frac{p}{2}}\mathbb{E}\left[\langle M, M\rangle_T^{\frac{p}{2}}\right] \\
&= (8p)^{\frac{p}{2}}\mathbb{E}\left[\left(\int_0^T e^{2\alpha s}\|\theta_s - \theta^*\|_2^2\right)^{\frac{p}{2}} ds\right] \\
&\overset{(ii)}{\le} (8p)^{\frac{p}{2}}\mathbb{E}\left[\left(\sup_{0\le t\le T} e^{\alpha t}\|\theta_t - \theta^*\|_2^2 \int_0^T e^{\alpha s} ds\right)^{\frac{p}{2}}\right] \\
&= (8p)^{\frac{p}{2}}\mathbb{E}\left[\left(\sup_{0\le t\le T} e^{\alpha t}\|\theta_t - \theta^*\|_2^2 \frac{(e^{\alpha T} - 1)}{\alpha}\right)^{\frac{p}{2}}\right] \\
&\overset{(iii)}{\le} \left(\frac{8pe^{\alpha T}}{\alpha}\right)^{\frac{p}{2}}\mathbb{E}\left[\left(\sup_{0\le t\le T} e^{\alpha t}\|\theta_t - \theta^*\|_2^2\right)^{\frac{p}{2}}\right]
\end{aligned}$$

Inequality (i) is a direct consequence of the Burkholder-Gundy-Davis inequality (Ren, 2008), (ii) follows by pulling out the supremum out of the integral, (iii) holds because $e^{\alpha T} - 1 \le e^{\alpha T}$.

Now, let us look at the moments of $V(\theta_t)$.

$$\begin{aligned}
\mathbb{E}\left[\left(\sup_{0\le t\le T} V(\theta_t)\right)^p\right]^{\frac{1}{p}} &\le \mathbb{E}\left[\left(\sup_{0\le t\le T}\left(\frac{\epsilon(n)^2}{m^2} + \frac{d}{\gamma n m} + \frac{\log B}{nm}\right) e^{\alpha t} + \frac{|M_t|}{\sqrt{\gamma n}}\right)^p\right]^{\frac{1}{p}} \\
&\le \mathbb{E}\left[\left(\sup_{0\le t\le T}\left(\frac{\epsilon(n)^2}{m^2} + \frac{d}{\gamma n m} + \frac{\log B}{nm}\right) e^{\alpha t} + \sup_{0\le t\le T}\frac{|M_t|}{\sqrt{\gamma n}}\right)^p\right]^{\frac{1}{p}}
\end{aligned}$$

Via the Minkowski Inequality, and the fact $\epsilon(n)$ is independent of t, we can expand the above as:

$$\mathbb{E}\left[\left(\sup_{0\le t\le T} V(\theta_t)\right)^p\right]^{\frac{1}{p}} \le \underbrace{\left(\frac{d}{\gamma nm}+\frac{\log B}{nm}\right)e^{\alpha T}}_{:=U_T} + \frac{e^{\alpha T}}{m^2}\mathbb{E}\left[\epsilon(n)^{2p}\right]^{\frac{1}{p}} + \mathbb{E}\left[\left(\sup_{0\le t\le T}\frac{|M_t|}{\sqrt{n}}\right)^p\right]^{\frac{1}{p}}$$

Since, from Proposition 1, we know that $\epsilon(n)$ is a $L\sqrt{\frac{d}{n\nu}}$-sub-Gaussian vector, we know that:

$$\mathbb{E}\left[\epsilon(n)^{2p}\right]^{\frac{1}{p}} \le \left(2L\sqrt{\frac{2dp}{n\nu}}\right)^2$$

Using our upper bound on the supremum of M_t gives:

$$\mathbb{E}\left[\left(\sup_{0\le t\le T} V(\theta_t)\right)^p\right]^{\frac{1}{p}} \le U_T + \frac{e^{\alpha T}8dL^2}{\nu m^2 n}p + \mathbb{E}\left[\left(\frac{8pe^{\alpha T}}{\gamma\alpha n}\right)^{\frac{p}{2}}\left(\sup_{0\le t\le T} e^{\alpha t}\|\theta_t-\theta^*\|_2^2\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} \tag{6}$$

We proceed by bounding the second term on the RHS of the expression above:

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{8pe^{\alpha T}}{\alpha n}\right)^{\frac{p}{2}}\left(\sup_{0\le t\le T} e^{\alpha t}\|\theta_t-\theta^*\|_2^2\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} &\overset{(i)}{\le} \mathbb{E}\left[\frac{2^{p-1}}{2}\left(\frac{8pe^{\alpha T}}{\alpha\gamma n}\right)^p + \frac{1}{2^p}\left(\sup_{0\le t\le T} e^{\alpha t}\|\theta_t-\theta^*\|_2^2\right)^p\right]^{\frac{1}{p}} \\
&\overset{(ii)}{\le} 2^{\frac{p-2}{p}}\mathbb{E}\left[\left(\frac{8pe^{\alpha T}}{\alpha\gamma n}\right)^p\right]^{\frac{1}{p}} + \frac{1}{2}\mathbb{E}\left[\left(\sup_{0\le t\le T} e^{\alpha t}\|\theta_t-\theta^*\|_2^2\right)^p\right]^{\frac{1}{p}} \\
&\overset{(iii)}{\le} 16\,\mathbb{E}\left[\left(\frac{pe^{\alpha T}}{\alpha\gamma n}\right)^p\right]^{\frac{1}{p}} + \underbrace{\frac{1}{2}\mathbb{E}\left[\left(\sup_{0\le t\le T} e^{\alpha t}\|\theta_t-\theta^*\|_2^2\right)^p\right]^{\frac{1}{p}}}_{I}
\end{aligned}$$

Inequality (i) follows from using Young's inequality for products on the term inside the expectation with constant 2^{p-1}, inequality (ii) is a consequence of Minkowski Inequality and (iii) because $2^{\frac{p-2}{p}} \le 2$. We note now that the second term I on the right hand side above is exactly:

$$\frac{1}{2}\mathbb{E}\left[\left(\sup_{0\le t\le T} V(\theta_t)\right)^p\right]^{\frac{1}{p}}$$

Plugging this into Equation 6 and rearranging gives:

$$\frac{1}{2}\mathbb{E}\left[\left(\sup_{0\le t\le T} V(\theta_t)\right)^p\right]^{\frac{1}{p}} \le U_T + \frac{16e^{\alpha T}}{\alpha\gamma n}p + \frac{e^{\alpha T}8dL^2}{\nu m^2 n}p,$$

which finally results in:

$$\mathbb{E}\left[\left(\sup_{0\le t\le T} V(\theta_t)\right)^p\right]^{\frac{1}{p}} \le \frac{2}{mn}\left(\frac{d}{\gamma} + \log B + \left(\frac{32}{\gamma} + \frac{8dL^2}{\nu m}\right)p\right)e^{\alpha T}. \tag{7}$$

Given this control on the moments of the supremum of $V(\theta_t)$ (recall $V(\theta) = \frac{1}{2}e^{\alpha t}\|\theta-\theta^*\|_2^2$), we finally construct the bound on the moments of $\|\theta_T - \theta^*\|$:

$$\begin{aligned}
\mathbb{E}[\|\theta_T - \theta^*\|^p]^{\frac{1}{p}} &= \mathbb{E}\left[e^{-\frac{p\alpha T}{2}} V(\theta_T)^{\frac{p}{2}}\right]^{\frac{1}{p}} \\
&\overset{(i)}{\leq} \mathbb{E}\left[e^{-\frac{p\alpha T}{2}} \left(\sup_{0\leq t\leq T} V(\theta_t)\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} \\
&= e^{-\frac{\alpha T}{2}} \left(\mathbb{E}\left[\left(\sup_{0\leq t\leq T} V(\theta_t)\right)^{\frac{p}{2}}\right]^{\frac{2}{p}}\right)^{\frac{1}{2}} \\
&\overset{(ii)}{\leq} e^{-\frac{\alpha T}{2}} \left(\frac{2}{mn}\left(\frac{d}{\gamma} + \log B + \left(\frac{16}{\gamma} + \frac{4dL^2}{\nu m}\right) p\right) e^{\alpha T}\right)^{\frac{1}{2}} \\
&= \sqrt{\frac{2}{mn}} \left(\frac{d}{\gamma} + \log B + \left(\frac{16}{\gamma} + \frac{4dL^2}{\nu m}\right) p\right)^{\frac{1}{2}}
\end{aligned}$$

Inequality (i) follows from taking the supremum of $V(\theta_t)$, inequality (ii) from plugging in the upper bound from Equation 7.

Taking the limit as $T \to \infty$ and using Fatou's Lemma, we therefore have that the moments of $\mathbb{E}[\|\theta - \theta^*\|^p]^{\frac{1}{p}}$, with probability at least $1 - \delta_1$, grow at a rate of $\sqrt{p}$:

$$\begin{aligned}
\mathbb{E}[\|\theta - \theta^*\|^p]^{\frac{1}{p}} &\leq \liminf_{T\to\infty} \mathbb{E}[\|\theta_T - \theta^*\|^p]^{\frac{1}{p}} &(8)\\
&= \sqrt{\frac{2}{mn}} \left(\frac{d}{\gamma} + \log B + \left(\frac{16}{\gamma} + \frac{4dL^2}{\nu m}\right) p\right)^{\frac{1}{2}}. &(9)
\end{aligned}$$

To simplify notation, let $D = \left(\frac{d}{\gamma} + \log B\right)$, and $\sigma = \left(\frac{16}{\gamma} + \frac{4dL^2}{\nu m}\right)$. Therefore we have:

$$\mathbb{E}[\|\theta - \theta^*\|^p]^{\frac{1}{p}} \leq \sqrt{\frac{2}{mn}(D + \sigma p)} \tag{10}$$

The result (10), guarantees us that the norm of the uncentered random variable $\theta - \theta^*$ has subgaussian tails. We make the parameters explicit via Markov's inequality:

$$\begin{aligned}
\mathbb{P}_{\theta\sim\mu_a^{(n)}}(\|\theta - \theta^*\| > \epsilon) &\leq \frac{\mathbb{E}[\|\theta - \theta^*\|^p]}{\epsilon^p} \\
&\leq \left(\frac{\sqrt{2(D + \sigma p)}}{\sqrt{mn}\epsilon}\right)^p.
\end{aligned}$$

Choosing $p = 2\log 1/\delta$ and letting

$$\epsilon = e^{\frac{1}{2}}\sqrt{\frac{2}{mn}(D + \sigma p)}$$

gives us our desired solution:

$$\mathbb{P}_{\theta\sim\mu_a^{(n)}[\gamma_a]}\left(\|\theta - \theta^*\|_2 > \sqrt{\frac{2e}{mn}\left(\frac{d}{\gamma} + \log B + \left(\frac{32}{\gamma} + \frac{8dL^2}{\nu m}\right)\log(1/\delta)\right)}\right) < \delta.$$

for $\delta \leq e^{-0.5}$. □

C. Introduction to the Langevin Algorithms

We refer to the stochastic process represented by the following stochastic differential equation as *continuous-time Langevin dynamics*:

$$\mathrm{d}\theta_t = -\nabla U(\theta_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t.$$

We have first encountered this continuous time Langevin dynamics in Eq. (2), where we have set $U(\theta) = -\gamma_a\left(nF_{n,a}(\theta) + \log\pi_a(\theta)\right) = -\gamma_a\sum_{i=1}^{n}\log p_a\left(x_{a,i}|\theta\right) - \gamma_a\log\pi_a(\theta)$ to prove posterior concentration of $\mu_a^{(n)}[\gamma_a]$.

One important feature of the Langevin dynamics is that its invariant distribution is proportional to $e^{-U(\theta)}$. We can therefore also use it to generate samples distributed according to the unscaled posterior distribution $\mu_a^{(n)}$. Via letting $U(\theta) = -\sum_{i=1}^{n}\log p_a\left(x_{a,i}|\theta\right) - \log\pi_a(\theta)$, we obtain a continuous time dynamics which generates trajectories that converge towards the posterior distribution $\mu_a^{(n)}$ exponentially fast. To obtain an implementable algorithm, we apply Euler-Maruyama discretization to the Langevin dynamics and arrive at the following ULA update:

$$\theta_{(i+1)h^{(n)}} \sim \mathcal{N}\left(\theta_{ih^{(n)}} - h^{(n)}\nabla U(\theta_{ih^{(n)}}), 2h^{(n)}\mathrm{I}\right).$$

Since $\nabla U(\theta) = -\sum_{i=1}^{n}\nabla\log p_a\left(x_{a,i}|\theta\right) - \nabla\log\pi_a(\theta)$ in the above update rule, the computation complexity within each iteration of the Langevin algorithm grows with the number of data being collected, n. To cope with the growing number of terms in $\nabla U(\theta)$, we take a stochastic gradient approach and define $\widehat{U}(\theta) = -\frac{n}{|\mathcal{S}|}\sum_{x_k\in\mathcal{S}}\nabla\log p_a(x_k|\theta) - \nabla\log\pi_a(\theta)$, where $\mathcal{S}$ is a subset of the dataset $\{x_{a,1},\cdots,x_{a,n}\}$. For simplicity, we form $\mathcal{S}$ via subsampling uniformly from $\{x_{a,1},\cdots,x_{a,n}\}$. Substituting the stochastic gradient $\nabla\widehat{U}$ for the full gradient ∇U in the above update rule results in the SGLD algorithm.

D. Proofs for Approximate MCMC Sampling

In this Appendix we supply the proofs of concentration for approximate samples from both the ULA and SGLD MCMC methods. We will quantify the computation complexity of generating samples which are distributed close enough to the posterior. We restate the assumptions required of the likelihood for the MCMC sampling methods to converge.

Assumption 1-Uniform (Assumption on the family $p_a(X|\theta_a)$: strengthened for approximate sampling)**.** *Assume that* $\log p_a(x|\theta_a)$ *is* L_a*-smooth and* m_a*-strongly concave over the parameter* θ_a*:*

$$\begin{aligned}
-\log p_a(x|\theta_a') - \nabla_\theta\log p_a(x|\theta_a')^\top(\theta_a-\theta_a') + \frac{m_a}{2}\|\theta_a-\theta_a'\|^2 &\le -\log p_a(x|\theta_a)\\
&\le -\log p_a(x|\theta_a') - \nabla_\theta\log p_a(x|\theta_a')^\top(\theta_a-\theta_a') + \frac{L_a}{2}\|\theta_a-\theta_a'\|^2, \quad \forall\theta_a,\theta_a'\in\mathbb{R}^{d_a}, x\in\mathbb{R}.
\end{aligned}$$

Assumption 3 (Assumptions on the prior distribution). For every $a\in\mathcal{A}$ assume that $\log\pi_a(\theta_a)$ is concave with L-Lipschitz gradients for all $\theta_a\in\mathbb{R}^{d_a}$:

$$\|\nabla_\theta\pi_a(\theta) - \nabla_\theta\pi_a(\theta')\| \le L_a\|\theta-\theta'\| \quad \forall\theta,\theta'\in\mathbb{R}^{d_a}$$

Assumption 4 (Joint Lipschitz smoothness of the family $\log p_a(X|\theta_a)$: for SGLD). Assume a joint Lipschitz smoothness condition, which strengthens Assumptions 1-Local and 2 to impose the Lipschitz smoothness on the entire bivariate function $\log p_a(x;\theta)$:

$$\|\nabla_\theta\log p_a(x|\theta_a) - \nabla_\theta\log p_a(x'|\theta_a)\| \le L_a\|\theta_a-\theta_a'\| + L_a^*\|x-x'\|, \quad \forall\theta_a,\theta_a'\in\mathbb{R}^{d_a}, x,x'\in\mathbb{R}.$$

We now begin by presenting the result for ULA.

D.1. Convergence of the unadjusted Langevin algorithm (ULA)

If function $\log p_a(x;\theta)$ satisfies the Lipschitz smoothness condition in Assumption 1-Local, then we can leverage gradient based MCMC algorithms to generate samples with convergence guarantees in the p-Wasserstein distance. As stated in Algorithm 2, we initialize ULA in the n-th round from the last iterate in the $(n-1)$-th round.

