

Appendix

Table of Contents

A	Proofs of the lemmas used in Theorem 1	12
A.1	Proof of Lemma 1	12
A.2	Proof of Lemma 2	13
A.3	Proof of Lemma 3	13
A.4	Different forms of the Gronwall inequality	14
B	Proof of Proposition 1	16
C	(Weakly) convex potentials: what is known and what we can hope for	16
D	Penalized Gradient Flow	17
D.1	Proof of Theorem 2	17
D.2	Proof of Lemma 6	18
D.3	Proof of Lemma 7	19
E	Examples of functions satisfying condition A(D, q)	20
E.1	Locally strongly convex functions	20
E.2	Cubic function $f(\mathbf{x}) = \ \mathbf{x} - \mathbf{x}_*\ _2^3$	20
E.3	Power function $f(\mathbf{x}) = \ \mathbf{x} - \mathbf{x}_*\ _2^a$	20

A Proofs of the lemmas used in Theorem 1

A.1 Proof of Lemma 1

We denote by $D_{\text{KL}}(\pi_\gamma \|\pi_{\tilde{\gamma}})$ the Kullback-Leibler divergence between the distributions π_γ and $\pi_{\tilde{\gamma}}$. Since $\pi_{\tilde{\gamma}}$ is $(m + \tilde{\gamma})$ -strongly log-concave, the transportation cost inequality (Gozlan and Léonard, 2010, Corollary 7.2) yields

$$W_2^2(\pi_{\tilde{\gamma}}, \pi_\gamma) \leq \frac{2}{m + \tilde{\gamma}} D_{\text{KL}}(\pi_\gamma \|\pi_{\tilde{\gamma}}). \quad (11)$$

Let us denote by c_γ the logarithm of the normalizing constant for π_γ so that $\pi_\gamma(\boldsymbol{\theta}) = \exp(-f(\boldsymbol{\theta}) - (1/2)\gamma\|\boldsymbol{\theta}\|_2^2 + c_\gamma)$. Similarly, we denote by $c_{\tilde{\gamma}}$ the logarithm of the normalizing constant of $\pi_{\tilde{\gamma}}$. This readily yields

$$\begin{aligned} D_{\text{KL}}(\pi_\gamma \|\pi_{\tilde{\gamma}}) &= \int_{\mathbb{R}^p} \pi_\gamma(\boldsymbol{\theta}) \log \left(\frac{\pi_\gamma(\boldsymbol{\theta})}{\pi_{\tilde{\gamma}}(\boldsymbol{\theta})} \right) d\boldsymbol{\theta} \\ &= \int_{\mathbb{R}^p} \pi_\gamma(\boldsymbol{\theta}) \left(1/2(\tilde{\gamma} - \gamma)\|\boldsymbol{\theta}\|_2^2 + c_\gamma - c_{\tilde{\gamma}} \right) d\boldsymbol{\theta} \\ &= 1/2(\tilde{\gamma} - \gamma)\mu_2(\pi_\gamma) + c_\gamma - c_{\tilde{\gamma}}. \end{aligned}$$

Using the inequality $e^{-x} \leq 1 - x + (1/2)x^2$ for all $x > 0$ implies the following upper bound on $c_\gamma - c_{\tilde{\gamma}}$:

$$\begin{aligned} c_\gamma - c_{\tilde{\gamma}} &= \log \left(\int_{\mathbb{R}^p} \pi_\gamma(\boldsymbol{\theta}) \exp(1/2(\gamma - \tilde{\gamma})\|\boldsymbol{\theta}\|_2^2) d\boldsymbol{\theta} \right) \\ &\leq \log \left(1 + 1/2(\gamma - \tilde{\gamma})\mu_2(\pi_\gamma) + 1/8(\gamma - \tilde{\gamma})^2\mu_4(\pi_\gamma) \right). \end{aligned}$$

Since $\log(1 + x) \leq x$ for $x > -1$ we get

$$D_{\text{KL}}(\pi_\gamma \|\pi_{\tilde{\gamma}}) \leq 1/8(\gamma - \tilde{\gamma})^2\mu_4(\pi_\gamma).$$

Since $m + \gamma \geq 0$, the distribution π_γ is log-concave. Thus, in view of (Dalalyan et al., 2019, Remark 3), we have the inequality $\mu_4(\pi_\gamma) \leq 442\mu_2^2(\pi_\gamma)$. Finally, combining these bounds with (11), we get

$$W_2(\pi_{\tilde{\gamma}}, \pi_\gamma) \leq \sqrt{\frac{2}{m + \tilde{\gamma}}} \times \frac{(\tilde{\gamma} - \gamma)\mu_4^{1/2}(\pi_\gamma)}{\sqrt{8}} \leq \frac{11\mu_2(\pi_\gamma)}{\sqrt{m + \tilde{\gamma}}}(\tilde{\gamma} - \gamma).$$

This completes the proof of the lemma.

A.2 Proof of Lemma 2

For $k \in \mathbb{N} \cup \{0\}$, define

$$h_k(\gamma) = \int_{\mathbb{R}^p} \|\boldsymbol{\theta}\|_2^k \exp(-f(\boldsymbol{\theta}) - \gamma\|\boldsymbol{\theta}\|_2^2/2) d\boldsymbol{\theta}.$$

If $\pi \in \mathcal{P}_k(\mathbb{R}^p)$ then the function h_k is continuous on $[0; +\infty)$. Indeed, if the sequence $\{\gamma_n\}_n$ converges γ_0 , when $n \rightarrow +\infty$, then the function $\|\boldsymbol{\theta}\|_2^k \exp(-f(\boldsymbol{\theta}) - (1/2)\gamma_n\|\boldsymbol{\theta}\|_2^2)$ is upper-bounded by $\|\boldsymbol{\theta}\|_2^k \exp(-f(\boldsymbol{\theta}))$. Thus in view of the dominated convergence theorem, we can interchange the limit and the integral. Since, by definition,

$$\mu_k(\pi_\gamma) = \frac{h_k(\gamma)}{h_0(\gamma)},$$

we get the continuity of $\mu_2(\pi_\gamma)$ and $\mu_4(\pi_\gamma)$. Let us now prove that $h_k(t)$ is continuously differentiable, when $\pi \in \mathcal{P}_{k+2}(\mathbb{R}^p)$. The integrand function in the definition of h_k is a continuously differentiable function with respect to t . In addition, its derivative is continuous and is as well integrable on \mathbb{R}^p , as we supposed that π has the $(k + 2)$ -th moment. Therefore, Leibniz integral rule yields the following

$$h'_k(\gamma) = -\frac{1}{2} \int_{\mathbb{R}^p} \|\boldsymbol{\theta}\|_2^{k+2} \exp(-f(\boldsymbol{\theta}) - \gamma\|\boldsymbol{\theta}\|_2^2/2) d\boldsymbol{\theta} = -\frac{1}{2} h_{k+2}(\gamma).$$

