A Proof of Lemma 1

To simplify the notation throughout this proof, for each $j \in \{1, \ldots, k\}$ denote $\phi^j = \phi_{\psi^j}$. We have

$$E_{\lambda,\tau}^{\boldsymbol{\nu},w}(\boldsymbol{\psi}^*) - E_{\lambda,\tau}^{\boldsymbol{\nu},w}(\boldsymbol{\psi}) = \sum_{j=1}^{k} w_j \mathbf{E}_{Y\sim\nu^j}\left[(\psi^*)^j(Y) - \psi^j(Y)\right] - \tau \log \frac{Z_{\boldsymbol{\psi}^*}}{Z_{\boldsymbol{\psi}}}. \tag{14}$$

Observe that for any $x \in \mathcal{X}$ it holds that

$$\frac{d\mu_{\boldsymbol{\psi}}}{d\mu_{\boldsymbol{\psi}^*}}(x) = \frac{Z_{\boldsymbol{\psi}^*}}{Z_{\boldsymbol{\psi}}} \exp\left(-\frac{\sum_{j=1}^{k} w_j(\phi^j(x) - (\phi^*)^j(x))}{\tau}\right).$$

Hence,

$$\begin{aligned}
\tau \log \frac{Z_{\boldsymbol{\psi}^*}}{Z_{\boldsymbol{\psi}}} &= \tau \log \mathbf{E}_{X\sim\mu_{\boldsymbol{\psi}^*}}\left[\frac{Z_{\boldsymbol{\psi}^*}}{Z_{\boldsymbol{\psi}}}\right] \\
&= \tau \log \mathbf{E}_{X\sim\mu_{\boldsymbol{\psi}^*}}\left[\frac{d\mu_{\boldsymbol{\psi}}}{d\mu_{\boldsymbol{\psi}^*}}(x) \exp\left(\frac{\sum_{j=1}^{k} w_j(\phi^j(x) - (\phi^*)^j(x))}{\tau}\right)\right] \\
&= \tau \log \mathbf{E}_{X\sim\mu_{\boldsymbol{\psi}}}\left[\exp\left(\frac{\sum_{j=1}^{k} w_j(\phi^j(x) - (\phi^*)^j(x))}{\tau}\right)\right] \\
&= \sup_{\mu \ll \mu_{\boldsymbol{\psi}}} \left\{\mathbf{E}_{X\sim\mu}\left[\sum_{j=1}^{k} w_j(\phi^j(x) - (\phi^*)^j(x))\right] - \tau \mathrm{KL}(\mu, \mu_{\boldsymbol{\psi}})\right\},
\end{aligned} \tag{15}$$

where in the final expression we have applied the Donsker-Varadhan variational principle (i.e., convex-conjugate duality between KL-divergence and cumulant generating functions); therein, the supremum runs over probability measures μ absolutely continuous with respect to $\mu_{\boldsymbol{\psi}}$, and it is attained by μ defined as

$$\begin{aligned}
\mu(dx) &\propto \exp\left(\frac{1}{\tau}\sum_{j=1}^{k} w_j(\phi^j(x) - (\phi^*)^j(x))\right) \mu_{\boldsymbol{\psi}}(dx) \\
&\propto \exp\left(\frac{1}{\tau}\sum_{j=1}^{k} w_j(\phi^j(x) - (\phi^*)^j(x))\right) \exp\left(-\frac{1}{\tau}\sum_{j=1}^{k} w_j \phi^j(x)\right) \pi_{\mathrm{ref}}(dx) \\
&\propto \exp\left(-\frac{1}{\tau}\sum_{j=1}^{k} w_j (\phi^*)^j(x)\right) \pi_{\mathrm{ref}}(dx) = \pi_{\boldsymbol{\psi}^*}(dx).
\end{aligned}$$

That is, the supremum in (15) is attained by $\mu = \mu_{\boldsymbol{\psi}^*}$. Hence, the identity (14) becomes

$$\begin{aligned}
&E_{\lambda,\tau}^{\boldsymbol{\nu},w}(\boldsymbol{\psi}^*) - E_{\lambda,\tau}^{\boldsymbol{\nu},w}(\boldsymbol{\psi}) \\
&= \sum_{j=1}^{k} w_j \mathbf{E}_{Y\sim\nu^j}\left[(\psi^*)^j(Y) - \psi^j(Y)\right] - \mathbf{E}_{X\sim\mu_{\boldsymbol{\psi}^*}}\left[\sum_{j=1}^{k} w_j(\phi^j(X) - (\phi^*)^j(X))\right] \\
&\qquad + \tau \mathrm{KL}(\mu_{\boldsymbol{\psi}^*}, \mu_{\boldsymbol{\psi}}) \\
&= \sum_{j=1}^{k} w_j \left(\mathbf{E}_{Y\sim\nu^j}\left[(\psi^*)^j(Y) - \psi^j(Y)\right] + \mathbf{E}_{X\sim\mu_{\boldsymbol{\psi}^*}}\left[(\phi^*)^j(X)) - \phi^j(X)\right]\right) + \tau \mathrm{KL}(\mu_{\boldsymbol{\psi}^*}, \mu_{\boldsymbol{\psi}}) \\
&\geq \tau \mathrm{KL}(\mu_{\boldsymbol{\psi}^*}, \mu_{\boldsymbol{\psi}}),
\end{aligned}$$

where the final inequality follows by noting that for each j the optimality of the pair $((\phi^*)^j, (\psi^*)^j)$ for the entropic optimal transport dual objective $E_{\lambda}^{\mu_{\boldsymbol{\psi}^*},\nu^j}$ implies that

$$\begin{aligned}
&\mathbf{E}_{Y\sim\nu^j}\left[(\psi^*)^j(Y) - \psi^j(Y)\right] + \mathbf{E}_{X\sim\mu_{\boldsymbol{\psi}^*}}\left[(\phi^*)^j(X) - \phi^j(X)\right] \\
&= E_{\lambda}^{\mu,\nu^j}((\phi^*)^j, (\psi^*)^j) - E_{\lambda}^{\mu,\nu^j}(\phi^j, \psi^j) \geq 0.
\end{aligned}$$

The proof of Lemma 1 is complete. $\square$

B Proof of Proposition 1

Recall that for any non-negative integer t we have

$$\mu_t(dx) = Z_t^{-1} \exp\left(-\frac{\sum_{j=1}^k w_j \phi_t^j(x)}{\tau}\right) \pi_{\text{ref}}(dx).$$

where Z_t is the normalizing constant defined by

$$Z_t = \int_{\mathcal{X}} \exp\left(-\frac{\sum_{j=1}^k w_j \phi_t^j(x)}{\tau}\right) \pi_{\text{ref}}(dx).$$

