A. Proof of Lemma 9

Before proving Lemma 9, we first prove some intermediate Lemmas.

Lemma 14. *Let μ, ν be any two distributions. Then, $\forall R > 0$, we have*

$$\begin{aligned}\mathrm{W}_2^2(\mu,\nu) \le & 4R^2\|\mu-\nu\|_{\mathrm{TV}} + 2\mathbb{E}_{X\sim\mu}\left[\|X\|_2^2 1_{\{\|X\|_2>R\}}\right] + 2R^2\mathbb{E}_{X\sim\mu}\left[1_{\{\|X\|_2>R\}}\right] \\ & + 2\mathbb{E}_{Y\sim\nu}\left[\|Y\|_2^2 1_{\{\|Y\|_2>R\}}\right] + 2R^2\mathbb{E}_{Y\sim\nu}\left[1_{\{\|Y\|_2>R\}}\right]\end{aligned}$$

where $1_{\{\|X\|_2>R\}}$ is the indicator function of the set $B(0,R)^c = \{x \in \mathbb{R}^d : \|x\|_2 > R\}$.

Proof. Let $X \sim \mu, Y \sim \nu$. W_2-distance between probability measures μ and ν can be interpreted as the most cost-efficient transport plan to transform μ into ν, defined as

$$\mathrm{W}_2^2(\mu,\nu) = \min_{(X,Y)\sim\gamma} \mathbb{E}\|X-Y\|_2^2, \tag{24}$$

where the minimization is over all probability measures γ that marginalize to μ, ν, namely,

$$\gamma(A \times \mathbb{R}^d) = \mu(A), \qquad \gamma(\mathbb{R}^d \times B) = \nu(B), \tag{25}$$

for any measurable sets $A, B \subseteq \mathbb{R}^d$. For a fixed such measure γ, let us decompose the right-hand side of (24) as

$$\mathbb{E}\|X-Y\|_2^2 = \mathbb{E}\left[\|X-Y\|_2^2 1_{E_R}\right] + \mathbb{E}\left[\|X-Y\|_2^2 1_{E_R^c}\right], \tag{26}$$

where 1_{E_R} stands for the indicator of the event $E_R = \{\|X\|_2 \le R, \|Y\|_2 \le R\}$. Above, E_R^c is the complement of E_R. For the first expectation on the right-hand side above, we write that

$$\begin{aligned}\mathbb{E}\left[\|X-Y\|_2^2 1_{E_R}\right] &\le 4R^2\mathbb{E}\left[1_{X\neq Y}1_{E_R}\right] \\ &\le 4R^2\mathbb{E}[1_{X\neq Y}].\end{aligned} \tag{27}$$

For the second expectation on the right-hand side of (26), we write that

$$\mathbb{E}\left[\|X-Y\|_2^2 1_{E_R^c}\right] \le 2\mathbb{E}\left[\|X\|_2^2 1_{E_R^c}\right] + 2\mathbb{E}\left[\|Y\|_2^2 1_{E_R^c}\right]. \qquad ((a+b)^2 \le 2a^2 + 2b^2) \tag{28}$$

Let us in turn focus on, say, the first expectation on the right-hand side of (28). Since

$$1_{E_R^c} = 1_{\{\|X\|_2>R\}} + 1_{\{\|X\|_2\le R\}}1_{\{\|Y\|_2>R\}},$$

we can write that

$$\begin{aligned}\mathbb{E}\left[\|X\|_2^2 1_{E_R^c}\right] &= \mathbb{E}\left[\|X\|_2^2 1_{\{\|X\|_2>R\}}\right] + \mathbb{E}\left[\|X\|_2^2 1_{\{\|X\|_2\le R\}}1_{\{\|Y\|_2>R\}}\right] \\ &\le \mathbb{E}\left[\|X\|_2^2 1_{\{\|X\|_2>R\}}\right] + R^2\mathbb{E}\left[1_{\{\|Y\|_2>R\}}\right].\end{aligned} \tag{29}$$

Bounding $\mathbb{E}\left[\|Y\|_2^2 1_{E_R^c}\right]$ similarly, we obtain

$$\begin{aligned}\mathbb{E}\|X-Y\|_2^2 \le & 4R^2\mathbb{E}[1_{X\neq Y}] + 2\mathbb{E}_{X\sim\mu}\left[\|X\|_2^2 1_{\{\|X\|_2>R\}}\right] + 2R^2\mathbb{E}_{X\sim\mu}\left[1_{\{\|X\|_2>R\}}\right] \\ & + 2\mathbb{E}_{Y\sim\nu}\left[\|Y\|_2^2 1_{\{\|Y\|_2>R\}}\right] + 2R^2\mathbb{E}_{Y\sim\nu}\left[1_{\{\|Y\|_2>R\}}\right]\end{aligned}$$

The result is then obtained by minimizing the above inequality over all coupling γ, and using the fact that $\|\mu-\nu\|_{\mathrm{TV}} = \min_{(X,Y)\sim\gamma}\mathbb{E}[1_{X\neq Y}]$ (Gibbs & Su, 2002). □

Lemma 15. *Suppose that μ, ν both satisfy Assumption 4 with $\eta, M_\eta > 0$ and such that $\mathbb{E}_{X\sim\mu}\left[\|X\|_2^2\right], \mathbb{E}_{Y\sim\nu}\left[\|Y\|_2^2\right] \le C^2$. Then, for any $R \ge C$,*

$$\mathrm{W}_2^2(\mu,\nu) \le 4R^2\|\mu-\nu\|_{\mathrm{TV}} + 8\left(R^2 + RC + C^2\right)e^{-\frac{R}{C}+1}. \tag{30}$$

Proof. We start from the result of Lemma 14. The goal is then to bound the each term on the right hand side using the tail property of log-concave distributions (Lemma 6).

