Taming the Interacting Particle Langevin Algorithm - the superlinear case

Tim Johnston[1], Nikolaos Makras[1], and Sotirios Sabanis[1,2,3]

[1]School of Mathematics, University of Edinburgh, UK
[2]The Alan Turing Institute, UK
[3]National Technical University of Athens, Greece

April 4, 2024

Abstract

Recent advances in stochastic optimization have yielded the interacting particle Langevin algorithm (IPLA), which leverages the notion of interacting particle systems (IPS) to efficiently sample from approximate posterior densities. This becomes particularly crucial within the framework of Expectation-Maximization (EM), where the E-step is computationally challenging or even intractable. Although prior research has focused on scenarios involving convex cases with gradients of log densities that grow at most linearly, our work extends this framework to include polynomial growth. Taming techniques are employed to produce an explicit discretization scheme that yields a new class of stable, under such non-linearities, algorithms which are called tamed interacting particle Langevin algorithms (tIPLA). We obtain non-asymptotic convergence error estimates in Wasserstein-2 distance for the new class under an optimal rate.

1 Introduction

The Expectation-Maximization (EM) algorithm is widely used for locating maximizers of posterior distributions. Applications span, but are not confined to, hyperparameter estimation, mixture models, hidden variable models, and variational inference [5]. At its essence, each iteration of the EM algorithm consists of two fundamental steps: the Expectation (E) and the Maximization (M) step. The algorithm is defined by alternating iteratively between these two steps. Given a data specification $p_\theta(x, y)$ parameterized by θ, where x represents the latent variable (interpreted as incomplete data) and y the observed data, our aim is to find θ^* that maximizes the marginal likelihood $q_\theta(x) = \int_{\mathbb{R}^{d_x}} p_\theta(x, y)dy$.

When the integral in the E-step is computationally challenging or intractable, Markov Chain Monte Carlo (MCMC) methods are often employed, traditionally using Metropolis-Hastings-type algorithms [4]. However, these approaches introduce scalability concerns and are susceptible to local mode entrapment.

In this landscape, a new method was introduced by [2], such that samples of the latent space variable were generated via an Unadjusted Langevin Algorithm (ULA) chain. Nevertheless, the general applicability of these results is curtailed by particular choices of step sizes. An alternative avenue was pioneered in [12], in which the study of the limiting behaviour of various gradient flows associated with appropriate free energy functionals led to an interacting particle system (IPS) that provides efficient estimates for maximum likelihood estimations. Further, [1] thoughtfully expanded on this framework by injecting noise into the dynamics of the parameter θ itself, thereby transitioning from deterministic to stochastic dynamics. This key modification provides a stochastic system with an invariant measure, allowing the establishment of non-asymptotic convergence to θ^*, the maximizer of the marginal likelihood.

In our work, while the convexity assumption is maintained, we address the challenge posed by superlinear growth exhibited by gradients of log densities, which makes other known algorithms, such as vanilla Langevin based algorithms, unstable. To counteract this, we implement taming techniques, initially researched for non-globally Lipschitz drifts for SDEs in [10] and subsequently in [18] and [19]. The latter approach has found applications in optimization and machine learning and led to the design of new MCMC algorithms as one typically deals with high nonlinear objective function, see e.g. [3],[16],[14],[20]. The underlying principle of all these algorithms is the rescaling of ULA's drift coefficient in such a way that maintains stability without significantly increasing computational complexity as in the case of implicit schemes, or by introducing additional constrains via adaptive stepsizes.

In this paper, we study two new algorithms from the tIPLA class, namely the coordinate wise version, known as tIPLAc, and the uniformly tamed version tIPLAu. Those algorithms are tamed versions of IPLA (developed in [1]) as explained in Subsection 2.2. The optimal rate of convergence is recovered for both algorithms and our estimates are explicit regarding their dependencies on dimension and the number of particles employed N.

1.1 Notation

We conclude this subsection with some basic notation. For $u, v \in \mathbb{R}^d$, define the scalar product $\langle u, v\rangle = \sum_{i=1}^d u_i v_i$ and the Euclidian norm $|u| = \langle u, u\rangle^{1/2}$. For all continuously differentiable functions f and we denote by ∇f it's gradient. The integer part of a real number x is denoted by $\lfloor x \rfloor$. For any $p \in \mathbb{N}$, we denote by $\mathcal{P}(\mathbb{R}^d)$ the set of probability measures on $\mathcal{B}(\mathbb{R}^d)$ and by $\mathcal{P}_p(\mathbb{R}^d) = \{\pi \in \mathcal{P} : \int_{\mathbb{R}^d} |x|_p^p d\pi(x) < \infty\}$ the set of all probability measures over $\mathcal{B}(\mathbb{R}^d)$ with finite p-th moment. For any two Borel probability measures μ and ν, we define the Wasserstein distance of order $p \geq 1$ as

$$W_p(\mu, \nu) = \left(\inf_{\zeta \in \prod(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p d\zeta(x, y) \right)^{1/p},$$

where by $\prod(\mu, \nu)$ we denote the set of transference plans of μ and ν. Moreover, for all $\mu,\ \nu \in \mathcal{P}_p(\mathbb{R}^d)$, there exists a transference plan $\zeta^* \in \prod(\mu, \nu)$ such that for any coupling (X, Y) distributed according to ζ^*, $W_p(\mu, \nu) = \mathbb{E}^{1/p}\left[|X - Y|^p\right]$.

2 Setting and Definitions

2.1 Initial setup

Let $p_\theta(x, \cdot)$ be the aforementioned joint probability density function of the latent variable x and for fixed (observed) data y. The goal of maximum marginal likelihood estimation (MMLE) is to find the parameter θ^* that maximises the marginal likelihood (Dempster et al.)[5]. To deal with the aforementioned log-density we define the negative log-likelihood for fixed $y \in \mathbb{R}^{d^y}$ as follows

$$U_y(\theta, x) := -\log p_\theta(x, y),$$

thus gaining the following notation for the quantity we are interested in maximizing $k_y(\theta) = q_\theta(y) = \int p_\theta(x, y)dx = \int e^{-U_y(\theta,x)}dx$. Lastly, for matters of clarity we also denote $h_y(v) = \nabla U_y(v)$ so that $h^x_y(v) = \nabla_x U_y(v)$ and $h^\theta_y(v) = \nabla_\theta U_y(v)$, where $v := (\theta, x)$. Henceforth, we drop the reference to the (fixed) data, i.e. to y, for reasons of brevity. Following the convention of [12], N particles $\mathcal{X}^{i,N}_t$ for $i \in \{1, \dots, N\}$ are used to estimate the gradient of q_θ, which are governed by following continuous-time dynamics:

$$d\vartheta^N_t = -\frac{1}{N}\sum_{j=1}^{N} \nabla_\theta U(\vartheta^N_t, \mathcal{X}^{j,N}_t) + \sqrt{\frac{2}{N}}dB^{0,N}_t, \tag{1}$$

$$d\mathcal{X}^{i,N}_t = -\nabla_x U(\vartheta^N_t, \mathcal{X}^{i,N}_t) + \sqrt{2}dB^{i,N}_t, \tag{2}$$

for $i = 1, .., N$, where $\{(B_t)^{i,N}_{t\geq 0}\}_{0\leq i\leq N}$ is a family of independent Brownian motions. The discrete time Markov chain associated with the above IPS (1)-(2) is obtained by the corresponding Euler-Maruyama discretization scheme of the given Langevin SDEs:

$$\begin{aligned} \theta^\lambda_0 =&\theta_0,\ \theta^\lambda_{n+1} = \theta^\lambda_n - \frac{\lambda}{N}\sum_{i=1}^{N} h^\theta(\theta^\lambda_n, X^{i,\lambda}_n) + \sqrt{\frac{2\lambda}{N}}\xi^{(0)}_{n+1}, \\ X^{i,\lambda}_0 =&x^i_0,\ X^{i,\lambda}_{n+1} = X^{i,\lambda}_n - \lambda h^x(\theta^\lambda_n, X^{i,\lambda}_n) + \sqrt{2\lambda}\xi^{(i)}_{n+1},\ \forall i \in \{1, ..., N\}, \end{aligned}$$

where $\theta_0 \in \mathbb{R}^{d^\theta}$, $x^i_0 \in \mathbb{R}^{d^x}$, $\lambda > 0$ the step-size parameter and for $i = 0$ and $\forall i \in \{1, \dots, N\}$, $(\xi^{(i)}_n)_{n\in\mathbb{N}}$ are i.i.d. standard d^θ and d^x respectively dimensional Gaussian variables. For reasons that become apparent after the introduction of the taming functions in Subsection 2.2.2 we also consider the following dynamics which are a different time-scaled version of the original equations (1)-(2):

$$d\vartheta_t^N = -\frac{1}{N^{p+1}} \sum_{j=1}^{N} \nabla_\theta U(\vartheta_t^N, \mathcal{X}_t^{j,N}) + \sqrt{\frac{2}{N^{p+1}}} dB_t^{0,N}, \tag{3}$$

$$d\mathcal{X}_t^{i,N} = -\frac{1}{N^p} \nabla_x U(\vartheta_t^N, \mathcal{X}_t^{i,N}) + \sqrt{\frac{2}{N^p}} dB_t^{i,N}, \tag{4}$$

where $p = 2\ell + 1$ is controlled by the order of polynomial growth ℓ in $\nabla U(\theta, x)$, see Remark 1. Such an adjustment allows the generalization of the results in [1] to the case of superlinear drift coefficients while keeping the underlying stationary distribution of the dynamics identical.

2.2 Taming approach

2.2.1 Introduction of the underlying taming technique

For a fixed $T > 0$, consider an SDE given by

$$dY(t) = b(t, Y(t))dt + \sigma(t, Y(t))dB_t, \ \forall t \in [0, T], \tag{5}$$

where $b(t, y)$ and $\sigma(t, y)$ are assumed to be $\mathcal{B}(\mathbb{R}_+) \times \mathcal{B}(\mathbb{R}^d)-$measurable functions. Then, the discrete time Markov chain associated with the ULA algorithm is obtained by the Euler-Maruyama discretization scheme of SDE(5) and is defined for all $n \in \mathbb{N}$ by:

$$Y(t_{n+1}) = Y(t_n) - \lambda b(t_n, Y(t_n)) + \sqrt{\lambda}\sigma(t_n, Y(t_n))\xi_{n+1},$$

where $\lambda > 0$ is the stepsize and $(\xi_n)_{n \geq 1}$ are i.i.d. standard Gaussian random variables. In the case where the drift coefficient b is superlinear, it is shown in [9] that ULA is unstable in the sense that any $p-$absolute moment of the algorithm ($p \geq 1$) diverges to infinity. In the SDE approximation literature, a new class of explicit numerical schemes has been introduced to study the case of non-globally Lipschitz conditions by modifying both the drift and diffusion coefficients in such a way that they grow at most linearly, for example see, [19], [21] and [11]. The efficiency of such schemes and their respective properties of $\mathcal{L}^p$ convergence create a strong incentive to extend those techniques to sampling and optimization. This adjustment is key to the development of algorithms that approximate non-linear systems while remaining computationally tractable. Typically tamed schemes are given by,

$$Y(t_{n+1}) = Y(t_n) - \lambda b_\lambda(t_n, Y(t_n)) + \sqrt{\lambda}\sigma_\lambda(t_n, Y(t_n))\xi_{n+1},$$

for an appropriate choice of taming functions $b_\lambda : \mathbb{R}^d \mapsto \mathbb{R}^d$ and $\sigma_\lambda : \mathbb{R}^d \mapsto \mathbb{R}^{d \times d}$.

2.2.2 Application in the IPLA framework

We extend this notion of taming to our setting when dealing with the superlinear nature of ∇U and a constant diffusion coefficient. To this end, we introduce a family of a taming functions $(h_\lambda)_{\lambda \geq 0}$ with $h_\lambda : \mathbb{R}^{d^\theta} \times \mathbb{R}^{d^x} \to \mathbb{R}^{d^\theta} \times \mathbb{R}^{d^x}$ which are close approximations of ∇U in a sense made precise below.

We suggest two such taming functions $h_\lambda(v)$, one of which is uniformly tamed and another by using a coordinate-wise approach:

$$h_{\lambda,u}(v) = \frac{h(v) - \mu v}{1 + \lambda^{1/2} N^{-p/2} |h(v) - \mu v|} + \mu v, \tag{6}$$

$$h_{\lambda,c}(v) = \left(\frac{h^{(i)}(v) - \mu v^{(i)}}{1 + \lambda^{1/2} |h^{(i)}(v) - \mu v^{(i)}|} + \mu v^{(i)} \right)_{i \in \{1,\dots,d^\theta + d^x\}}, \tag{7}$$

where μ is the strong convexity constant given in A2. In [3] it's experimentally established that the coordinate-wise version outperforms the uniform taming approach. This is in agreement with the observation that uniform taming cannot distinguish between the different levels of contribution each coordinate offers to the gradient. However, there is a trafeoff for using the coordinate-wise approach, as in order to obtain an appropriate dissipativity condition for the tamed function h_λ, one needs to require additional smoothness on the potential U.