With the notation introduced above, we have

$$E(\boldsymbol{\psi}_t) = \sum_{j=1}^k w_j \mathbf{E}_{Y\sim\nu^j}\left[\psi_t^j(Y)\right] - \tau \log Z_t.$$

Hence,

$$\begin{aligned}
E(\boldsymbol{\psi}_{t+1}) - E(\boldsymbol{\psi}_t) &= \sum_{j=1}^k w_j \mathbf{E}_{Y\sim\nu^j}\left[\psi_{t+1}^j(Y) - \psi_t^j(Y)\right] - \tau \log \frac{Z_{t+1}}{Z_t}. \\
&= \eta\lambda \sum_{j=1}^k w_j \mathbf{E}_{Y\sim\nu^j}\left[\log \frac{d\nu^j}{d\nu_t^j}(Y)\right] - \tau \log \frac{Z_{t+1}}{Z_t}. \\
&= \min(\lambda, \tau) \sum_{j=1}^k w_j \mathrm{KL}(\nu^j, \nu_t^j) - \tau \log \frac{Z_{t+1}}{Z_t}.
\end{aligned}$$

Therefore, to prove Proposition 1 it suffices to show that the inequality

$$\log \frac{Z_{t+1}}{Z_t} \le 0 \tag{16}$$

holds for any $t \ge 0$. We will complete the proof of Proposition 1 using the following lemma, the proof of which is deferred to the end of this section.

Lemma 3. *Let $(\boldsymbol{\psi}_t)_{t\ge 0}$ be any sequence of the form*

$$\psi_{t+1}^j = \psi_t^j + \eta\lambda \log(\Delta_t^j),$$

where for $j \in \{1, \ldots, k\}$, $(\Delta_t^j)_{t\ge 0}$ is an arbitrary sequence of strictly positive functions and $\eta = \min(1, \tau/\lambda)$. Then, for any $t \ge 0$ it holds that

$$\tau \log \frac{Z_{\boldsymbol{\psi}_{t+1}}}{Z_{\boldsymbol{\psi}_t}} \le \min(\lambda, \tau) \log \sum_{j=1}^k w_j \mathbf{E}_{Y\sim\nu^j_{\boldsymbol{\psi}_t}}\left[\Delta_t^j(Y)\right].$$

To complete the proof of Proposition 1, we will apply the above lemma with $\Delta_t^j = \log \frac{d\nu^j}{d\nu_t^j}$. Indeed, we have

$$\begin{aligned}
\tau \log \frac{Z_{t+1}}{Z_t} &\le \min(\lambda, \tau) \log \sum_{j=1}^k w_j \mathbf{E}_{Y\sim\nu_t^j}\left[\frac{d\nu^j}{d\nu_t^j}(Y)\right] \\
&= \min(\lambda, \tau) \log \sum_{j=1}^k w_j \mathbf{E}_{Y\sim\nu^j}\left[1\right] \\
&= 0.
\end{aligned}$$

By (16), the proof of Proposition 1 is complete. □

B.1 Proof of Lemma 3

We will break down the proof with the help of the following lemma, the proof of which can be found in Section B.2.

Lemma 4. *For any sequence* $(\boldsymbol{\psi}_t)_{t\geq 0}$ *and any* $t \geq 0$ *it holds that*

$$\log \frac{Z_{\boldsymbol{\psi}_{t+1}}}{Z_{\boldsymbol{\psi}t}} \leq \begin{cases} \frac{\lambda}{\tau} \log \sum_{j=1}^{k} w_j \mathbf{E}_{X\sim\mu_t}\left[\exp\left(\frac{-\phi_{t+1}^j(X)+\phi_t^j(X)}{\tau}\right)^{\tau/\lambda}\right] & \text{if } \tau \geq \lambda, \\ \log \sum_{j=1}^{k} w_j \mathbf{E}_{X\sim\mu_t}\left[\exp\left(\frac{-\phi_{t+1}^j(X)+\phi_t^j(X)}{\tau}\right)\right] & \text{if } \tau < \lambda, \end{cases}$$

where $\boldsymbol{\phi}_t = \boldsymbol{\phi}_{\boldsymbol{\psi}_t}$ *and* $\mu_t(dx) = Z_{\boldsymbol{\psi}_t}^{-1}\exp(-\sum_{j=1}^k w_j\phi_t^j(x)/\tau)\pi_{\mathrm{ref}}(dx)$.

Observe that the sequence $(\boldsymbol{\psi}_t)_{t\geq 0}$ of the form stated in Lemma 3 satisfies, for any for any $j \in \{1,\ldots,k\}$ and any $t \geq 0$,

$$\exp\left(\frac{-\phi_{t+1}^j + \phi_t^j}{\tau}\right) = \exp\left(-\frac{\lambda}{\tau}\log\frac{d\mu_t}{d\tilde{\mu}_t^j}\right) = \left(\frac{d\tilde{\mu}_t^j}{d\mu_t}\right)^{\lambda/\tau},$$

where

$$\begin{aligned}\frac{d\tilde{\mu}_t^j}{d\mu_t}(x) &= \int_{\mathcal{X}} \nu(dy)\exp\left(\frac{\psi_{t+1}^j(y) + \phi_t^j(x) - c(x,y)}{\lambda}\right) \\ &= \int_{\mathcal{X}} \nu^j(dy)\Delta_t^j(y)^\eta \exp\left(\frac{\psi_t^j(y) + \phi_t^j(x) - c(x,y)}{\lambda}\right).\end{aligned}$$

Hence, by Lemma 4 we have

$$\begin{aligned}&\log\frac{Z_{\boldsymbol{\psi}_{t+1}}}{Z_{\boldsymbol{\psi}_t}} \\ &\leq \frac{1}{\tau}\min(\lambda,\tau)\sum_{j=1}^k w_j \mathbf{E}_{X\sim\mu_t}\Bigg[\\ &\qquad \left(\int_{\mathcal{X}}\nu^j(dy)\Delta_t^j(y)^\eta\exp\left(\frac{\psi_t^j(y)+\phi_t^j(X)-c(X,y)}{\lambda}\right)\right)^{\max(1,\lambda/\tau)}\Bigg].\end{aligned} \tag{17}$$

We split the remaining proof into two cases: $\tau \geq \lambda$ and $\tau < \lambda$.