The latter yields the smoothness of h_k . Finally, in order to prove the monotony of $\mu_2(\pi_\gamma)$, we will simply calculate its derivative

$$\begin{aligned} (\mu_2(\pi_\gamma))' &= -\frac{1}{2h_0(\gamma)} h_4(\gamma) - \frac{h'_0(\gamma)}{h_0(\gamma)^2} h_2(\gamma) \\ &= -\frac{1}{2} \mu_4(\pi_\gamma) + \frac{h_2^2(\gamma)}{2h_0(\gamma)^2} \\ &= \frac{1}{2} (\mu_2^2(\pi_\gamma) - \mu_4(\pi_\gamma)). \end{aligned}$$

Since the latter is always negative, this completes the proof of the lemma.

A.3 Proof of Lemma 3

From the definition of Wasserstein distance, we have

$$W_2(\nu_{t+\delta}^{\text{PLD}}, \mathbf{Q}_{t,\delta}) \leq \|\tilde{\mathbf{L}}_{t+\delta} - \mathbf{L}_{t+\delta}^{\text{PLD}}\|_{\mathbb{L}_2}.$$

In view of the definition of the process $\tilde{\mathbf{L}}$, we can write

$$\tilde{\mathbf{L}}_{t+\delta} - \mathbf{L}_{t+\delta}^{\text{PLD}} = \int_t^{t+\delta} \left(\nabla f(\mathbf{L}_s^{\text{PLD}}) - \nabla f(\tilde{\mathbf{L}}_s) + \alpha(s)\mathbf{L}_s^{\text{PLD}} - \alpha(t)\tilde{\mathbf{L}}_s \right) ds.$$

Therefore we have

$$\|\tilde{\mathbf{L}}_{t+\delta} - \mathbf{L}_{t+\delta}^{\text{PLD}}\|_{\mathbb{L}_2} \leq \underbrace{\left\| \int_t^{t+\delta} \left(\nabla f(\mathbf{L}_s^{\text{PLD}}) - \nabla f(\tilde{\mathbf{L}}_s) \right) ds \right\|_{\mathbb{L}_2}}_{:=T_1} + \underbrace{\left\| \int_t^{t+\delta} \left(\alpha(s)\mathbf{L}_s^{\text{PLD}} - \alpha(t)\tilde{\mathbf{L}}_s \right) ds \right\|_{\mathbb{L}_2}}_{:=T_2}.$$

Now let us analyze these two terms separately. We start with T_1 :

$$\begin{aligned} \|T_1\|_{\mathbb{L}_2} &= \left\| \int_t^{t+\delta} \left(\nabla f(\mathbf{L}_s^{\text{PLD}}) - \nabla f(\tilde{\mathbf{L}}_s) \right) ds \right\|_{\mathbb{L}_2} \\ &\leq \int_t^{t+\delta} \left\| \nabla f(\mathbf{L}_s^{\text{PLD}}) - \nabla f(\tilde{\mathbf{L}}_s) \right\|_{\mathbb{L}_2} ds \\ &\leq M \int_t^{t+\delta} \left\| \mathbf{L}_s^{\text{PLD}} - \tilde{\mathbf{L}}_s \right\|_{\mathbb{L}_2} ds. \end{aligned}$$

These are due to the Minkowskii inequality and the Lipschitz continuity of the gradient. In order to bound the second term T_2 , we will add and subtract the term $\alpha(t+s)\tilde{\mathbf{L}}_{t+s}$. Similar to the case above, we get the following upper bound:

$$\begin{aligned} \|T_2\|_{\mathbb{L}_2} &\leq \int_t^{t+\delta} \alpha(s) \left\| \mathbf{L}_s^{\text{PLD}} - \tilde{\mathbf{L}}_s \right\|_{\mathbb{L}_2} ds + \int_t^{t+\delta} |\alpha(s) - \alpha(t)| \left\| \tilde{\mathbf{L}}_s \right\|_{\mathbb{L}_2} ds \\ &= \int_0^\delta \alpha(t+s) \left\| \mathbf{L}_{t+s}^{\text{PLD}} - \tilde{\mathbf{L}}_{t+s} \right\|_{\mathbb{L}_2} ds + \int_0^\delta |\alpha(t+s) - \alpha(t)| \left\| \tilde{\mathbf{L}}_{t+s} \right\|_{\mathbb{L}_2} ds. \end{aligned}$$

Recall that $\tilde{\mathbf{L}}_{t+s}$ is the solution of Langevin SDE with an $(m + \alpha(t))$ -strongly convex potential function, and $Q_{t,s}$ is its distribution on \mathbb{R}^p . Thus, the triangle inequality yields

$$\begin{aligned} \left\| \tilde{\mathbf{L}}_{t+s} \right\|_{\mathbb{L}_2} &= W_2(Q_{t,s}, \delta_0) \leq W_2(Q_{t,s}, \pi_{\alpha(t)}) + W_2(\pi_{\alpha(t)}, \delta_0) \\ &\leq W_2(\nu_t^{\text{PLD}}, \pi_{\alpha(t)}) \exp(-ms - \alpha(t)s) + \sqrt{\mu_2(\pi_{\alpha(t)})} \\ &\leq W_2(\nu_t^{\text{PLD}}, \pi_{\alpha(t)}) + \sqrt{\mu_2(\pi_{\alpha(t)})} := V_t. \end{aligned}$$