Theorem 5 (ULA Convergence)**.** *Assume that the likelihood $\log p_a(x;\theta)$ and prior π_a satisfy Assumption 1-Uniform and Assumption 3. We take step size $h^{(n)} = \frac{1}{32}\frac{m_a}{n\left(L_a+\frac{1}{n}L_a\right)^2} = \mathcal{O}\left(\frac{1}{nL_a\kappa_a}\right)$ and number of steps $N = 640\frac{\left(L_a+\frac{1}{n}L_a\right)^2}{m_a^2} = \mathcal{O}\left(\kappa_a^2\right)$ in Algorithm 2. If the posterior distribution satisfy the concentration inequality that $\mathbb{E}_{\theta\sim\mu_a^{(n)}}\left[\|\theta-\theta^*\|^p\right]^{\frac{1}{p}} \le \frac{1}{\sqrt{n}}\widetilde{D}$, then for any positive even integer p, we have convergence of the ULA algorithm in W_p distance to the posterior $\mu_a^{(n)}$: $W_p\left(\widehat{\mu}_a^{(n)},\mu_a^{(n)}\right) \le \frac{2}{\sqrt{n}}\widetilde{D}, \forall\widetilde{D} \ge \sqrt{\frac{32}{m_a}d_a p}$.*

Proof of Theorem 5. We use induction to prove this theorem.

- For $n=1$, we initialize at θ_0 which is within a $\sqrt{\frac{d_a}{m_a}}$-ball from the maximum of the target distribution, $\theta_p^* = \arg\max p_a(\theta|x_1)$, where $p_a(\theta|x_1) \propto p_a(x_1|\theta)\pi_a(\theta)$ and negative $\log p_a(\theta|x_1)$ is m_a-strongly convex and (L_a+L_a)-Lipschitz smooth. Invoking Lemma 10, we obtain that for $\mathrm{d}\mu_a^{(1)} = p_a(\theta|x_1)\mathrm{d}\theta$, Wasserstein-$p$ distance between the target distribution and the point mass at its mode: $W_p\left(\mu_a^{(1)},\delta\left(\theta_p^*\right)\right) \le 5\sqrt{\frac{1}{m_a}d_a p}$. Therefore, $W_p\left(\mu_a^{(1)},\delta\left(\theta_0\right)\right) \le W_p\left(\mu_a^{(1)},\delta\left(\theta_p^*\right)\right) + \left\|\theta_0-\theta_p^*\right\| \le 6\sqrt{\frac{1}{m_a}d_a p}$. We then invoke Lemma 6, with initial condition $\mu_0 = \delta\left(\theta_p^*\right)$, to obtain the convergence in the N-th iteration of Algorithm 2 after the first pull to arm a:

$$W_p^p\left(\mu_{Nh^{(1)}},\mu_a^{(1)}\right) \le \left(1-\frac{m_a}{8}h^{(1)}\right)^{p\cdot N} W_p^p\left(\delta\left(\theta_0\right),\mu_a^{(1)}\right) + 2^{5p}\frac{(L_a+L_a)^p}{m_a^p}(d_a p)^{p/2}\left(h^{(1)}\right)^{p/2},$$

  where we have substituted in the strong convexity m_a for $\widehat{m}$ and the Lipschitz smoothness (L_a+L_a) for $\widehat{L}$. Plugging in the step size $h^{(1)} = \frac{1}{32}\frac{m_a}{(L_a+L_a)^2} \le \min\left\{\frac{m_a}{32(L_a+L_a)^2}, \frac{1}{1024}\frac{m_a^2}{(L_a+L_a)^2}\frac{\widetilde{D}^2}{d_a p}\right\}$, and number of steps $N = \frac{20}{m_a}\frac{1}{h^{(1)}} = 640\frac{(L_a+L_a)^2}{m_a^2}$, $W_p^p\left(\widehat{\mu}_a^{(1)},\mu_a^{(1)}\right) = W_p^p\left(\mu_{Nh^{(1)}},\mu_a^{(1)}\right) \le 2\widetilde{D}^p$.

- Assume that after the $(n-1)$-th pull and before the n-th pull to the arm a, the ULA algorithm guarantees that $W_p\left(\widehat{\mu}_a^{(n-1)},\mu_a^{(n-1)}\right) \le \frac{2}{\sqrt{n-1}}\widetilde{D}$. We now prove that after the n-th pull and before the $(n+1)$-th pull, it is guaranteed that $W_p\left(\widehat{\mu}_a^{(n)},\mu_a^{(n)}\right) \le \frac{2}{\sqrt{n}}\widetilde{D}$. We first obtain from the assumed posterior concentration inequality:

$$W_p(\mu_a^{(n)},\delta\left(\theta^*\right)) \le \mathbb{E}_{\theta\sim\mu_a^{(n)}}\left[\|\theta-\theta^*\|^p\right]^{\frac{1}{p}} \le \frac{1}{\sqrt{n}}\widetilde{D}. \tag{11}$$

  Therefore, for $n\ge 2$,

$$W_p\left(\mu_a^{(n)},\mu_a^{(n-1)}\right) \le W_p(\mu_a^{(n)},\delta\left(\theta^*\right)) + W_p(\mu_a^{(n-1)},\delta\left(\theta^*\right)) \le \frac{3}{\sqrt{n}}\widetilde{D}.$$

  We combine this bound with the induction hypothesis and obtain that

$$W_p\left(\mu_a^{(n)},\widehat{\mu}_a^{(n-1)}\right) \le W_p\left(\mu_a^{(n)},\mu_a^{(n-1)}\right) + W_p\left(\mu_a^{(n-1)},\widehat{\mu}_a^{(n-1)}\right) \le \frac{8}{\sqrt{n}}\widetilde{D}.$$

  From Lemma 6, we know that for $\widehat{m} = n\cdot m_a$ and $\widehat{L} = n\cdot L_a + L_a$, with initial condition $\mu_0 = \widehat{\mu}_a^{(n-1)}$, with accurate gradient,

$$W_p^p\left(\mu_{ih^{(n)}},\mu_a^{(n)}\right) \le \left(1-\frac{\widehat{m}}{8}h^{(n)}\right)^{p\cdot i} W_p^p\left(\widehat{\mu}_a^{(n-1)},\mu_a^{(n)}\right) + 2^{5p}\frac{\widehat{L}^p}{\widehat{m}^p}(d_a p)^{p/2}\left(h^{(n)}\right)^{p/2}.$$

  If we take step size $h^{(n)} = \frac{1}{32}\frac{\widehat{m}}{\widehat{L}^2} \le \min\left\{\frac{\widehat{m}}{32\widehat{L}^2}, \frac{1}{1024}\frac{1}{n}\frac{\widehat{m}^2}{\widehat{L}^2}\frac{\widetilde{D}^2}{d_a p}\right\}$ and number of steps taken in the ULA algorithm from $(n-1)$-th pull till n-th pull to be: $\widehat{N} \ge \frac{20}{\widehat{m}}\frac{1}{h^{(n)}}$,

$$W_p^p\left(\widehat{\mu}_a^{(n)},\mu_a^{(n)}\right) = W_p^p\left(\mu_{\widehat{N}h^{(n)}},\mu_a^{(n)}\right) \le \left(1-\frac{\widehat{m}}{8}h^{(n)}\right)^{p\cdot\widehat{N}}\frac{8^p\widetilde{D}^p}{n^{p/2}} + 2^{5p}\frac{\widehat{L}^p}{\widehat{m}^p}(d_a p)^{p/2}\left(h^{(n)}\right)^{p/2} \le \frac{2\widetilde{D}^p}{n^{p/2}}, \tag{12}$$

leading to the result that $W_p\left(\widehat{\mu}_a^{(n)}, \mu_a^{(n)}\right) \leq \frac{2}{\sqrt{n}}\widetilde{D}$.

Since at least one round would have past from the $(n-1)$-th pull to the n-th pull to arm a, taking number of steps in each round t to be $N = \frac{20}{\widehat{m}}\frac{1}{h^{(n)}} = 640\frac{\left(L_a+\frac{1}{n}L_a\right)^2}{m_a^2}$ suffices.

Therefore, $N = 640\frac{\left(L_a+\frac{1}{n}L_a\right)^2}{m_a^2} = \mathcal{O}\left(\frac{L_a^2}{m_a^2}\right)$. □

D.2. Convergence of the stochastic gradient Langevin algorithm (SGLD)

If $\log p_a(x;\theta)$ satisfies a stronger joint Lipschitz smoothness condition in Assumption 4, similar guarantees can be obtained for stochastic gradient MCMC algorithms.

Theorem 6 (SGLD Convergence)**.** *Assume that the family $\log p_a(x;\theta)$ and prior π_a satisfy Assumption 1-Uniform, Assumption 3, and Assumption 4. We take number of data samples in the stochastic gradient estimate $k = 32\frac{(L_a^*)^2}{m_a\nu_a} = 32\kappa_a^2$, step size $h^{(n)} = \frac{1}{32}\frac{m_a}{n\left(L_a+\frac{1}{n}L_a\right)^2} = \mathcal{O}\left(\frac{1}{nL_a\kappa_a}\right)$ and number of steps $N = 1280\frac{\left(L_a+\frac{1}{n}L_a\right)^2}{m_a^2} = \mathcal{O}\left(\kappa_a^2\right)$ in Algorithm 2. If the posterior distribution satisfy the concentration inequality that $\mathbb{E}_{\theta\sim\mu_a^{(n)}}\left[\|\theta-\theta^*\|^p\right]^{\frac{1}{p}} \leq \frac{1}{\sqrt{n}}\widetilde{D}$, then for any positive even integer p, we have convergence of the ULA algorithm in W_p distance to the posterior $\mu_a^{(n)}$: $W_p\left(\widehat{\mu}_a^{(n)}, \mu_a^{(n)}\right) \leq \frac{2}{\sqrt{n}}\widetilde{D}$, $\forall \widetilde{D} \geq \sqrt{\frac{32}{m_a}d_a p}$.*

Proof of Theorem 6. Similar to Theorem 5, we use induction to prove this theorem. After the first pull to arm a, we take the same $640\frac{\left(L_a+\frac{1}{n}L_a\right)^2}{m_a^2}$ number of steps to converge to $W_p^p\left(\widehat{\mu}_a^{(1)}, \mu_a^{(1)}\right) \leq 2\widetilde{D}^p$.

Assume that after the $(n-1)$-th pull and before the n-th pull to the arm a, the SGLD algorithm guarantees that $W_p\left(\widehat{\mu}_a^{(n-1)}, \mu_a^{(n-1)}\right) \leq \frac{2}{\sqrt{n-1}}\widetilde{D}$. We prove that after the n-th pull and before the $(n+1)$-th pull, it is guaranteed that $W_p\left(\widehat{\mu}_a^{(n)}, \mu_a^{(n)}\right) \leq \frac{2}{\sqrt{n}}\widetilde{D}$. Following the proof of Theorem 5, we combine the assumed posterior concentration inequality and the induction hypothesis to obtain:

$$W_p\left(\mu_a^{(n)}, \widehat{\mu}_a^{(n-1)}\right) \leq W_p\left(\mu_a^{(n)}, \mu_a^{(n-1)}\right) + W_p\left(\mu_a^{(n-1)}, \widehat{\mu}_a^{(n-1)}\right) \leq \frac{8}{\sqrt{n}}\widetilde{D}.$$

Denote function U as the negative log-posterior density over parameter θ. From Lemma 6, we know that for $\widehat{m} = n \cdot m_a$ and $\widehat{L} = n \cdot L_a + L_a$, with initial condition that $\mu_0 = \widehat{\mu}_a^{(n-1)}$, if the difference between the stochastic gradient $\nabla\widehat{U}$ and the exact one ∇U is bounded as $\mathbb{E}\left[\left\|\nabla U(\theta) - \nabla\widehat{U}(\theta)\right\|^p \middle|\theta\right] \leq \Delta_p$, then

$$W_p^p\left(\mu_{ih^{(n)}}, \mu_a^{(n)}\right) \leq \left(1 - \frac{\widehat{m}}{8}h^{(n)}\right)^{p\cdot i} W_p^p\left(\widehat{\mu}_a^{(n-1)}, \mu_a^{(n)}\right) + 2^{5p}\frac{\widehat{L}^p}{\widehat{m}^p}\left(d_a p\right)^{p/2}\left(h^{(n)}\right)^{p/2} + 2^{2p+3}\frac{\Delta_p}{\widehat{m}^p}.$$

We demonstrate in the following Lemma 5 that

$$\Delta_p \leq 2\frac{n^{p/2}}{k^{p/2}}\left(\frac{\sqrt{d_a p}L_a^*}{\sqrt{\nu_a}}\right)^p.$$

Lemma 5. *Denote $\widehat{U}$ as the stochastic estimator of U. Then for stochastic gradient estimate with k data points,*

$$\mathbb{E}\left[\left\|\nabla\widehat{U}(\theta) - \nabla U(\theta)\right\|^p \middle|\theta\right] \leq 2\frac{n^{p/2}}{k^{p/2}}\left(\frac{\sqrt{d_a p}L_a^*}{\sqrt{\nu_a}}\right)^p.$$

If we take the number of samples in the stochastic gradient estimator $k = 32\frac{(L_a^*)^2}{m_a\nu_a}$, then $\Delta_p \leq \frac{2}{32^{p/2}}\left(n \cdot m_a\right)^{p/2} \cdot (p \cdot d_a)^{p/2} \leq 2^{-2p-5}\frac{\widehat{m}^p\widetilde{D}^p}{n^{p/2}}$ for any $p \geq 2$. Consequently, $2^{2p+3}\frac{\Delta_p}{\widehat{m}^p} \leq \frac{1}{4}\frac{\widetilde{D}^p}{n^{p/2}}$.

If we take step size $h^{(n)} = \frac{1}{32}\frac{\widehat{m}}{\widehat{L}^2} \le \min\left\{\frac{\widehat{m}}{32\widehat{L}^2}, \frac{1}{1024}\frac{1}{n}\frac{\widehat{m}^2}{\widehat{L}^2}\frac{\widetilde{D}^2}{d_a p}\right\}$ and number of steps taken in the SGLD algorithm from $(n-1)$-th pull till n-th pull to be: $\widehat{N} \ge \frac{40}{\widehat{m}}\frac{1}{h^{(n)}}$,

$$W_p^p\left(\widehat{\mu}_a^{(n)}, \mu_a^{(n)}\right) = W_p^p\left(\mu_{\widehat{N}h^{(n)}}, \mu_a^{(n)}\right) \le \left(1 - \frac{\widehat{m}}{8}h^{(n)}\right)^{p\cdot\widehat{N}} \frac{8^p\widetilde{D}^p}{n^{p/2}} + 2^{5p}\frac{\widehat{L}^p}{\widehat{m}^p}(d_a p)^{p/2}\left(h^{(n)}\right)^{p/2} + 2^{2p+3}\frac{\Delta_p}{\widehat{m}^p}$$
$$\le \frac{2\widetilde{D}^p}{n^{p/2}},$$

leading to the result that $W_p\left(\widehat{\mu}_a^{(n)}, \mu_a^{(n)}\right) \le \frac{2}{\sqrt{n}}\widetilde{D}$. Since at least one round would have past from the $(n-1)$-th pull to the n-th pull to arm a, taking number of steps in each round t to be $N = \frac{40}{\widehat{m}}\frac{1}{h^{(n)}}$ suffices.