We propose two different algorithms for the EM implementation within the tIPLA class, which are determined by the choice of the taming function and the stochastic dynamics. We find that when the uniform taming is used, then the time-scaled dynamics (3)-(4) are more suitable as the addition of the exponent in the number of particles N is essential to guarantee convergence. Hence (3)-(4) paired with (6) leads to the following scheme:

$$\theta_0^{\lambda,u} = \theta_0,\ \theta_{n+1}^{\lambda,u} = \theta_n^{\lambda,u} - \frac{\lambda}{N^{p+1}} \sum_{i=1}^{N} h^\theta_{\lambda,u}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u}) + \sqrt{\frac{2\lambda}{N^{p+1}}}\xi_{n+1}^{(0)}, \tag{8}$$

$$X_0^{i,\lambda,u} = x_0^i,\ X_{n+1}^{i,\lambda,u} = X_n^{i,\lambda,u} - \frac{\lambda}{N^p} h^x_{\lambda,u}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u}) + \sqrt{\frac{2\lambda}{N^p}}\xi_{n+1}^{(i)},\ \forall i \in \{1,...,N\}. \tag{9}$$

In the case of the coordinate-wise taming, the extra smoothness required obviates any modification to the original dynamics. Therefore one uses (1)-(2) with (7) to obtain the second algorithm:

$$\theta_0^{\lambda,c} = \theta_0,\ \theta_{n+1}^{\lambda,c} = \theta_n^{\lambda,c} - \frac{\lambda}{N} \sum_{i=1}^{N} h^\theta_{\lambda,c}(\theta_n^{\lambda,c}, X_n^{i,\lambda,c}) + \sqrt{\frac{2\lambda}{N}}\xi_{n+1}^{(0)}, \tag{10}$$

$$X_0^{i,\lambda,c} = x_0^i, X_{n+1}^{i,\lambda,c} = X_n^{i,\lambda,c} - \lambda h^x_{\lambda,c}(\theta_n^{\lambda,c}, X_n^{i,\lambda,c}) + \sqrt{2\lambda}\xi_{n+1}^{(i)},\ \forall i \in \{1,...,N\}. \tag{11}$$

It is important to note that despite the differences in their setup, both algorithms yield the same non-asymptotic convergence behavior and dependence on the dimension of the

Algorithm 1 Tamed Interacting Particle Langevin Algorithm (tIPLAu)

Require: N,λ,$\pi_{\text{init}} \in \mathcal{P}(\mathbb{R}^{d^\theta}) \times \mathcal{P}((\mathbb{R}^{d^x})^N)$
Draw $(\theta_0, \{X_0^{i,N}\}_{1\leq i\leq N})$ from π_{init}
for n = 0 to $n_T = \lfloor T/\lambda \rfloor$ **do**
$\theta_{n+1}^{\lambda,u} = \theta_n^{\lambda,u} - \frac{\lambda}{N^{p+1}} \sum_{i=1}^N h_{\lambda,u}^\theta(\theta_n^{\lambda,u}, X_n^{i,\lambda,u}) + \sqrt{\frac{2\lambda}{N^{p+1}}}\xi_{n+1}^{(0)}$
$X_{n+1}^{i,\lambda,u} = X_n^{i,\lambda,u} - \frac{\lambda}{N^p} h_{\lambda,u}^x(\theta_n^{\lambda,u}, X_n^{i,\lambda,u}) + \sqrt{\frac{2\lambda}{N^p}}\xi_{n+1}^{(i)},\ \forall i \in 1, ..., N$
end for
return θ_{n_T+1}

E optimization challenge. Both schemes are summarised in the algorithms (tIPLAu) and (tIPLAc) given below. It is apparent that the newly proposed algorithms, designated as tIPLAu and tIPLAc, incorporate foundational elements that exist in [12], which employed deterministic dynamics in the θ component. Their design profits also from the approach in [1] which introduced stochastic noise in the $\theta-$direction, allowing for the explicit calculation of the invariant measure. A pivotal enhancement that comes with this approach involves the application of a taming technique via the recalibration of the original drift terms, h_x and h_θ, into tamed counterparts, as described in (6) and (7). This adjustment is instrumental in deriving nonasymptotic results, particularly in cases characterized by superlinear gradients in the log-likelihood function.

Algorithm 2 Tamed Interacting Particle Langevin Algorithm - Coordinatewise (tIPLAc)

Require: N,λ,$\pi_{\text{init}} \in \mathcal{P}(\mathbb{R}^{d^\theta}) \times \mathcal{P}((\mathbb{R}^{d^x})^N)$
Draw $(\theta_0, \{X_0^{i,N}\}_{1\leq i\leq N})$ from π_{init}
for n = 0 to $n_T = \lfloor T/\lambda \rfloor$ **do**
$\theta_{n+1}^{\lambda,c} = \theta_n^{\lambda,c} - \frac{\lambda}{N} \sum_{i=1}^N h_{\lambda,c}^\theta(\theta_n^{\lambda,c}, X_n^{i,\lambda,c}) + \sqrt{\frac{2\lambda}{N}}\xi_{n+1}^{(0)}$
$X_{n+1}^{i,\lambda,c} = X_n^{i,\lambda,c} - \lambda h_{\lambda,c}^x(\theta_n^{\lambda,c}, X_n^{i,\lambda,c}) + \sqrt{2\lambda}\xi_{n+1}^{(i)},\ \forall i \in 1, ..., N$
end for
return θ_{n_T+1}

2.3 Essential quantities and proof strategy of main result

In this subsection, we lay out the analytical framework and convergence properties of the sequences $(\theta_n^{\lambda,\cdot})_{n\geq 0}$ of either tIPLAu $(\theta_n^{\lambda,u})_{n\geq 0}$ or tIPLAc $(\theta_n^{\lambda,c})_{n\geq 0}$, resulting in their convergence to the minimiser θ^*. A central concept of the technical analysis is the rescaling of the original dynamics of the following form

$$\mathcal{Z}_t^N = \left(\vartheta_t^N, N^{-1/q}\mathcal{X}_t^{1,N}, \ldots, N^{-1/q}\mathcal{X}_t^{N,N}\right), \tag{12}$$

along with the corresponding rescaling that is required for the algorithms and their continuous time interpolations, see more details in Appendix A. This rescaled version of the particle system is pivotal in bridging the dynamics represented by either (1)-(2) or (3)-(4) with the pairs $\mathcal{V}_t^{i,N} := (\vartheta_t^N, \mathcal{X}_t^{i,N})$ generated by the algorithms. The essence of this rescaling lies in its ability to equate the moments of $\mathcal{Z}_t^N$ with the averaged moments of the pairs $\mathcal{V}_t^{i,N}$ across any $\mathbb{L}^q$ norm. However, given our focus on Wasserstein-2 distance, we opt for a rescaling factor of $q = 2$. This choice yields the critical property:

$$|\mathcal{Z}_t|^2 = \frac{1}{N}\sum_{i=1}^{N} |\mathcal{V}_t^i|^2.$$

It is imperative to note that this convenient rescaling does not detract from the analysis objectives, as our primary interest is in the convergence of the $\theta-$component. The introduction of particles serves primarily to facilitate sampling, hence their specific scaling does not impact the core convergence analysis.

In the present analysis, the interacting particle systems defined in equations (1)-(2) and (3)-(4) are deliberately constructed to target the same invariant measure π_*^N characterized by the density $Z^{-1}\exp(-\sum_{i=1}^N U(\theta, x^i))$. This aspect is critical as it assigns the number of particles N a role akin to the inverse temperature parameter encountered in simulated annealing algorithms, thereby regulating the concentration of the $\theta-$marginal of the invariant measure π_Θ^N, towards the minimizer θ^*. This behavior, wherein $\pi_\Theta^N \to \delta_{\theta^*}$ as $N \to \infty$, is established in Proposition 2.

Moreover, the convergence rate of the Langevin diffusion to its invariant measure is a standard result, one could consult [7] and the references within. This is further generalized to particle systems in the research conducted by [1]. By integrating their findings into our framework, we arrive at Proposition 3 and 4. Equipped with these guarantees, we can now theoretically approximate the minimizer θ^*, and subsequently, our analysis focuses to the discretization errors inherent in the proposed schemes.

One expects a dimensional scaling of at least $\mathcal{O}(d^{1/2})$ for the Wasserstein-2 numerical error of Langevin based algorithms, for example [6]. In the context of the E optimization where N particles are generated, the true dimentionality of the problem is $d^\theta + N \cdot d^x$, which can be naively interpreted as the need to increase the number of iterations to achieve convergence as N increases. However, this dependency on N is effectively mitigated by making use of the implied symmetry within the dynamics of the interacting particle system by considering only the θ-marginal in the analysis. Such an approach enables us to attain the classical Euler-Maruyama convergence rate for the numerical solutions of stochastic differential equations, as demonstrated in Proposition 5 and 6.

In closing, one decomposes the global $\mathcal{L}^2$-error, denoted as $\mathbb{E}^{1/2}\left[|\theta^* - \theta_n^{\lambda,\cdot}|^2\right]$, into three distinct components:

$$\mathbb{E}^{1/2}\left[|\theta^* - \theta_n^{\lambda,\cdot}|^2\right] \le W_2(\delta_{\theta^*}, \pi_\Theta^N) + W_2(\pi_\Theta^N, \mathcal{L}(\vartheta_{n\lambda}^N)) + W_2(\mathcal{L}(\vartheta_{n\lambda}^N), \mathcal{L}(\theta_n^{\lambda,\cdot})),$$

where $\theta_n^{\lambda,\cdot}$ stands for the iterates of either of the algorithms, tIPLAu and tIPLAc alike. Moreover, $W_2(\delta_{\theta^*}, \pi_\Theta^N)$ quantifies the deviation of the invariant measure from the minimizer, $W_2(\pi_\Theta^N, \mathcal{L}(\vartheta_{n\lambda}^N))$ captures the discrepancy between the law of the dynamics and their invariant measure, and $W_2(\mathcal{L}(\vartheta^N_{n\lambda}), \mathcal{L}(\theta^{\lambda,\cdot}_n))$ encompasses the error induced by the discretization scheme, namely, the divergence between the law of the iterations of our algorithms and their continuous counterparts.

3 Main Assumptions and Results

In this section we provide the assumptions on the potential U and it's gradient ∇U that define the framework, under which the main results for the non-asymptotic behaviour of the newly proposed algorithms, tIPLAu and tIPLAc, are derived.
Let $d := d^\theta + d^x$, U be a $C^1\left(\mathbb{R}^d\right)$ function and recall that $h(v) := h(\theta, x) := \nabla U(\theta, x)$ is a locally Lipschitz function in θ and x.

A1. There exist $L > 0$ and $\ell > 0$ such that

$$|h(v) - h(v')| \leq L\left(1 + |v|^\ell + |v'|^\ell\right)|v - v'|,\ \forall v, v' \in \mathbb{R}^d.$$

Additionally we require U to be $\mu-$strongly convex.

A2. There exists $\mu > 0$ such that

$$\langle v - v', h(v) - h(v')\rangle \geq \mu|v - v'|^2,\ \forall v, v' \in \mathbb{R}^d.$$

Assumption A1 is a significant relaxation of the global Lipschitz condition which is widely used in the literature. It allows the gradient h to grow polynomially fast at infinity. Assumption A2 guarantees that U has a unique minimiser and also can be seen as a monotonicity type condition which is satisfied by the drift coefficients of the corresponding SDEs.

We also present a growth condition for the gradient ∇U, which is essential in the case where the coordinate-wise taming function (7) is used.

A3. For each i in $\{1, \ldots, d\}$ there exists $\mu > 0$ such that

$$h^{(i)}(v)v^{(i)} \geq \frac{\mu}{2}|v^{(i)}|^2 - \frac{1}{2\mu}|h_0^{(i)}|^2.$$

Assumption A3 is a coordinate-wise dissipativity type condition which plays a crucial role in establishing moment bounds for tIPLAc. Its need stems from the fact that Remark 2 does not guarantee that the taming function (7) preservers the dissipativity of the drift $h(v) = \nabla U(v)$.

Here, we emphasize that Assumption A3 is primarily influenced by the structural properties of the taming function. This assumption could potentially be relaxed to necessitate only argument-wise dissipativity, as described by the following mathematical formulation:

$$h^\theta(v)\theta \geq \frac{\mu}{2}|\theta|^2 - \frac{1}{2\mu}|h_0^\theta|^2 \text{ and } h^x(v)x \geq \frac{\mu}{2}|x|^2 - \frac{1}{2\mu}|h_0^x|^2.$$

This relaxation is applicable if one considers the taming function

$$h^w_{\lambda,a}(v) = \frac{h^w(v) - \mu w}{1 + \lambda^{1/2}|h^w(v) - \mu w|} + \mu w \text{ for } w = \theta,\ x,$$

which represents an intermediate scenario bridging the uniform and coordinate-wise approaches. Such consideration leads to an alternative formulation of the tIPLA algorithm. However, this variant of the tIPLA algorithm does not capture our interest, primarily due to its inherent limitations. Specifically, it fails to adequately discriminate between the components of the gradient that exhibit explosive behavior and those that do not. While simultaneously, this alternative formulation imposes more stringent requirements than those necessitated by the uniform case.

A4. The initial condition $z_0^N = \left(\theta_0, N^{-1/2}x_0^1, \ldots, N^{-1/2}x_0^N\right)$ is such that

$$\mathbb{E}\left[|z_0^N|^{2p_0}\right] < \infty,$$

where $p_0 = 2(\ell + 1)$ and ℓ being the order of polynomial growth of $\nabla U(\theta, x)$ as given in Assumption A1.

Remark 1. One notices that Assumption A1 allows for control of the growth of $h(v) = \nabla U(v)$, i.e., for every $v \in \mathbb{R}^d$,

$$h(v) \le K(1 + |v|^{\ell+1}),$$

where $K = 2L + |h_0|$.

Remark 2.In view of Assumption A2, strong convexity implies dissipativity, i.e, for every $v \in \mathbb{R}^d$,

$$\langle v, h(v)\rangle \ge \frac{\mu}{2}|v|^2 - b,$$

where $b = \dfrac{1}{2\mu}|h_0|^2$.

3.1 Taming functions and inherited properties

Important properties of the suggested taming functions, which act as drift coefficients in the corresponding numerical schemes, are established in this subsection.

Property 1. For all $\lambda > 0$ and $v \in \mathbb{R}^d$ one has

$$|h_{\lambda,u}(v)| \le \mu|v| + \lambda^{-1/2}N^{p/2}.$$

The original and tamed functions are sufficiently close for $\lambda > 0$ small enough.

Property 2.For all $\lambda > 0$ and $v \in \mathbb{R}^d$ one has

$$|h_{\lambda,u}(v) - h(v)| \leq C_1 \lambda^{1/2} N^{-p/2} (1 + |v|^{2(\ell+1)}),$$

where $C_1 = 2^{2(\ell+3/2)} \max\{K^2, \mu^2\}$.

The tamed function $h_{\lambda,u}(v)$ inherits the dissipativity condition established in Remark 2.

Property 3. For all $\lambda > 0$ and $v \in \mathbb{R}^d$ one has

$$\langle v, h_{\lambda,u}(v) \rangle \geq \frac{\mu}{2}|v|^2 - b,$$

where b is given in Remark 2. To see this consider the cases in which $\langle h(v) - \mu v, v\rangle$ is greater or less than 0 separately.
In the context of unified framework established in [15], the properties 1-3 therein are recovered for $\delta = 2$ and $\gamma = 1/2$ which is on par with THϵO POULA [13] and TUSLA [16] agloritнms.

Comment. Regarding tIPLAc, where the coordinate-wise taming function (7) is used, Properties 1-2 are also guaranteed without the N-factor and also a slightly worse coefficient in Property 1, that is: $|h_{\lambda,c}(v)| \leq \mu|v| + (d^\theta + d^x)\lambda^{-1/2}$. However for Property 3 to be recovered, we require the additional smoothness provided by Assumption A3.

3.2 Preliminary Theoretical Statements

Both interacting SDEs defined by (1)-(2) and (3)-(4) respectively, admit a strong solution under A1-A2. This follows since each of the drift coefficients of the SDEs are locally Lipschitz functions and dissipative in view of Remark 2 whereas the diffusion coefficients are constant, one could consult [17](Theorem 2.3.5) for more details. Moreover, they exhibit the same invariant measure, as one would expect from a standard Langevin diffusion.

Proposition 1. Let A1 and A2 hold. Then, the measure π_*^N characterized by the density $Z^{-1} \exp^{-\sum_{i=1}^N U(\theta, x^i)}$, with Z being the normalizing constant, is the invariant measure for both the interacting particles systems (1)-(2) and (3)-(4).