The case $\tau \geq \lambda$. When $\tau \geq \lambda$, we have $\max(1, \lambda/\tau) = 1$ and $\eta = \min(1, \tau/\lambda) = 1$. Thus, (17) yields

$$\begin{aligned}&\log\frac{Z_{\boldsymbol{\psi}_{t+1}}}{Z_{\boldsymbol{\psi}_t}} \\ &\leq \frac{1}{\tau}\min(\lambda,\tau)\log\sum_{j=1}^k w_j\mathbf{E}_{X\sim\mu_t}\left[\int_{\mathcal{X}}\nu^j(dy)\Delta_t^j(y)\exp\left(\frac{\psi_t^j(y)+\phi_t^j(X)-c(X,y)}{\lambda}\right)\right] \\ &= \frac{1}{\tau}\min(\lambda,\tau)\log\sum_{j=1}^k w_j\left[\int_{\mathcal{X}}\mu_t(dx)\int_{\mathcal{X}}\nu^j(dy)\Delta_t^j(y)\exp\left(\frac{\psi_t^j(y)+\phi_t^j(X)-c(X,y)}{\lambda}\right)\right] \\ &= \frac{1}{\tau}\min(\lambda,\tau)\log\sum_{j=1}^k w_j\left[\int_{\mathcal{X}}\Delta_t^j(y)\nu^j(dy)\int_{\mathcal{X}}\exp\left(\frac{\psi_t^j(y)+\phi_t^j(X)-c(X,y)}{\lambda}\right)\mu_t(dx)\right] \\ &= \frac{1}{\tau}\min(\lambda,\tau)\log\sum_{j=1}^k w_j\left[\int_{\mathcal{X}}\Delta_t^j(y)\nu^j(dy)\frac{d\nu_{\boldsymbol{\psi}_t}^j}{d\nu^j}(y)\right] \\ &= \frac{1}{\tau}\min(\lambda,\tau)\log\sum_{j=1}^k w_j\mathbf{E}_{Y\sim\nu_{\boldsymbol{\psi}_t}^j}\left[\Delta_t^j(y)\right].\end{aligned}$$

This completes the proof of Lemma 3 when $\tau \geq \lambda$.

The case $\tau < \lambda$. For $j \in \{1, \ldots, k\}$ and any $x \in \mathcal{X}$ define the measure ρ_x by

$$\rho_x^j(dy) = \nu^j(dy) \exp\left(\frac{\psi_t^j(y) + \phi_t^j(x) - c(x,y)}{\lambda}\right).$$

By the definition of ϕ_t^j, we have

$$\begin{aligned}
&\int_{\mathcal{X}} \rho_x^j(dy) \\
&= \int_{\mathcal{X}} \nu(dy) \exp\left(\frac{\psi_t^j(y) - c(x,y)}{\lambda}\right) \exp\left(\frac{\phi_t^j(x)}{\lambda}\right) \\
&= \int_{\mathcal{X}} \nu(dy) \exp\left(\frac{\psi_t^j(y) - c(x,y)}{\lambda}\right) \exp\left(-\log \int_{\mathcal{X}} \nu^j(dy') \exp\left(\frac{\psi_t^j(y') - c(x,y')}{\lambda}\right)\right) \\
&= 1
\end{aligned}$$

In particular, ρ_x is a probability measure. Hence, (17) can be rewritten as

$$\log \frac{Z_{\boldsymbol{\psi}_{t+1}}}{Z_{\boldsymbol{\psi}_t}} \leq \log \sum_{j=1}^k w_j \mathbf{E}_{X\sim\mu_t}\left[\mathbf{E}_{Y\sim\rho_X^j}\left[\Delta_t^j(Y)^\eta \middle| X\right]^{\lambda/\tau}\right]$$

Because $\lambda/\tau > 1$, the function $x \mapsto x^{\lambda/\tau}$ is convex. Applying Jensen's inequality to the conditional expectation and using the fact that $\eta\lambda/\tau = 1$, it follows that

$$\begin{aligned}
\log \frac{Z_{\boldsymbol{\psi}_{t+1}}}{Z_{\boldsymbol{\psi}_t}} &\leq \log \sum_{j=1}^k w_j \mathbf{E}_{X\sim\mu_t}\left[\mathbf{E}_{Y\sim\rho_X^j}\left[\Delta_t^j(Y) \middle| X\right]\right] \\
&= \log \sum_{j=1}^k w_j \int_{\mathcal{X}} \mu_t(dx) \int_{\mathcal{X}} \Delta_t^j(y) \exp\left(\frac{\psi_t^j(y) + \phi_t^j(x) - c(x,y)}{\lambda}\right) \nu(dy). \quad (18)
\end{aligned}$$

By the definition of $\nu^j_{\boldsymbol{\psi}_t}$ we have

$$\frac{d\nu^j_{\boldsymbol{\psi}_t}}{d\nu^j}(y) = \int_{\mathcal{X}} \exp\left(\frac{\psi_t^j(y) + \phi_t^j(x) - c(x,y)}{\lambda}\right) \mu_t(dx).$$

Interchanging the order of integration in (18) and plugging in the above equation yields

$$\begin{aligned}
\log \frac{Z_{\boldsymbol{\psi}_{t+1}}}{Z_{\boldsymbol{\psi}_t}} &\leq \log \sum_{j=1}^k w_j \int_{\mathcal{X}} \left[\int_{\mathcal{X}} \exp\left(\frac{\psi_t^j(y) + \phi_t^j(x) - c(x,y)}{\lambda}\right) \mu_t(dx)\right] \Delta_t^j(y) \nu^j(dy) \\
&= \log \sum_{j=1}^k w_j \int_{\mathcal{X}} \left[\frac{d\nu^j_{\boldsymbol{\psi}_t}}{d\nu^j}(y)\right] \Delta_t^j(y) \nu^j(dy) \\
&= \log \sum_{j=1}^k w_j \mathbf{E}_{Y\sim\nu^j_{\boldsymbol{\psi}_t}}\left[\Delta_t^j(Y)\right].
\end{aligned}$$

This completes the proof of Lemma 3. □

B.2 Proof of Lemma 4

To simplify the notation, denote $Z_t = Z_{\boldsymbol{\psi}_t}$. Let $x \in \mathcal{X}$ and $t \geq 0$. We have $\mu_t \ll \mu_{t+1}$ with the Radon-Nikodym derivative $d\mu_{t+1}/d\mu_t$ given by

$$\begin{aligned}
\frac{d\mu_{t+1}}{d\mu_t}(x) &= \frac{Z_t}{Z_{t+1}} \exp\left(\frac{-\sum_{j=1}^k w_k(\phi_{t+1}^j(x) - \phi_t^j(x))}{\tau}\right) \\
&= \frac{Z_t}{Z_{t+1}} \prod_{j=1}^k \exp\left(\frac{-\phi_{t+1}^j(x) + \phi_t^j(x)}{\tau}\right)^{w_j}.
\end{aligned}$$