Therefore, $N = 1280\frac{\left(L_a + \frac{1}{n}L_a\right)^2}{m_a^2} = \mathcal{O}\left(\frac{L_a^2}{m_a^2}\right)$. □

Proof of Lemma 5. We first develop the expression:

$$\mathbb{E}\left[\left\|\nabla U(\theta) - \nabla\widehat{U}(\theta)\right\|^p\right] = n^p\mathbb{E}\left[\left\|\frac{1}{n}\sum_{i=1}^n \nabla\log p(x_i|\theta_a) - \frac{1}{k}\sum_{j=1}^k \nabla\log p(x_j|\theta_a)\right\|^p\right]$$
$$= \frac{n^p}{k^p}\mathbb{E}\left[\left\|\sum_{j=1}^k\left(\frac{1}{n}\sum_{i=1}^n \nabla\log p(x_i|\theta_a) - \nabla\log p(x_j|\theta_a)\right)\right\|^p\right].$$

We note that

$$\nabla\log p(x_j|\theta_a) - \frac{1}{n}\sum_{i=1}^n\nabla\log p(x_i|\theta_a) = \frac{1}{n}\sum_{i\ne j}\left(\nabla\log p(x_j|\theta_a) - \nabla\log p(x_i|\theta_a)\right).$$

By the joint Lipschitz smoothness Assumption 4, we know that $\nabla\log p(x|\theta_a)$ is a Lipschitz function of x:

$$\|\nabla\log p(x_j|\theta_a) - \nabla\log p(x_i|\theta_a)\| \le L_a^*\|x_j - x_i\|.$$

On the other hand, the data x follows the true distribution $p(x;\theta^*)$, which by Assumption 2 is ν_a-strongly log-concave. Applying Theorem 3.16 in (Wainwright, 2019), we obtain that $(\nabla\log p(x_j|\theta_a) - \nabla\log p(x_i|\theta_a))$ is $\frac{2L_a^*}{\sqrt{\nu_a}}$-sub-Gaussian. Leveraging the Azuma-Hoeffding inequality for martingale difference sequences (Wainwright, 2019), we obtain that sum of the $(n-1)$ sub-Gaussian random variables:

$$\left(\nabla\log p(x_j|\theta_a) - \frac{1}{n}\sum_{i=1}^n\nabla\log p(x_i|\theta_a)\right),$$

is $\frac{2\sqrt{n-1}L_a^*}{n\sqrt{\nu_a}}$-sub-Gaussian. In the same vein, $\left(\sum_{j=1}^k\left(\frac{1}{n}\sum_{i=1}^n\nabla\log p(x_i|\theta_a) - \nabla\log p(x_j|\theta_a)\right)\right)$ is $\frac{2\sqrt{k(n-1)}L_a^*}{n\sqrt{\nu_a}}$-sub-Gaussian. We then invoke the $\frac{2\sqrt{d_a k(n-1)}L_a^*}{n\sqrt{\nu_a}}$-sub-Gaussianity of

$$\left\|\sum_{j=1}^k\left(\frac{1}{n}\sum_{i=1}^n\nabla\log p(x_i|\theta_a) - \nabla\log p(x_j|\theta_a)\right)\right\|$$

and have

$$\mathbb{E}\left[\left\|\sum_{j=1}^k\left(\frac{1}{n}\sum_{i=1}^n\nabla\log p(x_i|\theta_a) - \nabla\log p(x_j|\theta_a)\right)\right\|^p\right] \le 2\left(\frac{2\sqrt{d_a k(n-1)p}L_a^*}{en\sqrt{\nu_a}}\right)^p.$$

Therefore,

$$\mathbb{E}\left[\left\|\nabla U(\theta)-\nabla\widehat{U}(\theta)\right\|^p\right]=\frac{n^p}{k^p}\mathbb{E}\left[\left\|\sum_{j=1}^{k}\left(\frac{1}{n}\sum_{i=1}^{n}\nabla\log p(x_i|\theta_a)-\nabla\log p(x_j|\theta_a)\right)\right\|^p\right]$$
$$\leq 2\frac{n^{p/2}}{k^{p/2}}\left(\frac{2\sqrt{d_a p}L_a^*}{e\sqrt{\nu_a}}\right)^p\leq 2\frac{n^{p/2}}{k^{p/2}}\left(\frac{\sqrt{d_a p}L_a^*}{\sqrt{\nu_a}}\right)^p.$$

□

D.3. Convergence of (Stochastic Gradient) Langevin Algorithm within Each Round

In this section, we examine convergence of the (stochastic gradient) Langevin algorithm to the posterior distribution over a-th arm at the n-th round. Since only the a-th arm and n-th round are considered, we drop these two indices in the notation whenever suitable. We also define some notation that will only be used within this subsection. For example, we focus on the θ parameter and denote the posterior measure $\mathrm{d}\mu_a^{(n)}(x;\theta)=\mathrm{d}\mu^*(\theta)=\exp\left(-U(\theta)\right)\mathrm{d}\theta$ as the target distribution.

Symbol	Meaning
μ^*	posterior distribution, μ_a^n
U	potential (i.e., negative log posterior density)
θ_U^*	minimum of the potential U (or mode of the posterior μ^*)
θ_t	interpolation between $\theta_{ih^{(n)}}$ and $\theta_{(i+1)h^{(n)}}$, for $t\in[ih^{(n)},(i+1)h^{(n)}]$
μ_t	measure associated with θ_t
θ_t^*	an auxiliary stochastic process with initial distribution μ^* and follows dynamics (17)
$\widehat{m}$	strong convexity of the potential U, nm_a
$\widehat{L}$	Lipschitz smoothness of the potential U, nL_a+L_a

We also formally define the Wasserstein-p distance used in the main text. Given a pair of distributions μ and ν on $\mathbb{R}^d$, a *coupling* γ is a joint distribution over the product space $\mathbb{R}^d\times\mathbb{R}^d$ that has μ and ν as its marginal distributions. We let $\Gamma(\mu,\nu)$ denote the space of all possible couplings of μ and ν. With this notation, the Wasserstein-p distance is given by

$$W^p(\mu,\nu)=\inf_{\gamma\in\Gamma(\mu,\nu)}\int_{\mathbb{R}^d\times\mathbb{R}^d}\|x-y\|^p\,\mathrm{d}\gamma(x,y). \tag{13}$$

We use the following (stochastic gradient) Langevin algorithm to generate approximate samples from the posterior distribution $\mu_a^{(n)}(\theta)$ at n-th round. For $i=0,\cdots,T$,

$$\theta_{(i+1)h^{(n)}}\sim\mathcal{N}\left(\theta_{ih^{(n)}}-h^{(n)}\nabla\widehat{U}(\theta_{ih^{(n)}}),2h^{(n)}\mathrm{I}\right), \tag{14}$$

where $\nabla\widehat{U}(\theta_{ih^{(n)}})$ is a stochastic estimate of $\nabla U(\theta_{ih^{(n)}})$. We prove in the following Lemma 6 the convergence of this algorithm within n-th round.

Lemma 6. *Assume that the potential U is $\widehat{m}$-strongly convex and $\widehat{L}$-Lipschitz smooth. Further assume that the p-th moment between the true gradient and the stochastic one satisfies:*

$$\mathbb{E}\left[\left\|\nabla U(\theta_{ih^{(n)}})-\nabla\widehat{U}(\theta_{ih^{(n)}})\right\|^p\middle|\theta_{ih^{(n)}}\right]\leq\Delta_p.$$

Then at i-th step, for $\mu_{ih^{(n)}}$ following the (stochastic gradient) Langevin algorithm with $h\leq\frac{\widehat{m}}{32\widehat{L}^2}$,

$$W_p^p\left(\mu_{ih^{(n)}},\mu^*\right)\leq\left(1-\frac{\widehat{m}}{8}h^{(n)}\right)^{p\cdot i}W_p^p\left(\mu_0,\mu^*\right)+2^{5p}\frac{\widehat{L}^p}{\widehat{m}^p}(dp)^{p/2}\left(h^{(n)}\right)^{p/2}+2^{2p+3}\frac{\Delta_p}{\widehat{m}^p}. \tag{15}$$

Remark 2. *When $\Delta_p=0$, Lemma 6 provides convergence rate of the unadjusted Langevin algorithm (ULA) with the exact gradient.*

Proof of Lemma 6. We first interpolate a continuous time stochastic process, θ_t, between $\theta_{ih^{(n)}}$ and $\theta_{(i+1)h^{(n)}}$. For $t \in [ih^{(n)}, (i+1)h^{(n)}]$,

$$\mathrm{d}\theta_t = \nabla \widehat{U}(\theta_{ih^{(n)}})\mathrm{d}t + \sqrt{2}\mathrm{d}B_t, \tag{16}$$

where B_t is standard Brownian motion. This process connects $\theta_{ih^{(n)}}$ and $\theta_{(i+1)h^{(n)}}$ and approximates the following stochastic differential equation which maintains the exact posterior distribution:

$$\mathrm{d}\theta_t^* = \nabla U(\theta_t^*)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t. \tag{17}$$

For a θ_t^* initialized from μ^* and following equation (17), θ_t^* will always have distribution μ^*.

We therefore design a coupling between the two processes: θ_t and θ_t^*, where θ_t follows equation (16) (and thereby interpolates Algorithm 2) and θ_t^* initializes from μ^* and follows equation (17) (and thereby preserves μ^*). By studying the difference between the two processes, we will obtain the convergence rate in terms of the Wasserstein-p distance.

For $t = ih^{(n)}$, we let $\theta_{ih^{(n)}}$ to couple optimally with $\theta^*_{ih^{(n)}}$, so that for

$$\left(\theta_{ih^{(n)}}, \theta^*_{ih^{(n)}}\right) \sim \gamma^* \in \Gamma_{opt}\left(\mu_{ih^{(n)}}, \mu^*_{ih^{(n)}}\right),$$

$\mathbb{E}\left[\left\|\theta_{ih^{(n)}} - \theta^*_{ih^{(n)}}\right\|^p\right] = W_p^p\left(\mu_{ih^{(n)}}, \mu^*\right)$. For $t \in [ih^{(n)}, (i+1)h^{(n)}]$, we choose a synchronous coupling $\bar{\gamma}\left(\theta_t, \theta_t^* | \theta_{ih^{(n)}}, \theta^*_{ih^{(n)}}\right) \in \Gamma\left(\mu_t(\theta_t|\theta_{ih^{(n)}}), \mu_t^*(\theta_t^*|\theta_{ih^{(n)}})\right)$ for the laws of θ_t and θ_t^*. (A synchonous coupling simply means that we use the same Brownian motion B_t in defining θ_t and θ_t^*.) We then obtain that for any pair $(\theta_t, \theta_t^*) \sim \bar{\gamma}$,

$$\begin{aligned}
\frac{\mathrm{d}\|\theta_t - \theta_t^*\|^p}{\mathrm{d}t} &= \|\theta_t - \theta_t^*\|^{p-2}\left\langle \theta_t - \theta_t^*, \frac{\mathrm{d}\theta_t}{\mathrm{d}t} - \frac{\mathrm{d}\theta_t^*}{\mathrm{d}t}\right\rangle \\
&= p\|\theta_t - \theta_t^*\|^{p-2}\left\langle \theta_t - \theta_t^*, -\nabla U(\theta_t) + \nabla U(\theta_t^*)\right\rangle \\
&+ p\|\theta_t - \theta_t^*\|^{p-2}\left\langle \theta_t - \theta_t^*, \nabla U(\theta_t) - \nabla \widehat{U}(\theta_{ih^{(n)}})\right\rangle \\
&\leq -p\widehat{m}\|\theta_t - \theta_t^*\|^p + p\|\theta_t - \theta_t^*\|^{p-1}\left\|\nabla U(\theta_t) - \nabla \widehat{U}(\theta_{ih^{(n)}})\right\| &(18)\\
&\leq -p\widehat{m}\|\theta_t - \theta_t^*\|^p &(19)\\
&\quad + p\left(\frac{p-1}{p}\left(\frac{p\widehat{m}}{2(p-1)}\right)\|\theta_t - \theta_t^*\|^p + \frac{1}{p}\frac{1}{\left(\frac{p\widehat{m}}{2(p-1)}\right)^{p-1}}\left\|\nabla U(\theta_t) - \nabla \widehat{U}(\theta_{ih^{(n)}})\right\|^p\right) &(20)\\
&\leq -\frac{p\widehat{m}}{2}\|\theta_t - \theta_t^*\|^p + \frac{2^{p-1}}{\widehat{m}^{p-1}}\left\|\nabla U(\theta_t) - \nabla \widehat{U}(\theta_{ih^{(n)}})\right\|^p, &(21)
\end{aligned}$$

where equation (20) follows from Young's inequality.

Equivalently, we can obtain

$$\frac{\mathrm{d}e^{\frac{p\widehat{m}}{2}t}\|\theta_t - \theta_t^*\|^p}{\mathrm{d}t} \leq e^{\frac{p\widehat{m}}{2}t}\frac{2^{p-1}}{\widehat{m}^{p-1}}\left\|\nabla U(\theta_t) - \nabla \widehat{U}(\theta_{ih^{(n)}})\right\|^p.$$

By the fundamental theorem of calculus,

$$\|\theta_t - \theta_t^*\|^p \leq e^{-\frac{p\widehat{m}}{2}\left(t - ih^{(n)}\right)}\left\|\theta_{ih^{(n)}} - \theta^*_{ih^{(n)}}\right\|^p + \frac{2^{p-1}}{\widehat{m}^{p-1}}\int_{ih^{(n)}}^t e^{-\frac{p\widehat{m}}{2}(t-s)}\left\|\nabla U(\theta_s) - \nabla \widehat{U}(\theta_{ih^{(n)}})\right\|^p \mathrm{d}s. \tag{22}$$

Taking expectation on both sides, we obtain that

$$\begin{aligned}
\mathbb{E}\left[\|\theta_t - \theta_t^*\|^p\right] &= \mathbb{E}\left[\mathbb{E}\left[\|\theta_t - \theta_t^*\|^p \mid \theta_{ih^{(n)}}, \theta^*_{ih^{(n)}}\right]\right] \\
&\leq e^{-\frac{p\widehat{m}}{2}\left(t - ih^{(n)}\right)}\mathbb{E}\left[\left\|\theta_{ih^{(n)}} - \theta^*_{ih^{(n)}}\right\|^p\right] \\
&+ \frac{2^{p-1}}{\widehat{m}^{p-1}}\int_{ih^{(n)}}^t e^{-\frac{p\widehat{m}}{2}(t-s)}\mathbb{E}\left[\left\|\nabla U(\theta_s) - \nabla \widehat{U}(\theta_{ih^{(n)}})\right\|^p\right]\mathrm{d}s. &(23)
\end{aligned}$$

In the above expression, the integral and expectation are exchanged using Tonelli's theorem, since

$$\left\|\nabla U(\theta_s) - \nabla \widehat{U}(\theta_{ih^{(n)}})\right\|^p$$

is positive measurable.

We further expand the expected error $\mathbb{E}\left[\left\|\nabla U(\theta_s) - \nabla \widehat{U}(\theta_{ih^{(n)}})\right\|^p\right]$:

$$\begin{aligned}
&\mathbb{E}\left[\left\|\nabla U(\theta_s) - \nabla \widehat{U}(\theta_{ih^{(n)}})\right\|^p\right] \\
&= \mathbb{E}\left[\left\|\nabla U(\theta_s) - \nabla U(\theta_{ih^{(n)}}) + \nabla U(\theta_{ih^{(n)}}) - \nabla \widehat{U}(\theta_{ih^{(n)}})\right\|^p\right] \\
&\leq \frac{1}{2}\mathbb{E}\left[\left\|2\left(\nabla U(\theta_s) - \nabla U(\theta_{ih^{(n)}})\right)\right\|^p\right] + \frac{1}{2}\mathbb{E}\left[\left\|2\left(\nabla U(\theta_{ih^{(n)}}) - \nabla \widehat{U}(\theta_{ih^{(n)}})\right)\right\|^p\right] \\
&= 2^{p-1}\mathbb{E}\left[\left\|\nabla U(\theta_s) - \nabla U(\theta_{ih^{(n)}})\right\|^p\right] + 2^{p-1}\mathbb{E}\left[\mathbb{E}\left[\left\|\nabla U(\theta_{ih^{(n)}}) - \nabla \widehat{U}(\theta_{ih^{(n)}})\right\|^p \middle| \theta_{ih^{(n)}}\right]\right] \\
&\leq 2^{p-1}\widehat{L}^p \cdot \mathbb{E}\left[\left\|\theta_s - \theta_{ih^{(n)}}\right\|^p\right] + 2^{p-1}\Delta_p.
\end{aligned} \tag{24}$$

Plugging into equation (22), we have that

$$\begin{aligned}
&\mathbb{E}\left[\|\theta_t - \theta_t^*\|^p\right] \\
&\leq e^{-\frac{p\widehat{m}}{2}\left(t-ih^{(n)}\right)}\mathbb{E}\left[\left\|\theta_{ih^{(n)}} - \theta^*_{ih^{(n)}}\right\|^p\right] \\
&+ 2^{2p-2}\frac{\widehat{L}^p}{\widehat{m}^{p-1}}\int_{ih^{(n)}}^t e^{-\frac{p\widehat{m}}{2}(t-s)}\mathbb{E}\left[\left\|\theta_s - \theta_{ih^{(n)}}\right\|^p\right]\mathrm{d}s + 2^{2p-2}(t-ih^{(n)})\frac{\Delta_p}{\widehat{m}^{p-1}}.
\end{aligned} \tag{25}$$

We provide an upper bound for $\int_{ih^{(n)}}^t e^{-\frac{p\widehat{m}}{2}(t-s)}\mathbb{E}\left[\left\|\theta_s - \theta_{ih^{(n)}}\right\|^p\right]\mathrm{d}s$ in the following lemma.