Proof. The existence of an invariant measure is established by invoking the Krylov-Bogoliubov Theorem, as presented in Theorem 7.1 [8]. This theorem guarantees the existence of an invariant measure for a Markov process generated by it's corresponding semigroup, given a tightness condition on the sequence of probability measured associated with that process. According to Da Prato's Proposition 7.10 [8], the establishment of uniform moment bounds serves as a sufficient condition for tightness, which is provided in Lemma 1. The uniqueness of the invariant measure follows from Theorem 7.16(ii) [8], as Hypothesis 7.13 is satisfied due to A1 and A2. Furthermore, one can verify that π_*^N

is indeed the invariant measure in discussion by checking that for any $\phi \in C_c^\infty(\mathbb{R}^d)$ one has $\int_{\mathbb{R}^d} \mathcal{L}\phi \, d\pi_*^N = 0$, where $\mathcal{L}$ is the infinitesimal generator of the process. See the proof of Proposition 2 in [1] for more details. $\square$

Henceforward we consider exclusively the $\theta-$marginal of the invariant measure and we show that indeed the number of particles N plays the role of the inverse temperature parameter in temperature annealing algorithms in the sense that it grants us control on the marginal's concentration around the minimiser.

Proposition 2. Let π_Θ^N denote the $\theta-$marginal of the invariant measure and θ^* as the maximiser of $k(\theta)$. Then, under A2, for any $N \in \mathbb{N}$, one has the bound

$$W_2(\pi_\Theta^N, \delta_{\theta^*}) \leq \sqrt{\frac{2d^\theta}{\mu N}}.$$

Proof. Follows from Proposition 3 in [1]. $\square$

Lastly we need the following ergodicity result to obtain the convergence of the law of the IPS to the invariant measure.

Proposition 3. Let A1, A2 and A4 hold and consider the first component, namely ϑ_t^N, of the continuous dynamics (3)-(4). Then, for any $N \in \mathbb{N}$,

$$W_2\left(\mathcal{L}(\vartheta_t^N), \pi_\Theta^N\right) \leq e^{-\mu t}\left(\mathbb{E}^{1/2}\left[|z_0^N - z^*|^2\right] + \left(\frac{d^x N + d^\theta}{\mu N^p}\right)^{1/2}\right).$$

Proof. Follows closely from the proof of Proposition 4 in [1]. One can easily verify that although the corresponding A1 in [1] invokes global Lipschitz continuity rather than local, its need there is solely to ensure the existence of a unique global solution, a fact that is already established at the start of Subsection 3.2. A further difference is the additional factor N^{-p} instead of N^{-1} which appears in the last term of the upper bound. $\square$

Proposition 4. Let A1, A3 and A4 hold and consider the first component, namely ϑ_t^N, of the continuous dynamics (1)-(2). Then, for any $N \in \mathbb{N}$,

$$W_2\left(\mathcal{L}(\vartheta_t^N), \pi_\Theta^N\right) \leq e^{-\mu t}\left(\mathbb{E}^{1/2}\left[|z_0^N - z^*|^2\right] + \left(\frac{d^x N + d^\theta}{\mu N}\right)^{1/2}\right).$$

Proof. Follows also closely from the proof of Proposition 4 in [1] with the note that A3 implies A2. $\square$

3.3 Discretisation Error Estimates

We consider the continuous interpolation of the algorithm tIPLAu which approximates the time-scaled version of the continuous dynamics (3)-(4) denoted as $\overline{Z}_t^{\lambda,u}$. We note that the law of the discretized process and it's interpolation coincide at grid-points, i.e. $\mathcal{L}(Z_n^{\lambda,u}) = \mathcal{L}(\overline{Z}_n^{\lambda,u})$. The continuous-time interpolation of the tIPLAu can be defined as:

$$\overline{\theta}_0^{\lambda,u} =\theta_0,\ d\overline{\theta}_t^{\lambda,u} = -\frac{\lambda}{N^{p+1}}\sum_{i=1}^N h^\theta_{\lambda,u}(\theta^{\lambda,u}_{\lfloor t\rfloor}, X^{i,\lambda,u}_{\lfloor t\rfloor})dt + \sqrt{\frac{2\lambda}{N^{p+1}}}dB_t^{0,\lambda}, \tag{13}$$

$$\overline{X}_0^{i,\lambda,u} =x_0^i,\ d\overline{X}_t^{i,\lambda,u} = -\frac{\lambda}{N^p} h^x_{\lambda,u}(\theta^{\lambda,u}_{\lfloor t\rfloor}, X^{i,\lambda,u}_{\lfloor t\rfloor})dt + \sqrt{\frac{2\lambda}{N^p}}dB_t^{i,\lambda}, \tag{14}$$

for all $i \in \{1,\dots,N\}$, while the time-changed SDEs (3)-(4)are given by:

$$d\vartheta^N_{\lambda t} = -\frac{\lambda}{N^{p+1}}\sum_{i=1}^N h^\theta(\vartheta^N_{\lambda t}, \mathcal{X}^{i,N}_{\lambda t})dt + \sqrt{\frac{2\lambda}{N^{p+1}}}dB_t^{0,\lambda}, \tag{15}$$

$$d\mathcal{X}^{i,N}_{\lambda t} = -\frac{\lambda}{N^p} h^x(\vartheta^N_{\lambda t}, \mathcal{X}^{i,N}_{\lambda t})dt + \sqrt{\frac{2\lambda}{N^p}}dB_t^{i,\lambda}, \tag{16}$$

where in both cases $B_t^\lambda := B_{\lambda t}/\sqrt{\lambda}, t \geq 0$, is a Brownian motion under its completed natural filtration $\mathcal{F}_t^\lambda := \mathcal{F}_{\lambda t}$.

Proposition 5. Let A1, A2 and A4 hold and consider the process $\vartheta^N_{\lambda t}$ as given by the dynamics (15)-(16). Then, for every $\lambda_0 < N^{2\ell+1}/4\mu$, there exists a constant $C > 0$ independent of N, n, λ such that for any $\lambda \in (0, \lambda_0)$,

$$\mathbb{E}^{1/2}\left[|\overline{\theta}_n^{\lambda,u} - \vartheta^N_{n\lambda}|^2\right] \leq \lambda^{1/2} C_{|z_0|,\mu,b,\ell,L}(1 + d^\theta/N + d^x)^{\ell+1},$$

for all $n \in \mathbb{N}$.

Proof. The proof is postponed to Appendix A. □

Following the same lines we define the corresponding auxiliary processes for the iterates of tIPLAc. The continuous-time interpolation is given by:

$$\overline{\theta}_0^{\lambda,c} = \theta_0,\ d\overline{\theta}_t^{\lambda,c} = -\frac{\lambda}{N}\sum_{i=1}^N h^\theta_{\lambda,c}(\theta^{\lambda,c}_{\lfloor t\rfloor}, X^{i,\lambda,c}_{\lfloor t\rfloor})dt + \sqrt{\frac{2\lambda}{N}}dB_t^{0,\lambda}, \tag{17}$$

$$\overline{X}_0^{i,\lambda,c} = x_0^i,\ d\overline{X}_t^{i,\lambda,c} = -\lambda h^x_{\lambda,c}(\theta^{\lambda,c}_{\lfloor t\rfloor}, X^{i,\lambda,c}_{\lfloor t\rfloor})dt + \sqrt{2\lambda}dB_t^{i,\lambda}, \tag{18}$$

and the time changed SDEs of (1)-(2) by

$$d\vartheta^N_{\lambda t} = -\frac{\lambda}{N}\sum_{i=1}^N h^\theta(\vartheta^N_{\lambda t}, \mathcal{X}^{i,N}_{\lambda t})dt + \sqrt{\frac{2\lambda}{N}}dB_t^{0,\lambda}, \tag{19}$$

$$d\mathcal{X}^{i,N}_{\lambda t} = -\lambda h^x(\vartheta^N_{\lambda t}, \mathcal{X}^{i,N}_{\lambda t})dt + \sqrt{2\lambda}dB_t^{i,\lambda}, \tag{20}$$

where the Brownian motions are defined as in (15)-(16).

Proposition 6. Let A1, A3 and A4 hold and consider the process $\vartheta^N_{\lambda t}$ as given by the dynamics (19)-(20). Then, for every $\lambda_0 < 1/4\mu$, there exists a constant $C > 0$ independent of N, n, λ such that for any $\lambda \in (0, \lambda_0)$,

$$\mathbb{E}^{1/2}\left[|\overline{\theta}^{\lambda,c}_n - \vartheta^N_{n\lambda}|^2\right] \leq \lambda^{1/2} C_{|z_0|,\mu,b,\ell,L}(1 + d^\theta + d^x)^{\ell+1},$$

for all $n \in \mathbb{N}$.

Proof. The proof is postponed to Appendix A. ☐

3.4 Main results and global error

Theorem 1. Consider the iterates $(\theta^{\lambda,u}_n)_{n\geq 0}$ as given in tIPLAu, and let A1, A2 and A4 hold. Then, for every $\lambda_0 < N^{2\ell+1}/4\mu$, there exists a constant $C > 0$ independent of N, n, λ such that, for any $\lambda \in (0, \lambda_0)$,

$$\begin{aligned}\mathbb{E}^{1/2}\left[|\theta^* - \theta^{\lambda,u}_n|^2\right] \leq \sqrt{\frac{2d^\theta}{\mu N}} + e^{-\mu n \lambda / N^{2\ell+1}} &\left(\mathbb{E}^{1/2}\left[|z^N_0 - z^*|^2\right] + \left(\frac{d^x N + d^\theta}{\mu N^{2\ell+1}}\right)^{1/2}\right) \\ &+ \lambda^{1/2} C_{|z_0|,\mu,b,\ell,L}(1 + d^\theta/N + d^x)^{\ell+1},\end{aligned}$$

for all $n \in \mathbb{N}$.

Proof. Combining the results from Propositions 2, 3 and 5, we are able to decompose the expectation into a term describing the concentration of the π^N_Θ around θ^*, a term describing the convergence of the IPS to its invariant measure, and a term describing the error induced by the time discretisation:

$$\begin{aligned}\mathbb{E}^{1/2}\left[|\theta^* - \theta^{\lambda,\cdot}_n|^2\right] &= W_2(\delta_{\theta^*}, \mathcal{L}(\theta^{\lambda,\cdot}_n)) \\ &\leq W_2(\delta_{\theta^*}, \pi^N_\Theta) + W_2(\pi^N_\Theta, \mathcal{L}(\vartheta^N_{n\lambda})) + W_2(\mathcal{L}(\vartheta^N_{n\lambda}), \mathcal{L}(\theta^{\lambda,\cdot}_n)).\end{aligned} \tag{21}$$

Substituting the bounds from the aforementioned Propositions yields the final inequality. ☐

Theorem 2. Consider the iterates $(\theta^{\lambda,c}_n)_{n\geq 0}$ as given in tIPLAc, and let A1, A3 and A4 hold. Then, for every $\lambda_0 < 1/4\mu$, there exists a constant $C > 0$ independent of N, n, λ such that, for any $\lambda \in (0, \lambda_0)$,

$$\begin{aligned}\mathbb{E}^{1/2}\left[|\theta^* - \theta^{\lambda,c}_n|^2\right] \leq \sqrt{\frac{2d^\theta}{\mu N}} + e^{-\mu n \lambda} &\left(\mathbb{E}^{1/2}\left[|z^N_0 - z^*|^2\right] + \left(\frac{d^x N + d^\theta}{\mu N}\right)^{1/2}\right) \\ &+ \lambda^{1/2} C_{|z_0|,\mu,b,\ell,L}(1 + d^\theta + d^x)^{\ell+1},\end{aligned}$$

for all $n \in \mathbb{N}$.

Proof. To prove the above inequality, one replaces the bound for $W_2(\pi_{\Theta}^N, \mathcal{L}(\vartheta_{n\lambda}^N))$ in (21) given by Proposition 3 by the bound given by Proposition 4 and respectively the bound for $W_2(\mathcal{L}(\vartheta_{n\lambda}^N), \mathcal{L}(\theta_n^{\lambda,\cdot}))$ given by Proposition 5 by the bound given by Proposition 6. □

It is observed that the dependence on the dimension d^θ of the parameter space in Theorem 2 exhibits a slight deterioration compared to that presented in Theorem 1 and in Theorem 1 of [1]. Notably, the factor $1/N$ is absent in the third term of Theorem 2. This is interpreted as a compromise entailed by the use of the coordinate-wise taming function (7), which requires the direct derivation of moment bounds for the pairs $\mathcal{V}_t^{i,N}$, as discussed in Subsection 2.3. Consequently, this approach results in the loss of the symmetrical structure that is attained by considering the N-particle system $\mathcal{X}_t^{i,N}$ in its entirety.

References

[1] Ö. Deniz Akyildiz, Francesca Romana Crucinio, Mark Girolami, Tim Johnston, and Sotirios Sabanis. "Interacting Particle Langevin Algorithm for Maximum Marginal Likelihood Estimation". In: *ArXiv* abs/2303.13429 (2023).

[2] Valentin De Bortoli, Alain Durmus, Marcelo Pereyra, and Ana Fernandez Vidal. "Efficient stochastic optimisation by unadjusted Langevin Monte Carlo". In: *Statistics and Computing* 31 (2019).

[3] Nicolas Brosse, Alain Durmus, Éric Moulines, and Sotirios Sabanis. "The tamed unadjusted Langevin algorithm". In: *Stochastic Processes and their Applications* 129.10 (2019), pp. 3638–3663. ISSN: 0304-4149.

[4] Ngoc Huy Chau, Éric Moulines, Miklós Rásonyi, Sotirios Sabanis, and Ying Zhang. "On Stochastic Gradient Langevin Dynamics with Dependent Data Streams: The Fully Nonconvex Case". In: *SIAM Journal on Mathematics of Data Science* 3.3 (2021), pp. 959–986.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin. "Maximum Likelihood from Incomplete Data via the EM Algorithm". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 39.1 (1977), pp. 1–38. ISSN: 00359246.

[6] Alain Durmus and Éric Moulines. "High-dimensional Bayesian inference via the unadjusted Langevin algorithm". In: *Bernoulli* 25.4A (2019), pp. 2854–2882.

[7] Alain Durmus and Éric Moulines. "Nonasymptotic convergence analysis for the unadjusted Langevin algorithm". In: *The Annals of Applied Probability* 27.3 (2017), pp. 1551–1587.

[8] Da Prato G. *An Introduction to Infinite-Dimensional Analysis*. Springer, 2006.

[9] Martin Hutzenthaler, Arnulf Jentzen, and Peter E. Kloeden. "Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 467.2130 (Dec. 2010), pp. 1563–1576. ISSN: 1471-2946.

[10] Martin Hutzenthaler, Arnulf Jentzen, and Peter E. Kloeden. "Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients". In: *The Annals of Applied Probability* 22.4 (Aug. 2012). ISSN: 1050-5164.

[11] Chaman Kumar and Sotirios Sabanis. "On explicit approximations for Lévy driven SDEs with super-linear diffusion coefficients". In: *Electronic Journal of Probability* 22.none (2017), pp. 1–19.