Multiplying both sides by Z_{t+1}/Z_t and taking expectations with respect to μ_t yields

$$\begin{aligned}\frac{Z_{t+1}}{Z_t} &= \mathbf{E}_{X\sim\mu_{t+1}}\left[\frac{Z_{t+1}}{Z_t}\right] \\ &= \mathbf{E}_{X\sim\mu_t}\left[\frac{Z_{t+1}}{Z_t}\frac{d\mu_{t+1}}{d\mu_t}(X)\right] \\ &= \mathbf{E}_{X\sim\mu_t}\left[\prod_{j=1}^k \exp\left(\frac{-\phi_{t+1}^j(X)+\phi_t^j(X)}{\tau}\right)^{w_j}\right].\end{aligned}$$

In the case $\tau < \lambda$, the proof is complete by the Artihmetic-Geometric mean inequality (recall that $w_j > 0$ for $j = 1, \ldots, k$ and $\sum_{j=1}^k w_j = 1$). On the other hand, if $\tau \geq \lambda$ then $x \mapsto x^{\lambda/\tau}$ is concave. Hence, it follows that

$$\begin{aligned}\log\frac{Z_{t+1}}{Z_t} &= \log\mathbf{E}_{X\sim\mu_t}\left[\left(\prod_{j=1}^k \exp\left(\frac{-\phi_{t+1}^j(X)+\phi_t^j(X)}{\tau}\right)^{w_j\tau/\lambda}\right)^{\lambda/\tau}\right] \\ &\leq \log\mathbf{E}_{X\sim\mu_t}\left[\left(\prod_{j=1}^k \exp\left(\frac{-\phi_{t+1}^j(X)+\phi_t^j(X)}{\tau}\right)^{w_j\tau/\lambda}\right)\right]^{\lambda/\tau} \\ &= \frac{\lambda}{\tau}\log\mathbf{E}_{X\sim\mu_t}\left[\left(\prod_{j=1}^k \exp\left(\frac{-\phi_{t+1}^j(X)+\phi_t^j(X)}{\tau}\right)^{w_j\tau/\lambda}\right)\right] \\ &\leq \sum_{j=1}^k w_j\mathbf{E}_{X\sim\mu_t}\left[\exp\left(\frac{-\phi_{t+1}^j(X)+\phi_t^j(X)}{\tau}\right)^{\tau/\lambda}\right],\end{aligned}$$

where the final step follows via the Arithmetic-Geometric mean inequality. This completes the proof of Lemma 4. □

C Proof of Theorem 2

For every $t \geq 0$ and $j \in \{1, \ldots, k\}$, let $\widetilde{\nu}_t^j$ be the distribution returned by the approximate Sinkhorn oracle that satisfies the properties listed in Definition 1. We follow along the lines of proof of Theorem 1.

First, we will establish an upper bound on the oscillation norm of the iterates $\widetilde{\boldsymbol{\psi}}_t$. Indeed, by the property four in Definition 1 we have

$$\|\widetilde{\psi}_{t+1}^j\|_{\mathrm{osc}} \leq (1-\eta)\|\widetilde{\psi}_t^j\|_{\mathrm{osc}} + \eta c_\infty(\mathcal{X}).$$

Since $\widetilde{\psi}_0^j = 0$, for any $t \geq 0$ we have $\|\widetilde{\psi}_t^j\|_{\mathrm{osc}} \leq c_\infty(\mathcal{X})$.

Let $\tilde{\delta}_t = E_{\lambda,\tau}^{\boldsymbol{\nu},w}(\psi^*) - E_{\lambda,\tau}^{\boldsymbol{\nu},w}(\widetilde{\boldsymbol{\psi}}_t)$ be the suboptimality gap at time t. Using the concavity upper bound (10) and the property two in Definition 1 we have

$$\begin{aligned}\tilde{\delta}_t &\leq 2c_\infty(\mathcal{X})\sum_{j=1}^k w_j\|\nu^j - \nu_t^j\|_{\mathrm{TV}} \\ &\leq \varepsilon + 2c_\infty(\mathcal{X})\sum_{j=1}^k w_j\|\nu^j - \widetilde{\nu}_t^j\|_{\mathrm{TV}} \\ &\leq \varepsilon + \sqrt{2}c_\infty(\mathcal{X})\sum_{j=1}^k w_j\sqrt{\mathrm{KL}(\nu^j,\widetilde{\nu}_t^j)} \\ &\leq \varepsilon + \sqrt{2}c_\infty(\mathcal{X})\sqrt{\sum_{j=1}^k w_j\mathrm{KL}(\nu^j,\widetilde{\nu}_t^j)}.\end{aligned}$$

Combining the property three[1] stated in the Definition 1 with Lemma 3 we obtain

$$\begin{aligned}
\tilde{\delta}_t - \tilde{\delta}_{t+1} &\geq \min(\lambda,\tau)\sum_{j=1}^k w_j \mathrm{KL}(v^j, \widetilde{v}_t^j) - \min(\lambda,\tau)\log\left(\sum_{j=1}^k w_j \int_{\mathcal{X}} \frac{d\nu_t}{d\widetilde{\nu}_t}(y)\nu^j(dy)\right) \\
&\geq \sum_{j=1}^k w_j \mathrm{KL}(v^j, \widetilde{v}_t^j) - \min(\lambda,\tau)\log\left(1 + \varepsilon^2/(2c_\infty(\mathcal{X})^2)\right) \\
&\geq \min(\lambda,\tau)\sum_{j=1}^k w_j \mathrm{KL}(v^j, \widetilde{v}_t^j) - \frac{\min(\lambda,\tau)}{2c_\infty(\mathcal{X})^2}\varepsilon^2 \\
&\geq \frac{\min(\lambda,\tau)}{2c_\infty(\mathcal{X})^2}\max\left\{0, \tilde{\delta}_t - \varepsilon\right\}^2 - \frac{\min(\lambda,\tau)}{2c_\infty(\mathcal{X})^2}\varepsilon^2.
\end{aligned}$$

Provided that $\widetilde{\delta}_t \geq 2\varepsilon$ it holds that

$$(\tilde{\delta}_t - 2\varepsilon) - (\tilde{\delta}_{t+1} - 2\varepsilon) \geq \frac{\min(\lambda,\tau)}{2c_\infty(\mathcal{X})}(\tilde{\delta}_t - 2\varepsilon)^2.$$

Let T be the first index such that $\widetilde{\delta}_{T+1} < 2\varepsilon$ and set $T = \infty$ if no such index exists. Then, the above equation is valid for any $t \leq T$. In particular, repeating the proof of Theorem 1, for any $t \leq T$ we have

$$\widetilde{\delta}_t - 2\varepsilon \leq \frac{2c_\infty(\mathcal{X})^2}{\min(\lambda,\tau)}\frac{1}{t},$$

which completes the proof of this theorem. □

D Proof of Lemma 2

The first property – the positivity of the probability mass function of $\widetilde{\nu}^j$ – is immediate from its definition.