Lemma 7. *For $h^{(n)} \leq \frac{\widehat{m}}{32\widehat{L}^2}$, and for $t \in [ih^{(n)}, (i+1)h^{(n)}]$,*

$$\begin{aligned}
&\int_{ih^{(n)}}^t e^{-\frac{p\widehat{m}}{2}(t-s)}\mathbb{E}\left[\left\|\theta_s - \theta_{ih^{(n)}}\right\|^p \mathrm{d}s\right] \\
&\leq 2^{3p-3}\widehat{L}^p\left(t-ih^{(n)}\right)^{p+1}W_p^p\left(\mu_{ih^{(n)}}, \mu^*\right) + \frac{8^p}{2}\left(t-ih^{(n)}\right)^{p/2+1}(dp)^{p/2} + 2^{2p-2}(t-ih^{(n)})^{p+1}\cdot\Delta_p.
\end{aligned} \tag{26}$$

Applying this upper bound to equation (25), we obtain that for $h^{(n)} \leq \frac{\widehat{m}}{32\widehat{L}^2}$, and for $t \in [ih^{(n)}, (i+1)h^{(n)}]$,

$$\begin{aligned}
\mathbb{E}\left[\|\theta_t - \theta_t^*\|^p\right] &\leq e^{-\frac{p\widehat{m}}{2}\left(t-ih^{(n)}\right)}\mathbb{E}\left[\left\|\theta_{ih^{(n)}} - \theta^*_{ih^{(n)}}\right\|^p\right] + 2^{5p-5}\frac{\widehat{L}^{2p}}{\widehat{m}^{p-1}}\left(t-ih^{(n)}\right)^{p+1}W_p^p\left(\mu_{ih^{(n)}}, \mu^*\right) \\
&\quad + 2^{5p-3}\frac{\widehat{L}^p}{\widehat{m}^{p-1}}\left(t-ih^{(n)}\right)^{p/2+1}(dp)^{p/2} + 2^{4p-4}\frac{\widehat{L}^p}{\widehat{m}^{p-1}}(t-ih^{(n)})^{p+1}\cdot\Delta_p \\
&\qquad + 2^{2p-2}(t-ih^{(n)})\frac{\Delta_p}{\widehat{m}^{p-1}} \\
&\leq \left(1 - \frac{\widehat{m}}{4}\left(t-ih^{(n)}\right)\right)^p\mathbb{E}\left[\left\|\theta_{ih^{(n)}} - \theta^*_{ih^{(n)}}\right\|^p\right] + 2^{5p-5}\frac{\widehat{L}^{2p}}{\widehat{m}^{p-1}}\left(t-ih^{(n)}\right)^{p+1}W_p^p\left(\mu_{ih^{(n)}}, \mu^*\right) \\
&\quad + 2^{5p-3}\frac{\widehat{L}^p}{\widehat{m}^{p-1}}\left(t-ih^{(n)}\right)^{p/2+1}(dp)^{p/2} + 2^{2p}(t-ih^{(n)})\frac{\Delta_p}{\widehat{m}^{p-1}}.
\end{aligned}$$

Recognizing that $\widehat{\gamma}\left(\theta_t, \theta_t^*\right) = \mathbb{E}_{\left(\theta_{ih^{(n)}}, \theta^*_{ih^{(n)}}\right)\sim\gamma^*}\left[\bar{\gamma}\left(\theta_t, \theta_t^* | \theta_{ih^{(n)}}, \theta^*_{ih^{(n)}}\right)\right]$ is a coupling, we achieve the upper bound for

$W_p^p(\mu_t, \mu^*)$:

$$
\begin{aligned}
W_p^p(\mu_t, \mu^*) &\le \mathbb{E}_{(\theta_t, \theta_t^*)\sim\widehat{\gamma}}\left[\|\theta_t - \theta_t^*\|^p\right] \\
&\le \left(1 - \frac{\widehat{m}}{4}\left(t - ih^{(n)}\right)\right)^p \mathbb{E}_{\left(\theta_{ih^{(n)}}, \theta^*_{ih^{(n)}}\right)\sim\gamma^*}\left[\left\|\theta_{ih^{(n)}} - \theta^*_{ih^{(n)}}\right\|^p\right] \\
&\quad + 2^{5p-5}\frac{\widehat{L}^{2p}}{\widehat{m}^{p-1}}\left(t - ih^{(n)}\right)^{p+1} W_p^p(\mu_{ih^{(n)}}, \mu^*) + 2^{5p-3}\frac{\widehat{L}^p}{\widehat{m}^{p-1}}\left(t - ih^{(n)}\right)^{p/2+1}(dp)^{p/2} \\
&\quad + 2^{2p}(t - ih^{(n)})\frac{\Delta_p}{\widehat{m}^{p-1}}. \\
&\le \left(1 - \frac{\widehat{m}}{8}\left(t - ih^{(n)}\right)\right)^p W_p^p(\mu_{ih^{(n)}}, \mu^*) + 2^{5p-3}\frac{\widehat{L}^p}{\widehat{m}^{p-1}}\left(t - ih^{(n)}\right)^{p/2+1}(dp)^{p/2} \qquad (27) \\
&\quad + 2^{2p}(t - ih^{(n)})\frac{\Delta_p}{\widehat{m}^{p-1}}. \qquad (28)
\end{aligned}
$$

Taking $t = (i+1)h^{(n)}$, the recurring bound reads

$$
W_p^p\left(\mu_{(i+1)h^{(n)}}, \mu^*\right) \le \left(1 - \frac{\widehat{m}}{8}h^{(n)}\right)^p W_p^p(\mu_{ih^{(n)}}, \mu^*) + 2^{5p-3}\frac{\widehat{L}^p}{\widehat{m}^{p-1}}(dp)^{p/2}\left(h^{(n)}\right)^{p/2+1} + \frac{4^p}{\widehat{m}^{p-1}}h^{(n)}\Delta_p.
$$

We finish the proof by invoking the recursion i times:

$$
\begin{aligned}
W_p^p(\mu_{ih^{(n)}}, \mu^*) &\le \left(1 - \frac{\widehat{m}}{8}h^{(n)}\right)^p W_p^p\left(\mu_{(i-1)h^{(n)}}, \mu^*\right) + 2^{5p-3}\frac{\widehat{L}^p}{\widehat{m}^{p-1}}(dp)^{p/2}\left(h^{(n)}\right)^{p/2+1} + \frac{4^p}{\widehat{m}^{p-1}}h^{(n)}\Delta_p \\
&\le \left(1 - \frac{\widehat{m}}{8}h^{(n)}\right)^{p\cdot i} W_p^p(\mu_0, \mu^*) \\
&\quad + \sum_{k=0}^{i-1}\left(1 - \frac{\widehat{m}}{8}h^{(n)}\right)^{p\cdot k}\cdot\left(2^{5p-3}\frac{\widehat{L}^p}{\widehat{m}^{p-1}}(dp)^{p/2}\left(h^{(n)}\right)^{p/2+1} + \frac{4^p}{\widehat{m}^{p-1}}h^{(n)}\Delta_p\right) \\
&\le \left(1 - \frac{\widehat{m}}{8}h^{(n)}\right)^{p\cdot i} W_p^p(\mu_0, \mu^*) + 2^{5p}\frac{\widehat{L}^p}{\widehat{m}^p}(dp)^{p/2}\left(h^{(n)}\right)^{p/2} + 2^{2p+3}\frac{\Delta_p}{\widehat{m}^p}. \qquad (29)
\end{aligned}
$$

□

D.3.1. Supporting proofs for Lemma 6

Proof of Lemma 7. We use the update rule of ULA to develop $\int_{ih^{(n)}}^t e^{-\frac{p\widehat{m}}{2}(t-s)}\mathbb{E}\left[\|\theta_s - \theta_{ih^{(n)}}\|^p\right]\mathrm{d}s$:

$$
\begin{aligned}
&\int_{ih^{(n)}}^t e^{-\frac{p\widehat{m}}{2}(t-s)}\mathbb{E}\left[\|\theta_s - \theta_{ih^{(n)}}\|^p\,\mathrm{d}s\right] \\
&= \int_{ih^{(n)}}^t e^{-\frac{p\widehat{m}}{2}(t-s)}\mathbb{E}\left[\left\|-(s - ih^{(n)})\left(\nabla U(\theta_{ih^{(n)}}) - \left(\nabla U(\theta_{ih^{(n)}}) - \nabla\widehat{U}(\theta_{ih^{(n)}})\right)\right) + \sqrt{2}(B_s - B_{ih^{(n)}})\right\|^p\right]\mathrm{d}s \\
&\le 2^{2p-2}(t - ih^{(n)})^p\int_{ih^{(n)}}^t e^{-\frac{p\widehat{m}}{2}(t-s)}\mathbb{E}\left[\|\nabla U(\theta_{ih^{(n)}})\|^p\right]\mathrm{d}s \\
&+ 2^{3p/2-1}\int_{ih^{(n)}}^t e^{-\frac{p\widehat{m}}{2}(t-s)}\mathbb{E}\left[\|B_s - B_{ih^{(n)}}\|^p\right]\mathrm{d}s \\
&+ 2^{2p-2}(t - ih^{(n)})^p\int_{ih^{(n)}}^t e^{-\frac{p\widehat{m}}{2}(t-s)}\mathbb{E}\left[\left\|\nabla U(\theta_{ih^{(n)}}) - \nabla\widehat{U}(\theta_{ih^{(n)}})\right\|^p\right]\mathrm{d}s \\
&\le 2^{2p-2}\widehat{L}^p\left(t - ih^{(n)}\right)^{p+1}\mathbb{E}\left[\|\theta_{ih^{(n)}} - \theta_U^*\|^p\right] + 2^{3p/2-1}\int_{ih^{(n)}}^t\mathbb{E}\left[\|B_s - B_{ih^{(n)}}\|^p\right]\mathrm{d}s \\
&\qquad + 2^{2p-2}\left(t - ih^{(n)}\right)^{p+1}\Delta_p. \qquad (30)
\end{aligned}
$$

where θ_U^* is the fixed point of U. We then use the following lemma to simplify the above expression.

Lemma 8. *The integrated p-th moment of the Brownian motion can be bounded as:*

$$\int_{ih^{(n)}}^{t} \mathbb{E}\left\|B_s - B_{ih^{(n)}}\right\|^p \mathrm{d}s \le 2\left(\frac{dp}{e}\right)^{p/2}\left(t - ih^{(n)}\right)^{p/2+1}. \tag{31}$$

We also provide bound for the p-th moment of $\|\theta_{ih^{(n)}} - \theta_U^*\|$.

Lemma 9. *For $\theta_{ih^{(n)}} \sim \mu_{ih^{(n)}}$,*

$$\mathbb{E}\left\|\theta_{ih^{(n)}} - \theta_U^*\right\|^p \le 2^{p-1} W_p^p\left(\mu_{ih^n}, \mu^*\right) + \frac{10^p}{2}\left(\frac{dp}{\widehat{m}}\right)^{p/2}. \tag{32}$$

Plugging the results into equation (30), we obtain that for $h^{(n)} \le \frac{\widehat{m}}{32\widehat{L}^2}$, and for $t \in [ih^{(n)}, (i+1)h^{(n)}]$,

$$\begin{aligned}
&\int_{ih^{(n)}}^{t} e^{-\frac{p\widehat{m}}{2}(t-s)} \mathbb{E}\left[\left\|\theta_s - \theta_{ih^{(n)}}\right\|^p \mathrm{d}s\right] \\
&\le 2^{3p-3}\widehat{L}^p\left(t - ih^{(n)}\right)^{p+1} W_p^p\left(\mu_{ih^n}, \mu^*\right) + \frac{40^p}{8}\widehat{L}^p\left(t - ih^{(n)}\right)^{p+1}\left(\frac{dp}{\widehat{m}}\right)^{p/2} \\
&+ \left(\frac{8}{e}\right)^{p/2}(dp)^{p/2}\left(t - ih^{(n)}\right)^{p/2+1} + 2^{2p-2}(t - ih^{(n)})^{p+1}\cdot\Delta_p \\
&\le 2^{3p-3}\widehat{L}^p\left(t - ih^{(n)}\right)^{p+1} W_p^p\left(\mu_{ih^n}, \mu^*\right) + \frac{8^p}{2}\left(t - ih^{(n)}\right)^{p/2+1}(dp)^{p/2} + 2^{2p-2}(t - ih^{(n)})^{p+1}\Delta_p.
\end{aligned} \tag{33}$$

□

Proof of Lemma 8. The Brownian motion term can be upper bounded by higher moments of a normal random variable:

$$\int_{ih^{(n)}}^{t} \mathbb{E}\left\|B_s - B_{ih^{(n)}}\right\|^p \mathrm{d}s \le \left(t - ih^{(n)}\right)\mathbb{E}\left\|B_t - B_{ih^{(n)}}\right\|^p = \left(t - ih^{(n)}\right)^{p/2+1}\mathbb{E}\left\|v\right\|^p,$$

where v is a standard d-dimensional normal random variable. We then invoke the $\sqrt{d}$ sub-Gaussianity of $\|v\|$ and have (assuming p to be an even integer):

$$\mathbb{E}\left\|v\right\|^p \le \frac{p!}{2^{p/2}(p/2)!}d^{p/2} \le \frac{e^{1/12p}\sqrt{2\pi p}(p/e)^p}{2^{p/2}\sqrt{\pi p}(p/2e)^{p/2}}d^{p/2} \le 2\left(\frac{dp}{e}\right)^{p/2}.$$

□

Proof of Lemma 9. For the $\mathbb{E}\left\|\theta_{ih^{(n)}} - \theta_U^*\right\|^p$ term, we note that any coupling of a distribution with a delta measure is their product measure. Therefore, $\mathbb{E}\left\|\theta_{ih^{(n)}} - \theta_U^*\right\|^p$ relates to the p-Wasserstein distance between $\mu_{ih^{(n)}}$ and the delta measure at the fixed point θ_U^*, $\delta\left(\theta_U^*\right)$:

$$\begin{aligned}
\mathbb{E}\left\|\theta_{ih^{(n)}} - \theta_U^*\right\|^p = W_p^p\left(\mu_{ih^{(n)}}, \delta\left(\theta_U^*\right)\right) &\le \left(W_p\left(\mu_{ih^{(n)}}, \mu^*\right) + W_p\left(\mu^*, \delta\left(\theta_U^*\right)\right)\right)^p \\
&\le 2^{p-1}W_p^p\left(\mu_{ih^{(n)}}, \mu^*\right) + 2^{p-1}W_p^p\left(\mu^*, \delta\left(\theta_U^*\right)\right).
\end{aligned}$$

We then bound $W_p^p\left(\mu^*, \delta\left(\theta_U^*\right)\right)$ in the following lemma.

Lemma 10. *Assume the posterior μ^* is $\widehat{m}$-strongly log-concave. Then for $\theta_U^* = \arg\max \mu^*$,*

$$W_p^p\left(\mu^*, \delta\left(\theta_U^*\right)\right) \le 5^p\left(\frac{dp}{\widehat{m}}\right)^{p/2}. \tag{34}$$

Therefore,

$$\mathbb{E}\left\|\theta_{ih^{(n)}}^{(n)}-\theta_n^*\right\|^p \leq 2^{p-1}W_p^p\left(\mu_{ih^{(n)}},\mu^*\right)+\frac{10^p}{2}\left(\frac{dp}{\widehat{m}}\right)^{p/2}.$$

□

Proof of Lemma 10. We first decompose $W_p\left(\mu^*,\delta\left(\theta_U^*\right)\right)$ into two terms:

$$W_p\left(\mu^*,\delta\left(\theta_U^*\right)\right) \leq W_p\left(\mu^*,\delta\left(\mathbb{E}_{\theta\sim\mu^*}\left[\theta\right]\right)\right)+\left\|\theta_U^*-\mathbb{E}_{\theta\sim\mu^*}\left[\theta\right]\right\|.$$

By the celebrate relation between mean and mode for 1-unimodal distributions (see, e.g., Basu and DasGupta, 1996, Theorem 7), we can first bound the difference between mean and mode:

$$\left(\theta_U^*-\mathbb{E}_{\theta\sim\mu^*}\left[\theta\right]\right)^{\mathrm{T}}\Sigma^{-1}\left(\theta_U^*-\mathbb{E}_{\theta\sim\mu^*}\left[\theta\right]\right)\leq 3.$$

where Σ is the covariance matrix of μ^*. Therefore,

$$\left\|\theta_U^*-\mathbb{E}_{\theta\sim\mu^*}\left[\theta\right]\right\|^2 \leq \frac{3}{\widehat{m}}. \tag{35}$$

We then bound $W_p\left(\mu^*,\delta\left(\mathbb{E}_{\theta\sim\mu^*}\left[\theta\right]\right)\right)$. Since the coupling between μ^* and the delta measure $\delta\left(\mathbb{E}_{\theta\sim\mu^*}\left[\theta\right]\right)$ is their product measure, we can directly obtain that the p-Wasserstein distance is the p-th moments of μ^*:

$$W_p^p\left(\mu^*,\delta\left(\mathbb{E}_{\theta\sim\mu^*}\left[\theta\right]\right)\right)=\int\left\|\theta-\mathbb{E}_{\theta\sim\mu^*}\left[\theta\right]\right\|^p\mathrm{d}\mu^*(\theta).$$

We invoke the Herbst argument (see, e.g., Ledoux, 1999) to obtain the p-th moment bound. We first note that for an $\widehat{m}$-strongly log-concave distribution, it has a log Sobolev constant of $\widehat{m}$. Then using the Herbst argument, we know that $x\sim\mu^*$ is a sub-Gaussian random vector with parameter $\sigma^2=\frac{1}{2\widehat{m}}$:

$$\int e^{\lambda u^{\mathrm{T}}\left(\theta-\mathbb{E}_{\theta\sim\mu^*}\left[\theta\right]\right)}\mathrm{d}\mu^*(\theta)\leq e^{\frac{\lambda^2}{4\widehat{m}}},\quad \forall\left\|u\right\|=1.$$