[12] Juan Kuntz, Jen Ning Lim, and Adam M. Johansen. "Particle algorithms for maximum likelihood training of latent variable models". In: *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*. Ed. by Francisco Ruiz, Jennifer Dy, and Jan-Willem van de Meent. Vol. 206. Proceedings of Machine Learning Research. PMLR, 2023, pp. 5134–5180.

[13] Dong-Young Lim, Ariel Neufeld, Sotirios Sabanis, and Ying Zhang. "Langevin dynamics based algorithm e-THϵO POULA for stochastic optimization problems with discontinuous stochastic gradient". In: *ArXiv* abs/2210.13193 (2022).

[14] Dong-Young Lim, Ariel Neufeld, Sotirios Sabanis, and Ying Zhang. "Non-asymptotic estimates for TUSLA algorithm for non-convex learning with applications to neural networks with ReLU activation function". In: *IMA Journal of Numerical Analysis* (June 2023). ISSN: 0272-4979.

[15] Dong-Young Lim and Sotirios Sabanis. "Polygonal Unadjusted Langevin Algorithms: Creating stable and efficient adaptive algorithms for neural networks". In: *To appear, Journal of Machine Learning Research* (2024).

[16] Attila Lovas, Iosif Lytras, Miklós Rásonyi, and Sotirios Sabanis. "Taming Neural Networks with TUSLA: Nonconvex Learning via Adaptive Stochastic Gradient Langevin Algorithms". In: *SIAM Journal on Mathematics of Data Science* 5.2 (2023), pp. 323–345.

[17] Xuerong Mao. *Stochastic Differential Equations and Applications*. Elsevier, 2007.

[18] Sotirios Sabanis. "A note on tamed Euler approximations". In: *Electronic Communications in Probability* 18.none (2013), pp. 1–10.

[19] Sotirios Sabanis. "Euler approximations with varying coefficients: The case of superlinearly growing diffusion coefficients". In: *The Annals of Applied Probability* 26.4 (Aug. 2016). ISSN: 1050-5164.

[20] Sotirios Sabanis and Ying Zhang. "Higher order Langevin Monte Carlo algorithm". In: *Electronic Journal of Statistics* 13.2 (2019), pp. 3805–3850.

[21] Sotirios Sabanis and Ying Zhang. "On explicit order 1.5 approximations with varying coefficients: The case of super-linear diffusion coefficients". In: *Journal of Complexity* 50 (Feb. 2019), pp. 84–115. ISSN: 0885-064X.

A Proofs of subsection 3

A.1 Uniform moments bounds

Lemma 1. Consider either the dynamics given by (1)-(2) or (3)-(4) and let Assumptions A1-A2 hold, then there exists a constant $C > 0$ such that

$$\sup_{t\geq 0} \mathbb{E}\left[|\mathcal{Z}_t^N|^2\right] \leq C.$$

Proof. Consider the rescaling dynamics as given in (12) for the original system of equations (1)-(2):

$$|\mathcal{Z}_t^N|^2 = |\vartheta_t^N|^2 + \frac{1}{N}\sum_{i=1}^{N} |\mathcal{X}_t^{i,N}|^2.$$

By applying standard arguments using stopping times, Grönwall's lemma and Fatou's lemma, one obtains that there is a constant c, which depends on time, such that $\sup_{0\leq t\leq T} \mathbb{E}\left[|\mathcal{Z}_t^N|^2\right] \leq c$ for any $T > 0$. Moreover, through the use of Itô's formula one derives

$$\begin{aligned}
|\mathcal{Z}_t^N|^2 \leq{}& |\vartheta_0^N|^2 - 2\int_0^t \langle \vartheta_s^N, \frac{1}{N}\sum_{i=1}^{N} h^\theta(\mathcal{V}_s^{i,N})\rangle ds + \frac{2d^\theta t}{N} + 2\sqrt{\frac{2}{N}}\int_0^t \vartheta_s^N dB_s^{0,N} \\
&+ \frac{1}{N}\sum_{i=1}^{N} |\mathcal{X}_0^{i,N}|^2 - \frac{2}{N}\sum_{i=1}^{N}\int_0^t \langle \mathcal{X}_s^{i,N}, h^x(\mathcal{V}_s^{i,N})\rangle ds + 2d^x t + \frac{2\sqrt{2}}{N}\sum_{i=1}^{N}\int_0^t \mathcal{X}_s^{i,N} dB_s^{i,N} ds \\
\leq{}& |\mathcal{Z}_0^N|^2 - 2\int_0^t \frac{1}{N}\sum_{i=1}^{N} \langle \mathcal{V}_s^{i,N}, h(\mathcal{V}_s^{i,N})\rangle ds + 2(d^\theta/N + d^x)t \\
&+ 2\sqrt{\frac{2}{N}}\int_0^t \vartheta_s^N dB_s^{0,N} + \frac{2\sqrt{2}}{N}\sum_{i=1}^{N}\int_0^t \mathcal{X}_s^{i,N} dB_s^{i,N} ds \\
\leq{}& |\mathcal{Z}_0^N|^2 - \mu\int_0^t |\mathcal{Z}_s^N|^2 ds + 2bt + 2(d^\theta/N + d^x)t \\
&+ 2\sqrt{\frac{2}{N}}\int_0^t \vartheta_s^N dB_s^{0,N} + \frac{2\sqrt{2}}{N}\sum_{i=1}^{N}\int_0^t \mathcal{X}_s^{i,N} dB_s^{i,N} ds.
\end{aligned}$$

Taking the expectation on both sides results in

$$\mathbb{E}\left[|\mathcal{Z}_t^N|^2\right] \leq \mathbb{E}\left[|\mathcal{Z}_0^N|^2\right] - \mu \int_0^t \mathbb{E}\left[|\mathcal{Z}_s^N|^2\right] ds + 2(b + d^\theta/N + d^x)t.$$

One notices that

$$\frac{d}{dt} e^{\mu t} \mathbb{E}\left[|\mathcal{Z}_t^N|^2\right] \leq e^{\mu t} C,$$

which via integrating implies

$$\mathbb{E}\left[|\mathcal{Z}_t^N|^2\right] \leq C/\mu \Rightarrow \sup_{t \geq 0} \mathbb{E}\left[|\mathcal{Z}_t^N|^2\right] \leq C,$$

as C is a constant independent of time. The corresponding result regarding equations (3)-(4) follows by going through the same steps, where the calculations differ only up to a constant in the SDEs coefficients. □

A.2 Uniformly tamed scheme - tIPLAu

A.2.1 Key quantities for the proof of the main Lemmas.

The following definitions that refer to the rescaled dynamics (13)-(14) of the algorithm tIPLAu and its continuous time interpolations (15)-(16) are given as:

$$Z_n^{\lambda,u} = \left(\theta_{n+1}^{\lambda,u}, N^{-1/2} X_{n+1}^{1,\lambda,u}, \ldots, N^{-1/2} X_{n+1}^{N,\lambda,u}\right), \tag{22}$$

$$\overline{Z}_t^{\lambda,u} = \left(\overline{\theta}_t^{\lambda,u}, N^{-1/2} \overline{X}_t^{1,\lambda,u}, \ldots, N^{-1/2} \overline{X}_t^{N,\lambda,u}\right), \tag{23}$$

$$\mathcal{Z}_{\lambda t}^N = \left(\vartheta_{\lambda t}^N, N^{-1/2} \mathcal{X}_{\lambda t}^{1,N}, \ldots, N^{-1/2} \mathcal{X}_{\lambda t}^{N,N}\right). \tag{24}$$

A.2.2 Moment and increment bounds

Lemma 2. Let A1, A2 and A4 hold. Then, for any $0 \leq \lambda < N^p/4\mu$, it holds that

$$\mathbb{E}\left[|Z_n^{\lambda,u}|^2\right] \leq C_{|z_0|,\mu,b}(1 + d^\theta/N + d^x),$$

for a constant $C > 0$ independent of N, n, λ, d^x and d^θ, which is given in the proof.

Proof. Consider the rescaled iterates as described in (22):

$$
\begin{aligned}
\left|Z_{n+1}^{\lambda,u}\right|^2 &= \left|\theta_{n+1}^{\lambda,u}\right|^2 + \frac{1}{N}\sum_{i=1}^{N}\left|X_{n+1}^{i,\lambda,u}\right|^2 \\
&= \left|\theta_n^{\lambda,u} - \frac{\lambda}{N^{p+1}}\sum_{i=1}^{N} h_{\lambda,u}^{\theta}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})\right|^2 + \frac{2\lambda}{N^{p+1}}|\xi_{n+1}^{(0)}|^2 \\
&+2\sqrt{\frac{2\lambda}{N^{p+1}}}\left\langle \theta_n^{\lambda,u} - \frac{\lambda}{N^{p+1}}\sum_{i=1}^{N} h_{\lambda,u}^{\theta}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u}), \xi_{n+1}^{(0)}\right\rangle \\
&+\frac{1}{N}\sum_{i=1}^{N}\left(\left|X_n^{i,\lambda,u} - \frac{\lambda}{N^p}h_{\lambda,u}^{x}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})\right|^2 + \frac{2\lambda}{N^p}|\xi_{n+1}^{(i)}|^2\right. \\
&\left. + 2\sqrt{\frac{2\lambda}{N^p}}\left\langle X_n^{i,\lambda,u} - \frac{\lambda}{N^p}h_{\lambda,u}^{x}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u}), \xi_{n+1}^{(i)}\right\rangle\right).
\end{aligned}
$$

Taking the conditional expectation on both sides with respect to the filtration generated by $Z_n^{\lambda,u}$ the cross terms are vanishing to 0 due to the independence between the $\xi_{n+1}^{(i)}$'s and $Z_n^{\lambda,u}$, yielding

$$
\begin{aligned}
\mathbb{E}\left[\left|Z_{n+1}^{\lambda,u}\right|^2 |Z_n^{\lambda,u}\right] &= \mathbb{E}\left[\left|\theta_n^{\lambda,u}\right|^2 |Z_n^{\lambda,u}\right] - \frac{2\lambda}{N^{p+1}}\sum_{i=1}^{N}\mathbb{E}\left[\langle\theta_n^{\lambda,u}, h_{\lambda,u}^{\theta}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})\rangle|Z_n^{\lambda,u}\right] \\
&+ \frac{\lambda^2}{N^{2(p+1)}}\mathbb{E}\left[\left|\sum_{i=1}^{N} h_{\lambda,u}^{\theta}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})\right|^2 |Z_n^{\lambda,u}\right] + \frac{2\lambda d^{\theta}}{N^{p+1}} \\
&+ \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\left|X_n^{i,\lambda,u}\right|^2 |Z_n^{\lambda,u}\right] \\
&- \frac{2\lambda}{N^{p+1}}\sum_{i=1}^{N}\mathbb{E}\left[\langle X_n^{i,\lambda}, h_{\lambda,u}^{x}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})\rangle|Z_n^{\lambda,u}\right] \\
&+ \frac{\lambda^2}{N^{2p+1}}\sum_{i=1}^{N}\mathbb{E}\left[|h_{\lambda,u}^{x}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})|^2|Z_n^{\lambda,u}\right] + \frac{2\lambda d^x}{N^p}.
\end{aligned}
$$

Furthermore by using the elementary inequality $(t_1 + \ldots + t_m)^p \leq m^{p-1}(t_1^p + \ldots, t_m^p)$ and the fact that all of the expressions within the conditional expectations are measurable we obtain

$$
\begin{aligned}
\mathbb{E}\left[\left|Z_{n+1}^{\lambda,u}\right|^2 | Z_n^{\lambda,u}\right] \le |\theta_n^{\lambda,u}| &+ \frac{1}{N}\sum_{i=1}^N |X_n^{i,\lambda,u}|^2 + \frac{2\lambda d^\theta}{N^{p+1}} + + \frac{2\lambda d^x}{N^p} \\
&- \frac{2\lambda}{N^{p+1}}\sum_{i=1}^N \langle \theta_n^\lambda, h_{\lambda,u}^\theta(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})\rangle \\
&- \frac{2\lambda}{N^{p+1}}\sum_{i=1}^N \langle X_n^{i,\lambda,u}, h_{\lambda,u}^x(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})\rangle \\
&+ \frac{\lambda^2}{N^{2p+1}}\sum_{i=1}^N |h_\lambda^\theta(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})|^2 + \frac{\lambda^2}{N^{2p+1}}\sum_{i=1}^N |h_{\lambda,u}^x(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})|^2.
\end{aligned}
$$

Recalling that $(h_{\lambda,u}^\theta, h_{\lambda,u}^x) = h_{\lambda,u}$, further leads to

$$
\begin{aligned}
\mathbb{E}\left[\left|Z_{n+1}^{\lambda,u}\right|^2 | Z_n^{\lambda,u}\right] \le \left|Z_n^{\lambda,u}\right|^2 &- \frac{2\lambda}{N^{p+1}}\sum_{i=1}^N \langle V_n^{i,\lambda,u}, h_{\lambda,u}(V_n^{i,\lambda,u})\rangle + \frac{\lambda^2}{N^{2p+1}}\sum_{i=1}^N |h_{\lambda,u}(V_n^{i,\lambda,u})|^2 \\
&+ \frac{2\lambda}{N^p}(d^\theta/N + d^x).
\end{aligned}
$$

Now using Properties 1 and 3 of the taming function, we get

$$
\begin{aligned}
\mathbb{E}\left[\left|Z_{n+1}^{\lambda,u}\right|^2 | Z_n^{\lambda,u}\right] &\le \left|Z_n^{\lambda,u}\right|^2 - \frac{\mu\lambda}{N^{p+1}}\sum_{i=1}^N \left|V_n^{i,\lambda,u}\right|^2 + \frac{2\lambda b}{N^p} \\
&\quad + \frac{2\lambda^2\mu^2}{N^{2p+1}}\sum_{i=1}^N \left|V_n^{i,\lambda,u}\right|^2 + \frac{2\lambda^2\lambda^{-1}N^p}{N^{2p}} + \frac{2\lambda}{N^p}(d^\theta/N + d^x) \\
&\le \left(1 - \frac{\lambda\mu}{N^p} + \frac{2\lambda^2\mu^2}{N^{2p}}\right)\left|Z_n^{\lambda,u}\right|^2 + \frac{2\lambda}{N^p}\left(b + 1 + d^\theta/N + d^x\right).
\end{aligned}
$$

By considering the restriction $\lambda < \dfrac{N^p}{4\mu}$ and iterating the above bound, we conclude with

$$
\begin{aligned}
\mathbb{E}\left[\left|Z_{n+1}^{\lambda,u}\right|^2\right] &\le \left(1 - \frac{\mu\lambda}{2N^p}\right)^n \mathbb{E}\left[\left|Z_0^{\lambda,u}\right|^2\right] + \frac{1 - (1 - \mu\lambda/2)^n}{\mu\lambda/2N^p}\frac{2\lambda}{N^p}(b + 1 + d^\theta/N + d^x) \\
&\le \mathbb{E}\left[\left|Z_0^{\lambda,u}\right|^2\right] + \frac{4}{\mu}(b + 1 + d^\theta/N + d^x) \\
&\le \left(\mathbb{E}\left[\left|Z_0^{\lambda,u}\right|^2\right] + \frac{4}{\mu}(b+1)\right)(1 + d^\theta/N + d^x) \\
&\le C_{|z_0|,\mu,b}(1 + d^\theta/N + d^x).
\end{aligned}
$$

□

Lemma 3. Let A1, A2 and A4 hold. Then, for any $0 \le \lambda < N^p/4\mu$ and $q \in [2, 2\ell+1) \cap \mathbb{N}$, it holds that,

$$\mathbb{E}\left[|Z_n^{\lambda,u}|^{2q}\right] \le C_{|z_0|,b,q,\mu}(1 + d^\theta/N + d^x)^q,$$

for a constant $C > 0$ independent of N, n, λ, d^x and d^θ.