We have

$$\begin{aligned}
\mathbb{E}\left[\|X\|_2^2 1_{\{\|X\|_2>R\}}\right] &= 2\int_{\|x\|_2>R}\int_{z\in\mathbb{R}} 1_{\{\|x\|_2\geq z\}} z dz d\mu(x) \\
&= 2\int_{z\in\mathbb{R}} z dz \int_{\|x\|_2\geq \max(R,z)} d\mu(x) \\
&= 2\int_{z\in\mathbb{R}} z \Pr\left[\|X\|_2 \geq \max(R,z)\right] dz \\
&= 2\Pr[\|X\|_2\geq R]\int_0^R z dz + 2\int_R^\infty z\Pr[\|X\|_2\geq z] dz \\
&\leq R^2 e^{-\frac{R}{C}+1} + 2\int_R^\infty z e^{-\frac{z}{C}+1} dz \\
&\leq \left(R^2+2CR+2C^2\right) e^{-\frac{R}{C}+1}.
\end{aligned} \tag{31}$$

Similarly, we have

$$\mathbb{E}[1_{\{\|X\|_2>R\}}] = \Pr[\|X|_2 > R] \leq e^{-\frac{R}{C}+1}. \tag{32}$$

Doing the same calculation for Y and replacing the terms in Lemma 14 provides the result. □

Using the previous Lemma, it is now easy to prove the result of Lemma 9.

Proof of Lemma 9. Let us apply Lemma 15 using

$$R = C\max\left(\log\left(\frac{1}{\|\mu-\nu\|_{\mathrm{TV}}}\right), 1\right).$$

With this choice of R and if $\|\mu-\nu\|_{\mathrm{TV}}\leq 1$, note that

$$e^{-\frac{R}{C}} = \|\mu-\nu\|_{\mathrm{TV}}. \tag{33}$$

On the other hand, if $\|\mu-\nu\|_{\mathrm{TV}} > 1$, then

$$e^{-\frac{R}{C}} \leq 1 \leq \|\mu-\nu\|_{\mathrm{TV}}. \tag{34}$$

Thus, Lemma 15 gives

$$\begin{aligned}
\mathrm{W}_2^2(\mu,\nu) &\leq 4C^2\max\left(\log^2\left(\frac{1}{\|\mu-\nu\|_{\mathrm{TV}}}\right),1\right)\|\mu-\nu\|_{\mathrm{TV}} + 8C^2\left(1+\max\left(\log\left(\frac{1}{\|\mu-\nu\|_{\mathrm{TV}}}\right),1\right)\right)^2\|\mu-\nu\|_{\mathrm{TV}} \\
&\leq 20C^2\max\left(\log^2\left(\frac{1}{\|\mu-\nu\|_{\mathrm{TV}}}\right),1\right)\|\mu-\nu\|_{\mathrm{TV}}.
\end{aligned} \tag{35}$$

Lemma 9 then follows from taking the square root of (35) and using $C^2 = \frac{d(d+1)}{\eta^2} + M_\eta$ according to Lemma 5.

B. Proof of Theorem 10

We start by showing the following result in the case where the target distribution μ^* satisfies $\mathbb{E}_{X\sim\mu^*}\left[\|X\|_2^2\right]\leq 1$.

Theorem 16. (iteration complexity of DL-ULA) *Let μ^* be a L-smooth log-concave distribution such that $\mathbb{E}_{X\sim\mu^*}\left[\|X\|_2^2\right]\leq 1$. Suppose that μ_0 also satisfies $\mathbb{E}_{X\sim\mu_0}\left[\|X\|_2^2\right]\leq 1$. For every $k\geq 1$, let*

$$n_k = Ldk^2e^{3k} \tag{36}$$

$$\gamma_k = \frac{1}{Ld} e^{-2k} \tag{37}$$

$$\tau_k = k. \tag{38}$$

Then, $\forall \epsilon > 0$, we have:

- *After $N^{\mathrm{KL}} = \tilde{\mathcal{O}}(Ld\epsilon^{-\frac{3}{2}})$ total iterations, we obtain $\mathrm{KL}(\tilde{\mu}_k; \mu^*) \leq \epsilon$ where $\tilde{\mu}_k$ is the distribution associated to the iterates of outer iteration k just* before *the projection step.*
- *After $N^{\mathrm{TV}} = \tilde{\mathcal{O}}(Ld\epsilon^{-3})$ total iterations, we obtain $\|\tilde{\mu}_k - \mu^*\|_{\mathrm{TV}} \leq \epsilon$.*
- *After $N^{\mathrm{W}_2} = \tilde{\mathcal{O}}(Ld\epsilon^{-6})$ total iterations, we obtain $\mathrm{W}_2(\tilde{\mu}_k, \mu^*) \leq \epsilon$.*

Proof. Recall that in Algorithm 1, we denote as $\bar{\mu}_k$ the average of the distributions associated to the iterates of outer iteration k just *before* the projection step, i.e., just before the projection step, $x_k \sim \bar{\mu}_k$. We also denote as $\tilde{\mu}_k$ the same distribution, but *after* the projection step, i.e. the iterate that will be used as a warm start for the next outer iteration.