Hence θ is $2\sqrt{\frac{d}{\widehat{m}}}$ norm-sub-Gaussian, which implies that

$$\left(\mathbb{E}_{\theta\sim\mu^*}\left[\left\|\theta-\mathbb{E}_{\theta\sim\mu^*}\left[\theta\right]\right\|^p\right]\right)^{1/p}\leq 2e^{1/e}\sqrt{\frac{dp}{\widehat{m}}}. \tag{36}$$

Combining equations (35) and (36), we obtain the final result that

$$\begin{aligned}W_p^p\left(\mu^*,\delta\left(\theta_U^*\right)\right) &\leq \left(2e^{1/e}\sqrt{\frac{dp}{\widehat{m}}}+\sqrt{\frac{3}{\widehat{m}}}\right)^p\\ &\leq 5^p\left(\frac{dp}{\widehat{m}}\right)^{p/2}.\end{aligned}$$

□

Lemma 11. *Assume that the likelihood $\log p_a(x;\theta)$, prior distribution, and true distributions satisfy Assumptions 1-3, and that arm a has been chosen $n=T_a(t)$ times up to iteration t of the Thompson sampling algorithm. Further, assume that we choose the stepsize step size $h^{(n)}=\frac{1}{32}\frac{m_a}{n\left(L_a+\frac{1}{n}L_a\right)^2}=\mathcal{O}\left(\frac{m_a}{nL_a^2}\right)$, and number of steps $N=640\frac{\left(L_a+\frac{1}{n}L_a\right)^2}{m_a^2}=\mathcal{O}\left(\frac{L_a^2}{m_a^2}\right)$ in Algorithm 2 then for $\delta_2\in(0,e^{-1/2})$:*

$$\mathbb{P}_{\theta_{a,t}\sim\bar{\mu}_a^{(n)}[\gamma_a]}\left(\left\|\theta_{a,t}-\theta_a^*\right\|_2>\sqrt{\frac{36e}{m_a n}\left(d_a+\log B_a+2\sigma\log 1/\delta_1+2\left(\sigma_a+\frac{m_a d_a}{18L_a\gamma_a}\right)\log 1/\delta_2\right)}\,\middle|\,Z_{n-1}\right)<\delta_2.$$

where $Z_{t-1} = \{\|\theta_{a,t-1} - \theta_a^*\| \le C(n)\}$ *for:*

$$C(n) = \sqrt{\frac{18e}{nm_a}} \left(d_a + \log B_a + 2\sigma \log 1/\delta_1\right)^{\frac{1}{2}},$$

$\sigma = 16 + \frac{4d_a L_a^2}{\nu_a m_a}$, *and where* $\theta_{a,t-1}$ *is the sample from the previous round of the Thompson sampling algorithm for arm* a.

Proof. We begin as in the proof of Theorem 3, except that we now take $\mu_0 = \delta_{\theta_{a,t-1}}$, where $\theta_{a,t-1}$ is the sample from the previous step of the algorithm:

$$W_p^p\left(\mu_{ih^{(n)}}, \mu_a^{(n)}\right) \le \left(1 - \frac{\widehat{m}}{8} h^{(n)}\right)^{p\cdot i} W_p^p\left(\delta(\theta_{a,t-1}), \mu_a^{(n)}\right) + \frac{80^p}{2} \frac{\widehat{L}^p}{\widehat{m}^p} (dp)^{p/2} \left(h^{(n)}\right)^{p/2}.$$

We first use the triangle inequality on the first term on the RHS:

$$\begin{aligned} W_p\left(\delta(\theta_{a,t-1}), \mu_a^{(n)}\right) &\le W_p\left(\delta(\theta_{a,t-1}), \delta_{\theta_a^*}\right) + W_p\left(\delta(\theta_a^*), \mu_a^{(n)}\right) \\ &= \|\theta_a^* - \theta_{a,t-1}\| + + W_p\left(\delta(\theta_a^*), \mu_a^{(n)}\right) \\ &\le C(n) + \frac{\tilde{D}}{\sqrt{n}} \end{aligned}$$

where we have used the fact that $\|\theta_a^* - \theta_{a,t-1}\| \le C(n)$ by assumption, and the definition of $\tilde{D}$ from the proof of Theorem 5: $\widetilde{D} = \sqrt{\frac{2}{m_a}} \left(d_a + \log B_a + \sigma p\right)^{\frac{1}{2}}$.

Since:

$$C(n) = \sqrt{\frac{18e}{m_a}} \left(d_a + \log B_a + 2\sigma \log 1/\delta_1\right)^{\frac{1}{2}},$$

We can further develop this upper bound:

$$\begin{aligned} W_p\left(\delta_{\theta_{a,t-1}}, \mu_a^{(n)}\right) &\le \frac{\tilde{D}}{\sqrt{n}} + C(n) \\ &\le 8\sqrt{\frac{2}{m_a n}} \left(d_a + \log B_a + 2\sigma \log 1/\delta_1 + \sigma p\right)^{\frac{1}{2}}, \end{aligned}$$

where to derive this result we have used the fact that $\sqrt{2(x+y)} \ge \sqrt{x} + \sqrt{y}$.

Letting $\bar{D} = \sqrt{\frac{2}{m_a n}} \left(d_a + \log B_a + 2\sigma \log 1/\delta_1 + \sigma p\right)^{\frac{1}{2}}$, we see that our final result is:

$$W_p\left(\delta_{\theta_{a,t-1}}, \mu_a^{(n)}\right) \le \frac{8}{\sqrt{n}} \bar{D},$$

where $\tilde{D} < \bar{D}$. Using the same choice of $h^{(n)}$ and number of steps N as in the proof or Theorem 5 guarantees us that:

$$W_p^p\left(\mu_{ih^{(n)}}, \mu_a^{(n)}\right) \le 2\left(\frac{\bar{D}}{\sqrt{n}}\right)^p$$

Further combining this with the triangle inequality, and the fact that $\tilde{D} < \bar{D}$ gives us that:

$$W_p\left(\mu_{ih^{(n)}}, \delta_{\theta^*}\right) \leq \frac{\tilde{D}}{\sqrt{n}} + \frac{\bar{D}}{\sqrt{n}} \leq 3\frac{\bar{D}}{\sqrt{n}},$$

Now, since the sample returned by the Langevin algorithm is given by:

$$\theta_a = \theta_N + Z, \tag{37}$$

where $Z \sim \mathcal{N}\left(0, \frac{1}{nL_a\gamma_a}I\right)$, it remains to bound the distance between the approximate posterior $\hat{\mu}_a^{(n)}$ of θ_a and the distribution of $\theta_{Nh^{(n)}}$. Since $\theta_a - \theta_{Nh^{(n)}} = Z$, for any even integer p,

$$\begin{aligned} W_p^p\left(\bar{\mu}_a^{(n)}, \bar{\mu}_a^{(n)}[\gamma_a]\right) = \left(\inf_{\gamma\in\Gamma\left(\bar{\mu}_a^{(n)}, \bar{\mu}_a^{(n)}[\gamma_a]\right)} \int \|\theta_a - \theta_N\|^p \,\mathrm{d}\theta_a \mathrm{d}\theta_N\right)^{1/p} &\leq \mathbb{E}[\|Z\|^p]^{\frac{1}{p}} \\ &\leq \sqrt{\frac{d}{nL_a\gamma_a}}\left(\frac{2^{p/2}\mathbf{\Gamma}(\frac{p+1}{2})}{\sqrt{\pi}}\right)^{1/p} \\ &\leq \sqrt{\frac{d}{nL_a\gamma_a}}\left(2^{p/2}\left(\frac{p}{2}\right)^{p/2}\right)^{1/p} \\ &\leq \sqrt{\frac{dp}{nL_a\gamma_a}}, \end{aligned}$$

where we have used upper bound of the Stirling type for the Gamma function $\mathbf{\Gamma}(\cdot)$ in the second last inequality.

Thus, we have, via the triangle inequality once again, that:

$$\begin{aligned} W_p\left(\bar{\mu}_a^{(n)[\gamma_a]}, \delta_{\theta^*}\right) &\leq 3\frac{\bar{D}}{\sqrt{n}} + \sqrt{\frac{dp}{nL_a\gamma_a}} \\ &\leq \sqrt{\frac{36}{m_a n}}\left(d_a + \log B_a + 2\sigma_a \log 1/\delta_1 + \left(\sigma_a + \frac{d_a}{18L_a\gamma_a}\right)p\right)^{\frac{1}{2}}, \end{aligned}$$

which, by the same derivation as in the proof of Theorem 1, gives us that:

$$\mathbb{P}_{\theta_{a,t}\sim\bar{\mu}_a^{(n)}[\gamma_a]}\left(\|\theta_{a,t} - \theta_a^*\|_2 > \sqrt{\frac{36e}{m_a n}\left(d_a + \log B_a + 2\sigma \log 1/\delta_1 + 2\left(\sigma_a + \frac{m_a d_a}{18L_a\gamma_a}\right)\log 1/\delta_2\right)}\,\middle|\, Z_{n-1}\right) < \delta_2.$$

for $\delta_2 \in (0, e^{-1/2})$:. ☐

We remark that via an identical argument, the following Lemma holds as well:

Lemma 12. *Assume that the family $\log p_a(x;\theta)$ and the prior π_a satisfy Assumptions 1-4 and that arm a has been chosen $n = T_a(t)$ times up to iteration t of the Thompson sampling algorithm. If we take number of data samples in the stochastic gradient estimate $k = 32\frac{(L_a^*)^2}{m_a\nu_a}$, step size $h^{(n)} = \frac{1}{32}\frac{m_a}{n\left(L_a + \frac{1}{n}L_a\right)^2} = \mathcal{O}\left(\frac{m_a}{nL_a^2}\right)$ and number of steps $N = 1280\frac{\left(L_a + \frac{1}{n}L_a\right)^2}{m_a^2} = \mathcal{O}\left(\frac{L_a^2}{m_a^2}\right)$ in Algorithm 2, then for $\delta_2 \in (0, e^{-1/2})$::*

$$\mathbb{P}_{\theta_{a,t}\sim\bar{\mu}_a^{(n)}[\gamma_a]}\left(\|\theta_{a,t} - \theta_a^*\|_2 > \sqrt{\frac{36e}{m_a n}\left(d_a + \log B_a + 2\sigma \log 1/\delta_1 + 2\left(\sigma_a + \frac{m_a d_a}{18L_a\gamma_a}\right)\log 1/\delta_2\right)}\,\middle|\, Z_{n-1}\right) < \delta_2.,$$

where $Z_{t-1} = \{\|\theta_{a,t-1} - \theta_a^*\| \leq C(n)\}$ for the parameters:

$$C(n) = \sqrt{\frac{18e}{nm_a}}\left(d_a + \log B_a + 2\sigma \log 1/\delta_1\right)^{\frac{1}{2}}, \qquad \sigma = 16 + \frac{4d_a L_a^2}{\nu_a m_a},$$

and $\theta_{a,t-1}$ being the sample from the previous round of the Thompson sampling algorithm over arm a.

E. Regret Proofs

We now present the proof of logarithmic regret of Thompson sampling under our assumptions with samples from the true posterior and from the approximate sampling schemes discussed in Section 4. To provide the regret guarantees for Thompson sampling with samples from the *true* posterior and from approximations to the posterior, we proceed as is common in regret proofs for multi-armed bandits by upper-bounding the number of times a sub-optimal arm $a \in \mathcal{A}$ is pulled up to time T, denoted $T_a(T)$. Without loss of generality we assume throughout this section that arm 1 is the optimal arm, and define the filtration associated with a run of the algorithm as $\mathcal{F}_t = \{A_1, X_1, A_2, X_2, ..., A_t, X_t\}$.

To upper bound the expected number of times a sub-optimal arm is pulled up to time T, we first define the event $E_a(t) = \{r_{a,t}(T_a(t)) \geq \bar{r}_1 - \epsilon\}$ for some $\epsilon > 0$. This captures the event that the mean calculated from the value of θ_a sampled from the posterior at time $t \leq T$, $r_{a,t}(T_a(t))$, is greater than $\bar{r}_1 - \epsilon$ (recall $\bar{r}_1$ is the optimal arm's mean). Given these events, we proceed to decompose the expected number of pulls of a sub-optimal arm $a \in \mathcal{A}$ as:

$$\mathbb{E}[T_a(T)] = \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{I}(A_t = a)\right] = \underbrace{\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{I}(A_t = a, E_a^c(t))\right]}_{I} + \underbrace{\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{I}(A_t = a, E_a(t))\right]}_{II}. \tag{38}$$

In Lemma 13 we upper bound (I), and then bound term (II) in Lemmas 14.

We note that this proof follows a similar structure to that of the regret bound for Thompson sampling for Bernoulli bandits and bounded rewards in (Agrawal and Goyal, 2012). However, to give the regret guarantees that incorporate the quality of the priors as well as the potential errors and lack of independence resulting from the approximate sampling methods we discuss in Section 4 the proof is more complex.

Lemma 13 (Bounding I). *For a sub-optimal arm $a \in \mathcal{A}$, we have that:*

$$I = \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{I}(A_t = a, E_a^c(t))\right] \leq \mathbb{E}\left[\sum_{s=1}^{T-1} \frac{1}{p_{1,s}} - 1\right].$$

where $p_{a,s} = \mathbb{P}(r_{a,t}(s) > \bar{r}_1 - \epsilon | \mathcal{F}_{t-1})$, for some $\epsilon > 0$.

Proof. To bound term I of (3), we first recall A_t is the arm achieving the largest sample reward mean at round t. Further, we define A_t' to be the arm achieving the maximum sample mean value among all the suboptimal arms:

$$A_t' = \operatorname*{argmax}_{a \in \mathcal{A}, a \neq 1} r_a(t, T_a(t)).$$

Since $\mathbb{E}\left[\mathbb{I}(A_t = a, E_a^c(t))\right] = \mathbb{P}\left(A_t = a, E_a^c(t)\right)$, we aim to bound $\mathbb{P}(A_t = a, E_a^c(t)|\mathcal{F}_{t-1})$. We note that the following inequality holds:

$$\begin{aligned}\mathbb{P}(A_t = a, E_a^c(t)|\mathcal{F}_{t-1}) &\leq \mathbb{P}(A_t' = a, E_a^c(t)|\mathcal{F}_{t-1})(\mathbb{P}(r_1(t, T_1(t)) \leq \bar{r}_1 - \epsilon|\mathcal{F}_{t-1})) \\ &= \mathbb{P}(A_t' = a, E_a^c(t)|\mathcal{F}_{t-1})(1 - \mathbb{P}(E_1(t)|\mathcal{F}_{t-1})).\end{aligned} \tag{39}$$

We also note that the term $\mathbb{P}(A_t' = a, E_a^c(t)|\mathcal{F}_{t-1})$ can be bounded as follows:

$$\begin{aligned}\mathbb{P}(A_t = 1, E_a^c(t)|\mathcal{F}_{t-1}) &\stackrel{(i)}{\geq} \mathbb{P}(A_t' = a, E_a^c(t), E_1(t)|\mathcal{F}_{t-1}) \\ &= \mathbb{P}(A_t' = a, E_a^c(t)|\mathcal{F}_{t-1})\mathbb{P}(E_1(t).|\mathcal{F}_{t-1})\end{aligned} \tag{40}$$

Inequality (i) holds because $\{A_t' = a, E_a^c(t), E_1(t)\} \subseteq \{A_t = 1, E_a^c(t), E_1(t)\}$. The equality is a consequence of the conditional independence of $E_1(t)$ and $\{A_t' = a, E_a^c(t)\}$ (conditioned on $\mathcal{F}_{t-1}$). [6]

[6]The conditional independence property holds for all of our sampling mechanisms because the sample distributions for the two distinct arms $(a, 1)$ are always conditionally independent on $\mathcal{F}_{t-1}$

Assuming $\mathbb{P}(E_1(t)|\mathcal{F}_{t-1}) > 0$ and[7] putting inequalities 39 and 40 together gives the following upper bound for $\mathbb{P}(A_t = a, E_a^c(t)|\mathcal{F}_{t-1})$:

$$\mathbb{P}(A_t = a, E_a^c(t)|\mathcal{F}_{t-1}) \leq \mathbb{P}(A_t = 1, E_a^c(t)|\mathcal{F}_{t-1}) \left(\frac{1 - \mathbb{P}(E_1(t)|\mathcal{F}_{t-1})}{\mathbb{P}(E_1(t)|\mathcal{F}_{t-1})} \right).$$

Letting $P(E_1(t)|\mathcal{F}_{t-1}) := p_{1,T_1(t)}$ and noting that$\{A_t = 1, E_a^c(t)\} \subseteq \{A_t = 1\}$:

$$\mathbb{P}(A_t = a, E_a^c(t)|\mathcal{F}_{t-1}) \leq \mathbb{P}(A_t = 1|\mathcal{F}_{t-1}) \left(\frac{1}{p_{1,T_1(t)}} - 1 \right). \tag{41}$$

Now, we use this to give an upper bound on the term of interest:

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{I}(A_t = a, E_a^c(t))\right] &\overset{(i)}{=} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}\left[\mathbb{I}\left(A_t = a, E_a^c(t)\right)|\mathcal{F}_{t-1}\right]\right] \\
&\overset{(ii)}{=} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{P}\left(A_t = a, E_a^c(t)|\mathcal{F}_{t-1}\right)\right] \\
&\overset{(iii)}{\leq} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{P}(A_t = 1|\mathcal{F}_{t-1}) \left(\frac{1}{p_{1,T_1(t)}} - 1\right)\right] \\
&\overset{(iv)}{=} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}\left[\mathbb{I}(A_t = 1)|\mathcal{F}_{t-1}\right] \left(\frac{1}{p_{1,T_1(t)}} - 1\right)\right] \\
&\overset{(v)}{=} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{I}(A_t = 1) \left(\frac{1}{p_{1,T_1(t)}} - 1\right)\right] \\
&\overset{(vi)}{\leq} \mathbb{E}\left[\sum_{s=1}^{T-1} \frac{1}{p_{1,s}} - 1\right].
\end{aligned}$$

Here the equality (i) is a consequence of the tower property, and equality (ii) by noting that $\mathbb{E}\left[\mathbb{I}\left(A_t = a, E_a^c(t)\right)|\mathcal{F}_{t-1}\right] = \mathbb{P}\left(A_t = a, E_a^c(t)|\mathcal{F}_{t-1}\right)$. Inequality (iii) follows by from Equation 41, and equality (iv) follows by definition. Finally, equality (v) follows by the tower property and the last line each the fact that $T_1(t) = s$ and $A_t = 1$ can only happen once for every $s = 1, ..., T$. This completes the proof. □

Given the bound on (I) from (3), we now present the tighter of two bounds on (II) which is used to provide regret guarantees for Thompson sampling with exact samples from the posteriors.