Proof. To make the forthcoming calculations more readable we define the following auxiliary processes:

$$\Delta_n^{\lambda,\theta} = \theta_n^{\lambda,u} - \frac{\lambda}{N^{p+1}} \sum_{i=1}^{N} h_{\lambda,u}^{\theta}\left(\theta_n^{\lambda}, X_n^{i,\lambda}\right), \ G_n^{\lambda,\theta} = \sqrt{\frac{2\lambda}{N^{p+1}}}\xi_{n+1}^{(0)},$$
$$\Delta_n^{\lambda,x,i} = X_n^{i,\lambda,u} - \frac{\lambda}{N^p} h_{\lambda,u}^{x}\left(\theta_n^{\lambda}, X_n^{i,\lambda}\right), \ G_n^{\lambda,x,i} = \sqrt{\frac{2\lambda}{N^p}}\xi_{n+1}^{(i)},$$

but from this point onwards till the completion of the proof we denote λ/N^p by just λ to make the computations less hectic. Recall that have already established via Lemma 2 the bound

$$A_n^{\lambda} := |\Delta_n^{\lambda,\theta}|^2 + \frac{1}{N}\sum_{i=1}^{N} |\Delta_n^{\lambda,i,x}|^2 \le \left(1 - \frac{\lambda\mu}{2}\right)|Z_n^{\lambda,u}|^2 + 2\lambda C,$$

and lastly let us define the quantity

$$B_n^{\lambda} = 2\langle \Delta_n^{\theta}, G_n^{\lambda,\theta}\rangle + \frac{2}{N}\sum_{i=1}^{N}\langle \Delta_n^{\lambda,x,i}, G_n^{\lambda,x,i}\rangle + |G_n^{\lambda,\theta}|^2 + \frac{1}{N}\sum_{i=1}^{N}|G_n^{\lambda,x,i}|^2.$$

Regarding the $2q$-th moment one writes

$$\begin{aligned}|Z_{n+1}^{\lambda,u}|^{2q} &= \left(A_n^{\lambda} + B_n^{\lambda}\right)^q \\ &\le (A_n^{\lambda})^q + 2q(A_n^{\lambda})^{q-1}B_n^{\lambda} + \sum_{k=2}^{q}\binom{q}{k}|A_n^{\lambda}|^{q-k}|B_n^{\lambda}|^k. \end{aligned} \tag{25}$$

We shall deal with each term separately

$$\begin{aligned}\mathbb{E}\left[(A_n^{\lambda})^q | Z_n^{\lambda,u}\right] = (A_n^{\lambda})^q &\le \left(\left(1 - \frac{\lambda\mu}{2}\right)|Z_n^{\lambda,u}|^2 + 2\lambda C\right)^q \\ &\le \left(1 + \frac{\lambda\mu}{4}\right)^{q-1}\left(1 - \frac{\lambda\mu}{2}\right)^q |Z_n^{\lambda,u}|^{2q} + \left(1 + \frac{4}{\lambda\mu}\right)^{q-1} 2^q\lambda^q C^q \\ &\le \left(1 - \frac{\lambda\mu}{4}\right)^{q-1}\left(1 - \frac{\lambda\mu}{2}\right)|Z_n^{\lambda,u}|^{2q} + \left(\lambda + \frac{4}{\mu}\right)^{q-1}\lambda(2C)^q \\ &\le r_q^{\lambda}|Z_n^{\lambda,u}|^{2q} + w_q^{\lambda}. \end{aligned} \tag{26}$$

where $r_q^{\lambda} = (1-\lambda\mu/4)^{q-1}(1-\lambda\mu/2)$ and $w_q^{\lambda} = (\lambda+4/\mu)^{q-1}\lambda(2C)^q$. Notice that we made use of the elementary equation $(r+s)^q \le (1+\epsilon)^{q-1}r^q + (1+1/\epsilon)^{q-1}s^q$ for the choice

$\epsilon = \lambda\mu/4$. Additionally for the shake of simplicity, let us denote $d = d^\theta/N + d^x$. Now on a similar note with the previous term

$$\begin{aligned}\mathbb{E}\left[2q(A_n^\lambda)^{q-1}B_n^\lambda|Z_n^{\lambda,u}\right] &= 2q(A_n^\lambda)^{q-1}\mathbb{E}\left[B_n^\lambda|Z_n^{\lambda,u}\right] = 4q\lambda\left(d^\theta/N + d^x\right)(A_n^\lambda)^{q-1} \\ &\leq 4q\lambda\left(d^\theta/N + d^x\right)\left(r_{q-1}^\lambda|Z_n^{\lambda,u}|^{2(q-1)} + w_{q-1}^\lambda\right) \\ &\leq 4q\lambda d\left(r_{q-1}^\lambda|Z_n^{\lambda,u}|^{2(q-1)} + w_{q-1}^\lambda\right). \end{aligned} \tag{27}$$

The 3rd term on (25) can be further expanded to

$$\begin{aligned}\sum_{k=2}^{q}\binom{q}{k}|A_n^\lambda|^{q-k}|B_n^\lambda|^k &= \sum_{m=0}^{q-2}\binom{q}{m+2}|A_n^\lambda|^{q-2-m}|B_n^\lambda|^{m+2} \\ &= \frac{q}{m+2}\frac{q-1}{m+1}\sum_{m=0}^{q-2}\binom{q-2}{m}|A_n^\lambda|^{q-2-m}|B_n^\lambda|^m|B_n^\lambda|^2 \\ &\leq q(q-1)\left(|A_n^\lambda| + |B_n^\lambda|\right)^{q-2}|B_n^\lambda|^2 \\ &\leq q(q-1)2^{q-3}|A_n^\lambda|^{q-2}|B_n^\lambda|^2 + q(q-1)2^{q-3}|B_n^\lambda|^q \\ &:= D + F. \end{aligned} \tag{28}$$

Taking the expectation of the above terms yields

$$\mathbb{E}[D|Z_n^{\lambda,u}] = q(q-1)2^{q-3}|A_n^\lambda|^{q-2}\mathbb{E}\left[|B_n^\lambda|^2|Z_n^{\lambda,u}\right],$$

where

$$\begin{aligned}\mathbb{E}\left[|B_n^\lambda|^2|Z_n^{\lambda,u}\right] &= \mathbb{E}\left[\left|2\langle\Delta_n^\theta, G_n^{\lambda,\theta}\rangle + \frac{2}{N}\sum_{i=1}^{N}\langle\Delta_n^{\lambda,x,i}, G_n^{\lambda,x,i}\rangle|G_n^{\lambda,\theta}|^2 + \frac{1}{N}\sum_{i=1}^{N}|G_n^{\lambda,x,i}|^2\right|^2|Z_n^{\lambda,u}\right] \\ &\leq 4\left(\mathbb{E}\left[4|\Delta_n^{\lambda,\theta}|^2|G_n^{\lambda,\theta}|^2|Z_n^{\lambda,u}\right] + \mathbb{E}\left[\frac{4}{N}\sum_{i=1}^{N}|\Delta_n^{\lambda,x,i}|^2|G_n^{\lambda,x,i}|^2|Z_n^{\lambda,u}\right]\right. \\ &\quad\left. + \mathbb{E}\left[|G_n^{\lambda,x,i}|^4|Z_n^{\lambda,u}\right] + \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}|G_n^{\lambda,x,i}|^4|Z_n^{\lambda,u}\right]\right) \\ &\leq 4\left(4|\Delta_n^{\lambda,\theta}|^2\mathbb{E}\left[|G_n^{\lambda,\theta}|^2|Z_n^{\lambda,u}\right] + \frac{4}{N}\sum_{i=1}^{N}|\Delta_n^{\lambda,x,i}|^2\mathbb{E}\left[|G_n^{\lambda,x,i}|^2|Z_n^{\lambda,u}\right]\right. \\ &\quad\left. + \mathbb{E}\left[|G_n^{\lambda,x,i}|^4|Z_n^{\lambda,u}\right] + \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}|G_n^{\lambda,x,i}|^4|Z_n^{\lambda,u}\right]\right). \end{aligned}$$

Recalling that each $G_n^{\lambda,\cdot,i}$ is distributed according to a Gaussian distribution, we subsequently derive

$$\begin{aligned}
\mathbb{E}\left[|B_n^\lambda|^2|Z_n^{\lambda,u}\right] &\leq 4\left(4(2\lambda d^\theta/N)|\Delta_n^{\lambda,\theta}|^2+\frac{4}{N}\sum_{i=1}^N(2\lambda d^x)|\Delta_n^{\lambda,x,i}|^2+3(2\lambda d^\theta/N)^2+3(2\lambda d^x)^2\right)\\
&\leq 16(2\lambda d)\left(|\Delta_n^{\lambda,\theta}|^2+\frac{1}{N}\sum_{i=1}^N|\Delta_n^{\lambda,x,i}|^2\right)+24(2\lambda d)^2\\
&\leq 16(2\lambda d)A_n^\lambda+24(2\lambda d)^2.
\end{aligned}$$

By substituting this result, the term D can be expressed as follows

$$\begin{aligned}
\mathbb{E}[D|Z_n^{\lambda,u}] &= q(q-1)2^{q-3}|A_n^\lambda|^{q-2}\left(16(2\lambda d)A_n^\lambda+24(2\lambda d)^2\right)\\
&= 16(2\lambda d)q(q-1)2^{q-3}\left(r_{q-1}^\lambda|Z_n^{\lambda,u}|^{2(q-1)}+w_{q-1}^\lambda\right)\\
&\quad+24(2\lambda d)^2q(q-1)2^{q-3}\left(r_{q-2}^\lambda|Z_n^{\lambda,u}|^{2(q-2)}+w_{q-2}^\lambda\right).
\end{aligned}\tag{29}$$

Additionally,

$$\begin{aligned}
\mathbb{E}[F|Z_n^{\lambda,u}] &= q(q-1)2^{q-3}|B_n^\lambda|^q\\
&= q(q-1)2^{q-3}\mathbb{E}\left[\left|2\langle\Delta_n^\theta,G_n^{\lambda,\theta}\rangle+\frac{2}{N}\sum_{i=1}^N\langle\Delta_n^{\lambda,x,i},G_n^{\lambda,x,i}\rangle\right.\right.\\
&\quad\left.\left.+|G_n^{\lambda,\theta}|^2+\frac{1}{N}\sum_{i=1}^N|G_n^{\lambda,x,i}|^2\right|^q|Z_n^{\lambda,u}\right]\\
&\leq q(q-1)2^{q-3}4^{q-1}\left(\mathbb{E}\left[2^q|\Delta_n^{\lambda,\theta}|^q|G_n^{\lambda,\theta}|^q|Z_n^{\lambda,u}\right]\right.\\
&\quad+\mathbb{E}\left[\frac{2^q}{N^q}\left|\sum_{i=1}^N|\Delta_n^{\lambda,x,i}||G_n^{\lambda,x,i}|\right|^q|Z_n^{\lambda,u}\right]+\mathbb{E}\left[|G_n^{\lambda,\theta}|^{2q}|Z_n^{\lambda,u}\right]\\
&\quad\left.+\mathbb{E}\left[\frac{1}{N^q}N^{q-1}\sum_{i=1}^N|G_n^{\lambda,x,i}|^{2q}|Z_n^{\lambda,u}\right]\right).
\end{aligned}\tag{30}$$

We handle the second term in (30) by applying the multinomial expansion:

$$\left|\sum_{i=1}^N|\Delta_n^{\lambda,x,i}||G_n^{\lambda,x,i}|\right|^q=\sum_{k_1+\ldots+k_N=q}\binom{q}{k_1,\ldots,k_N}\prod_{i=1}^N\left(|\Delta_n^{\lambda,x,i}||G_n^{\lambda,x,i}|\right)^{k_i},$$

and using the fact that $\Delta_n^{\lambda,x,i}$ are $Z_n^{\lambda,u}-$measurable, hence

$$
\begin{aligned}
&\frac{2^q}{N^q}\mathbb{E}\left[\left|\sum_{i=1}^N |\Delta_n^{\lambda,x,i}|G_n^{\lambda,x,i}|\right|^q |Z_n^{\lambda,u}\right] \\
&= \frac{2^q}{N^q}\mathbb{E}\left[\sum_{k_1+\ldots+k_N=q}\binom{q}{k_1,\ldots,k_N}\prod_{i=1}^N\left(|\Delta_n^{\lambda,x,i}||G_n^{\lambda,x,i}|\right)^{k_i}|Z_n^{\lambda,u}\right] \\
&= \frac{2^q}{N^q}\sum_{k_1+\ldots+k_N=q}\binom{q}{k_1,\ldots,k_N}\prod_{i=1}^N|\Delta_n^{\lambda,x,i}|^{k_i}\prod_{i=1}^N\mathbb{E}\left[|G_n^{\lambda,x,i}|^{k_i}|Z_n^{\lambda,u}\right] \\
&= \frac{2^q}{N^q}\sum_{k_1+\ldots+k_N=q}\binom{q}{k_1,\ldots,k_N}\prod_{i=1}^N|\Delta_n^{\lambda,x,i}|^{k_i}\prod_{i=1}^N\left((2\lambda d^x)^{k_i/2}k_i!!\right) \\
&= \frac{2^q}{N^q}\sum_{k_1+\ldots+k_N=q}\binom{q}{k_1,\ldots,k_N}\prod_{i=1}^N|\Delta_n^{\lambda,x,i}|^{k_i}(2\lambda d^x)^{\sum_{i=1}^N k_i/2}k_i!! \\
&\leq \frac{2^q q!!(2\lambda d^x)^{q/2}}{N^q}\sum_{k_1+\ldots+k_N=q}\binom{q}{k_1,\ldots,k_N}\prod_{i=1}^N|\Delta_n^{\lambda,x,i}|^{k_i} \\
&= \frac{2^q q!!(2\lambda d^x)^{q/2}}{N^q}\left(\sum_{i=1}^N|\Delta_n^{\lambda,x,i}|\right)^q \leq \frac{2^q q!!(2\lambda d^x)^{q/2}}{N^{q-1}}\left(\sum_{i=1}^N|\Delta_n^{\lambda,x,i}|^2\right)^{q/2} \\
&\leq 2^q q!!(2\lambda d^x)^{q/2}\left(\frac{1}{N}\sum_{i=1}^N|\Delta_n^{\lambda,x,i}|^2\right)^{q/2}.
\end{aligned}
$$