In order to show the result, we will show by induction that $\forall k \geq 1$,

$$\|\tilde{\mu}_k - \mu^*\|_{\mathrm{TV}} \leq u_k e^{-k} \tag{39}$$

where $\{u_k\}_{k\geq 1}$ is a real-valued sequence defined as $u_1 = \min(2\sqrt{e}W_2(\mu_0, \mu^*) + 1 + 2\sqrt{2}, 2e)$ and $u_k = 4\sqrt{eu_{k-1}} + 9 + 2\sqrt{2}$.

Let us fix $k \geq 2$. Thanks to the inequality (10),

$$\begin{aligned}
\|\bar{\mu}_k - \mu^*\|_{\mathrm{TV}} &\leq \sqrt{2KL(\bar{\mu}_k; \mu^*)} \qquad \text{(Pinsker's inequality)} \\
&\leq \sqrt{\frac{W_2^2(\tilde{\mu}_{k-1}, \mu^*)}{\gamma_k n_k} + 2Ld\gamma_k} \\
&\leq \frac{W_2(\tilde{\mu}_{k-1}, \mu^*)}{\sqrt{\gamma_k n_k}} + \sqrt{2Ld\gamma_k}
\end{aligned} \tag{40}$$

In order to use a recursion argument, we need to bound $W_2(\tilde{\mu}_{k-1}, \mu^*)$ by $\|\tilde{\mu}_{k-1} - \mu^*\|_{\mathrm{TV}}$. Note that the projection step for $\tilde{\mu}_{k-1}$ with $\tau_{k-1} = (k-1)$ ensures that $\mathrm{Pr}_{X\sim\tilde{\mu}_{k-1}}(\|X\|_2 \geq k-1) = 0$. Knowing that $\mathbb{E}_{X\sim\mu^*}\left[\|X\|_2^2\right] \leq 1$, we can apply Lemma 15 on $W_2(\tilde{\mu}_{k-1}, \mu^*)$ using $R = k$. Also, by replacing the values for γ_k, n_k, we get

$$W_2^2(\tilde{\mu}_{k-1}, \mu^*) \leq 4k^2\|\tilde{\mu}_{k-1} - \mu_{k-1}\|_{\mathrm{TV}} + 16ek^2e^{-k}.$$

Thus,

$$\|\bar{\mu}_k - \mu^*\|_{\mathrm{TV}} \leq \frac{2k\|\tilde{\mu}_{k-1} - \mu^*\|_{\mathrm{TV}} + 4\sqrt{e}ke^{-\frac{k}{2}}}{ke^{\frac{k}{2}}} + \sqrt{2}e^{-k}$$

Now, by using the recursion hypothesis, i.e. that $\|\tilde{\mu}_{k-1} - \mu^*\|_{\mathrm{TV}} \leq u_{k-1}e^{-k+1}$, we have:

$$\|\bar{\mu}_k - \mu^*\|_{\mathrm{TV}} \leq \left(2\sqrt{eu_{k-1}} + 4\sqrt{e} + \sqrt{2}\right) e^{-k} \tag{41}$$

Then, by taking into account the projection step at the end of outer iteration k, we obtain

$$\begin{aligned}
\|\tilde{\mu}_k - \mu_k\|_{\mathrm{TV}} &\leq \|\tilde{\mu}_k - \bar{\mu}_k\|_{\mathrm{TV}} + \|\bar{\mu}_k - \mu^*\|_{\mathrm{TV}} \qquad \text{(triangle inequality)} \\
&= \Pr_{X\sim\bar{\mu}_k}[\|X\|_2 > \tau_k] + \|\bar{\mu}_k - \mu^*\|_{\mathrm{TV}},
\end{aligned} \tag{42}$$

where the last line above follows because the projection step ensures $\Pr_{X\sim\tilde{\mu}_k}[\|X\|_2 > \tau_k] = 0$. In turn, to compute the probability in the last line above, we write that

$$\begin{aligned}\Pr_{X\sim\bar{\mu}_k}[\|X\|_2 \geq \tau_k] &\leq \Pr_{X\sim\mu^*}[\|X\|_2 \geq \tau_k] + |\bar{\mu}_k([\tau_k,\infty]) - \mu^*([\tau_k,\infty])| \qquad \text{(triangle inequality)}\\ &\leq e^{-k} + \|\bar{\mu}_k - \mu^*\|_{\mathrm{TV}}, \end{aligned} \tag{43}$$

By combining (41), (42) and (43), we finally obtain

$$\begin{aligned}\|\tilde{\mu}_k - \mu^*\|_{\mathrm{TV}} &\leq 2\|\bar{\mu}_k - \mu^*\|_{\mathrm{TV}} + e^{-k}\\ &\leq \left(4\sqrt{eu_{k-1}} + 9 + 2\sqrt{2}\right)e^{-k}\\ &= u_k e^{-k}\end{aligned}$$

Finally, using equations (40), (42) and (43) applied at $k=1$, we can also apply Lemma 15 and we get:

$$\|\tilde{\mu}_1 - \mu_1\|_{\mathrm{TV}} \leq \left(2W_2(\mu_0,\mu^*) + 2\sqrt{2} + 1\right)e^{-1} \tag{44}$$

which proves the result for the initial case. We thus showed that equation (39) holds for all $k \geq 1$.