Lemma 14 (Bounding II - exact posterior)**.** *For a sub-optimal arm $a \in \mathcal{A}$, we have that:*

$$II = \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{I}(A_t = a, E_a(t))\right] \leq 1 + \mathbb{E}\left[\sum_{s=1}^{T} \mathbb{I}\left(p_{a,s} > \frac{1}{T}\right)\right].$$

where $p_{a,s} = \mathbb{P}(r_{a,t}(s) > \bar{r}_1 - \epsilon|\mathcal{F}_{t-1})$, for some $\epsilon > 0$.

Proof. The upper bound for term II in (3) follows the exact same proof as in (Agrawal and Goyal, 2012), and we recreate it for completeness below. Let $\mathcal{T} = \{t : p_{a,T_a(t)} > \frac{1}{T}\}$, then:

[7]In all the cases we consider, including approximate sampling schemes, this property holds. In that case, since the Gaussian noise in the Langevin diffusion ensures all sets of the form (a, b) have nonzero probability mass.

$$\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{I}(A_t=a,E_a(t))\right]\leq\underbrace{\mathbb{E}\left[\sum_{t\in\mathcal{T}}\mathbb{I}(A_t=a)\right]}_{I}+\underbrace{\mathbb{E}\left[\sum_{t\notin\mathcal{T}}\mathbb{I}(E_a(t))\right]}_{II} \tag{42}$$

By definition, term I in (42) satisfies:

$$\sum_{t\in\mathcal{T}}\mathbb{I}(A_t=a)=\sum_{t\in\mathcal{T}}\mathbb{I}\left(A_t=a,p_{a,T_a(t)}>\frac{1}{T}\right)\leq\sum_{s=1}^{T}\mathbb{I}\left(p_{a,s}>\frac{1}{T}\right)$$

To address term II in (42), we note that, by definition: $\mathbb{E}[\mathbb{I}(E_a(t))|\mathcal{F}_{t-1}]=p_{a,T_a(t)}$. Therefore, using the definition of the set of times $\mathcal{T}$, we can construct this simple upper bound:

$$\begin{aligned}\mathbb{E}\left[\sum_{t\notin\mathcal{T}}\mathbb{I}(E_a(t))\right]&=\mathbb{E}\left[\sum_{t\notin\mathcal{T}}\mathbb{E}\left[\mathbb{I}\left(E_a(t)\right)|\mathcal{F}_{t-1}\right]\right]\\&=\mathbb{E}\left[\sum_{t\notin\mathcal{T}}p_{a,t}\right]\\&\leq\sum_{t\notin\mathcal{T}}\frac{1}{T}\\&\leq 1\end{aligned}$$

Using the two upper bounds for terms I and II in (42) gives out desired result:

$$\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{I}(A_t=a,E_a(t))\right]\leq 1+\mathbb{E}\left[\sum_{s=1}^{T}\mathbb{I}\left(p_{a,s}>\frac{1}{T}\right)\right]$$

□

E.1. Regret of Exact Thompson Sampling

We now present two technical lemmas for use in the proof of the regret of exact Thompson sampling. The first technical lemma, provides a lower bound on the probability of an arm begin optimistic in terms of the quality of the prior:

Lemma 15. *Suppose the likelihood and reward distributions satisfy Assumptions 1-3, then for all $n=1,...,T$ and $\gamma_1=\frac{\nu_1 m_1^2}{8d_1L_1^3}$:*

$$\mathbb{E}\left[\frac{1}{p_{1,n}}\right]\leq 64\sqrt{\frac{L_1}{m_1}B_1}$$

Proof. Throughout this proof we drop the dependence on the arm to simplify notation (unless necessary). We first analyze $\|\theta^*-\theta_u\|^2$ where θ_u is the mode of the posterior of arm 1 after having received n samples from the arm which satisfies:

$$\frac{1}{n}\nabla\log\pi_1(\theta_u)+\nabla F_{1,n}(\theta_u)=0$$

Given this definition, and letting $\hat{\theta}=\theta_u-\theta^*$ we have that:

$$\begin{aligned}\hat{\theta}^T\left(\nabla F_n(\theta^*)-\nabla F_n(\theta_u)\right)-\frac{1}{n}\hat{\theta}^T\nabla\log\pi(\theta_u)=\hat{\theta}^T\nabla F_n(\theta^*)\\m\|\hat{\theta}\|^2\leq\frac{m}{2}\|\hat{\theta}\|^2+\frac{1}{2m}\|\nabla F_n(\theta^*)\|^2+\frac{\log B_1}{n}\\\|\hat{\theta}\|^2\leq\frac{1}{m^2}\|\nabla F_n(\theta^*)\|^2+\frac{2\log B_1}{mn}\end{aligned}$$

Noting that $|a^T(\theta^* - \theta_u)| \le \sqrt{A^2 \|\hat{\theta}\|^2}$ we find that:

$$
\begin{aligned}
p_{1,s} &= Pr\left(\alpha^T\left(\theta - \theta_u\right) \ge \alpha^T\left(\theta^* - \theta_u\right) - \epsilon\right) \\
&\ge Pr\left(\alpha^T(\theta - \theta_u) \ge \underbrace{\sqrt{\frac{2A^2 \log B_1}{nm} + \frac{A^2}{m^2}\|\nabla F_n(\theta^*)\|^2}}_{=t}\right),
\end{aligned}
$$

where we note that $\|F_n(\theta^*)\|$ in Proposition 1 is a 1-dimensional $\frac{dL_a}{\sqrt{n\nu}}$ subgaussian random variable.

Now, since we know that the posterior over θ is $\gamma(n+1)L$-smooth and γmn-strongly log concave, with mode θ_u, we know from e.g (Saumard and Wellner, 2014) Theorem 3.8 that the marginal density of $\alpha^T\theta$ is $\frac{\gamma(n+1)L}{A^2}$-smooth and $\frac{\gamma mn}{A^2}$-strongly log-concave.

Thus we have that:

$$
Pr\left(\alpha^T\left(\theta - \theta_u\right) \ge t\right) \ge \sqrt{\frac{nm}{(n+1)L}} Pr(Z \ge t)
$$

where $Z \sim \mathcal{N}\left(0, \frac{A^2}{\gamma(n+1)L}\right)$.

Now using a lower bound on the cumulative density function of a Gaussian random variable, we find that, for $\sigma^2 = \frac{A^2}{\gamma(n+1)L}$:

$$
p_{1,s} \ge \sqrt{\frac{nm}{2\pi(n+1)L}} \begin{cases} \frac{\sigma t}{t^2+\sigma^2} e^{-\frac{t^2}{2\sigma^2}} & : t > \frac{A}{\sqrt{\gamma(n+1)L}} \\ 0.34 & : t \le \frac{A}{\sqrt{\gamma(n+1)L}} \end{cases}
$$

Thus we have that:

$$
\begin{aligned}
\frac{1}{p_{1,s}} &\le \sqrt{\frac{2\pi(n+1)L}{nm}} \begin{cases} \frac{t^2+\sigma^2}{\sigma t} e^{\frac{t^2}{2\sigma^2}} & : t > \frac{A}{\sqrt{\gamma(n+1)L}} \\ \frac{1}{0.34} & : t \le \frac{A}{\sqrt{\gamma(n+1)L}} \end{cases} \\
&\le \sqrt{\frac{2\pi(n+1)L}{nm}} \begin{cases} \left(\frac{t}{\sigma} + 1\right) e^{\frac{t^2}{2\sigma^2}} & : t > \frac{A}{\sqrt{\gamma(n+1)L}} \\ 3 & : t \le \frac{A}{\sqrt{\gamma(n+1)L}} \end{cases}
\end{aligned}
$$

Taking the expectation of both sides with respect to the samples $X_1, ..., X_n$, letting $\kappa = L/m$, and using the fact that $\frac{n+1}{n} \le 2$ for $n \ge 1$ we find that:

$$
\mathbb{E}\left[\frac{1}{p_{1,s}}\right] \le 6\sqrt{\pi\kappa} + 2\sqrt{\pi\kappa}\mathbb{E}\left[\left(\frac{\sqrt{\frac{2A^2 \log B_1}{nm} + \frac{A^2}{m^2}\|\nabla F_n(\theta^*)\|^2}}{\sigma} + 1\right) e^{\frac{t^2}{2\sigma^2}}\right]
$$

Noting that $\sqrt{\frac{2A^2 \log B_1}{nm} + \frac{A^2}{m^2}\|\nabla F_n(\theta^*)^T\|^2} \le A\sqrt{\frac{2 \log B_1}{nm}} + \frac{A}{m}\|\nabla F_n(\theta^*)\|$, and letting $Y = \|\nabla F_n(\theta^*)\|$ to simplify notation, this further simplifies:

$$
\mathbb{E}\left[\frac{1}{p_{1,s}}\right] \le 6\sqrt{\pi\kappa} + 2\sqrt{\pi\kappa}\mathbb{E}\left[\left(\sqrt{4\gamma\kappa \log B_1} + \frac{A}{m\sigma} Y\right) e^{2\gamma\kappa \log B_1 + \frac{(n+1)\gamma L}{2m^2} Y^2}\right]
$$

Via Cauchy-Schwartz we can further develop this upper bound and find that:

$$\mathbb{E}\left[\frac{1}{p_{1,s}}\right] \leq 6\sqrt{\pi\kappa} + 2\sqrt{\pi\kappa}e^{2\gamma\kappa\log B_1}\left(\sqrt{4\gamma\kappa\log B_1}\mathbb{E}\left[e^{\frac{(n+1)\gamma L}{2m^2}Y^2}\right] + \frac{A}{m\sigma}\sqrt{\mathbb{E}\left[Y^2\right]}\sqrt{\mathbb{E}\left[e^{\frac{(n+1)\gamma L}{m^2}Y^2}\right]}\right)$$

Since Y is sub-Gaussian, Y^2 is sub-exponential such that:

$$\mathbb{E}\left[e^{\lambda Y^2}\right] \leq e \quad \text{and} \quad \mathbb{E}\left[Y^2\right] \leq 2\frac{dL^2}{\nu n}$$

for $\lambda < \frac{n\nu}{4dL^2}$. Therefore if :

$$\gamma = \frac{\nu m^2}{8dL^3}$$

Simplifying the bound further gives:

$$\begin{aligned}\mathbb{E}\left[\frac{1}{p_{1,s}}\right] &\leq 6\sqrt{\pi\kappa} + 2\sqrt{\pi\kappa}e^{2\gamma\kappa\log B_1}\left(\sqrt{4\gamma\kappa\log B_1}e + 2\sqrt{\frac{e\gamma(n+1)L}{m^2}\frac{dL^2}{\nu n}}\right)\\ &\leq 6\sqrt{\pi\kappa} + 2\sqrt{\pi\kappa}e^{\frac{\log B_1}{4}}\left(\sqrt{\frac{\log B_1}{2}}e + 2\sqrt{e}\right)\end{aligned}$$

where we have used the fact that $\kappa, d \geq 1$ and the fact that we can assume without loss of generality that $L/\nu \geq 1$. Thus, this bound simplifies to:

$$\begin{aligned}\mathbb{E}\left[\frac{1}{p_{1,s}}\right] &\leq 6\sqrt{\pi\kappa} + 2\sqrt{\pi\kappa}e^{2\gamma\kappa\log B_1}\left(\sqrt{4\gamma\kappa\log B_1}e + 2\sqrt{\frac{e\gamma(n+1)L}{m^2}\frac{dL^2}{\nu n}}\right)\\ &\leq 2\sqrt{\pi\kappa}\left(B_1\right)^{\frac{1}{4}}\left(\sqrt{\frac{\log B_1}{2}}e + 7\right)\\ &\leq 4\sqrt{\pi\kappa}\left(B_1\right)^{\frac{1}{4}}\left(\sqrt{\log B_1} + 4\right)\\ &\leq 64\sqrt{\kappa B_1}\end{aligned}$$

where we used the fact that $x^{1/4}(\sqrt{\log x} + 4) \leq 8\sqrt{x}$ for $x \geq 1$ and $\sqrt{\pi} < 2$ to simplify our bound. □

The last technical lemma upper bounds the two terms defined in Lemma 1.