Summing up, we have

$$\left\| \mathbf{L}_{t+\delta}^{\text{PLD}} - \tilde{\mathbf{L}}_{t+\delta} \right\|_{\mathbb{L}_2} \leq \int_0^\delta (M + \alpha(t+s)) \left\| \mathbf{L}_{t+s}^{\text{PLD}} - \tilde{\mathbf{L}}_{t+s} \right\|_{\mathbb{L}_2} ds + \tilde{\alpha}_t(\delta) V_t,$$

where $\tilde{\alpha}_t(\delta)$ is an auxiliary function defined as

$$\tilde{\alpha}_t(\delta) := \int_0^\delta |\alpha(t+s) - \alpha(t)| ds.$$

Now let us define $\Phi(s) = \left\| \mathbf{L}_{t+s}^{\text{PLD}} - \tilde{\mathbf{L}}_{t+s} \right\|_{\mathbb{L}_2}$. The last inequality can be rewritten as

$$\Phi(\delta) \leq \int_0^\delta (M + \alpha(t+s)) \Phi(s) ds + \tilde{\alpha}_t(\delta) V_t.$$

The (integral form of the) Gronwall inequality, lemma 5, implies that

$$\begin{aligned} \Phi(\delta) &\leq V_t \int_0^\delta \tilde{\alpha}_t(s) (M + \alpha(t+s)) e^{\int_s^\delta (M + \alpha(t+u)) du} ds + \tilde{\alpha}_t(\delta) V_t \\ &= V_t \int_0^\delta \tilde{\alpha}_t'(s) e^{\int_s^\delta (M + \alpha(t+u)) du} ds \\ &\leq V_t \tilde{\alpha}_t(\delta) \exp \left\{ M\delta + \int_0^\delta \alpha(t+u) du \right\}. \end{aligned}$$

This completes the proof.

A.4 Different forms of the Gronwall inequality

In this section, we provide two forms of the Gronwall inequality that are used in the present work. For the sake of the self-containedness, the proofs of these inequalities are also provided.

Lemma 4 (Differential form). *Let $A : [a, b] \rightarrow \mathbb{R}$ and $B : [a, b] \rightarrow \mathbb{R}$ be two functions. If the function $\Phi : [a, b] \rightarrow \mathbb{R}$ satisfies the recursive differential inequality*

$$\Phi'(x) \leq A(x)\Phi(x) + B(x), \quad \forall x \in [a, b], \quad (12)$$

then it also satisfies the inequality

$$\Phi(x) \leq \Phi(a) \exp \left\{ \int_a^x A(z) dz \right\} + \int_a^x B(s) \exp \left\{ \int_s^x A(z) dz \right\} ds, \quad \forall x \in [a, b].$$

Proof. To ease notation, we set $E(x) = \exp\{-\int_a^x A(z) dz\}$. By multiplying both sides of (12) by $E(x)$, we get

$$\left(\Phi(x)E(x) \right)' \leq B(x)E(x), \quad \forall x \in [a, b].$$

Integrating this inequality, we arrive at

$$\Phi(x)E(x) \leq \Phi(a)E(a) + \int_a^x B(s)E(s) ds.$$

Dividing both sides of this inequality by $E(x) > 0$ and taking into account that $E(a) = 1$, we get the claim of the lemma. \square

Lemma 5 (Integral form). *Let $A : [a, b] \rightarrow [0, +\infty)$ and $B : [a, b] \rightarrow \mathbb{R}$ be two functions. If the function $\Phi : [a, b] \rightarrow \mathbb{R}$ satisfies the recursive integral inequality*

$$\Phi(x) \leq \int_a^x A(s)\Phi(s) ds + B(x), \quad \forall x \in [a, b],$$

then it also satisfies the inequality

$$\Phi(x) \leq \int_a^x A(s)B(s) \exp \left\{ \int_s^x A(z) dz \right\} ds + B(x), \quad \forall x \in [a, b]. \quad (13)$$

Proof. We set

$$\Psi(x) = \exp \left\{ - \int_a^x A(z) dz \right\} \int_a^x A(s)\Phi(s) ds.$$

We have

$$\begin{aligned} \Psi'(x) &= -A(x)\Psi(x) + \exp \left\{ - \int_a^x A(z) dz \right\} A(x)\Phi(x) \\ &\leq -A(x)\Psi(x) + \exp \left\{ - \int_a^x A(z) dz \right\} A(x) \left(\int_a^x A(s)\Phi(s) ds + B(x) \right) \\ &= -A(x)\Psi(x) + A(x)\Psi(x) + A(x)B(x) \exp \left\{ - \int_a^x A(z) dz \right\}. \end{aligned}$$

Therefore,

$$\Psi(x) \leq \Psi(a) + \int_a^x A(s)B(s) \exp \left\{ - \int_a^s A(z) dz \right\} ds.$$

Replacing Ψ by its expression and using the fact that $\Psi(a) = 0$, we get

$$\exp \left\{ - \int_a^x A(z) dz \right\} \int_a^x A(s)\Phi(s) ds \leq \int_a^x A(s)B(s) \exp \left\{ - \int_a^s A(z) dz \right\} ds.$$

This implies that

$$\int_a^x A(s)\Phi(s) ds \leq \int_a^x A(s)B(s) \exp \left\{ \int_s^x A(z) dz \right\} ds.$$

Combining this inequality with (13), we get the claim of the lemma. \square

B Proof of Proposition 1

For the penalty factor $\alpha(t) = 1/(A + 2t)$, we get $\beta(t) = \int_0^t \alpha(s) ds = (1/2) \log(1 + (2/A)t)$. This implies that

$$\sqrt{\mu_2(\pi)} e^{-\beta(t)} + 11\mu_2(\pi)\sqrt{\alpha(t)} = \frac{\sqrt{A\mu_2(\pi)} + 11\mu_2(\pi)}{\sqrt{A + 2t}}.$$

Finally, the middle term in the right hand side of (2) takes the form

$$\begin{aligned} 11\mu_2(\pi) \int_0^t \frac{|\alpha'(s)|}{\sqrt{\alpha(s)}} e^{\beta(s)-\beta(t)} ds &= \frac{11\mu_2(\pi)}{\sqrt{A + 2t}} \int_0^t \frac{2}{A + 2s} ds \\ &= \frac{11\mu_2(\pi)}{\sqrt{A + 2t}} \log(1 + (2/A)t). \end{aligned}$$

Combining these relations, we get the claim of the proposition.

C (Weakly) convex potentials: what is known and what we can hope for

Many recent papers investigated the case of strongly convex potential; this case is now rather well understood. Let us briefly summarize here some facts and conjectures that can shed some light on the broader case of weakly convex potential. This might help to understand what can be expected to be proved in the framework studied in this work.