It is easy to verify that the sequence $\{u_k\}_{k\geq 1}$ converges, and is upper bounded by $U = \max(u_1, u^*)$ where $u^* = \lim_{k\to\infty} u_k$. Moreover, since $\mathbb{E}_{X\sim\mu^*}\left[\|X\|_2^2\right], \mathbb{E}_{X\sim\mu_0}\left[\|X\|_2^2\right] \leq 1$ we have that $W_2(\mu_0,\mu^*) \leq 2$, and thus U is dimension independent.

After each outer iteration k, we thus have $\|\tilde{\mu}_k - \mu^*\|_{\mathrm{TV}} \leq Ue^{-k}$. Therefore, after $K^{\mathrm{TV}} = \log(\frac{U}{\epsilon})$ iterations, we have $\|\tilde{\mu}_k - \mu^*\|_{\mathrm{TV}} \leq \epsilon$. The total number of iterations required is

$$\begin{aligned}N^{\mathrm{TV}} &= \sum_{k=1}^{K^{\mathrm{TV}}} n_k\\ &\leq LdK^2 \sum_{k=1}^{K^{\mathrm{TV}}} e^{3k}\\ &= \frac{1}{1-e^{-3}} Ld\log^2\left(\frac{U}{\epsilon}\right)U^3\epsilon^{-3}\end{aligned}$$

Similarly, we also have $\mathrm{W}_2^2(\tilde{\mu}_k,\mu^*) \leq 4k^2\|\tilde{\mu}_k - \mu^*\|_{\mathrm{TV}} + 16ek^2e^{-k} \leq (4U+16e)k^2e^{-k}$. Thus, after $K^{\mathrm{W}_2} = \log(\frac{4U+16e}{\epsilon^2})$ iterations, we have $\mathrm{W}_2^2(\tilde{\mu}_k,\mu^*) \leq \epsilon\log(\frac{4U+16e}{\epsilon^2})$. The total number of iterations required is $N^{\mathrm{W}_2} = \mathcal{O}(Ld\epsilon^{-6})$.

Finally, we have $\mathrm{KL}(\bar{\mu}_k;\mu^*) \leq \frac{W_2^2(\tilde{\mu}_{k-1},\mu^*)}{2\gamma_k n_k} + Ld\gamma_k \leq 2\|\tilde{\mu}_{k-1} - \mu^*\|_{\mathrm{TV}}e^{-k} + e^{-2k} \leq (U+1)e^{-2k}$. Therefore, after $K^{\mathrm{KL}} = \frac{1}{2}\log(\frac{U+1}{\epsilon})$ iterations, we have $\mathrm{KL}(\bar{\mu}_k;\mu^*) \leq \epsilon$. The total number of iterations required is $N^{\mathrm{KL}} = \mathcal{O}(Ld\epsilon^{-\frac{3}{2}})$.

□

In order to show the more general theorem 10, we must get rid of the assumption that $\mathbb{E}_{X\sim\mu^*}\left[\|X\|_2^2\right] \leq 1$. To this end, we will suppose that we apply DL-ULA to a contracted version of μ^*, for which theorem 10 applies. Then, we will dilate the obtained sample in order to recover samples from the desired measure μ^* and bound the error induced by this dilatation in order to obtain the final convergence result.

Let us first recall the notion of push-forward measure.

Definition 17. *Let $h : \mathbb{R}^d \to \mathbb{R}$ be a strongly convex function whose gradient is denoted as $\nabla h : \mathbb{R}^d \to \mathbb{R}^d$. We say that ν is the* push-forward measure *of μ under ∇h, and we write $\nu = \nabla h \# \mu$, if ν is the distribution obtained by sampling from μ, and then applying the map ∇h to the samples.*

More precisely, it means that for every Borel set E on $\mathbb{R}^d$, we have $\nu(E) = \mu(\nabla h^{-1}(E))$.

Lemma 18. *Let $\mathrm{d}\mu = e^{-f(x)}\,\mathrm{d}x$ and $\mathrm{d}\nu = e^{-g(x)}\,\mathrm{d}x$ be such that $\nu = \nabla h \# \mu$ for some strongly convex function h. Then, the triplet (μ, ν, h) must satisfy the Monge-Ampère equation:*

$$e^{-f} = e^{-g \circ \nabla h} \det \nabla^2 h.$$

Let $\mathrm{d}\mu^* = e^{-f(x)}\,\mathrm{d}x$ be an L-smooth log-concave target distribution such that $\mathbb{E}_{X \sim \mu^*}\left[\|X\|_2^2\right] \leq M^2$. Instead of directly sample from μ^*, suppose that we sample from the shrunk distribution $\nu^* = \nabla h \# \mu^*$ with $h(x) = \frac{1}{2M}\|x\|_2^2$ for some $M \geq 0$, i.e., $\nabla h(x) = \frac{x}{M}$. In this particular case, we have that $\det \nabla^2 h(x)$ is independent of x. Therefore, we have according the Lemma 18 that $\mathrm{d}\nu^* \propto e^{-f(Mx)}\,\mathrm{d}x$.

This means that ν^* is the same distribution as μ^*, after the samples have been divided by M. It is easy to see that this scaling procedure implies that $\mathbb{E}_{X\sim\nu^*}\left[\|X\|_2\right] = \frac{1}{M}\mathbb{E}_{X\sim\mu^*}\left[\|X\|_2\right] \leq 1$.