Lemma 16. *Suppose the likelihood, true reward distributions, and priors satisfy Assumptions 1-3, then for* $\gamma_a = \frac{\nu_a m_a^2}{8d_a L_a^3}$*:*

$$\sum_{s=1}^{T-1}\mathbb{E}\left[\frac{1}{p_{1,s}} - 1\right] \leq 64\sqrt{\frac{L_1}{m_1}B_1}\left[\frac{8eA_1^2}{m\Delta_a^2}(D_1 + 4\sigma_1\log 2)\right] + 1 \tag{43}$$

$$\sum_{s=1}^{T}\mathbb{E}\left[\mathbb{I}\left(p_{a,s} > \frac{1}{T}\right)\right] \leq \frac{8eA_a^2}{m\Delta_a^2}(D_a + 2\sigma_a\log(T)) \tag{44}$$

Where for $a \in \mathcal{A}$*,* D_a *is given by:*

$$D_a = \log B_a + \frac{8d_a^2L_a^3}{m_a^2\nu_a} \qquad \sigma_a = \frac{256 d_a L_a^3}{m_a^2\nu_a} + \frac{8d_aL_a^2}{m_a\nu_a}$$

Proof. We begin by showing that (43) holds. To do so, we first note that, by definition $p_{1,s}$ satisfies:

$$p_{1,s} = \mathbb{P}(r_{1,t}(s) > \bar{r}_1 - \epsilon | \mathcal{F}_{t-1}) \tag{45}$$
$$= 1 - \mathbb{P}(r_{1,t}(s) - \bar{r}_1 < -\epsilon | \mathcal{F}_{t-1}) \tag{46}$$
$$\geq 1 - \mathbb{P}(|r_{1,t}(s) - \bar{r}_1| > \epsilon | \mathcal{F}_{t-1}) \tag{47}$$
$$\geq 1 - \mathbb{P}_{\theta \sim \mu_1^{(s)}} \left(\|\theta - \theta^*\| > \frac{\epsilon}{A_1} \right) \tag{48}$$

where the last inequality follows from the fact that $r_{1,t}(s)$ and $\bar{r}_1$ are A_a-Lipschitz functions of $\theta \sim \mu_1^{(s)}$ and θ^* respectively. We then use the fact that the posterior distribution $\mathbb{P}_{\theta \sim \mu_1^{(s)}}$ satisfies the concentration bound from Theorem 1 for $\delta \in (0, e^{-1/2})$. Therefore, we have that:

$$\mathbb{P}_{\theta \sim \mu_1^{(s)}} \left(\|\theta - \theta^*\| > \frac{\epsilon}{A_1} \right) \leq \exp\left(-\frac{1}{2\sigma_1} \left(\frac{mn\epsilon^2}{2eA_1^2} - D_1 \right) \right), \tag{49}$$

where we use the constant D_1 and σ_1 defined in the proof of Theorem 1 to simplify notation. We remark that this bound is not useful unless:

$$n > \frac{2eA_1^2}{\epsilon^2 m} D_1.$$

Thus, choosing $\epsilon = (\bar{r}_1 - \bar{r}_a)/2 = \Delta_a/2$ and ℓ as:

$$\ell = \left\lceil \frac{8eA_1^2}{m\Delta_a^2} (D_1 + 2\sigma_1 \log 2) \right\rceil.$$

we proceed as follows:

$$\sum_{s=\ell}^{T-1} \mathbb{E}\left[\frac{1}{p_{1,s}} - 1 \right] \leq \sum_{s=0}^{T-1} \frac{1}{1 - \frac{1}{2}\delta(s)} - 1$$
$$\leq \int_{s=1}^{\infty} \frac{1}{1 - \frac{1}{2}\delta(s)} - 1 ds$$

where:

$$\frac{1}{2}\delta(s) = \frac{1}{2} \exp\left(-\frac{1}{2\sigma_1} \left(\frac{m\epsilon^2}{2eA_1^2} s \right) \right) \leq e^{-1/2}, \forall s \geq \ell$$

and the first inequality follows from our choice of ℓ and the second by upper bounding the sum by an integral. To finish, we write $\delta(s) = exp(-c * s)$, and solve the integral to find that :

$$\int_{s=1}^{\infty} \frac{1}{1 - \frac{1}{2}\delta(s)} - 1 ds = \frac{\log 2 - \log(2e^c - 1)}{c} + 1 \leq \frac{\log 2}{c} + 1.$$

plugging in for c gives:

$$\sum_{s=1}^{T-1} \mathbb{E}\left[\frac{1}{p_{1,s}} - 1 \right] \leq \sum_{s=1}^{\ell-1} \mathbb{E}\left[\frac{1}{p_{1,s}} - 1 \right] + \frac{8eA_1^2}{m\Delta_a^2} 2\sigma_1 \log 2 + 1$$
$$\leq 64 \sqrt{\frac{L_1}{m_1}} B_1 \left\lceil \frac{8eA_1^2}{m\Delta_a^2} (D_1 + 4\sigma_1 \log 2) \right\rceil + 1$$

To show that (44) holds, we do a similar derivation as in (48):

$$\begin{aligned}
\sum_{s=1}^{T} \mathbb{E}\left[\mathbb{I}\left(p_{a,s} > \frac{1}{T}\right)\right] &= \sum_{s=1}^{T} \mathbb{E}\left[\mathbb{I}\left(\mathbb{P}(r_{a,t}(s) - \bar{r}_a > \Delta_a - \epsilon|\mathcal{F}_{t-1}) > \frac{1}{T}\right)\right] \\
&= \sum_{s=1}^{T} \mathbb{E}\left[\mathbb{I}\left(\mathbb{P}(r_{a,t}(s) - \bar{r}_a > \frac{\Delta_a}{2}|\mathcal{F}_{t-1}) > \frac{1}{T}\right)\right] \\
&\leq \sum_{s=1}^{T} \mathbb{E}\left[\mathbb{I}\left(\mathbb{P}\left(|r_{a,t}(s) - \bar{r}_a| > \frac{\Delta_a}{2}\middle|\mathcal{F}_{t-1}\right) > \frac{1}{T}\right)\right] \\
&\leq \sum_{s=1}^{T} \mathbb{E}\left[\mathbb{I}\left(\mathbb{P}_{\theta\sim\mu_a^{(s)}[\gamma_a]}\left(\|\theta - \theta^*\| > \frac{\Delta_a}{2A_a}\right) > \frac{1}{T}\right)\right].
\end{aligned}$$

Using the posterior concentration result from Theorem 1 we upper bound the number of pulls $\bar{n}$ of arm a such that for all $n \geq \bar{n}$:

$$\mathbb{P}_{\theta\sim\mu_a^{(n)}[\gamma_a]}\left(\|\theta - \theta^*\| > \frac{\Delta_a}{2A_a}\right) \leq \frac{1}{T}.$$

Since the posterior for arm a after n pulls of arm a has the same form as in (49), and $1/T \leq e^{-0.5}$ we can choose $\bar{n}$ as:

$$\bar{n} = \frac{8eA_a^2}{m\Delta_a^2}(D_a + 2\sigma_a \log(T)).$$

This completes the proof.

□

Given these lemma's the proof of Theorem 2 is straightforwards. For clarity, we restate the theorem below:

Theorem E.1. *When the likelihood and true reward distributions satisfy Assumptions 1-3 and $\gamma_a = \frac{\nu_a m_a^2}{8d_a L_a^3}$ we have that the expected regret after $T > 0$ rounds of Thompson sampling with exact sampling satisfies:*

$$\mathbb{E}[R(T)] \leq \sum_{a>1} \frac{CA_a^2}{m_a\Delta_a}\left(\log B_a + d_a^2\kappa_a^3 + d_a\kappa_a^3\log(T)\right) + \sqrt{\kappa_1 B_1}\frac{CA_1^2}{m_1\Delta_a}\left(1 + \log B_1 + d_1^2\kappa_1^3\right) + \Delta_a$$

Where C is a universal constant independent of problem-dependent parameters.

Proof. We invoke Lemmas 13 and 14, to find that:

$$\mathbb{E}\left[T_a(T)\right] \leq \underbrace{\sum_{s=1}^{T-1} \mathbb{E}\left[\frac{1}{p_{1,s}} - 1\right]}_{(I)} + \underbrace{\sum_{s=1}^{T} \mathbb{E}\left[\mathbb{I}\left(1 - p_{a,s} > \frac{1}{T}\right)\right]}_{(II)} \tag{50}$$

Now, invoking Lemma 16, we use the upper bounds for terms (I) and (II) in the regret decomposition and expanding D_a and D_1 to give that:

$$\begin{aligned}
\mathbb{E}[R(T)] \leq \sum_{a>1} \frac{8eA_a^2}{m_a\Delta_a}\left(\log B_a + 8d_a\kappa_a^3\left(d_a + 66\log(T)\right)\right) \\
+ \sqrt{\kappa_1 B_1}\frac{512eA_a^2}{m_1\Delta_a^2}\left(1 + \log B_1 + 8d_1\kappa_1^3\left(d_1 + 132\log(2)\right)\right) + \Delta_a \\
\leq \sum_{a>1} \frac{CA_a^2}{m_a\Delta_a}\left(\log B_a + d_a^2\kappa_a^3 + d_a\kappa_a^3\log(T)\right) \\
+ \sqrt{\kappa_1 B_1}\frac{CA_1^2}{m_1\Delta_a}\left(1 + \log B_1 + d_1^2\kappa_1^3\right) + \Delta_a
\end{aligned}$$

□

E.2. Regret of Approximate Sampling

For the proof of Theorem 4, we proceed similarly as for the proof of Theorem 2, but require another intermediate lemma to deal with the fact that the samples from the arms are no longer conditionally independent given the filtration (due to the fact that we use the last sample as the initialization of the filtration). To do so, we first define the event:

$$Z_a(T) = \cap_{t=1}^{T-1} Z_{a,t},$$

where:

$$Z_{a,t} = \left\{\|\theta_{a,t} - \theta_a^*\| < \sqrt{\frac{18e}{nm_a}}\left(d_a + \log B_a + 2\left(16 + \frac{4dL_a^2}{\nu_a m_a}\right)\log 1/\delta_1\right)^{\frac{1}{2}}\right\},$$

Lemma 17. *Suppose the likelihood and reward distributions satisfy Assumptions 1-4, Then the regret of a Thompson sampling algorithm with approximate sampling can be decomposed as:*

$$\mathbb{E}[R(T)] \leq \sum_{a>1} \Delta_a \mathbb{E}\left[T_a(T) \,\middle|\, Z_a(T) \cap Z_1(T)\right] + 2\Delta_a \tag{51}$$

Proof. We begin by conditioning on the event $Z_a(T) \cap Z_1(T)$ for each $a \in \mathcal{A}$, where we note that by construction $p_Z = \mathbb{P}((Z_a(T)^c \cup Z_1(T)^c)) \leq \mathbb{P}(Z_1(T)^c) + \mathbb{P}(Z_a(T)^c) = 2T\delta_1)$ (since via Lemma 3, the probability of each event in $Z_a(T)^c$ and $Z_1(T)^c$ is less than δ_1).

Therefore, we must have that:

$$\begin{aligned}
\mathbb{E}[T_a(T)] &\leq \mathbb{E}\left[T_a(T) \,\middle|\, Z_a(T) \cap Z_1(T)\right] + \mathbb{E}\left[T_a(T) \,\middle|\, (Z_a(T)^c \cup Z_1(T)^c)\right] p_Z \\
&\leq \mathbb{E}\left[T_a(T) \,\middle|\, Z_a(T) \cap Z_1(T)\right] + 2T\delta_3\mathbb{E}\left[T_a(T) \,\middle|\, (Z_a(T)^c \cup Z_1(T)^c)\right] \\
&\leq \mathbb{E}\left[T_a(T) \,\middle|\, Z_a(T) \cap Z_1(T)\right] + 2\delta_3 T^2,
\end{aligned}$$

where in the first line we use the fact that $1 - p_Z \leq 1$ and in the last line we used the fact that $T_a(T)$ is trivially less than T. Choosing $\delta_1 = 1/T^2 \leq e^{-1/2}$ completes the proof. □

With this decomposition in hand, we can now proceed as in Lemma 15 to provide anti-concentration guarantees for the approximate posteriors.

Lemma 18. *Suppose the likelihood and true reward distributions satisfy Assumptions 1-4: then if* $\gamma_1 = \frac{\nu m^2}{32(16L\nu m+4dL^3)}$, *for all* $n = 1, ..., T$ *all samples from the the (stochastic gradient) ULA method with the hyperparameters and runtime as described in Theorem 3 satisfy:*

$$\mathbb{E}\left[\frac{1}{p_{1,n}}\right] \leq 27\sqrt{B_1}$$

Proof. We begin by using the last step of our Langevin Dynamics and show that it exhibits the desired anti-concentration properties. In particular, we know that $\theta_{1,t} \sim \mathcal{N}(\theta_{1,Nh}, \frac{1}{\gamma}I)$, such that:

$$\begin{aligned} p_{1,s} &= Pr\left(\alpha^T\left(\theta - \theta_{1,Nh}\right) \geq \alpha^T\left(\theta^* - \theta_{1,Nh}\right) - \epsilon\right) \\ &\geq Pr\left(Z \geq \underbrace{A\|\theta_{1,Nh} - \theta_*\|}_{:=t}\right) \end{aligned}$$

where $Z \sim \mathcal{N}(0, \frac{A^2}{nL\gamma}I)$ by construction.

Now using a lower bound on the cumulative density function of a Gaussian random variable, we find that, for $\sigma^2 = \frac{A^2}{nL\gamma}$:

$$p_{1,s} \geq \sqrt{\frac{1}{2\pi}}\begin{cases} \frac{\sigma t}{t^2+\sigma^2}e^{-\frac{t^2}{2\sigma^2}} & : t > \frac{A}{\sqrt{nL\gamma}} \\ 0.34 & : t \leq \frac{A}{\sqrt{nL\gamma}} \end{cases}$$

Thus we have that:

$$\frac{1}{p_{1,s}} \leq \sqrt{2\pi}\begin{cases} \left(\frac{t}{\sigma}+1\right)e^{\frac{t^2}{2\sigma^2}} & : t > \frac{A}{\sqrt{nL\gamma}} \\ 3 & : t \leq \frac{A}{\sqrt{nL\gamma}} \end{cases}$$

Taking the expectation of both sides with respect to the samples $X_1, ..., X_n$, we find that:

$$\begin{aligned} \mathbb{E}\left[\frac{1}{p_{1,s}}\right] &\leq 3\sqrt{2\pi} + \sqrt{2\pi}\mathbb{E}\left[\left(\sqrt{nL\gamma}\|\theta_{1,Nh} - \theta_*\| + 1\right)e^{nL\gamma\|\theta_{1,Nh}-\theta_*\|^2}\right] \\ &\leq 3\sqrt{2\pi} + \sqrt{2\pi nL\gamma}\sqrt{\mathbb{E}\left[\|\theta_{1,Nh} - \theta_*\|^2\right]}\sqrt{\mathbb{E}\left[e^{nL\gamma\|\theta_{1,Nh}-\theta_*\|^2}\right]} + \sqrt{2\pi}\mathbb{E}\left[e^{\frac{nL\gamma}{2}\|\theta_{1,Nh}-\theta_*\|^2}\right] \end{aligned}$$

Now, we remark that, from Theorems 5 and 6, we have that for both approximate sampling schemes:

$$\mathbb{E}\left[\|\theta_{1,Nh} - \theta^*\|^2\right] \leq \frac{18}{mn}\left(d + \log B + 32 + \frac{8dL^2}{\nu m}\right)$$

Further, we note that $\|\theta_{1,Nh} - \theta^*\|^2$ is a sub-exponential random variable. To see this, we analyze its moment generating function:

$$\mathbb{E}[e^{nL\gamma\|\theta_{1,Nh}-\theta^*\|^2}] = 1 + \sum_{i=1}^{\infty}\mathbb{E}\left[\frac{(nL\gamma)^i\|\theta_{1,Nh} - \theta_*\|^{2i}}{i!}\right]$$

Borrowing the notation from the proof of Theorem 1, we know that

$$\mathbb{E}\left[\|\theta_{1,Nh}-\theta_*\|^{2p}\right] \leq 3\left(\frac{2D}{mn}+\frac{4\sigma p}{mn}\right)^p$$

where:

$$D = d + \log B \quad \text{and} \quad \sigma = 16 + \frac{4dL^2}{\nu m}$$

Plugging this in above gives:

$$\begin{aligned}
\mathbb{E}[e^{\gamma\|\theta_{1,Nh}-\theta_*\|^2}] &\leq 1 + 3\sum_{i=1}^{\infty}\frac{\left(\frac{2nL\gamma D + 4nL\gamma\sigma i}{mn}\right)^i}{i!} \\
&\leq 1 + \frac{3}{2}\sum_{i=1}^{\infty}\frac{1}{i!}\left(\frac{4nL\gamma D}{mn}\right)^i + \frac{3}{2}\sum_{i=1}^{\infty}\frac{1}{i!}\left(\frac{8nL\gamma\sigma i}{nm}\right)^i \\
&\leq \frac{3}{2}e^{\frac{4nL\gamma D}{mn}} + \frac{3}{2}\sum_{i=1}^{\infty}\left(\frac{8nL\gamma e\sigma}{nm}\right)^i
\end{aligned}$$

where, we have use the identities $(x+y)^i \leq 2^{i-1}(x^i+y^i)$ for $i \geq 1$, and $i! \geq (i/e)^i$ to simplify the bound.
If $\gamma \leq \frac{m}{32L\sigma}$, then we have that:

$$\mathbb{E}[e^{nL\gamma\|\theta_{1,Nh}-\theta_*\|^2}] \leq \frac{3}{2}\left(e^{\frac{4nL\gamma D}{m}} + 2.5\right)$$

which, together with the upper bound on γ gives:

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{p_{1,s}}\right] &\leq 3\sqrt{2\pi} + \frac{3}{2}\sqrt{\frac{16\pi nL\gamma}{m}(D+2\sigma)}\left(e^{\frac{2nL\gamma D}{m}} + 2\right) + \frac{3}{2}\sqrt{2\pi}\left(e^{\frac{4nL\gamma D}{m}} + 7.5\right) \\
&\leq 3\sqrt{2\pi} + \frac{3}{2}\left(\sqrt{\frac{\pi(d+\log B)}{2\sigma}} + \sqrt{\pi}\right)\left(e^{\frac{d+\log B}{16\sigma}} + 2\right) + \frac{3}{2}\sqrt{2\pi}\left(e^{\frac{d+\log B}{8\sigma}} + 2.5\right)
\end{aligned}$$

where we used the sub-additivity of $\sqrt{x}$, the fact that $\sqrt{\frac{3}{2}} < \frac{3}{2}$, $sqrt2.5 < 2$ and substituted in the values for σ and D to simplify the boung. Finally since$\frac{L^2}{m\nu} > 1$, we find that $\sigma > max(4d, 1)$, allowing us to simplify the bound further to:

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{p_{1,s}}\right] &\leq 3\sqrt{2\pi} + \frac{3}{2}\sqrt{\frac{\pi}{8} + \frac{\log B}{2}}\left(2B^{1/16} + 2\right) + \frac{3}{2}\sqrt{2\pi}\left(2B^{1/8} + 2.5\right) \\
&\leq 18 + \frac{3}{\sqrt{2}}\underbrace{\left(B^{1/16} + B^{1/16}\sqrt{\log B} + \log B + 2B^{1/8}\right)}_{I} \\
&\leq 18 + 12/\sqrt{2}\sqrt{B} \leq 27\sqrt{B}
\end{aligned}$$

where to simplify the bound we used the fact that $\sqrt{\pi} < 2$ and $I \leq 4\sqrt{B}$ and that $18 + 12/\sqrt{2}x \leq 27x$ for $x \geq 1$.