The ergodicity properties of the Langevin process are closely related to such notions of functional analysis as the spectral gap, the Poincaré and the log-Sobolev inequalities. Thus, the generator of a Markov semi-group associated with an m -strongly convex potential has a spectral-gap \mathcal{C}_{SG} at least equal to m . This property was exploited by Dalalyan (2017) to derive guarantees on the LMC algorithm. It is known that the spectral gap exists if and only if the invariant density satisfies the Poincaré inequality. Furthermore, the spectral gap is equal to the inverse of the Poincaré constant \mathcal{C}_{P} . Furthermore, distributions associated to m -strongly convex potentials satisfy the log-Sobolev inequality with the constant $\mathcal{C}_{\text{LS}} \leq 1/m$. This property was used by Durmus and Moulines (2019) to extend the guarantees to the Wasserstein-2 distance.

Note that the log-Sobolev inequality is stronger than the Poincaré inequality and $\mathcal{C}_{\text{P}} \leq \mathcal{C}_{\text{LS}}$. For m -strongly convex potentials, we have $\mathcal{C}_{\text{SG}}^{-1} = \mathcal{C}_{\text{P}} \leq \mathcal{C}_{\text{LS}} \leq 1/m$. Results in (Dalalyan, 2017; Durmus and Moulines, 2019) imply that in order to get a Wasserstein distance smaller than $\varepsilon\sqrt{p/m}$, it suffices to perform a number of LMC iterations proportional to $(M/m)^2\varepsilon^{-2}$, up to logarithmic factors. A formal proof of the fact that the same result holds for the densities satisfying the log-Sobolev inequality with constant $1/m$ (but which are not necessarily m -strongly log-concave) was given in (Vempala and Wibisono, 2019).

On the other hand, it was established by (Bobkov, 1999) that any log-concave distribution satisfies the Poincaré inequality. However, the Poincaré constant might depend on the dimension. In (Kannan et al., 1995), the authors conjectured that there is a universal constant $\mathcal{C}_{\text{KLS}} > 0$ such that for any log-concave distribution π on \mathbb{R}^p ,

$$\mathcal{C}_{\text{P}} \leq \mathcal{C}_{\text{KLS}} \|\mathbf{E}_{\pi}[\mathbf{X}\mathbf{X}^{\top}]\|_{\text{op}} := \mathcal{C}_{\text{KLS}}\mu_{\text{op}}(\pi). \quad (\text{KLS})$$

Despite important efforts made in recent years (see (Alonso-Gutiérrez and Bastero, 2015; Cattiaux and Guillin, 2018)), this conjecture is still unproved. Finally, in the recent paper (Chewi et al., 2020), Corollary 4 establishes that $W_2(\mu_t^{\text{LD}}, \pi) \leq \sqrt{2\mathcal{C}_{\text{P}}\chi^2(\nu_0|\pi)} e^{-t/\mathcal{C}_{\text{P}}}$. While the exponential in t convergence to zero is a very appealing property of this result, it comes with two shortcomings. To the previously mentioned difficulty of assessing the Poincaré constant, one has to add the challenging problem of finding a meaningful upper bound on the χ^2 -divergence between the initial distribution and the target.

What can we hope for in the light of the previous discussion? As shown in (Dalalyan, 2017, Lemma 5), for f satisfying (m, M) -SCGL, choosing $\nu_0 = \mathcal{N}(\mathbf{x}_*, M^{-1}\mathbf{I}_p)$ yields $\chi^2(\nu_0|\pi) \leq (M/m)^{p/2}$. In the case $m = 0$, it might be possible to replace m by $1/\mathcal{C}_{\text{P}}$ in this result. If in addition, we admit

inequality [KLS](#), then we get

$$\begin{aligned} W_2^2(\mu_t^{\text{LD}}, \pi) &\leq 2\mathcal{C}_P (M\mathcal{C}_P)^{p/2} e^{-2t/\mathcal{C}_P} \\ &\leq 2\mathcal{C}_{\text{KLS}}\mu_{\text{op}}(\pi) (M\mathcal{C}_{\text{KLS}}\mu_{\text{op}}(\pi))^{p/2} e^{-2t/\mathcal{C}_{\text{KLS}}\mu_{\text{op}}(\pi)}. \end{aligned}$$

This is, probably, the best upper bound one could hope for in the general log-concave setting by Langevin diffusion based algorithms. We see that it has three drawbacks as compared to our result stated in [Proposition 1](#). First, it requires the knowledge of a minimizer \mathbf{x}^* . Second, it involves the Lipschitz constant M of the gradient. Third, it is heavily based on \mathcal{C}_{KLS} , which might be very large.

D Penalized Gradient Flow

D.1 Proof of [Theorem 2](#)

We recall that for every $\gamma \in \mathbb{R}$, is given by $f_\gamma(\cdot) := f(\cdot) + \gamma\|\cdot\|_2^2/2$. We define \mathbf{x}_γ the minimum point of f_γ . In particular, $\mathbf{x}_0 = \mathbf{x}^*$. The triangle inequality yields

$$\|\mathbf{X}_t^{\text{PGF}} - \mathbf{x}_*\|_2 \leq \|\mathbf{X}_t^{\text{PGF}} - \mathbf{x}_{\alpha(t)}\|_2 + \|\mathbf{x}_{\alpha(t)} - \mathbf{x}_0\|_2 \quad (14)$$

for every $t > 0$. We will bound these two terms separately. [A\(A, q\)](#) for $\gamma = 0$ and $\tilde{\gamma} = \alpha(t)$ yields the following bound on the second term:

$$\|\mathbf{x}_{\alpha(t)} - \mathbf{x}_0\|_2 \leq D\alpha(t)^{1-q}\|\mathbf{x}_*\|_2^{1-q}.$$