Plugging these results into (30) we get

$$
\begin{aligned}
\mathbb{E}[F|Z_n^{\lambda,u}] \leq\ & q(q-1)2^{3q-5}\Big(2^q q!!(2\lambda d^\theta/N)^{q/2}\left(|\Delta_n^{\lambda,\theta}|^2\right)^{q/2} \\
&+2^q q!!(2\lambda d^x)^{q/2}\left(\frac{1}{N}\sum_{i=1}^N|\Delta_n^{\lambda,x,i}|^2\right)^{q/2}+(2q)!!(2\lambda d^\theta/N)^q+(2q)!!(2\lambda d^x)^q\Big) \\
&\leq q(q-1)2^{3q-5}\left(2^q q!!(2\lambda d)^{q/2}\left(|\Delta_n^{\lambda,\theta}|^2+\frac{1}{N}\sum_{i=1}^N|\Delta_n^{\lambda,x,i}|^2\right)^{q/2}+2(2q)!!(2\lambda d)^q\right) \\
&\leq q(q-1)2^{4q-5}(2q)!!\left((2\lambda d)^{q/2}\left|A_n^\lambda\right|^{q/2}+(2\lambda d)^q\right) \\
&\leq q(q-1)2^{4q-5}(2q)!!(2\lambda d)^{q/2}\left(r_{q/2}^\lambda|Z_n^{\lambda,u}|^q+w_{q/2}^\lambda\right) \\
&+q(q-1)2^{4q-5}(2q)!!(2\lambda d)^q.
\end{aligned}
\tag{31}
$$

Combining (29),(31) yields the following bound for (28),

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{k=2}^{q}\binom{q}{k}|A_n^\lambda|^{q-k}|B_n^\lambda|^k\right] \\
&\leq 16(2\lambda d)q(q-1)2^{q-3}\left(r_{q-1}^\lambda|Z_n^{\lambda,u}|^{2(q-1)}+w_{q-1}^\lambda\right) \\
&+24(2\lambda d)^2q(q-1)2^{q-3}\left(r_{q-2}^\lambda|Z_n^{\lambda,u}|^{2(q-2)}+w_{q-2}^\lambda\right) \\
&+(2q)!!(2\lambda d)^{q/2}q(q-1)2^{4q-5}\left(r_{q/2}^\lambda|Z_n^{\lambda,u}|^{q}+w_{q/2}^\lambda\right) \\
&+(2q)!!(2\lambda d)^{q}q(q-1)2^{4q-5}. \qquad (32)
\end{aligned}
$$

Hence one obtains for (25) via (26),(27) and (32),

$$
\begin{aligned}
\mathbb{E}\left[|Z_{n+1}|^{2q}|Z_n\right] &\leq r_q^\lambda|Z_n^{\lambda,u}|^{2q}+w_q^\lambda \\
&+4q\left(2\lambda d\right)\left(1+4(q-1)2^{q-3}\right)\left(r_{q-1}^\lambda|Z_n^{\lambda,u}|^{2(q-1)}+w_{q-1}^\lambda\right) \\
&+24(2\lambda d)^2q(q-1)2^{q-3}\left(r_{q-2}^\lambda|Z_n^{\lambda,u}|^{2(q-2)}+w_{q-2}^\lambda\right) \\
&+(2q)!!(2\lambda d)^{q/2}q(q-1)2^{4q-5}\left(r_{q/2}^\lambda|Z_n^{\lambda,u}|^{q}+w_{q/2}^\lambda\right) \\
&+(2q)!!(2\lambda d)^{q}q(q-1)2^{4q-5}. \qquad (33)
\end{aligned}
$$

Consider $|Z_n^{\lambda,u}|\geq\sqrt{\frac{2d}{\mu/4}}\left\{(2q)!!q(q-1)2^{4q-5}\right\}^{1/2}\geq\sqrt{\frac{2d}{\mu/4}}\left\{(2q)!!q(q-1)2^{4q-5}\right\}^{1/q}$. Then one observes

$$
\begin{aligned}
\mathbb{E}\left[|Z_{n+1}^{\lambda,u}|^{2q}|Z_n\right] &\leq r_q^\lambda|Z_n^{\lambda,u}|^{2q}+w_q^\lambda \\
&+\frac{\lambda\mu}{2\cdot 4}|Z_n^{\lambda,u}|^2\left(r_{q-1}^\lambda|Z_n^{\lambda,u}|^{2(q-1)}\right)+\frac{(\lambda\mu)^2}{2\cdot 4^2}|Z_n^{\lambda,u}|^4\left(r_{q-2}^\lambda|Z_n^{\lambda,u}|^{2(q-2)}\right) \\
&+\frac{(\lambda\mu)^{q/2}}{2\cdot 4^{q/2}}|Z_n^{\lambda,u}|^q\left(r_{q/2}^\lambda|Z_n^{\lambda,u}|^{q}\right)+(2q)!!q(q-1)2^{4q-5}(2\lambda d)^q \\
&+(2q)!!q(q-1)2^{4q-5}\left((2\lambda d)w_{q-1}^\lambda+(2\lambda d)^2w_{q-2}^\lambda+(2\lambda d)^{q/2}w_{q/2}^\lambda\right) \\
&\leq r_{q/2}^\lambda\left(\left(1-\frac{\lambda\mu}{4}\right)^{q/2}+\frac{1}{2}\left(\frac{\lambda\mu}{4}\right)\left(1-\frac{\lambda\mu}{4}\right)^{q/2-1}\right. \\
&\left.+\frac{1}{2}\left(\frac{\lambda\mu}{4}\right)^2\left(1-\frac{\lambda\mu}{4}\right)^{q/2-2}+\frac{1}{2}\left(\frac{\lambda\mu}{4}\right)^{q/2}\right)|Z_n^{\lambda,u}|^{2q} \\
&+w_q^\lambda+(2q)!!q(q-1)2^{4q-5}(2\lambda d)^q \\
&+(2q)!!q(q-1)2^{4q-5}\left((2\lambda d)w_{q-1}^\lambda+(2\lambda d)^2w_{q-2}^\lambda+(2\lambda d)^{q/2}w_{q/2}^\lambda\right).
\end{aligned}
$$

Using the fact that $\lambda\mu \leq 1$ we get

$$\begin{aligned}
\mathbb{E}\left[|Z_{n+1}^{\lambda,u}|^{2q}|Z_n\right] \leq & \left(1-\frac{\lambda\mu}{2}\right)|Z_n^{\lambda,u}|^{2q} \\
& + w_q^\lambda + (2q)!!q(q-1)2^{4q-5}(2\lambda d)^q \\
& + (2q)!!q(q-1)2^{4q-5}\left((2\lambda d)w_{q-1}^\lambda + (2\lambda d)^2 w_{q-2}^\lambda + (2\lambda d)^{q/2}w_{q/2}^\lambda\right) \\
\leq & \left(1-\frac{\lambda\mu}{2}\right)|Z_n^{\lambda,u}|^{2q} + (2q)!!q(q-1)2^{4q-2}(8/\mu)^{q-1}\lambda(2C)^q.
\end{aligned}$$

Consequently, on $\{|Z_n^{\lambda,u}| \leq \sqrt{\dfrac{2d}{\mu/4}}\{(2q)!!q(q-1)2^{4q-5}\}^{1/2}\}$, we have

$$\begin{aligned}
\mathbb{E}\left[|Z_{n+1}^{\lambda,u}|^{2q}|Z_n\right] \leq & \left(1-\frac{\lambda\mu}{2}\right)|Z_n^{\lambda,u}|^{2q} \\
& + (2q)!!q(q-1)2^{4q-2}(8/\mu)^{q-1}\lambda(2C)^q \\
& + \left((2q)!!q(q-1)2^{4q-2}\right)^{q/2}(8/\mu)^{q-1}\lambda(2C)^q.
\end{aligned}$$

So all in all (33) results to

$$\begin{aligned}
\mathbb{E}\left[|Z_{n+1}^{\lambda,u}|^{2q}\right] & \leq \left(1-\frac{\lambda\mu}{2}\right)\mathbb{E}\left[|Z_n^{\lambda,u}|^{2q}\right] + M_q(8/\mu)^q C^q \\
& \leq C_{|z_0|,b,q,\mu}(1+d^\theta/N+d^x)^q,
\end{aligned}$$

where $C_{|z_0|,b,q,\mu} = \mathbb{E}\left[|Z_0^\lambda|^{2q}\right]M_q(8/\mu)^q(b+1)^q$ and $M_q = (2q)!!q(q-1)2^{6q-5}$. □

Lemma 4. Let A1, A2 and A4 hold. Then, for every $\lambda_0 < N^p/4\mu$, there exists a constant $C > 0$ independent of N, n, λ such that for any $\lambda \in (0, \lambda_0)$ one has

$$\mathbb{E}\left[|Z_{n+1}^{\lambda,u} - Z_n^{\lambda,u}|^4\right] \leq \lambda^2 N^{-2p} C_{|z_0|,\mu,b}(1+d^\theta/N+d^x)^2.$$

Proof.

$$\begin{aligned}
|Z_{n+1}^{\lambda,u} - Z_n^{\lambda,u}|^4 &= \left(|\theta_{n+1}^{\lambda,u} - \theta_n^{\lambda,u}|^2 + \frac{1}{N}\sum_{i=1}^N |X_{n+1}^{i,\lambda,u} - X_n^{i,\lambda,u}|^2\right)^2 \\
&= \left(\left|\frac{-\lambda}{N^{p+1}}\sum_{i=1}^N h_{\lambda,u}^\theta(\theta_n^{\lambda,u}, X_n^{i,\lambda,u}) + \sqrt{\frac{2\lambda}{N^{p+1}}}\xi_{n+1}^{(0)}\right|^2\right. \\
&\quad \left. + \frac{1}{N}\sum_{i=1}^N\left|-\frac{\lambda}{N^p}h_\lambda^x(\theta_n^{\lambda,u}, X_n^{i,\lambda,u}) + \sqrt{\frac{2\lambda}{N^p}}\xi_{n+1}^{(i)}\right|^2\right)^2
\end{aligned}$$

Moreover

$$
\begin{aligned}
|Z_{n+1}^{\lambda,u} - Z_n^{\lambda,u}|^4 \le \Bigg(& \frac{\lambda^2}{N^{2(p+1)}} \Big| \sum_{i=1}^N h_{\lambda,u}^{\theta}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})\Big|^2 + \frac{\lambda^2}{N^{2p+1}} \sum_{i=1}^N |h_{\lambda}^{x}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})|^2 \\
& -2\langle \frac{\lambda}{N^{p+1}} \sum_{i=1}^N h_{\lambda,u}^{\theta}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u}), \sqrt{\frac{2\lambda}{N^{p+1}}}\xi_{n+1}^{(0)}\rangle \\
-2\frac{\lambda}{N^{p+1}} \sum_{i=1}^N \langle h_{\lambda,u}^{x}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u}), & \sqrt{\frac{2\lambda}{N^{p}}}\xi_{n+1}^{(i)}\rangle + \frac{2\lambda}{N^{p+1}}|\xi_{n+1}^{(0)}|^2 + \frac{2\lambda}{N^{p+1}} \sum_{i=1}^N |\xi_{n+1}^{(i)}|^2 \Bigg)^2 \\
\le \frac{\lambda^4}{N^{4p+2}} \left(\sum_{i=1}^N |h_{\lambda,u}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})|^2 \right)^2 & + 2 \left(\frac{\lambda^2}{N^{2p+1}} \sum_{i=1}^N |h_{\lambda,u}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})|^2 \right) \\
& \times \Bigg(-2\langle \frac{\lambda}{N^{p+1}} \sum_{i=1}^N h_{\lambda,u}^{\theta}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u}), \sqrt{\frac{2\lambda}{N^{p+1}}}\xi_{n+1}^{(0)}\rangle \\
-2\frac{\lambda}{N^{p+1}} \sum_{i=1}^N \langle h_{\lambda,u}^{x}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u}), & \sqrt{\frac{2\lambda}{N^{p}}}\xi_{n+1}^{(i)}\rangle + \frac{2\lambda}{N^{p+1}}|\xi_{n+1}^{(0)}|^2 + \frac{2\lambda}{N^{p+1}} \sum_{i=1}^N |\xi_{n+1}^{(i)}|^2 \Bigg) \\
& + \Bigg| -2\langle \frac{\lambda}{N^{p+1}} \sum_{i=1}^N h_{\lambda,u}^{\theta}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u}), \sqrt{\frac{2\lambda}{N^{p+1}}}\xi_{n+1}^{(0)}\rangle \\
-2\frac{\lambda}{N^{p+1}} \sum_{i=1}^N \langle h_{\lambda,u}^{x}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u}), & \sqrt{\frac{2\lambda}{N^{p}}}\xi_{n+1}^{(i)}\rangle + \frac{2\lambda}{N^{p+1}}|\xi_{n+1}^{(0)}|^2 + \frac{2\lambda}{N^{p+1}} \sum_{i=1}^N |\xi_{n+1}^{(i)}|^2 \Bigg|^2 . \qquad (34)
\end{aligned}
$$

The first term in (34) is immediately bounded by the moments of $Z_n^{\lambda,u}$,

$$
\begin{aligned}
\mathbb{E}\left[\frac{\lambda^4}{N^{4p+2}} \left(\sum_{i=1}^N |h_{\lambda,u}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})|^2 \right)^2 |Z_n^{\lambda,u} \right] & \le \frac{\lambda^4}{N^{4p+2}} \left(2\mu^2 \sum_{i=1}^N |V_n^{i,\lambda,u}|^2 + 2\lambda^{-1} N^p \right)^2 \\
& \le 4\mu^4 \frac{\lambda^4}{N^{4p}} \frac{1}{N^2} \left(\sum_{i=1}^N |V_n^{i,\lambda,u}|^2 \right)^2 + 4\frac{\lambda^2}{N^{2p}} \\
& \le \mu^2 \frac{\lambda^2}{N^{2p}} |Z_n^{\lambda,u}|^4 + 4\frac{\lambda^2}{N^{2p}}. \qquad (35)
\end{aligned}
$$

where the last inequality follows due to the restriction $\lambda < N^p/4\mu$.