To simplify the notation, denote in what follows

$$K^j(x,y) = \exp\left(\frac{\phi_{\psi^j}(x) + \psi^j(y) - c(x,y)}{\lambda}\right).$$

With this notation, recall that

$$\widehat{\nu}_{\boldsymbol{\psi}}^j(y_l^j) = \frac{1}{n}\sum_{i=1}^n \nu^j(y_l^j)K(X_i, y_l^j).$$

The above is a sum of n non-negative random variables bounded by one with expectation

$$(\nu')^j(y_l^j) = \mathbf{E}_{X\sim\mu'_{\boldsymbol{\psi}}}\left[\nu^j(y_l^j)\right]$$

It follows by Hoeffding's inequality and the union bound that with probability at least $1-\delta$ the following holds for any $j \in \{1,\ldots,k\}$ and any $l \in \{1,\ldots,m_j\}$:

$$\left|\widehat{\nu}_{\boldsymbol{\psi}}(y_l^j) - (\nu')^j(y_l^j)\right| \leq \sqrt{\frac{2\log\left(\frac{2m}{\delta}\right)}{n}}.$$

In particular, the above implies that

$$\begin{aligned}
\|\widetilde{\nu}_{\boldsymbol{\psi}}^j - \nu_{\boldsymbol{\psi}}^j\|_{\mathrm{TV}} &\leq 2\zeta + (1-\zeta)\|\widetilde{\nu}_{\boldsymbol{\psi}}^j - \nu_{\boldsymbol{\psi}}^j\|_{\mathrm{TV}} \\
&\leq 2\zeta + (1-\zeta)\|\widetilde{\nu}_{\boldsymbol{\psi}}^j - (\nu')^j\|_{\mathrm{TV}} + (1-\zeta)\|(\nu')^j - \nu_{\boldsymbol{\psi}}^j\|_{\mathrm{TV}} \\
&\leq 2\zeta + \|\widetilde{\nu}_{\boldsymbol{\psi}}^j - (\nu')^j\|_{\mathrm{TV}} + \|(\nu')^j - \nu_{\boldsymbol{\psi}}^j\|_{\mathrm{TV}} \\
&\leq 2\zeta + m_j\varepsilon_\mu + m_j\sqrt{\frac{2\log\left(\frac{2m}{\delta}\right)}{n}}.
\end{aligned}$$

[1]The third property, unlike claimed in the main text, should read as: $\mathbf{E}_{Y\sim\nu^j}[\frac{d\nu_{\boldsymbol{\psi}}^j}{d\widetilde{\nu}_{\boldsymbol{\psi}}^j}(Y)] \leq 1+\varepsilon^2/(2c_\infty(\mathcal{X})^2)$.

Notice that the above bound can be made arbitrarily close to $m_j\varepsilon_\mu$ by taking a large enough n and a small enough ζ. This proves the second property of Definition 1.

To prove the third property[2], observe that

$$\begin{aligned}
\mathbf{E}_{Y\sim\nu^j}\left[\frac{\nu^j_{\boldsymbol{\psi}}(Y)}{\widetilde{\nu}^j_{\boldsymbol{\psi}}(Y)}\right] &= \mathbf{E}_{Y\sim\nu^j}\left[\frac{\widehat{\nu}^j_{\boldsymbol{\psi}}(Y)}{\widetilde{\nu}^j_{\boldsymbol{\psi}}(Y)} + \frac{\nu^j_{\boldsymbol{\psi}}(Y) - \widehat{\nu}^j_{\boldsymbol{\psi}}(Y)}{\widetilde{\nu}^j_{\boldsymbol{\psi}}(Y)}\right] \\
&\le \mathbf{E}_{Y\sim\nu^j}\left[\frac{1}{1-\zeta} + \frac{\nu^j_{\boldsymbol{\psi}}(Y) - \widehat{\nu}^j_{\boldsymbol{\psi}}(Y)}{\widetilde{\nu}^j_{\boldsymbol{\psi}}(Y)}\right] \\
&\le \mathbf{E}_{Y\sim\nu^j}\left[1 + \frac{\zeta}{1-\zeta} + \frac{\left|\nu^j_{\boldsymbol{\psi}}(Y) - \widehat{\nu}^j_{\boldsymbol{\psi}}(Y)\right|}{\widetilde{\nu}^j_{\boldsymbol{\psi}}(Y)}\right] \\
&\le 1 + \frac{\zeta}{1-\zeta} + \frac{1}{\zeta}\|\nu^j_{\boldsymbol{\psi}}(Y) - \widehat{\nu}^j_{\boldsymbol{\psi}}(Y)\|_{\mathrm{TV}} \\
&\le 1 + 2\zeta + \frac{1}{\zeta}\left(m_j\varepsilon_\mu + m_j\sqrt{\frac{2\log\left(\frac{2m}{\delta}\right)}{n}}\right).
\end{aligned}$$

This concludes the proof of the third property.