To bound the first term of [\(14\)](#), we aim at obtaining a Gronwall-type inequality for the function

$$\phi(t) := \|\mathbf{X}_t^{\text{PGF}} - \mathbf{x}_{\alpha(t)}\|_2.$$

To this end, we consider an auxiliary stochastic process $\{\tilde{\mathbf{X}}_u : u \geq t\}$, defined as a solution of the following differential equation

$$d\tilde{\mathbf{X}}_u = -(\nabla f(\tilde{\mathbf{X}}_u) + \alpha(t)\tilde{\mathbf{X}}_u)du,$$

with the starting point $\tilde{\mathbf{X}}_t = \mathbf{X}_t$. This is in fact the gradient flow corresponding to the strongly-convex potential $f_{\alpha(t)}$. The triangle inequality yields

$$\phi(t+\delta) \leq \left\| \mathbf{X}_{t+\delta}^{\text{PGF}} - \tilde{\mathbf{X}}_{t+\delta} \right\|_2 + \left\| \tilde{\mathbf{X}}_{t+\delta} - \mathbf{x}_{\alpha(t)} \right\|_2 + \|\mathbf{x}_{\alpha(t)} - \mathbf{x}_{\alpha(t+\delta)}\|_2.$$

From the linear convergence of the gradient flow of an $\alpha(t)$ -strongly convex function, we get the following:

$$\left\| \mathbf{X}_{t+\delta}^{\text{PGF}} - \mathbf{x}_{\alpha(t)} \right\|_2 \leq \exp(-\delta\alpha(t)) \left\| \tilde{\mathbf{X}}_t - \mathbf{x}_{\alpha(t)} \right\|_2 = \exp(-\delta\alpha(t)) \phi(t).$$

In order to bound the distance between $\mathbf{x}_{\alpha(t)}$ and $\mathbf{x}_{\alpha(t+\delta)}$, we use again [A\(A, q\)](#) condition, thus

$$\|\mathbf{x}_{\alpha(t)} - \mathbf{x}_{\alpha(t+\delta)}\|_2 \leq \frac{D}{\alpha^q(t)}(\alpha(t) - \alpha(t+\delta))\|\mathbf{x}_*\|_2^{1-q}.$$

Thus we obtain a bound for $\phi(t+\delta)$, that depends linearly on $\phi(t)$:

$$\phi(t+\delta) \leq \left\| \mathbf{X}_{t+\delta}^{\text{PGF}} - \tilde{\mathbf{X}}_{t+\delta} \right\|_2 + e^{-\delta\alpha(t)}\phi(t) + \frac{D}{\alpha^q(t)}(\alpha(t) - \alpha(t+\delta))\|\mathbf{x}_*\|_2^{1-q}. \quad (15)$$

Let us subtract $\phi(t)$ from both sides of [\(15\)](#) and divide by δ :

$$\begin{aligned} \frac{\phi(t+\delta) - \phi(t)}{\delta} &\leq \frac{1}{\delta} \cdot \left\| \mathbf{X}_{t+\delta}^{\text{PGF}} - \tilde{\mathbf{X}}_{t+\delta} \right\|_2 + \frac{\exp(-\delta\alpha(t)) - 1}{\delta} \cdot \phi(t) \\ &\quad + \frac{D(\alpha(t) - \alpha(t+\delta))}{\delta\alpha^q(t)}\|\mathbf{x}_*\|_2^{1-q}. \end{aligned} \quad (16)$$

The next lemma provides an upper bound on $\left\| \mathbf{X}_{t+\delta}^{\text{PGF}} - \tilde{\mathbf{X}}_{t+\delta} \right\|_2$ showing that it is $o(\delta)$, when $\delta \rightarrow 0$.

Lemma 6. *Suppose f satisfies (m, M) -SCGL with $m = 0$. Then for every $t, \delta > 0$, and for every integrable function $\alpha : [t, t + \delta] \rightarrow [0, \infty)$,*

$$\|\widetilde{\mathbf{X}}_{t+\delta} - \mathbf{X}_{t+\delta}^{\text{PGF}}\|_2 \leq (\phi(t) + \|\mathbf{x}_{\alpha(t)}\|_2) \exp \left\{ M\delta + \int_0^\delta \alpha(t+u) du \right\} \int_0^\delta |\alpha(t+s) - \alpha(t)| ds.$$

The proof can be found in the Appendix D.2. When δ tends to 0, according to Lemma 6, the first term of the right-hand side of (16) vanishes. Thus, after passing to the limit, we are left with the following Gronwall-type inequality:

$$\phi'(t) \leq -\alpha(t)\phi(t) - \frac{D\alpha'(t)}{\alpha^q(t)} \cdot \|\mathbf{x}_*\|_2^{1-q}. \quad (17)$$

Here we tacitly used the fact that $\|\mathbf{x}_{\alpha(t+\delta)}\|_2 \leq \|\mathbf{x}_0\|_2$. Recalling that the function $\beta(t)$ is given by $\beta(t) = \int_0^t \alpha(s) ds$, one can rewrite (17) as

$$(\phi(t)e^{\beta(t)})' \leq -\frac{D\alpha'(t)e^{\beta(t)}}{\alpha^q(t)} \|\mathbf{x}_*\|_2^{1-q}.$$

Therefore we infer the following bound on $\phi(t)$:

$$\phi(t) \leq \phi(0)e^{-\beta(t)} - D\|\mathbf{x}_*\|_2^{1-q} \int_0^t \frac{\alpha'(s)}{\alpha^q(s)} e^{\beta(s)-\beta(t)} ds.$$

Combining this bound with (7), we obtain the inequality

$$\|\mathbf{X}_t^{\text{PGF}} - \mathbf{x}_0\|_2 \leq \|\mathbf{X}_0^{\text{PGF}} - \mathbf{x}_{\alpha(0)}\|_2 e^{-\beta(t)} - D\|\mathbf{x}_*\|_2^{1-q} \int_0^t \frac{\alpha'(s)}{\alpha^q(s)} e^{\beta(s)-\beta(t)} ds + D\|\mathbf{x}_*\|_2^{1-q} \alpha(t)^{1-q}.$$

Since the process $\mathbf{X}_t^{\text{PGF}}$ starts at point 0, $\|\mathbf{X}_0^{\text{PGF}} - \mathbf{x}_{\alpha(0)}\|_2 = \|\mathbf{x}_{\alpha(0)}\|_2$. The next lemma bounds $\|\mathbf{x}_{\alpha(0)}\|_2$.

Lemma 7. *The function $\gamma \mapsto \|\mathbf{x}_\gamma\|_2$ is a non-increasing continuous function on the interval $[0, \infty)$.*

Therefore, $\|\mathbf{x}_{\alpha(0)}\|_2 \leq \|\mathbf{x}_0\|_2 = \|\mathbf{x}_*\|_2$, which completes the proof of Theorem 2.