Thus, if we apply DL-ULA for sampling from ν^*, then we can apply the convergence result provided by theorem 16. Note that this push-forward implies that ν^* is M^2L-smooth, i.e., the Lipschitz constant has been multiplied by M^2. Indeed, if $g(x) = f(Mx)$ and f is L-smooth, then,

$$\begin{aligned}\|\nabla g(y) - \nabla g(x)\|_2 &= M\|\nabla f(My) - \nabla f(Mx)\|_2 \\ &\leq M^2\|y - x\|_2.\end{aligned}$$

Let $\tilde{\nu}$ be the approximated distribution obtained using DL-ULA on ν with $n_k = LM^2dk^2e^{3k}$, $\gamma_k = \frac{1}{LM^2d}e^{-2k}$ and $\tau_k = k$. Then, according to Theorem 16, we have the following convergence results:

- After $N^{\mathrm{KL}} = \tilde{\mathcal{O}}(LM^2d\epsilon^{-\frac{3}{2}})$ total iterations, we obtain $\mathrm{KL}(\tilde{\nu} - \nu^*) \leq \epsilon$.
- After $N^{\mathrm{TV}} = \tilde{\mathcal{O}}(LM^2d\epsilon^{-3})$ total iterations, we obtain $\|\tilde{\nu} - \nu^*\|_{\mathrm{TV}} \leq \epsilon$.
- After $N^{\mathrm{W}_2} = \tilde{\mathcal{O}}(LM^2d\epsilon^{-6})$ total iterations, we obtain $\mathrm{W}_2(\tilde{\nu}, \nu^*) \leq \epsilon$.

By applying the inverse mapping $\nabla h^{-1}(x) = Mx$, we obtain samples from $\tilde{\mu} = \nabla h^{-1} \# \tilde{\nu}$. Interestingly, it can be shown that applying the same push-forward on two measures does not change their TV-distance not their KL divergence (Hsieh et al., 2018):

$$\begin{aligned}\|\tilde{\nu} - \nu^*\|_{\mathrm{TV}} &= \|\nabla h^{-1}\#\tilde{\nu} - \nabla h^{-1}\#\nu^*\|_{\mathrm{TV}} = \|\tilde{\mu} - \mu^*\|_{\mathrm{TV}}, \\ \mathrm{KL}(\tilde{\nu}; \nu^*) &= \mathrm{KL}(\nabla h^{-1}\#\tilde{\nu}; \nabla h^{-1}\#\nu^*) = \mathrm{KL}(\tilde{\mu}; \mu^*).\end{aligned}$$

In terms of W_2-distance, when applying the same mapping ∇h^{-1} to two measures, it can be shown that

$$\mathrm{W}_2(\tilde{\mu}; \mu^*) \leq M\,\mathrm{W}_2(\nabla h\#\tilde{\mu}; \nabla h\#\mu^*) = M\,\mathrm{W}_2(\tilde{\nu}; \nu^*).$$

Therefore, by sampling from ν^*, and then multiplying the obtained samples by M, we obtain the following convergence results:

- After $N^{\mathrm{KL}} = \tilde{\mathcal{O}}(LM^2d\epsilon^{-\frac{3}{2}})$ total iterations, we obtain $\mathrm{KL}(\tilde{\mu} - \mu^*) \leq \epsilon$.
- After $N^{\mathrm{TV}} = \tilde{\mathcal{O}}(LM^2d\epsilon^{-3})$ total iterations, we obtain $\|\tilde{\mu} - \mu^*\|_{\mathrm{TV}} \leq \epsilon$.
- After $N^{\mathrm{W}_2} = \tilde{\mathcal{O}}(LM^2d\left(\frac{\epsilon}{M}\right)^{-6}) = \tilde{\mathcal{O}}(LM^8d\epsilon^{-6})$ total iterations, we obtain $\mathrm{W}_2(\tilde{\mu}, \mu^*) \leq \epsilon$.

Finally, we make the following important observation. By modifying the parameters γ_k, τ_k, it is possible to mimic the above procedure by directly applying DL-ULA to μ^*. Suppose that we apply DL-ULA for sampling from $\mathrm{d}\nu^* = e^{g(y)}\,\mathrm{d}y$, where $g(y) = f(My)$, using parameters γ_k, n_k, τ_k. Let y_i be the iterates of some arbitrary outer iteration k, and let $x_i = My_i$ be their scaled version. The ULA iterates are:

$$\begin{cases} y_{i+1} = y_i + \gamma_i \nabla g(y_i) + \sqrt{2\tilde{\gamma}_i}g_i \\ x_{i+1} = My_{i+1} \end{cases}$$

Since $\nabla g(y_i) = M\nabla f(My_i)$, we can rewrite this scheme only in terms of $\{x_i\}$:

$$x_{i+1} = x_i + M^2\gamma_i \nabla f(x_i) + \sqrt{2M^2\gamma_i} g_i$$

Moreover, applying the projection step to y_i with parameter τ_k is the same as applying this projection to x_i with parameter $M\tau_k$.

Therefore, applying DL-ULA to ν^* using parameters n_k, γ_k, τ_k, and then multiplying the iterates by M is the same as directly applying DL-ULA to μ^* using parameters $n_k, M^2\gamma_k, M\tau_k$.

Overall, if we apply DL-ULA to a distribution μ^* such that $\mathbb{E}_{X\sim\mu^*}\left[\|X\|_2^2\right] \leq M^2$ using $n_k = LM^2dk^2e^{3k}$, $\gamma_k = \frac{1}{Ld}e^{-k}$ and $\tau_k = Mk$, then we can guarantee convergence rates of $\tilde{\mathcal{O}}(LM^2d\epsilon^{-\frac{3}{2}})$, $\tilde{\mathcal{O}}(LM^2d\epsilon^{-3})$ and $\tilde{\mathcal{O}}(LM^8d\epsilon^{-6})$ in KL divergence, TV-distance and W_2-distance respectively.