It remains to prove the fourth property of Definition 1. Observe that for any y, y' we have

$$\begin{aligned}
&\left(\psi^j(y) - \eta\lambda\log\frac{\widetilde{\nu}^j(y)}{\nu^j(y)}\right) - \left(\psi^j(y') - \eta\lambda\log\frac{\widetilde{\nu}^j(y')}{\nu^j(y')}\right) \\
&= \left(\psi^j(y) - \psi^j(y')\right) + \eta\lambda\log\left(\frac{\zeta + (1-\zeta)\frac{1}{n}\sum_{i=1}^n K^j(X_i, y')}{\zeta + (1-\zeta)\frac{1}{n}\sum_{i=1}^n K^j(X_i, y)}\right) \\
&= \left(\psi^j(y) - \psi^j(y')\right) + \eta\lambda\log\left(\frac{\frac{\zeta}{1-\zeta} + \frac{1}{n}\sum_{i=1}^n K^j(X_i, y')}{\frac{\zeta}{1-\zeta} + \frac{1}{n}\sum_{i=1}^n K^j(X_i, y)}\right) \\
&\le \left(\psi^j(y) - \psi^j(y')\right) + \eta\lambda\log\left(\frac{\frac{\zeta}{1-\zeta} + \exp\left(\frac{c_\infty(\mathcal{X}) + \psi^j(y') - \psi^j(y)}{\lambda}\right)\frac{1}{n}\sum_{i=1}^n K^j(X_i, y)}{\frac{\zeta}{1-\zeta} + \frac{1}{n}\sum_{i=1}^n K^j(X_i, y)}\right).
\end{aligned}$$

Now observe that for any $a, b > 0$ the function $g : [0,\infty) \to (0,\infty)$ defined by $g(x) = (x+a)/(x+b)$ is increasing if $a < b$ and decreasing if $a \ge b$. Thus, g is maximized either at zero or at infinity. It thus follows that

$$\begin{aligned}
&\eta\lambda\log\left(\frac{\frac{\zeta}{1-\zeta} + \exp\left(\frac{c_\infty(\mathcal{X})}{\lambda}\right)\frac{1}{n}\sum_{i=1}^n K^j(X_i, y)}{\frac{\zeta}{1-\zeta} + \frac{1}{n}\sum_{i=1}^n K^j(X_i, y)}\right) \\
&\le \begin{cases} \eta c_\infty(\mathcal{X}) - \eta(\psi^j(y) - \psi^j(y')) & \text{if } \exp\left(\frac{c_\infty(\mathcal{X})}{\lambda}\right) \ge 1 \\ 0 & \text{otherwise.}\end{cases}
\end{aligned}$$

This proves the claim and completes the proof of this lemma. □

E Approximate Sampling From $\mu_{\boldsymbol{\psi}}$ via Langevin Monte Carlo

The purpose of this section is to show how sampling via Langevin Monte Carlo algorithm yields the first provable convergence guarantees for computing barycenters in the free-support setup (cf. the discussion at the end of Section 2.2). In particular, we provide computational guarantees for implementing Algorithm 2.

[2] The third property, unlike claimed in the main text, should read as: $\mathbf{E}_{Y\sim\nu^j}[\frac{d\nu^j_{\boldsymbol{\psi}}}{d\widetilde{\nu}^j_{\boldsymbol{\psi}}}(Y)] \le 1 + \varepsilon^2/(2c_\infty(\mathcal{X})^2)$.

A measure μ is said to satisfy the logarithmic Sobolev inequality (LSI) with constant C if for all sufficiently smooth functions f it holds that

$$\mathbf{E}_{\mu}[f^2 \log f^2] - \mathbf{E}_{\mu}[f^2] \log \mathbf{E}_{\mu}[g^2] \leq 2C \mathbf{E}_{\mu}[\|\nabla f\|_2^2].$$

To sample from a measure $\mu(dx) = \exp(-f(x))dx$ supported on $\mathbb{R}^d$, the unadjusted Langevin Monte Carlo algorithm is defined via the following recursive update rule:

$$x_{k+1} = x_k - \eta \nabla f(x_k) + \sqrt{2\eta} Z_k, \quad \text{where} \quad Z_k \sim \mathcal{N}(0, I_d). \tag{19}$$

The following Theorem is due to Vempala and Wibisono [55, Theorem 3].

Theorem 3. *Let $\mu(dx) = \exp(-f(x))dx$ be a measure on $\mathbb{R}^d$. Suppose that μ satisfies LSI a with constant C and that f has L-Lipschitz gradient with respect to the Euclidean norm. Consider the sequence of iterates $(x_k)_{k \geq 0}$ defined via* (19) *and let let ρ_k be the distribution of x_k. Then, for any $\varepsilon > 0$, any $\eta \leq \frac{1}{8L^2C} \min\{1, \frac{\varepsilon}{4d}\}$, and any $k \geq \frac{2C}{\eta} \log \frac{2\mathrm{KL}(\rho_0, \mu)}{\varepsilon}$, it holds that*

$$\mathrm{KL}(\rho_k, \mu) \leq \varepsilon.$$

Thus, LSI on the measure μ provides convergence guarantees on $\mathrm{KL}(\rho_k, \mu)$. It is shown in [55, Lemma 1] how to initialize the iterate x_0 so that $\mathrm{KL}(\rho_0, \mu)$ scales linearly with the ambient dimension d up to some additional terms.

To implement the approximate Sinkhorn oracle described in Definition 1, we can combine Lemma 2 with approximate sampling via Langevin Monte Carlo; note that by Pinsker's inequality, Kullback-Leibler divergence guarantees provide total variation guarantees which are sufficient for the application of Lemma 2. Therefore, providing provable convergence guarantees for Algorithm 2, the inexact version of Algorithm 1, amounts to proving that we can do arbitrarily accurate approximate sampling from distributions of the form

$$\mu_{\boldsymbol{\psi}}(dx) \propto \mathbb{1}_{\mathcal{X}}(x) \exp(-V_{\boldsymbol{\psi}}(x)/\tau) dx, \quad \text{where} \quad V_{\boldsymbol{\psi}}(x) = \sum_{j=1}^{k} w_j \phi^j_{\psi^j}(x).$$

Here $\mathbb{1}_{\mathcal{X}}$ is the indicator function of $\mathcal{X}$, $\boldsymbol{\psi}$ is an arbitrary iterate generated by Algorithm 2, and we consider the free-support setup characterized via the choice $\pi_{\mathrm{ref}}(dx) = dx$.

Notice that we cannot apply Theorem 3 directly because the measure $\mu_{\boldsymbol{\psi}}$ defined above has constrained support while Theorem 3 only applies for measures supported on all of $\mathbb{R}^d$. Nevertheless, we will show that the compactly supported measure $\mu_{\boldsymbol{\psi}}$ can be approximated by a measure $\mu_{\boldsymbol{\psi},\sigma}$, where parameter σ will trade-off LSI constant of $\mu_{\boldsymbol{\psi},\sigma}$ against the total variation norm between the two measures. To this end, define

$$\mu_{\boldsymbol{\psi},\sigma} =\propto \exp(-V_{\boldsymbol{\psi}}(x)/\tau - \mathrm{dist}(x, \mathcal{X})^2/(2\sigma^2))dx, \quad \text{where} \quad \mathrm{dist}(x, \mathcal{X}) = \inf_{y \in \mathcal{X}} \|x - y\|_2. \tag{20}$$

The argument presented below works for any cost function c such that $c(\cdot, y)$ is Lipschitz on $\mathcal{X}$ and grows quadratically at infinity. However, to not cloud the whole picture with technical details, we shall simply take $c(x, y) = \|x - y\|_2^2$. The exact problem setup is formalized below.