D.2 Proof of Lemma 6

From the definition of $\widetilde{\mathbf{X}}$, we can write

$$\widetilde{\mathbf{X}}_{t+\delta} - \mathbf{X}_{t+\delta}^{\text{PGF}} = \int_t^{t+\delta} \left(\nabla f(\mathbf{X}_s^{\text{PGF}}) - \nabla f(\widetilde{\mathbf{X}}_s) + \alpha(s)\mathbf{X}_s^{\text{PGF}} - \alpha(t)\widetilde{\mathbf{X}}_s \right) ds.$$

Therefore we have

$$\|\widetilde{\mathbf{X}}_{t+\delta} - \mathbf{X}_{t+\delta}^{\text{PGF}}\|_2 \leq \underbrace{\left\| \int_t^{t+\delta} \left(\nabla f(\mathbf{X}_s^{\text{PGF}}) - \nabla f(\widetilde{\mathbf{X}}_s) \right) ds \right\|_2}_{:=T_1} + \underbrace{\left\| \int_t^{t+\delta} \left(\alpha(s)\mathbf{X}_s^{\text{PGF}} - \alpha(t)\widetilde{\mathbf{X}}_s \right) ds \right\|_2}_{:=T_2}.$$

Now let us analyze these two terms separately. We start with T_1 :

$$\begin{aligned} \|T_1\|_2 &= \left\| \int_t^{t+\delta} \left(\nabla f(\mathbf{X}_s^{\text{PGF}}) - \nabla f(\widetilde{\mathbf{X}}_s) \right) ds \right\|_2 \\ &\leq \int_t^{t+\delta} \left\| \nabla f(\mathbf{X}_s^{\text{PGF}}) - \nabla f(\widetilde{\mathbf{X}}_s) \right\|_2 ds \\ &\leq M \int_t^{t+\delta} \|\mathbf{X}_s^{\text{PGF}} - \widetilde{\mathbf{X}}_s\|_2 ds. \end{aligned}$$

These are due to the Minkowskii inequality and the Lipschitz continuity of the gradient. In order to bound the second term T_2 , we will add and subtract the term $\alpha(t+s)\widetilde{\mathbf{X}}_{t+s}$. Similar to the case above, we get the following upper bound:

$$\begin{aligned}\|T_2\|_2 &\leq \int_t^{t+\delta} \alpha(s) \|\mathbf{X}_s^{\text{PGF}} - \widetilde{\mathbf{X}}_s\|_2 ds + \int_t^{t+\delta} |\alpha(s) - \alpha(t)| \|\widetilde{\mathbf{X}}_s\|_2 ds \\ &= \int_0^\delta \alpha(t+s) \|\mathbf{X}_{t+s}^{\text{PGF}} - \widetilde{\mathbf{X}}_{t+s}\|_2 ds + \int_0^\delta |\alpha(t+s) - \alpha(t)| \|\widetilde{\mathbf{X}}_{t+s}\|_2 ds.\end{aligned}$$

Recall that $\widetilde{\mathbf{X}}_{t+s}$ is the gradient flow of an $(m + \alpha(t))$ -strongly convex potential function. Thus, the triangle inequality yields

$$\begin{aligned}\|\widetilde{\mathbf{X}}_{t+s}\|_2 &\leq \|\widetilde{\mathbf{X}}_{t+s} - \mathbf{x}(t)\|_2 + \|\mathbf{x}(t)\|_2 \\ &\leq \|\widetilde{\mathbf{X}}_t - \mathbf{x}_t\|_2 \exp(-ms - \alpha(t)s) + \|\mathbf{x}_t\|_2 \\ &\leq \|\mathbf{X}_t^{\text{PGF}} - \mathbf{x}_t\|_2 + \|\mathbf{x}_t\|_2 := V_t.\end{aligned}$$

Summing up, we have

$$\|\mathbf{X}_{t+\delta}^{\text{PGF}} - \widetilde{\mathbf{X}}_{t+\delta}\|_2 \leq \int_0^\delta (M + \alpha(t+s)) \|\mathbf{X}_{t+s}^{\text{PGF}} - \widetilde{\mathbf{X}}_{t+s}\|_2 ds + \tilde{\alpha}_t(\delta) V_t,$$

where $\tilde{\alpha}_t(\delta)$ is an auxiliary function defined as

$$\tilde{\alpha}_t(\delta) := \int_0^\delta |\alpha(t+s) - \alpha(t)| ds.$$

Now let us define $\Phi(s)L := \|\mathbf{X}_{t+s}^{\text{PGF}} - \widetilde{\mathbf{X}}_{t+s}\|_2$. The last inequality can be rewritten as

$$\Phi(\delta) \leq \int_0^\delta (M + \alpha(t+s)) \Phi(s) ds + \tilde{\alpha}_t(\delta) V_t.$$

The (integral form of the) Gronwall inequality implies that

$$\begin{aligned}\Phi(\delta) &\leq V_t \int_0^\delta \tilde{\alpha}_t(s) (M + \alpha(t+s)) e^{\int_s^\delta (M + \alpha(t+u)) du} ds + \tilde{\alpha}_t(\delta) V_t \\ &= V_t \int_0^\delta \tilde{\alpha}_t'(s) e^{\int_s^\delta (M + \alpha(t+u)) du} ds \\ &\leq V_t \tilde{\alpha}_t(\delta) \exp \left\{ M\delta + \int_0^\delta \alpha(t+u) du \right\}.\end{aligned}$$

This completes the proof.