□

With this lemma in hand, we can now proceed as in Lemma 16 to finalize the proof of Theorem 4.

Lemma 19. *Suppose the likelihood, true reward distributions, and priors satisfy Assumptions 1-4, the samples are generated from the sampling schemes described in Theorem 6 and Theorem 5, and $\gamma_a = \frac{m_a}{32L_a\sigma_a}$ then:*

$$\sum_{s=1}^{T-1} \mathbb{E}\left[\frac{1}{\widehat{p}_{1,s}} - 1 \middle| Z_1(T)\right] \leq 27\sqrt{B_1}\left\lceil \frac{144eA_1^2}{m\Delta_a^2}(d_1 + \log B_1 + 4\sigma_1 \log T + 12d_1\sigma_1 \log 2)\right\rceil + 1 \tag{52}$$

$$\sum_{s=1}^{T} \mathbb{E}\left[\mathbb{I}\left(\widehat{p}_{a,s} > \frac{1}{T}\right) \middle| Z_a(T)\right] \leq \frac{144eA_a^2}{m\Delta_a^2}(d_a + \log B_a + 10d_a\sigma_a \log(T)), \tag{53}$$

where $\widehat{p}_{a,s}$ is the distribution of a sample from the approximate posterior $\widehat{\mu}_a$ after s samples have been collected, and for $a \in \mathcal{A}$, σ_a is given by:

$$\sigma_a = 16 + \frac{4d_aL_a^2}{m_a\nu_a}.$$

Proof. We begin by showing that (52) holds. To do so, we proceed identically as in the proof of Lemma 16 to note that, by definition $\widehat{p}_{1,s}$ satisfies:

$$\begin{aligned}
\widehat{p}_{1,s} &= \mathbb{P}(r_{1,t}(s) > \bar{r}_1 - \epsilon|\mathcal{F}_{t-1}) && (54)\\
&= 1 - \mathbb{P}(r_{1,t}(s) - \bar{r}_1 < -\epsilon|\mathcal{F}_{t-1}) && (55)\\
&\geq 1 - \mathbb{P}(|r_{1,t}(s) - \bar{r}_1| > \epsilon|\mathcal{F}_{t-1}) && (56)\\
&\geq 1 - \mathbb{P}_{\theta\sim\widehat{\mu}_1^{(s)}}\left(\|\theta - \theta^*\| > \frac{\epsilon}{A_1}\right), && (57)
\end{aligned}$$

where the last inequality follows from the fact that $r_{1,t}(s)$ and $\bar{r}_1$ are A_a-Lipschitz functions of $\theta \sim \mu_1^{(s)}$ and θ^* respectively. We then use the fact that conditioned on $Z_1(T)$, the approximate posterior distribution $\mathbb{P}_{\theta\sim\widehat{\mu}_1^{(s)}}$ satisfies the identical concentration bounds from Lemmas 12 and Lemma 11. Substituting in the assumed value of γ_1, and simplifying, we have that the distribution of the samples conditioned on $Z_1(T)$ satisfy:

$$\mathbb{P}_{\theta_{1,t}\sim\bar{\mu}_1^{(s)}[\gamma_1]}\left(\|\theta_{1,t} - \theta_1^*\|_2 > \sqrt{\frac{36e}{m_1 n}\left(d_1 + \log B_1 + 4\sigma_1 \log T + 6d_1\sigma_1 \log 1/\delta_2\right)} \middle| Z_{n-1}\right) < \delta_2.,$$

Equivalently, we have that:

$$\mathbb{P}_{\theta\sim\bar{\mu}_1^{(s)}}[\gamma_1]\left(\|\theta - \theta^*\| > \frac{\epsilon}{A_1}\right) \leq exp\left(-\frac{1}{6d_1\sigma_1}\left(\frac{m_1 n\epsilon^2}{36eA_1^2} - \bar{D}_1\right)\right), \tag{58}$$

where we define $\bar{D}_1 = d_1 + \log B_1 + 4\sigma \log T$, to simplify notation. We remark that this bound is not useful unless:

$$n > \frac{16eA_1^2}{\epsilon^2 m_1}\bar{D}_1.$$

Thus, choosing $\epsilon = (\bar{r}_1 - \bar{r}_a)/2 = \Delta_a/2$, we can choose ℓ as:

$$\ell = \left\lceil \frac{144eA_1^2}{m\Delta_a^2}(\bar{D}_1 + 6d_1\sigma_1 \log 2)\right\rceil.$$

With this choice of ℓ, we proceed exactly as in the proof of Lemma 16 to find that :

$$\begin{aligned}
\sum_{s=1}^{T-1} \mathbb{E}\left[\frac{1}{\widehat{p}_{1,s}} - 1 \middle| Z_1(T)\right] &\leq 27\sqrt{B_1}\ell + \sum_{s=\ell}^{T-1} \mathbb{E}\left[\frac{1}{p_{1,s}} - 1 \middle| Z_1(T)\right]\\
&\leq 27\sqrt{B_1}\left\lceil \frac{144eA_1^2}{m\Delta_a^2}(\bar{D}_1 + 12d_1\sigma_1 \log 2)\right\rceil + 1,
\end{aligned}$$

where we used the upper bound from Lemma 18 to bound the first ℓ terms in the first inequality.

To show that (53) holds, we use a similar derivation as in (57):

$$\sum_{s=1}^{T}\mathbb{E}\left[\mathbb{I}\left(p_{a,s}>\frac{1}{T}\right)\Big|Z_a(T)\right]\leq\sum_{s=1}^{T}\mathbb{E}\left[\mathbb{I}\left(\mathbb{P}_{\theta\sim\bar{\mu}_a^{(s)}[\gamma_a]}\left(\|\theta-\theta^*\|>\frac{\Delta_a}{2A_a}\right)>\frac{1}{T}\right)\Big|Z_a(T)\right]$$

Since on the event $Z_a(T)$, the posterior concentration result from Lemmas 12 and Lemma 11 holds, it remains to upper bound the number of pulls $\bar{n}$ of arm a such that for all $n\geq\bar{n}$:

$$\mathbb{P}_{\theta\sim\bar{\mu}_a^{(n)}[\gamma_a]}\left(\|\theta-\theta^*\|>\frac{\Delta_a}{2A_a}\right)\leq\frac{1}{T}.$$

Since the posterior for arm a after n pulls of arm a has the same form as in (49), we can choose $\bar{n}$ as:

$$\bar{n}=\frac{144eA_a^2}{m\Delta_a^2}(\bar{D}_a+6d_a\sigma_a\log(T)).$$

Using the fact that $d_a >\geq 1$ to simplify the bound completes the proof. □

Putting the results of Lemmas 17 and 19 together gives us our final theorem:

Theorem E.2 (Regret of Thompson sampling with (stochastic gradient) Langevin algorithm)**.** *When the likelihood and true reward distributions satisfy Assumptions 1-4: we have that the expected regret after $T>0$ rounds of Thompson sampling with the (stochastic gradient) ULA method with the hyper-parameters and runtime as described in Lemmas 11 (and 12 respectively), and $\gamma_a=\frac{\nu_a m_a^2}{32(16L_a\nu_a m_a+4d_aL_a^3)}=O\left(\frac{1}{d_a\kappa_a^3}\right)$ satisfies:*

$$\begin{aligned}\mathbb{E}[R(T)]\leq\sum_{a>1}\frac{CA_a^2}{m_a\Delta_a}&\left(d_a+\log B_a+d_a^2\kappa_a^2\log T\right)\\&+\frac{C\sqrt{B_1}A_1^2}{m_1\Delta_a}\left(1+\log B_1+d_1\kappa_1^2\log T+d_1^2\kappa_1^2\right)+3\Delta_a.\end{aligned}$$

where C is a universal constant that is independent of problem dependent parameters and $\kappa_a=L_a/m_a$.

Proof. To begin, we invoke Lemma 17, which shows that we only need to bound the number of times a suboptimal arm $a\in\mathcal{A}$ is chosen on the 'nice' event $Z_1(T)\cap Z_a(T)$ where the gradient of the log likelihood has concentrated and the approximate samples have been in high probability regions of the posteriors. We then invoke Lemmas 13 and 14, to find that:

$$\mathbb{E}\left[T_a(T)\Big|Z_1(T)\cap Z_a(T)\right]\leq 1+\ell \tag{59}$$

$$+\underbrace{\sum_{s=\ell}^{T-1}\mathbb{E}\left[\frac{1}{p_{1,s}}-1\Big|Z_1(T)\right]}_{(I)}+\underbrace{\sum_{s=1}^{T}\mathbb{E}\left[\mathbb{I}\left(1-p_{a,s}>\frac{1}{T}\right)\Big|Z_a(T)\right]}_{(II)} \tag{60}$$

Now, invoking Lemma 16, we use the upper bounds for terms (I) and (II) in the regret decomposition, use our choice of both δ_1 and $\delta_3=1/T^2$, expanding D_a and D_1, and use the fact that $\lceil x\rceil\leq x+1$ to give that:

$$\mathbb{E}[R(T)] \leq \sum_{a>1} \frac{144eA_a^2}{m_a\Delta_a}\left(d_a + \log B_a + 10d_a\left(16 + \frac{4d_aL_a^2}{\nu_a m_a}\right)\log(T)\right)$$
$$+ 27\sqrt{B_1}\frac{144eA_1^2}{m_1\Delta_a}\left(1 + d_1 + \log B_1 + 4\left(16 + \frac{4d_1L_a^2}{\nu_1 m_1}\right)(\log T + 3d_1\log 2)\right) + 3\Delta_a.$$
$$\leq \sum_{a>1}\frac{CA_a^2}{m_a\Delta_a}\left(d_a + \log B_a + d_a^2\kappa_a^2\log T\right)$$
$$+ \frac{C\sqrt{B_1}A_1^2}{m_1\Delta_a}\left(1 + \log B_1 + d_1\kappa_1^2\log T + d_1^2\kappa_1^2\right) + 3\Delta_a.$$

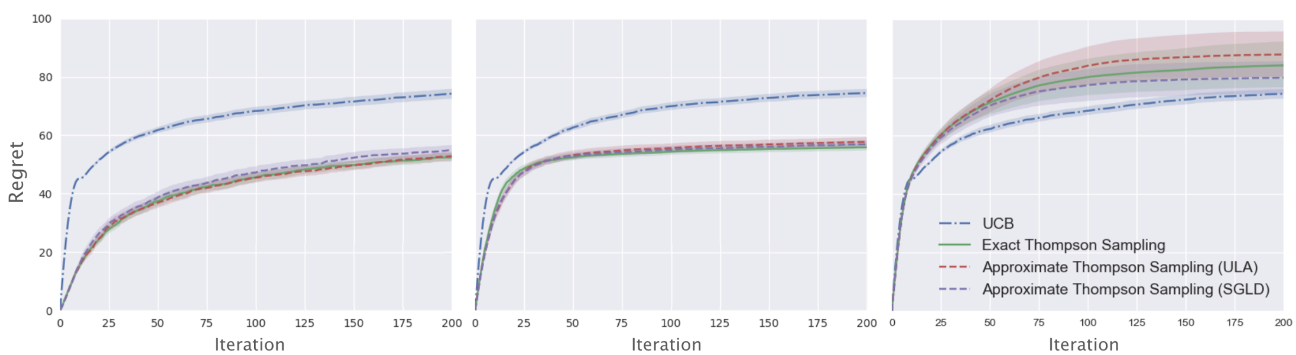
Using the fact that $\kappa_a \geq 1$ and that $d_1 \geq 1$ allows us to simplify to get our desired result.

□

F. Numerical Experiments

We empirically corroborate our theoretical results with numerical experiments of approximate Thompson sampling in log-concave multi-armed bandit instances. We benchmark against both UCB and exact Thompson sampling across three different multi-armed bandit instances, where in the first instance, the priors reflect correct ordering of the mean rewards for all arms; in the second instance, the priors are agnostic of the ordering; in the third instance, the priors reflects the complete opposite ordering. See Appendix G for details of the experimental settings.

As suggested in our theoretical analysis in Section 4, we use a constant number of steps for both ULA and SGLD (with constant number of data points in the stochastic gradient evaluation) to generate samples from the approximate posteriors. The regret of the three algorithms averaged across 100 runs is displayed in Figure 1, where we see approximate Thompson sampling with samples generated by ULA and SGLD perform competitively against both exact Thompson sampling and UCB across all three instances.



a. Good Priors b. Uniform Priors c. Adversarial Priors

Figure 1. Performance of exact and approximate Thompson sampling vs UCB on Gaussian bandits with (a) "good priors" (priors reflecting the correct ordering of the arms' means), (b) the same priors on all the arms' means, and (c) "bad priors" (priors reflecting the exact opposite ordering of the arms' means). The shaded regions represent the 95% confidence interval around the mean regret across 100 runs of the algorithm.

We observe significant performance gains from the (approximate) Thompson sampling approach over the deterministic UCB algorithm when the priors are suggestive or even non-informative of the appealing arms. When the priors are adversarial to the algorithm, the UCB algorithm outperforms the Thompson sampling approach as expected. (This case corresponds to the constant B_a in the Theorems 2 and 4 being large). Also as the theory predicts, we observe little difference between the exact and the approximate Thompson sampling methods in terms of the regret. If we zoom in and scrutinize further, we can see that SGLD slightly outperforms the exact Thompson sampling method in the adversarial prior case. This might be due to the added stochasticity from the approximate sampling techniques, which improves the robustness against bad priors.

G. Details in the Numerical Experiments

We benchmark the effectiveness of approximate Thompson sampling against both UCB and exact Thompson sampling across three different Gaussian multi-armed bandit instances with 10 arms. We remark that the use of Gaussian bandit instances is due to the fact that the closed form for the posteriors allows for us to properly benchmark against exact Thompson sampling and UCB, though our theory applies to a broader family of prior/likelihood pairs.

In all three instances we keep the reward distributions for each arm fixed such that their means are evenly spaced from 0 to 10 ($\bar{r}_1 = 1$, $\bar{r}_2 = 2$, and so on), and their variances are all 1. In each instance we use different priors over the means of the arms to analyze whether the approximate Thompson sampling algorithms preserve the performance of exact Thompson sampling. In the first instance, the priors reflect the correct orderings of the means. We use Gaussian priors with variance 4, and means evenly spaced between 5 and 10 such that $\mathbb{E}_{\pi_1}[X] = 5$, and $\mathbb{E}_{\pi_{10}}[X] = 10$. In the second instance, the prior for each arm is a Gaussian with mean 7.5 and variance 4. Finally, the third instance is 'adversarial' in the sense that the priors reflects the complete opposite ordering of the means. In particular, the priors are still Gaussians such that their means are evenly spaced between 5 and 10 with variance 4, but this time $\mathbb{E}_{\pi_1}[X] = 10$, and $\mathbb{E}_{\pi_{10}}[X] = 5$.

As suggested in our theoretical analysis in Section 4, we use a constant number of steps for both ULA and SGLD to generate samples from the approximate posteriors. In particular, for ULA, we take $N = 100$ and double that number for SGLD $N = 200$. We also choose the stepsize for both algorithms to be $\frac{1}{32T_a(t)}$. For SGLD, we use a batch size of $\min(T_a(t), 32)$. Further, since $d_a = \kappa_a = 1$ since this is a Gaussian family, we take the scaling to be $\gamma_a = 1$. The regret is calculated as $\sum_{t=1}^{T} \bar{r}_{10} - \bar{r}_{A_t}$ for the three algorithms and is averaged across 100 runs. Finally, for the implementation of UCB, we used the time-horizon tuned UCB (Lattimore and Szepesvári, 2020) and the known variance, σ^2 of the arms in the upper confidence bounds (to maintain a level playing field between algorithms):

$$UCB_a(t) = \frac{1}{T_a(t)} \sum_{i=1}^{t-1} X_{A_i} \mathbb{I}\{A_i = a\} + \sqrt{\frac{4\sigma^2 \log 2T}{T_a(t)}}.$$