Finally, thanks to Lemma 5, we know that we can choose $M = \sqrt{\frac{2d(d+1)}{\eta^2} + M_\eta^2} = \mathcal{O}(d)$. Thus, plugging this value inside the convergence results above concludes the theorem.

C. Proof of Lemma 12

Proof. A similar result has been shown in (Brosse et al., 2017) (Proposition 5) for W_1 distance, and it is only a matter of trivial technicalities to extend their result to W_2 distance. Since the full proof requires to introduce several concepts that are out of the scope of this paper, we only present the required modifications that allow us to extend the result from W_1- to W_2-distance.

Using (Villani, 2009), Theorem 6.15, we have:

$$W_2^2(\mu_\lambda, \mu^*) \leq 2\int_{\mathbb{R}^d} \|x\|_2^2 |\mu^*(x) - \mu_\lambda(x)| dx = A + B \tag{45}$$

where

$$A = \int_{K^c} \|x\|_2^2 \mu_\lambda(x) dx \; , \; B = \left(1 - \frac{\int_K e^{-f}}{\int_{\mathbb{R}^d} e^{-f_\lambda}}\right) \int_K \|x\|_2^2 \mu^*(x) dx \tag{46}$$

Following very closely the proof in (Brosse et al., 2017) (equations 48 to 51), we can easily obtain:

$$A \leq \Delta_1^{-1} \sum_{i=0}^{d-1} \left(\frac{d}{r}\sqrt{\frac{\pi\lambda}{2}}\right)^{d-i} \left(R^2 + 2R\sqrt{\lambda(d-i+2)} + \lambda(d-i+2)\right). \tag{47}$$

Therefore, for $\lambda \leq \frac{r^2}{2\pi d^2}$,

$$A \leq \Delta_1^{-1}\sqrt{2\pi\lambda} d r^{-1} \left(R^2 + 2Rr\sqrt{\frac{3}{2d\pi}} + r^2\frac{3}{2d\pi}\right). \tag{48}$$

Moreover, it is also shown in (Brosse et al., 2017) (equations 17, 30, 42) that $\left(1 - \frac{\int_K e^{-f}}{\int_{\mathbb{R}^d} e^{-f_\lambda}}\right) \leq \Delta_1^{-1} 2\pi\lambda d r^{-1}$, which implies:

$$B \leq \Delta_1^{-1}\sqrt{2\pi\lambda} d r^{-1} R^2 \tag{49}$$

We thus showed that $W_2(\mu_\lambda, \mu^*) \leq C\sqrt{d}\lambda^{\frac{1}{4}}$ for some $C > 0$ depending on D, r, Δ_1.

□

D. Convergence rate of HULA for sampling from a distribution over a bounded domain

The proof of Theorem 13 is very similar to the one for DL-ULA. Before presenting it, we will need an auxiliary Lemma, showing the light tail property of the distributions μ_λ.

Lemma 19. *For $\lambda \leq \frac{r^2}{8d^2}$, the distribution μ_λ as defined in equation* (18) *satisfies*

$$Pr_{X\sim\mu_\lambda}(\|X\|_2 \geq R) \leq \sigma e^{-\frac{R}{D}}$$

for some scalar $\sigma > 0$ and any $R > 0$, where D is the diameter of the constraint set Ω.

Proof. Suppose first that $R \geq 2D$. Then,

$$\begin{aligned}
\Pr_{X\sim\mu_\lambda}[\|X\|_2 \geq R] &= \frac{\int_{\mathrm{B}(0,R)^c} e^{-f(x)-\frac{1}{2\lambda}\|x-\mathrm{proj}_\Omega(x)\|_2^2}\,\mathrm{d}x}{\int_\Omega e^{-f(x)}\,\mathrm{d}x + \int_{\Omega^c} e^{-f(x)-\frac{1}{2\lambda}\|x-\mathrm{proj}_\Omega(x)\|_2^2}\,\mathrm{d}x} \\
&\leq \Delta_1 \frac{\int_{\mathrm{B}(0,R)^c} e^{-\frac{1}{2\lambda}(\|x\|_2-D)^2}\,\mathrm{d}x}{\mathrm{Vol}(\Omega)} \\
&\leq \Delta_1 \mathrm{Vol}(\Omega)^{-1} \int_R^\infty u^{d-1} e^{-\frac{1}{2\lambda}(u-D)^2}\,\mathrm{d}u \\
&= \Delta_1 \mathrm{Vol}(\Omega)^{-1} d\mathrm{Vol}(B(0,1)) \int_R^\infty u^{d-1} e^{-\frac{1}{2\lambda}(u-D)^2}\,\mathrm{d}u \\
&\leq \Delta_1 d \frac{\mathrm{Vol}(B(0,1))}{\mathrm{Vol}(B(0,r))} D^{d-1} \int_{R-D}^\infty (u+D)^{d-1} e^{-\frac{1}{2\lambda}u^2}\,\mathrm{d}u \\
&\leq \Delta_1 d \frac{1}{r^d} \int_{R-D}^\infty (2u)^{d-1} e^{-\frac{1}{2\lambda}u^2}\,\mathrm{d}u \qquad \text{since } u \geq R-D \geq D \\
&\leq \Delta_1 d \frac{1}{r^d} 2^{d-1} \int_{\frac{1}{2\lambda}(R-D)^2}^\infty (2v\lambda)^{\frac{d-1}{2}} e^{-v} \sqrt{\frac{\lambda}{2v}}\,\mathrm{d}u \qquad (v = \frac{1}{2\lambda}u^2) \\
&\leq \Delta_1 d \frac{2^{\frac{3}{2}d-3}\lambda^{\frac{d}{2}}}{r^d} \Gamma\left(\frac{d}{2}; \frac{1}{2\lambda}(R-D)^2\right) \qquad \text{where } \Gamma(s;x) \text{ is the incomplete Gamma function} \\
&\leq \Delta_1 d \frac{2^{-3}}{d^d} \frac{d}{2} \left(\frac{1}{2\lambda}(R-D)^2\right)^{\frac{d}{2}} e^{-\frac{1}{2\lambda}(R-D)^2} \qquad \text{since for } x \geq s,\, \Gamma(s;x) \leq s x^s e^{-x},\, \lambda \leq \frac{r^2}{8d^2} \\
&\leq \left(\Delta_1^{\frac{1}{d^2}} 2^{\frac{-4}{d^2}} d^{\frac{2}{d^2}} \left(\frac{(R-D)^2}{2\lambda d^2}\right)^{\frac{1}{2d}} e^{-\frac{1}{2\lambda d^2}(R-D)^2}\right)^{d^2} \\
&\leq \left(c_d e^{-\frac{1}{\sqrt{2\lambda}d}(R-D)}\right)^{d^2} \qquad \text{since } xe^{-x^2} \leq e^{-x}\ \forall x \geq 0 \text{ and } \frac{1}{2\lambda d^2}(R-D)^2 \geq 1
\end{aligned}$$