D.3 Proof of Lemma 7

Suppose that $\gamma_1 < \gamma_2$. We want to show that $\|\mathbf{x}_{\gamma_1}\|_2 > \|\mathbf{x}_{\gamma_2}\|_2$. Let us consider the function f_{γ_2} . We have that

$$f_{\gamma_2}(\mathbf{x}_{\gamma_2}) \leq f_{\gamma_2}(\mathbf{x}_{\gamma_1}) = f(\mathbf{x}_{\gamma_1}) + \gamma_2 \|\mathbf{x}_{\gamma_1}\|_2 / 2.$$

The definition of f_{γ_1} yields

$$\begin{aligned}f_{\gamma_2}(\mathbf{x}_{\gamma_2}) &\leq f_{\gamma_1}(\mathbf{x}_{\gamma_1}) + (\gamma_2 - \gamma_1) \|\mathbf{x}_{\gamma_1}\|_2 / 2 \\ &\leq f_{\gamma_1}(\mathbf{x}_{\gamma_2}) + (\gamma_2 - \gamma_1) \|\mathbf{x}_{\gamma_1}\|_2 / 2 \\ &\leq f_{\gamma_2}(\mathbf{x}_{\gamma_2}) + (\gamma_2 - \gamma_1) (\|\mathbf{x}_{\gamma_1}\|_2 - \|\mathbf{x}_{\gamma_2}\|_2) / 2.\end{aligned}$$

Here the second passage is valid, as \mathbf{x}_{γ_1} is the minimum point of f_{γ_1} . Since $\gamma_2 > \gamma_1$, the difference $\|\mathbf{x}_{\gamma_1}\|_2 - \|\mathbf{x}_{\gamma_2}\|_2$ is positive. Thus the monotony is proved.

To prove the continuity of the function we take a sequence γ_n that tends to γ_0 and show that $\mathbf{x}_{\gamma_n} \rightarrow \mathbf{x}_{\gamma_0}$. Assumption **A**(D, q) yields

$$\|\mathbf{x}_{\gamma_n} - \mathbf{x}_{\gamma_0}\|_2 \leq \frac{D}{\max(\gamma_n, \gamma_0)^q} |\gamma_n - \gamma_0| \|\mathbf{x}_*\|_2, \quad \forall n \in \mathbb{N}.$$

Since $q < 1$, the ratio of $|\gamma_n - \gamma_0|$ and $\max(\gamma_n, \gamma_0)^q$ tends to zero, when $n \rightarrow \infty$. This concludes the proof.

E Examples of functions satisfying condition $\mathbf{A}(D, q)$

In this section we consider several functions that are convex but not strongly convex and satisfy $\mathbf{A}(D, q)$ condition presented in Section 3.

E.1 Locally strongly convex functions

We prove that locally strongly convex functions satisfy $\mathbf{A}(D, 0)$. Recalling Lemma 7 we get that $\|\mathbf{x}_\gamma\|_2 \leq \|\mathbf{x}_*\|_2$. Thus we can consider the function only on $\mathcal{B}(0, \|\mathbf{x}_*\|_2)$. Since f is locally strongly convex, there exists m_* such that it is m_* -strongly convex in the ball $\mathcal{B}(0, \|\mathbf{x}_*\|_2)$. The latter means, that $f_{\tilde{\gamma}}$ is $(m_* + \tilde{\gamma})$ -strongly convex. Therefore (Nesterov, 2004)[Theorem 2.1.9] yields the following:

$$\|\mathbf{x}_\gamma - \mathbf{x}_{\tilde{\gamma}}\|_2 \leq \frac{1}{m_* + \tilde{\gamma}} \|\nabla f_{\tilde{\gamma}}(\mathbf{x}_\gamma) - \nabla f_{\tilde{\gamma}}(\mathbf{x}_{\tilde{\gamma}})\|_2.$$

Using the optimality condition for differentiable functions one gets $\nabla f_{\tilde{\gamma}}(\mathbf{x}_\gamma) = (\tilde{\gamma} - \gamma)\mathbf{x}_\gamma$ for all $\gamma \geq 0$. Therefore, for every $0 \leq \gamma < \tilde{\gamma}$, we obtain

$$\|\mathbf{x}_\gamma - \mathbf{x}_{\tilde{\gamma}}\|_2 \leq \frac{1}{m_* + \tilde{\gamma}} \|(\tilde{\gamma} - \gamma)\mathbf{x}_\gamma\|_2 \leq \frac{\tilde{\gamma} - \gamma}{m_*} \|\mathbf{x}_\gamma\|_2.$$

The latter is true due to Lemma 7. Thus f satisfies $\mathbf{A}(1/m_*, 0)$.

E.2 Cubic function $f(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}_*\|_2^3$

In this section we show that the cubic function satisfies $\mathbf{A}(1/\sqrt{3\|\mathbf{x}_*\|_2}, 1/2)$. It is straightforward to verify that the function f is convex. f_γ is strongly convex and the optimality condition for \mathbf{x}_γ yields the following equality:

$$\nabla f(\mathbf{x}_\gamma) + \gamma\mathbf{x}_\gamma = 3\|\mathbf{x}_\gamma - \mathbf{x}_*\|_2(\mathbf{x}_\gamma - \mathbf{x}_*) + \gamma\mathbf{x}_\gamma = 0. \quad (18)$$

In the case when $\mathbf{x}_* = 0$, the penalized minimum point \mathbf{x}_γ equals 0, for every γ , thus we suppose in the following that $\mathbf{x}_* \neq 0$. Since the norm is scalar, (18) yields that the vectors $\mathbf{x}_\gamma - \mathbf{x}_*$ and \mathbf{x}_γ are co-linear. Therefore there exists a real number λ_γ such that $\mathbf{x}_\gamma = \lambda_\gamma\mathbf{x}_*$. Lemma 7 implies that $|\lambda_\gamma| \leq 1$, thus the following quadratic equality is true:

$$-3\|\mathbf{x}_*\|_2(\lambda_\gamma - 1)^2\mathbf{x}_* + \gamma\lambda_\gamma\mathbf{x}_* = 0. \quad (19)$$

As said in the beginning, $\mathbf{x}_* \neq 0$, therefore it its coefficient that is equal to zero. Solving the quadratic equation with respect to λ_γ , we get the following formula:

$$\lambda_\gamma = 1 - \frac{\gamma}{\gamma/2 + \sqrt{3\gamma\|\mathbf{x}_*\|_2 + \gamma^2/4}}.$$

According to Lemma 7, for every $\tilde{\gamma} > \gamma$, we have $|\lambda_\gamma| > |\lambda_{\tilde{\gamma}}|$. On the other hand, from (19) one deduces that $\lambda_\gamma > 0$, for every $\gamma > 0$. Thus, inserting the found value for λ_γ , we obtain the following inequality:

$$\begin{aligned} \|\mathbf{x}_\gamma - \mathbf{x}_{\tilde{\gamma}}\|_2 &= \|\mathbf{x}_*\|_2 \left(\frac{\tilde{\gamma}}{\tilde{\gamma}/2 + \sqrt{3\tilde{\gamma}\|\mathbf{x}_*\|_2 + \tilde{\gamma}^2/4}} - \frac{\gamma}{\gamma/2 + \sqrt{3\gamma\|\mathbf{x}_*\|_2 + \gamma^2/4}} \right) \\ &\leq \frac{(\tilde{\gamma} - \gamma)\|\mathbf{x}_*\|_2}{\tilde{\gamma}/2 + \sqrt{3\tilde{\gamma}\|\mathbf{x}_*\|_2 + \tilde{\gamma}^2/4}} \leq \frac{(\tilde{\gamma} - \gamma)\|\mathbf{x}_*\|_2^{1/2}}{\sqrt{3\tilde{\gamma}}}. \end{aligned}$$

Therefore f satisfies $\mathbf{A}(1/\sqrt{3}, 1/2)$.