Problem Setting 1. We consider the setting described in Section 4.1. In addition, suppose that

1. the reference measure $\pi_{\mathrm{ref}}(dx)$ is the Lebesgue measure (free-support setup);
2. $\mathcal{X} \subseteq \mathcal{B}_R = \{x \in \mathbb{R}^d : \|x\|_2 \leq R\}$ for some constant $R < \infty$;
3. $c : \mathbb{R}^d \times \mathbb{R}^d \to [0, \infty)$ is defined by $c(x, y) = \|x - y\|_2^2$;
4. for any $\boldsymbol{\psi}$ generated by Algorithm 1 we have access to a stationary point $x_{\boldsymbol{\psi}}$ of $V_{\boldsymbol{\psi}}$ over $\mathcal{X}$.

The final condition can be implemented in polynomial time using a first order gradient method. The implication of this condition is that by [55, Lemma 1], for any $\sigma > 0$, the initialization scheme $x_0 \sim \mathcal{N}(x_{\boldsymbol{\psi}}, I_d)$ for the Langevin algorithm (19) satisfies

$$\mathrm{KL}(\rho_0, \mu_{\boldsymbol{\psi},\sigma}) \leq \frac{c_\infty(\mathcal{X})}{\tau} + \frac{d}{2} \log \frac{L_\sigma}{2\pi},$$

where L_σ is the smoothness constant of $V_{\boldsymbol{\psi}}/\tau + \mathrm{dist}(x, \mathcal{X})/(2\sigma^2)$ (see Lemma 5).

The following properties are satisfied by the measure $\mu_{\boldsymbol{\psi},\sigma}$.

Lemma 5. *Consider the setup described in Problem Setting 1. Let ψ be any iterate generated by Algorithm 2 and let $\mu_{\psi,\sigma}$ be the distribution defined in* (20)*. Then, the measure $\mu_{\psi,\sigma}$ satisfies the following properties:*

1. *For any $\sigma \in (0, 1/4]$ it holds that*

$$\|\mu_{\psi} - \mu_{\psi,\sigma}\|_{\mathrm{TV}} \leq 2\sigma \exp\left(\frac{8R^2}{\tau}\right)\left[\left(4Rd^{-1/4}\right)^{d-1} + 1\right].$$

2. *Let $V_\sigma(x) = \exp(-V_{\psi}(x)/\tau - \mathrm{dist}(x, \mathcal{X})^2/(2\sigma^2))$; thus $\mu_{\psi,\sigma}(dx) = \exp(-V_\sigma(x))dx$. The function V_σ has L_σ-Lipschitz gradient where*

$$L_\sigma = \frac{1}{\tau} + \frac{1}{\tau\lambda}4R^2 \max_j m_j + \frac{1}{\sigma^2}.$$

3. *The measure $\mu_{\psi,\sigma}$ satisfies LSI with a constant $C_\sigma = \mathrm{poly}(R, \exp(R^2/\tau), L_\sigma)$.*

Above, the notation $C = \mathrm{poly}(x, y, z)$ denotes a constant that depends polynomially on x, y and z.

Before proving this lemma, let us state and prove the main result of this section.

Corollary 1. *Consider the setup described in Problem Setting 1. Then, for any confidence parameter $\delta \in (0, 1)$ and any accuracy parameter $\varepsilon > 0$, we can simulate a step of Algorithm 2 with success probability at least $1 - \delta$ in time polynomial in*

$$\varepsilon^{-1}, d, R, \exp(R^2/\tau), (Rd^{-1/4})^d, \tau^{-1}, \lambda^{-1}, d, m, \log(m/\delta).$$

Comparing the above guarantee with the discussion at the end of Section 4.1, we see an additional polynomial dependence on $(Rd^{-1/4})^d$. We believe this term to be an artefact of our analysis, which appears due to the total variation norm approximation bound in Lemma 5. Ignoring this term (or considering the setup with $R \leq d^{1/4}$), the running time of our algorithm depends exponentially in R^2/τ. We conclude with the following two observations. First, because approximating Wasserstein barycenters is NP-hard in general [4], an algorithm with polynomial dependence on all problem parameters does not exist (unless $\mathrm{P} = \mathrm{NP}$). Second, combining the above corollary with Theorem 2, obtaining an ε approximation of (λ, τ)-Barycenter can be done in time polynomial in ε^{-1}. This should be contrasted with numerical schemes based on discretizations of the set $\mathcal{X}$, which would, in general, result in computational complexity of order $(R/\varepsilon)^d$ to reach the same accuracy.

Proof. Let ψ be an arbitrary iterate generated via Algorithm 2. We can simulate a step of approximate Sinkhorn oracle with accuracy ε via Lemma 2 (with $\zeta = \varepsilon/4$) in time $\mathrm{poly}(n, m, d)$ provided access to $n = \mathrm{poly}(\varepsilon^{-1}, m, \log(m/\delta))$ samples from any distribution μ'_{ψ} such that

$$\|\mu'_{\psi} - \mu_{\psi}\|_{\mathrm{TV}} \leq \frac{\varepsilon^2}{16m}. \tag{21}$$

To find a choice of μ'_{ψ} satisfying the above bound, consider the distribution

$$\mu_{\psi,\sigma} \quad \text{with} \quad \sigma = \frac{\varepsilon^2}{32m} \cdot \left(2\exp\left(\frac{8R^2}{\tau}\right)\left[\left(4Rd^{-1/4}\right)^{d-1} + 1\right]\right)^{-1}.$$

Let C_σ and L_σ be the LSI and smoothness constants of the distribution $\mu_{\psi,\sigma}$ provided in Lemma 5. By Theorem 3, it suffices to run the Langevin algorithm (19) for $\mathrm{poly}(\varepsilon^{-1}, m, d, C_\sigma, L_\sigma)$ number of iterations to obtain a sample from a distribution $\widetilde{\mu}_{\psi,\sigma}$ such that

$$\|\widetilde{\mu}_{\psi,\sigma} - \mu_{\psi,\sigma}\|_{\mathrm{TV}} \leq \frac{\varepsilon^2}{32m}.$$

In particular, by the triangle inequality for the total variation norm, the choice $\mu'_{\psi} = \widetilde{\mu}_{\psi,\sigma}$ satisfies (21). This finishes the proof. □

E.1 Proof of Lemma 5

To simplify the notation, denote $\mu = \mu_{\psi}, \mu_\sigma = \mu_{\psi,\sigma}$, $V(x) = V_{\psi}(x)/\tau$, and $V_\sigma(x) = V(x)/\tau + \mathrm{dist}(x, \mathcal{X})^2/(2\sigma^2)$.