where in the last line, $c_d = \Delta_1^{\frac{1}{d^2}} 2^{\frac{-4}{d^2}} d^{\frac{2}{d^2}}$. If $c_d e^{-\frac{\sqrt{\frac{1}{2\lambda}}}{d}(R-D)} \geq 1$, then, this does not provide a useful bound, and we can always write $\Pr_{X\sim\mu_\lambda}[\|X\|_2 \geq R] \leq 1 \leq c_d e^{-\frac{\sqrt{\frac{1}{2\lambda}}}{d}(R-D)}$. On the other hand, if $c_d e^{-\frac{\sqrt{\frac{1}{2\lambda}}}{d}(R-D)} \leq 1$, then we have $\Pr_{X\sim\mu_\lambda}[\|X\|_2 \geq R] \leq \left(c_d e^{-\frac{\sqrt{\frac{1}{2\lambda}}}{d}(R-D)}\right)^{d^2} \leq c_d e^{-\frac{\sqrt{\frac{1}{2\lambda}}}{d}(R-D)}$.

Therefore, we can write:

$$\begin{aligned}
\Pr_{X\sim\mu_\lambda}[\|X\|_2 \geq R] &\leq c_d e^{-\frac{\sqrt{\frac{1}{2\lambda}}}{d}(R-D)} \\
&\leq c_d e^{-2(\frac{R}{D}-1)} \qquad \text{since } \lambda \leq \frac{r^2}{8d^2} \leq \frac{D^2}{8d^2} \\
&\leq \max(1, c_d) e^2 e^{-\frac{R}{D}}.
\end{aligned}$$

Moreover, in the case $R \leq 2D$, we have $\max(1, c_d)e^2 e^{-\frac{R}{D}} \geq 1 \geq \Pr_{X\sim\mu_\lambda}[\|X\|_2 \geq R]$. We thus showed the result with $\sigma = \max(1, c_d)e^2$. Note that although c_d depends on d, it is bounded and converges to 1 as $d \to \infty$, thus it does not involve any asymptotic dependence in d.

□

Using this Lemma, we can now prove our convergence result for DL-MYULA (Theorem 13).

Proof. Let denote $\mu_k \equiv \mu_{\lambda_k}$ the target distributions of the ULA iterations at outer iteration $k \geq 1$, and μ_{init} the initial distribution. It is straightforward to show that the distributions μ_k are L_k-smooth with $L_k = L + \frac{1}{\lambda_k}$.

The proof goes exactly the same way as for Theorem 10. We will show by induction that $\forall k \geq 1$,

$$\|\tilde{\mu}_k - \mu_k\|_{\mathrm{TV}} \leq u_k e^{-k} + \sqrt{2 + \frac{16d^2}{Lr^2}} e^{-2k}$$

where $\{u_k\}_{k\geq 1}$ is defined $u_1 = \sqrt{e}\left(\mathrm{W}_2(\mu_{init}, \mu^* + C_\Omega d^{\frac{1}{4}})\right)$ and the recurrence relation

$$u_k = 4D\sqrt{eu_{k-1}} + 4D\sqrt{\sigma} + \frac{2C_\Omega d^{\frac{1}{4}}(\sqrt{e}+1)}{k^2} + \frac{2\sqrt{2}d^{\frac{1}{2}}}{L} + \sigma.$$