Notice that in the second term the inner products $\langle \cdot, \cdot \rangle$ vanish under expectation

$$\begin{aligned}
&2\left(\frac{\lambda^2}{N^{2p+1}}\sum_{i=1}^{N}|h_{\lambda,u}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})|^2\right)\left(\frac{2\lambda}{N^{p+1}}\mathbb{E}\left[|\xi_{n+1}^{(0)}|^2|Z_n^{\lambda,u}\right] + \frac{2\lambda}{N^{p+1}}\sum_{i=1}^{N}\mathbb{E}\left[|\xi_{n+1}^{(i)}|^2|Z_n^{\lambda,u}\right]\right) \\
&\quad \leq 2\frac{\lambda^2}{N^{2p}}\left(2\mu^2|Z_n^{\lambda,u}|^2 + 2\lambda^{-1}N^p\right)\left(2\frac{\lambda}{N^p}d^\theta/N + 2\frac{\lambda}{N^p}d^x\right) \leq 8\frac{\lambda^3}{N^{3p}}d(\mu^2|Z_n^{\lambda,u}|^2 + \lambda^{-1}N^p) \\
&\qquad \leq 2d\mu\frac{\lambda^2}{N^{2p}}|Z_n^{\lambda,u}|^2 + 8d\frac{\lambda^2}{N^{2p}}. \qquad (36)
\end{aligned}$$

Using the usual elementary inequality and the Cauchy-Schwartz on the last term of (34) we get

$$\begin{aligned}
&\frac{16\lambda^2}{N^{2p+2}}\left|\sum_{i=1}^{N}h^\theta_{\lambda,u}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})\right|^2 \mathbb{E}\left[|\sqrt{\frac{2\lambda}{N^{p+1}}}\xi_{n+1}^{(0)}|^2\right] \\
&+ \frac{16\lambda^2}{N^{2p+1}}\sum_{i=1}^{N}|h^x_\lambda(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})|^2\mathbb{E}\left[|\sqrt{\frac{2\lambda}{N^p}}\xi_{n+1}^{(i)}|^2\right] \\
&+ \frac{16\lambda^2}{N^{2p+2}}\mathbb{E}\left[|\xi_{n+1}^{(0)}|^4\right] + \frac{16\lambda^2}{N^{2p+1}}\sum_{i=1}^{N}\mathbb{E}\left[|\xi_{n+1}^{(i)}|^4\right] \\
&\leq 16\frac{\lambda^2}{N^{2p}}\left(2\frac{\lambda}{N^p}d^\theta/N\right)\frac{1}{N}\sum_{i=1}^{N}|h^\theta_{\lambda,u}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})|^2 \\
&+ 16\frac{\lambda^2}{N^{2p}}\left(2\frac{\lambda}{N^p}d^x\right)\frac{1}{N}\sum_{i=1}^{N}|h^x_\lambda(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})|^2 \\
&+ 16\frac{\lambda^2}{N^{2p}}(d^\theta/N)^2 + 16\frac{\lambda^2}{N^{2p}}(d^x)^2
\end{aligned}$$

which by using Property 1 of the taming function, can be further written as

$$\begin{aligned}
&\frac{16\lambda^2}{N^{2p+2}}\left|\sum_{i=1}^{N}h^\theta_{\lambda,u}(\theta_n^{\lambda,u}, X_n^{i,\lambda,u})\right|^2 \mathbb{E}\left[|\sqrt{\frac{2\lambda}{N^{p+1}}}\xi_{n+1}^{(0)}|^2\right] \\
&\leq 16\frac{\lambda^3}{N^{3p}}(2d)\frac{1}{N}\sum_{i=1}^{N}\left(2\mu^2|V_n^{i,\lambda,u}|^2 + 2\lambda^{-1}N^p\right) + 16\frac{\lambda^2}{N^{2p}}d^2 \\
&\leq 16\cdot 4\frac{\lambda^3}{N^{3p}}\mu^2 d|Z_n^{\lambda,u}|^2 + 16\cdot 4\frac{\lambda^2}{N^{2p}}d + 16\frac{\lambda^2}{N^{2p}}d^2 \\
&\leq 16d\mu\frac{\lambda^2}{N^{2p}}|Z_n^{\lambda,u}|^2 + 64\frac{\lambda^2}{N^{2p}}d + 16\frac{\lambda^2}{N^{2p}}d^2. \qquad (37)
\end{aligned}$$

Hence (34) in view of (35),(36) and (37) can be further bounded as

$$\mathbb{E}\left[|Z^{\lambda,u}_{n+1} - Z^{\lambda,u}_n|^4 |Z^{\lambda,u}_n\right] \leq \mu^2 \frac{\lambda^2}{N^{2p}} |Z^{\lambda,u}_n|^4 + 18d\mu \frac{\lambda^2}{N^{2p}} |Z^{\lambda,u}_n|^2 + 4\frac{\lambda^2}{N^{2p}}(4 + 18d + d^2)$$
$$\leq \frac{\lambda^2}{N^{2p}} \left(\mu^2 |Z^{\lambda,u}_n|^4 + 18d\mu |Z^{\lambda,u}_n|^2 + 4(4 + 18d + d^2)\right).$$

The final result follows by taking the expectation and using the Lemmas 2-3

$$\mathbb{E}\left[|Z^{\lambda,u}_{n+1} - Z^{\lambda,u}_n|^4\right] \leq \frac{\lambda^2}{N^{2p}} 81 \left(\mu^2 C^2_{|z_0|,\mu,b} + 18\mu C_{|z_0|,b,2,\mu} + 4\right) (1 + d^\theta/N + d^x)^2.$$

□

A.2.3 Proof of Proposition 5

Proof. Consider the usual split on the difference between the interpolation and the scaled dynamics and apply the Itô's formula for $x \to |x|^2$

$$|\overline{Z}^{\lambda,u}_t - \mathcal{Z}_{\lambda t}|^2 = |\overline{\theta}^{\lambda,u}_t - \vartheta_{\lambda t}|^2 + \frac{1}{N}\sum_{i=1}^N |\overline{X}^{i,\lambda,u}_t - \mathcal{X}^i_{\lambda t}|^2$$
$$= -2\lambda \int_0^t \left\langle \frac{1}{N^{p+1}} \sum_{i=1}^N h^\theta_{\lambda,u}(\theta^{\lambda,u}_{\lfloor t \rfloor}, X^{i,\lambda,u}_{\lfloor t \rfloor}) - \frac{1}{N^{p+1}} \sum_{i=1}^N h^\theta(\vartheta^\lambda_{\lambda t}, \mathcal{X}^i_{\lambda t}), \overline{\theta}^{\lambda,u}_t - \vartheta_{\lambda t} \right\rangle ds$$
$$- \frac{2\lambda}{N^{p+1}} \sum_{i=1}^N \int_0^t \left\langle h^x_{\lambda,u}(\theta^{\lambda,u}_{\lfloor t \rfloor}, X^{i,\lambda,u}_{\lfloor t \rfloor}) - h^x(\vartheta^\lambda_{\lambda t}, \mathcal{X}^i_{\lambda t}), \overline{X}^{i,\lambda,u}_t - \mathcal{X}^i_{\lambda t} \right\rangle ds$$
$$= -\frac{2\lambda}{N^{p+1}} \sum_{i=1}^N \int_0^t \langle h_{\lambda,u}(V^{i,\lambda,u}_{\lfloor t \rfloor}) - h(\mathcal{V}^i_{\lambda t}), \overline{V}^{i,\lambda,u}_t - \mathcal{V}^i_{\lambda t} \rangle ds.$$

In order to make the forthcoming calculations more readable we define the following quantities:

$$e^i_t = \overline{V}^{i,\lambda,u}_t - \mathcal{V}^i_{\lambda t} \text{ and } e_t = \overline{Z}^{\lambda,u}_t - \mathcal{Z}_{\lambda t},$$

so that it follows: $1/N \sum_{i=1}^N |e^i_t|^2 = |e_t|^2$. Then by taking the derivative of the above expansion we obtain

$$\frac{d}{dt}|e_t|^2 = -\frac{2\lambda}{N^{p+1}} \sum_{i=1}^N \langle h_{\lambda,u}(V^{i,\lambda,u}_{\lfloor t \rfloor}) - h(\mathcal{V}^i_{\lambda t}), e^i_t \rangle$$
$$= -\frac{2\lambda}{N^{p+1}} \sum_{i=1}^N \langle h(\overline{V}^{i,\lambda,u}_t) - h(\mathcal{V}^i_{\lambda t}), e^i_t \rangle - \frac{2\lambda}{N^{p+1}} \sum_{i=1}^N \langle h_{\lambda,u}(V^{i,\lambda,u}_{\lfloor t \rfloor}) - h(V^{i,\lambda,u}_{\lfloor t \rfloor}), e^i_t \rangle$$
$$- \frac{2\lambda}{N^{p+1}} \sum_{i=1}^N \langle h(V^{i,\lambda,u}_{\lfloor t \rfloor}) - h(\overline{V}^{i,\lambda,u}_t), e^i_t \rangle$$
$$:= k_1(t) + k_2(t) + k_3(t). \tag{38}$$

The first term $k_1(t)$ is controlled through the convexity of the initial potential, the taming error $k_2(t)$ can be controlled via the properties of the taming function and the additional moment bounds that we have established, which are also used to control the discretisation error $k_3(t)$. In particular ,

$$\begin{aligned} k_1(t) &= -\frac{2\lambda}{N^{p+1}}\sum_{i=1}^{N}\langle h(\overline{V}_t^{i,\lambda,u}) - h(\mathcal{V}_{\lambda t}^i), e_t^i\rangle \\ &\leq -\frac{2\lambda\mu}{N^{p+1}}\sum_{i=1}^{N}|e_t^i|^2 = -\frac{2\lambda\mu}{N^p}|e_t|^2. \end{aligned} \tag{39}$$

Using the Cauchy-Schwarz inequality and the ϵ-Young inequality for $\epsilon = \mu/2$ yields

$$\begin{aligned} k_2(t) &= -\frac{2\lambda}{N^{p+1}}\sum_{i=1}^{N}\langle h_{\lambda,u}(V_{\lfloor t\rfloor}^{i,\lambda,u}) - h(V_{\lfloor t\rfloor}^{i,\lambda,u}), e_t^i\rangle \\ &\leq \frac{2\lambda}{N^{p+1}}\sum_{i=1}^{N}\left(\frac{\mu}{4}|e_t^i|^2 + \frac{1}{\mu}|h_{\lambda,u}(V_{\lfloor t\rfloor}^{i,\lambda,u}) - h(V_{\lfloor t\rfloor}^{i,\lambda,u})|^2\right) \\ &\leq \frac{2\lambda\mu}{4N^p}|e_t|^2 + \frac{2\lambda}{\mu N^{p+1}}\sum_{i=1}^{N}\frac{\lambda C_1}{N^p}(1 + |V_{\lfloor t\rfloor}^{i,\lambda,u}|^{4(\ell+1)}) \\ &\leq \frac{\lambda\mu}{2N^p}|e_t|^2 + \frac{2\lambda^2 C_1}{\mu N^{2p}} + \frac{2\lambda^2 C_1}{\mu N^p}\left(\frac{1}{N}\sum_{i=1}^{N}|V_{\lfloor t\rfloor}^{i,\lambda,u}|^2\right)^{2(\ell+1)} \\ &\leq \frac{\lambda\mu}{2N^p}|e_t|^2 + \frac{2\lambda^2 C_1}{\mu N^p} + \frac{2\lambda^2 C_1}{\mu N^p}|Z_{\lfloor t\rfloor}^{\lambda,u}|^{4(\ell+1)}. \end{aligned} \tag{40}$$

Note that the constant C_1 was introduced in Property 2 and that we choosed $p = 2\ell + 1$ in order to get an expression with a power of $|Z_{\lfloor t\rfloor}^{\lambda,u}|$. Similarly for the last term we get

$$\begin{aligned} k_3(t) &= -\frac{2\lambda}{N^{p+1}}\sum_{i=1}^{N}\langle h(V_{\lfloor t\rfloor}^{i,\lambda,u}) - h(\overline{V}_t^{i,\lambda,u}), e_t^i\rangle \\ &\leq \frac{2\lambda}{N^{p+1}}\sum_{i=1}^{N}\left(\frac{\mu}{4}|e_t^i|^2 + \frac{1}{\mu}|h(V_{\lfloor t\rfloor}^{i,\lambda,u}) - h(\overline{V}_t^{i,\lambda,u})|^2\right) \\ &\leq \frac{\lambda\mu}{2N^p}|e_t|^2 + \frac{2\lambda C_2}{\mu N^{p+1}}\sum_{i=1}^{N}\left((1 + |V_{\lfloor t\rfloor}^{i,\lambda,u}|^{2\ell} + |\overline{V}_t^{i,\lambda,u}|^{2\ell})|V_{\lfloor t\rfloor}^{i,\lambda,u} - \overline{V}_t^{i,\lambda,u}|^2\right) \\ &\leq \frac{\lambda\mu}{2N^p}|e_t|^2 + \frac{2\lambda C_2}{\mu N^{p+1}}\sum_{i=1}^{N}\left(1 + |V_{\lfloor t\rfloor}^{i,\lambda,u}|^{2\ell} + |\overline{V}_t^{i,\lambda,u}|^{2\ell}\right)\sum_{i=1}^{N}|V_{\lfloor t\rfloor}^{i,\lambda,u} - \overline{V}_t^{i,\lambda,u}|^2. \end{aligned}$$

By reordering, taking the expectation and using the Holder's inequality we further get:

$$\begin{aligned}
\mathbb{E}\left[k_3(t)\right] &\leq \frac{\lambda\mu}{2N^p}\mathbb{E}\left[|e_t|^2\right] \\
&\quad + \frac{2\lambda C_2}{\mu N^{(p+1)/2}}\mathbb{E}\left[\frac{1}{N^\ell}\sum_{i=1}^{N}\left(1+|V_{\lfloor t\rfloor}^{i,\lambda,u}|^{2\ell}+|\overline{V}_t^{i,\lambda,u}|^{2\ell}\right)\frac{1}{N}\sum_{i=1}^{N}|V_{\lfloor t\rfloor}^{i,\lambda,u}-\overline{V}_t^{i,\lambda,u}|^2\right] \\
&\leq \frac{\lambda\mu}{2N^p}\mathbb{E}\left[|e_t|^2\right] + \frac{2\lambda C_2}{\mu}\mathbb{E}\left[\left(1+|Z_{\lfloor t\rfloor}^{\lambda,u}|^{2\ell}+|\overline{Z}_t^{\lambda,u}|^{2\ell}\right)|Z_{\lfloor t\rfloor}^{\lambda,u}-\overline{Z}_t^{\lambda,u}|^2\right] \\
&\leq \frac{\lambda\mu}{2N^p}\mathbb{E}\left[|e_t|^2\right] + \frac{6\lambda C_2}{\mu}\mathbb{E}^{1/2}\left[\left(1+|Z_{\lfloor t\rfloor}^{\lambda,u}|^{4\ell}+|\overline{Z}_t^{\lambda,u}|^{4\ell}\right)\right]\mathbb{E}^{1/2}\left[|Z_{\lfloor t\rfloor}^{\lambda,u}-\overline{Z}_t^{\lambda,u}|^4\right] \\
&\leq \frac{\lambda\mu}{2N^p}\mathbb{E}\left[|e_t|^2\right] + \frac{6\lambda^2 C_2 C_3}{\mu N^p}\mathbb{E}^{1/2}\left[\left(1+|Z_{\lfloor t\rfloor}^{\lambda,u}|^{4\ell}+|\overline{Z}_t^{\lambda,u}|^{4\ell}\right)\right](1+d^\theta/N+d^x) \\
&\leq \frac{\lambda\mu}{2N^p}\mathbb{E}\left[|e_t|^2\right] + \frac{6\lambda^2 C_2 C_3 C_4}{\mu N^p}(1+d^\theta/N+d^x)^{\ell+1}. \qquad (41)
\end{aligned}$$

Taking the expectation for the rest of the terms in (39) and (40) respectively and combining them with (41), one obtains in view of (38),

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}\left[|e_t|^2\right] &\leq -\frac{\mu\lambda}{N^p}\mathbb{E}\left[|e_t|^2\right] + \frac{2\lambda^2 C_1}{\mu N^p} + \frac{2\lambda^2 C_1 C_5}{\mu N^p}(1+d^\theta/N+d^x)^{2(\ell+1)} \\
&\quad + \frac{6\lambda^2 C_2 C_3 C_4}{\mu N^p}(1+d^\theta/N+d^x)^{\ell+1} \\
&\leq -\frac{\mu\lambda}{N^p}\mathbb{E}\left[|e_t|^2\right] + \frac{\lambda^2 C_6}{\mu N^p}(1+d^\theta/N+d^x)^{2(\ell+1)},
\end{aligned}$$

by multiplying both sides by the integrating factor $e^{\lambda\mu t/N^p}$ and rearranging the term one gets

$$\begin{aligned}
\frac{d}{dt}\left(e^{\lambda\mu t/N^p}\mathbb{E}\left[|e_t|^2\right]\right) &\leq \frac{\lambda^2 C_6}{\mu N^p}(1+d^\theta/N+d^x)^{2(\ell+1)}e^{\lambda\mu t/N^p} \\
\mathbb{E}\left[|e_t|^2\right] &\leq \lambda\frac{C_6}{\mu^2}(1+d^\theta/N+d^x)^{2(\ell+1)}.
\end{aligned}$$

□

A.3 Coordinate-wise tamed scheme - tIPLAc

A.3.1 Key quantities for the proof of the main Lemmas.