Total variation norm bound. With the above shorthand notation, we have

$$\mu(dx) = \mathbb{1}_{\mathcal{X}} Z^{-1} \exp(-V(x))dx, \quad \text{where} \quad Z = \int_{\mathcal{X}} \exp(-V(x))dx$$

and

$$\mu_\sigma(dx) = (Z + Z_\sigma)^{-1} \exp(-V_\sigma(x))dx, \quad \text{where} \quad Z_\sigma = \int_{\mathbb{R}^d \setminus \mathcal{X}} \exp(-V_\sigma(x))dx.$$

We have

$$\begin{aligned} \|\mu - \mu_\sigma\|_{\mathrm{TV}} &= \int_{\mathbb{R}^d \setminus \mathcal{X}} (Z + Z_\sigma)^{-1} \exp(-V_\sigma(x))dx + \int_{\mathcal{X}} |(Z + Z_\sigma)^{-1} - Z^{-1}| \exp(-V(x))dx \\ &= \frac{2Z_\sigma}{Z + Z_\sigma} \leq \frac{2Z_\sigma}{Z} \leq 2\exp\left(\frac{c_\infty(\mathcal{X})}{\tau}\right) Z_\sigma \leq 2\exp\left(\frac{4R^2}{\tau}\right) Z_\sigma. \end{aligned}$$

We thus need to upper bound Z_σ. Let $\mathrm{Vol}(A)$ be the Lebesgue measure of the set A, let ∂A denote the boundary of A, and let $A + B = \{a + b : a \in A, b \in B\}$ be the Minkowski sum of sets A and B. Using the facts that for each $j \in \{1, \ldots, k\}$ we have $\sup_{y \in \mathcal{X}} \psi^j(y) \leq c_\infty(\mathcal{X}) \leq 4R^2$ and that $\mathcal{X} \subseteq \mathcal{B}_R = \{x : \|x\|_2 \leq R\}$ we have

$$\begin{aligned} Z_\sigma &= \int_{\mathbb{R}^d \setminus \mathcal{X}} \exp(-V_\sigma(x))dx \\ &\leq \exp\left(\frac{4R^2}{\tau}\right) \int_{\mathbb{R}^d \setminus \mathcal{X}} \exp\left(-\frac{\mathrm{dist}(x, \mathcal{X})}{2\sigma^2}\right) dx \\ &= \exp\left(\frac{4R^2}{\tau}\right) \int_0^\infty \mathrm{Vol}(\partial(\mathcal{X} + \mathcal{B}_x)) \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \\ &\leq \exp\left(\frac{4R^2}{\tau}\right) \int_0^\infty \mathrm{Vol}(\partial \mathcal{B}_{R+x}) \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \\ &= \exp\left(\frac{4R^2}{\tau}\right) \frac{\pi^{d/2}}{\Gamma(d/2)} \int_0^\infty (R + x)^{d-1} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx. \end{aligned}$$

Bounding $(R + x)^{d-1} \leq 2^{d-1}R^{d-1} + 2^{d-1}x^{d-1}$ and computing the integrals results in

$$\begin{aligned} \|\mu - \mu_\sigma\|_{\mathrm{TV}} &\leq 2\exp\left(\frac{8R^2}{\tau}\right) \frac{\pi^{d/2}}{\Gamma(d/2)} 2^{d-1} \left[R^{d-1}\sigma\frac{\sqrt{\pi}}{2} + 2^{d/2-1}\Gamma(d/2)\sigma^d\right] \\ &\leq 2\sigma \exp\left(\frac{8R^2}{\tau}\right) \left[\frac{(2R)^{d-1}}{\Gamma(d/2)} + (4\sigma)^{d-1}\right]. \end{aligned}$$

Using the assumption $\sigma \leq 1/4$ and using the bound $\Gamma(d) \geq (d/2)^{d/2}$ we can further simplify the above bound to

$$\|\mu - \mu_\sigma\|_{\mathrm{TV}} \leq 2\sigma \exp\left(\frac{8R^2}{\tau}\right) \left[\left(4Rd^{-1/4}\right)^{d-1} + 1\right],$$

which completes the proof of the total variation bound.

Lipschitz constant of the gradient. Recall that for any any $j \in \{1, \ldots, d\}$ we have

$$\phi^j(x) - \frac{1}{2}\|x\|_2^2 = -\lambda \log\left(\sum_{l=1}^{n_j} \exp\left(\frac{\psi^j(y_l^j) - \frac{\|y_l^j\|_2^2}{2} + \langle x, y_l^j \rangle}{\lambda}\right) \nu^j(y_l^j)\right).$$

Denote $\widetilde{\phi}^j(x) = \phi^j(x) - \frac{1}{2}\|x\|_2^2$. Fix any x, x' and define $g(t) = \widetilde{\phi}^j(x + (x' - x)t)$. Then, for any $t \in [0, 1]$ we have

$$g''(s) = -\frac{1}{\lambda}\mathrm{Var}_{L \sim \rho_t}\left[(Y^j(x' - x))_L\right] \geq -\frac{1}{\lambda}\|x - x'\|_2^2 m_j 4R^2, \tag{22}$$

where

$$\rho_t(l) \propto \nu(y_l^j) \exp\left(\frac{\psi^j(y_l^j) - \frac{\|y_l^j\|_2^2}{2} + \langle x + t(x' - x), y_l^j \rangle}{\lambda}\right)$$

and $Y^j \in \mathbb{R}^{d \times m_j}$ is the matrix whose l-th column is equal to the vector y_l^j.

Because $\widetilde{\psi}^j$ is concave, the bound (22) shows that ϕ^j is $1 + \frac{1}{\lambda} m_j 4R^2$-smooth.

Combining the above with the fact that the convex function $\mathrm{dist}(x, \mathcal{X})$ has 1-Lipschitz gradient [6, Proposition 12.30] proves the desired smoothness bound on the function V_σ.

LSI Constant bound. The result follows, for example, by applying the sufficient log-Sobolev inequality criterion stated in [13, Corollary 2.1, Equation (2.3)], combined with the bound (22). The exact constant appearing in the log-Sobolev inequality can be traced from [13, Equation (3.10)].