E.3 Power function $f(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}_*\|_2^a$

For $a \geq 2$, we consider the function $f(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}_*\|_2^a$. We show here that f satisfies $\mathbf{A}((1/a)^{1/(a-1)}, (a-2)/(a-1))$. Since f_γ is a differentiable strongly-convex function, we get the following equation for \mathbf{x}_γ :

$$a\|\mathbf{x}_\gamma - \mathbf{x}_*\|_2^{a-2}(\mathbf{x}_\gamma - \mathbf{x}_*) + \gamma\mathbf{x}_\gamma = 0. \quad (20)$$

Similar to the previous case, we notice that $\mathbf{x}_\gamma - \mathbf{x}_*$ and \mathbf{x}_γ are co-linear. Thus, there exists λ_γ such that $\mathbf{x}_\gamma = (1 - \lambda_\gamma)\mathbf{x}_*$. Since \mathbf{x}_* is assumed to be non-zero, in order to calculate \mathbf{x}_γ , one needs to solve the following equation:

$$|\lambda_\gamma|^{a-2}\lambda_\gamma = \frac{\gamma(1 - \lambda_\gamma)}{a\|\mathbf{x}_*\|^{a-2}}. \quad (21)$$

Thus the p -dimensional equation (20) reduces to equation (21) involving a one-dimensional unknown. Lemma 7 yields $\lambda_{\tilde{\gamma}} > \lambda_\gamma > 0$ for every $\tilde{\gamma} > \gamma \geq 0$. In addition, from (21), we have that $\lambda_\gamma \leq 1$ for every $\gamma > 0$. It is straightforward to verify that for every $\gamma \geq 0$, (21) has exactly one solution satisfying these conditions.

Lemma 8. *Let $\alpha \geq 1$. If $(\lambda_s : s \in (0, 1))$ satisfies $\lambda_s^\alpha = s(1 - \lambda_s)$ for every $s \in (0, 1)$, then*

$$|\lambda_s - \lambda_{s'}| \leq \frac{|s - s'|}{(s \vee s')^{(\alpha-1)/\alpha}}, \quad \forall s', s \in (0, 1).$$

Proof. Without loss of generality, we assume that $s' \leq s$. Computing the derivative of both sides of the identity $\lambda_s^\alpha = s(1 - \lambda_s)$, we get

$$\lambda_s' = \frac{1 - \lambda_s}{\alpha\lambda_s^{\alpha-1} + s} \geq 0.$$

This implies that $\lambda_{s'} \leq \lambda_s$. In addition,

$$\begin{aligned} \lambda_s - \lambda_{s'} &\leq \frac{\lambda_s^\alpha - \lambda_{s'}^\alpha}{\lambda_s^{\alpha-1}} \\ &= \frac{s(1 - \lambda_s) - s'(1 - \lambda_{s'})}{\lambda_s^{\alpha-1}} \\ &= \frac{(s - s')(1 - \lambda_{s'})}{\lambda_s^{\alpha-1}} - \frac{s(\lambda_s - \lambda_{s'})}{\lambda_s^{\alpha-1}}. \end{aligned}$$

Rearranging the terms, we arrive at

$$\begin{aligned} \lambda_s - \lambda_{s'} &\leq \frac{(s - s')(1 - \lambda_{s'})}{\lambda_s^{\alpha-1}\left(1 + \frac{s}{\lambda_s^{\alpha-1}}\right)} \\ &= \frac{(s - s')(1 - \lambda_{s'})}{\lambda_s^{\alpha-1} + s} \end{aligned}$$

In the last fraction, the numerator is bounded by $s - s'$, while the denominator satisfies

$$\begin{aligned} \lambda_s^{\alpha-1} + s &= (s(1 - \lambda_s))^{(\alpha-1)/\alpha} + s \\ &\geq (s(1 - s^{1/\alpha}))^{(\alpha-1)/\alpha} + s \\ &\geq s^{(\alpha-1)/\alpha}(1 - s^{1/\alpha}) + s = s^{(\alpha-1)/\alpha}. \end{aligned}$$

This completes the proof of the lemma. \square

Applying this lemma to (21), we get

$$\lambda_{\tilde{\gamma}} - \lambda_\gamma \leq \frac{\tilde{\gamma} - \gamma}{a\|\mathbf{x}_*\|_2^{a-2}(\tilde{\gamma}/a\|\mathbf{x}_*\|_2^{a-2})^{(a-2)/(a-1)}} = \frac{\tilde{\gamma} - \gamma}{a^{1/(a-1)}\|\mathbf{x}_*\|_2^{(a-2)/(a-1)}\tilde{\gamma}^{(a-2)/(a-1)}},$$

for all $\gamma, \tilde{\gamma}$ satisfying $0 \leq \gamma \leq \tilde{\gamma} \leq a\|\mathbf{x}_*\|_2^{a-2}$. In conclusion, we get

$$\begin{aligned} \|\mathbf{x}_{\tilde{\gamma}} - \mathbf{x}_\gamma\|_2 &\leq \frac{\tilde{\gamma} - \gamma}{a^{1/(a-1)}\|\mathbf{x}_*\|_2^{(a-2)/(a-1)}\tilde{\gamma}^{(a-2)/(a-1)}}\|\mathbf{x}_*\|_2 \\ &\leq \frac{\tilde{\gamma} - \gamma}{\tilde{\gamma}^{(a-2)/(a-1)}}(\|\mathbf{x}_*\|_2/a)^{1/(a-1)}. \end{aligned}$$

This concludes the proof.