For any $k \geq 1$, we have:

$$\begin{aligned}
\|\bar{\mu}_k - \mu_k\|_{\mathrm{TV}} &\leq \sqrt{2\,\mathrm{KL}(\bar{\mu}_k; \mu_k)} \qquad \text{(Pinsker's inequality)} \\
&\leq \sqrt{\frac{W_2^2(\tilde{\mu}_{k-1}, \mu_k)}{\gamma_k n_k} + 2L_k d\gamma_k} \\
&\leq \frac{W_2(\tilde{\mu}_{k-1}, \mu_k)}{\sqrt{\gamma_k n_k}} + \sqrt{2L_k d\gamma_k} \\
&\leq \frac{W_2(\tilde{\mu}_{k-1}, \mu_{k-1})}{\sqrt{\gamma_k n_k}} + \frac{W_2(\mu_{k-1}, \mu^*)}{\sqrt{\gamma_k n_k}} + \frac{W_2(\mu_k, \mu^*)}{\sqrt{\gamma_k n_k}} + \sqrt{2L_k d\gamma_k}
\end{aligned} \tag{50}$$

For the second and third term, we can use Lemma 12 and the values of λ_k to show that $\forall k \geq 1$,

$$W_2(\mu_k, \mu^*) \leq C_\Omega d^{\frac{1}{4}} e^{-\frac{k}{2}} \tag{51}$$

For the first term, we use Lemma 14 with $R = Dk$ together with the fact that $\Pr_{X\sim\tilde{\mu}_{k-1}}(\|X\|_2 \geq Dk) = 0$ thanks to the projection step, and the light tail property of μ_k to obtain

$$W_2^2(\tilde{\mu}_{k-1}, \mu_{k-1}) \leq 4D^2k^2\|\tilde{\mu}_{k-1} - \mu_{k-1}\|_{\mathrm{TV}} + 4D^2k^2\sigma e^{-k+1}. \tag{52}$$

By replacing (51) and (52) in (50), and using the recursion hypothesis for $\|\tilde{\mu}_{k-1} - \mu_{k-1}\|_{\mathrm{TV}}$, we obtain

$$\|\bar{\mu}_k - \mu_k\|_{\mathrm{TV}} \leq \left(2D\sqrt{eu_{k-1}} + 2D\sqrt{\sigma} + \frac{C_\Omega d^{\frac{1}{4}}(\sqrt{e}+1)}{k^2} + \frac{\sqrt{2}d^{\frac{1}{2}}}{L}\right) e^{-k} + \sqrt{2 + \frac{16d^2}{Lr^2}} e^{-2k} \tag{53}$$

Similarly as for DL-ULA, and using Lemma 19 we can show that

$$\|\tilde{\mu}_k - \mu_k\|_{\mathrm{TV}} \leq 2\|\bar{\mu}_k - \mu_k\|_{\mathrm{TV}} + \sigma e^{-k}$$

Thus, using the recurrence relation for u_k, we have

$$\|\tilde{\mu}_k - \mu_k\|_{\mathrm{TV}} \leq u_k e^{-k} + \sqrt{2 + \frac{16d^2}{Lr^2}} e^{-2k} \tag{54}$$

as required to show the induction property. The case for $k = 1$ is shown analogous to DL-ULA.

Finally, in order to relate $\tilde{\mu}_k$ to the target distribution μ^*, we use the result shown in (Bubeck et al., 2018) that $\|\mu_\lambda - \mu^*\|_{\mathrm{TV}} \leq C' d\sqrt{\lambda}$ for some constant $C' > 0$ and $\forall \lambda < \frac{r^2}{8d^2}$.

We can easily show that the sequence $\{u_k\}_{k\geq 1}$ increasingly converges to the following limit:

$$U = 8eD^2 + 4D\sqrt{\sigma} + \frac{2C_\Omega d^{\frac{1}{4}}(\sqrt{e}+1)}{k^2} + \frac{2\sqrt{2}d^{\frac{1}{2}}}{L} + \sigma + 4D\sqrt{4eD^2 + 2D\sqrt{\sigma} + \frac{C_\Omega d^{\frac{1}{4}}(\sqrt{e}+1)}{k^2} + \frac{\sqrt{2}d^{\frac{1}{2}}}{L} + \frac{\sigma}{2}}$$
$$= \mathcal{O}(\sqrt{d}).$$

We thus have for all $k \geq 1$:

$$\|\tilde{\mu}_k - \mu^*\|_{\mathrm{TV}} \leq (U + C'\sqrt{d})e^{-k} + \sqrt{2 + \frac{16d^2}{Lr^2}}e^{-2k}$$

Therefore, after $K^{\mathrm{TV}} = \log\left(\frac{2\max\left(U + C'\sqrt{d}, \left(2 + \frac{16d^2}{Lr^2}\right)^{\frac{1}{4}}\right)}{\epsilon}\right)$ iterations, we have $\|\tilde{\mu}_k - \mu^*\|_{\mathrm{TV}} \leq \epsilon$. The total number of iterations required is $N^{\mathrm{TV}} = \tilde{\mathcal{O}}(Ld^{3.5}\epsilon^{-5})$.

Finally, using $\mathrm{W}_2^2(\tilde{\mu}_k, \mu^*) \leq 4D^2k^2\|\tilde{\mu}_k - \mu^*\|_{\mathrm{TV}}$, we can obtain a similar convergence result, i.e., after $K^{\mathrm{W}_2} = \log\left(\frac{8D^2\max\left(U + C'\sqrt{d}, \left(2 + \frac{16d^2}{Lr^2}\right)^{\frac{1}{4}}\right)}{\epsilon}\right)$ iterations, we have $\mathrm{W}_2(\tilde{\mu}_k, \mu^*) \leq \epsilon \log^2(K)$. The total number of iterations required is $N^{\mathrm{W}_2} = \tilde{\mathcal{O}}(Ld^{3.5}\epsilon^{-10})$.

$\square$