The following definitions that refer to the rescaled dynamics (17)-(18) of the algorithm tIPLAc and its continuous time interpolations (19)-(20) are given as:

$$Z_n^{\lambda,c} = \left(\theta_{n+1}^{\lambda,c}, N^{-1/2}X_{n+1}^{1,\lambda,c}, \ldots, N^{-1/2}X_{n+1}^{N,\lambda,c}\right), \qquad (42)$$

$$\overline{Z}_t^{\lambda,c} = \left(\overline{\theta}_t^{\lambda,c}, N^{-1/2}\overline{X}_t^{1,\lambda,c}, \ldots, N^{-1/2}\overline{X}_t^{N,\lambda,c}\right), \tag{43}$$

$$\mathcal{Z}_{\lambda t}^N = \left(\vartheta_{\lambda t}^N, N^{-1/2}\mathcal{X}_{\lambda t}^{1,N}, \ldots, N^{-1/2}\mathcal{X}_{\lambda t}^{N,N}\right). \tag{44}$$

A.3.2 Moment and increment bounds

Lemma 5. Let A1, A3 and A4 hold. Then, for any $0 \leq \lambda < 1/4\mu$, it holds that,

$$\mathbb{E}\left[|V_n^{i,\lambda,c}|^2\right] \leq C_{|z_0|,\mu,b}(1 + d^\theta + d^x),$$

for a constant $C > 0$ independent of N, n, λ, d^x and d^θ, $\forall i \in \{1, \ldots, N\}$.

Proof.

$$\begin{aligned}
\left|V_{n+1}^{i,\lambda,c}\right|^2 &= \left|\theta_{n+1}^{\lambda,c}\right|^2 + \left|X_{n+1}^{i,\lambda,c}\right|^2 \\
&= \left|\theta_n^{\lambda,c} - \frac{\lambda}{N}\sum_{i=1}^N h_{\lambda,c}^\theta(\theta_n^{\lambda,c}, X_n^{i,\lambda,c})\right|^2 + \frac{2\lambda}{N}|\xi_{n+1}^{(0)}|^2 \\
&+ 2\sqrt{\frac{2\lambda}{N}}\left\langle \theta_n^{\lambda,c} - \frac{\lambda}{N}\sum_{i=1}^N h_{\lambda,c}^\theta(\theta_n^{\lambda,c}, X_n^{i,\lambda,c}), \xi_{n+1}^{(0)}\right\rangle \\
&+ \left|X_n^{i,\lambda,c} - \lambda h_{\lambda,c}^x(\theta_n^{\lambda,c}, X_n^{i,\lambda,c})\right|^2 + 2\lambda|\xi_{n+1}^{(i)}|^2 \\
&+ 2\sqrt{2\lambda}\left\langle X_n^{i,\lambda,c} - \lambda h_{\lambda,c}^x(\theta_n^{\lambda,c}, X_n^{i,\lambda,c}), \xi_{n+1}^{(i)}\right\rangle.
\end{aligned}$$

Taking the conditional expectation on both sides with respect to the filtration generated by $V_n^{i,\lambda,c}$ the cross terms are vanishing to 0, yielding

$$\begin{aligned}
\mathbb{E}\left[\left|V_{n+1}^{i,\lambda,c}\right|^2 |V_n^{i,\lambda,c}\right] &= \mathbb{E}\left[\left|\theta_n^{\lambda,c}\right|^2 |V_n^{i,\lambda,c}\right] - \frac{2\lambda}{N}\sum_{i=1}^N \mathbb{E}\left[\langle\theta_n^{\lambda,c}, h_{\lambda,c}^\theta(\theta_n^{\lambda,c}, X_n^{i,\lambda,c})\rangle|V_n^{i,\lambda,c}\right] \\
&+ \frac{\lambda^2}{N^2}\mathbb{E}\left[\left|\sum_{i=1}^N h_{\lambda,c}^\theta(\theta_n^{\lambda,c}, X_n^{i,\lambda,c})\right|^2 |V_n^{i,\lambda,c}\right] + \frac{2\lambda d^\theta}{N} + \mathbb{E}\left[\left|X_n^{i,\lambda,c}\right|^2 |V_n^{i,\lambda,c}\right] \\
&- 2\lambda\mathbb{E}\left[\langle X_n^{i,\lambda,c}, h_{\lambda,c}^x(\theta_n^{\lambda,c}, X_n^{i,\lambda,c})\rangle|V_n^{i,\lambda,c}\right] \\
&+ \lambda^2\mathbb{E}\left[|h_{\lambda,c}^x(\theta_n^{\lambda,c}, X_n^{i,\lambda,c})|^2|V_n^{i,\lambda,c}\right] + 2\lambda d^x.
\end{aligned}$$

Furthermore by using the elementary inequality $(t_1 + \ldots + t_m)^p \leq m^{p-1}(t_1^p + \ldots, t_m^p)$ and the fact that all of the expressions within the conditional expectations are measurable we obtain

$$
\begin{aligned}
\mathbb{E}\left[\left|V_{n+1}^{i,\lambda,c}\right|^2 | V_n^{i,\lambda,c}\right] &\leq |\theta_n^{\lambda,c}| - \frac{2\lambda}{N}\sum_{i=1}^N \langle \theta_n^{\lambda,c}, h_{\lambda,c}^\theta(\theta_n^{\lambda,c}, X_n^{i,\lambda,c})\rangle + \frac{\lambda^2}{N}\sum_{i=1}^N |h_\lambda^\theta(\theta_n^{\lambda,c}, X_n^{i,\lambda,c})|^2 \\
&\quad + \frac{2\lambda d^\theta}{N} + |X_n^{i,\lambda,c}|^2 - 2\lambda\langle X_n^{i,\lambda,c}, h_{\lambda,c}^x(\theta_n^{\lambda,c}, X_n^{i,\lambda,c})\rangle \\
&\quad + \lambda^2 |h_{\lambda,c}^x(\theta_n^{\lambda,c}, X_n^{i,\lambda,c})|^2 + 2\lambda d^x \\
&\leq \left|V_n^{i,\lambda,c}\right|^2 - \frac{2\lambda}{N}\sum_{i=1}^N\sum_{j=1}^{d^\theta} \theta_{n,j}^\lambda \cdot h_{\lambda,c}^{\theta_j}(\theta_n^{\lambda,c}, X_n^{i,\lambda,c}) \\
&\quad + \frac{\lambda^2}{N}\sum_{i=1}^N\sum_{j=1}^{d^\theta} |h_{\lambda,c}^{\theta_j}(\theta_n^{\lambda,c}, X_n^{i,\lambda,c})|^2 - 2\lambda\sum_{j=1}^{d^x} X_{n,j}^{i,\lambda}\cdot h_{\lambda,c}^{x_j}(\theta_n^{\lambda,c}, X_n^{i,\lambda,c}) \\
&\quad + \lambda^2\sum_{j=1}^{d^x} |h_{\lambda,c}^{x_j}(\theta_n^{\lambda,c}, X_n^{i,\lambda,c})|^2 + 2\lambda(d^\theta/N + d^x) \\
&\leq \left|V_n^{i,\lambda,c}\right|^2 - \frac{2\lambda}{N}\sum_{i=1}^N\sum_{j=1}^{d^\theta}\left(\frac{\mu}{2}|\theta_{n,j}^{\lambda,c}|^2 - b\right) \\
&\quad + \frac{\lambda^2}{N}\sum_{i=1}^N\sum_{j=1}^{d^\theta}\left(2\mu^2|\theta_{n,j}^{\lambda,c}|^2 + 2\lambda^{-1}\right) - 2\lambda\sum_{j=1}^{d^x}\left(\frac{\mu}{2}|X_{n,j}^{i,\lambda,c}|^2 - b\right) \\
&\quad + \lambda^2\sum_{j=1}^{d^x}\left(2\mu^2|X_{n,j}^{i,\lambda,c}|^2 + 2\lambda^{-1}\right) + 2\lambda(d^\theta/N + d^x).
\end{aligned}
$$

Executing the summations yields

$$
\begin{aligned}
\mathbb{E}\left[\left|V_{n+1}^{i,\lambda,c}\right|^2 | V_n^{i,\lambda,c}\right] &\leq \left|V_n^{i,\lambda,c}\right|^2 - \lambda\mu|\theta_n^{\lambda,c}|^2 - \lambda\mu|X_n^{i,\lambda,c}|^2 + 2\lambda^2\mu^2|\theta_n^{\lambda,c}|^2 + 2\lambda^2\mu^2|X_n^{i,\lambda,c}|^2 \\
&\quad + 2\lambda b(d^\theta + d^x) + 2\lambda(d^\theta + d^x) + 2\lambda(d^\theta/N + d^x) \\
&\leq (1 - \lambda\mu + 2\lambda^2\mu^2)\left|V_n^{i,\lambda,c}\right|^2 + 2\lambda(b+1)(d^\theta + d^x) + 2\lambda(d^\theta/N + d^x).
\end{aligned}
$$

Consider the restriction $\lambda\mu < 1/4$, then this implies that

$$
\mathbb{E}\left[\left|V_{n+1}^{i,\lambda,c}\right|^2\right] \leq (1 - \lambda\mu/2)\mathbb{E}\left[\left|V_n^{i,\lambda,c}\right|^2\right] + 2\lambda(b+1)(d^\theta + d^x) + 2\lambda(d^\theta/N + d^x).
$$

Iterating the above bound finally yields

$$
\begin{aligned}
\mathbb{E}\left[\left|V_{n+1}^{i,\lambda,c}\right|^2\right] &\leq (1 - \lambda\mu/2)^n\left|V_0^{i,\lambda,c}\right|^2 + \frac{1 - (1 - \lambda\mu/2)^n}{\lambda\mu/2} 4\lambda(b+1)(d^\theta + d^x) \\
&\leq \left(|V_0|^2 + (8/\mu)(b+1)\right)(1 + d^\theta + d^x).
\end{aligned}
$$

$\square$

Lemma 6. Let A1, A3 and A4 hold. Then, for any $0 \le \lambda < 1/4\mu$ and $q \in [2, 2\ell+1) \cap \mathbb{N}$, it holds that,

$$\mathbb{E}\left[|V_n^{i,\lambda,c}|^{2q}\right] \le C_{|z_0|,\mu,b,p}(1 + d^\theta + d^x)^q,$$

for a constant $C > 0$ independent of N, n, λ, d^x and d^θ, $\forall i \in \{1, \ldots, d^\theta + Nd^x\}$.

Proof. Similarly as in the proof of Lemma 3 we define the following auxiliary quantities:

$$A_n^\lambda := |\Delta_n^{\lambda,\theta}|^2 + |\Delta_n^{\lambda,i,x}|^2,$$
$$B_n^\lambda = 2\langle \Delta_n^\theta, G_n^{\lambda,\theta}\rangle + 2\langle \Delta_n^{\lambda,x,i}, G_n^{\lambda,x,i}\rangle + |G_n^{\lambda,\theta}|^2 + |G_n^{\lambda,x,i}|^2.$$

Now for the 2q-th Moment one writes:

$$\begin{aligned}|V_{n+1}^{i,\lambda,c}|^{2q} &= \left(A_n^\lambda + B_n^\lambda\right)^q \\ &\le (A_n^\lambda)^q + 2q(A_n^\lambda)^{q-1}B_n^\lambda + \sum_{k=2}^{q}\binom{q}{k}|A_n^\lambda|^{q-k}|B_n^\lambda|^k \end{aligned} \tag{45}$$

We shall deal with each term separately,

$$\begin{aligned}\mathbb{E}\left[|A_n^\lambda|^q|Z_n^\lambda\right] = (A_n^\lambda)^q &\le \Big((1-\lambda\mu/2)\left|V_n^{i,\lambda,c}\right|^2 + 2\lambda(b+1)(d^\theta + d^x)\Big)^q \\ &\le \left(1+\frac{\lambda\mu}{4}\right)^{q-1}\left(1-\frac{\lambda\mu}{2}\right)^q|V_n^{i,\lambda,c}|^{2q} + \left(1+\frac{4}{\lambda\mu}\right)^{q-1}2^q\lambda^qC^q \\ &\le \left(1-\frac{\lambda\mu}{4}\right)^{q-1}\left(1-\frac{\lambda\mu}{2}\right)|V_n^{i,\lambda,c}|^{2q} + \left(\lambda+\frac{4}{\mu}\right)^{q-1}\lambda(2C)^q \\ &\le r_q^\lambda|V_n^{i,\lambda,c}|^{2q} + w_q^\lambda. \end{aligned} \tag{46}$$

where $r_q^\lambda = (1-\lambda\mu/4)^{q-1}(1-\lambda\mu/2)$ and $w_q^\lambda = (\lambda + 4/\mu)^{q-1}\lambda(2C)^q$.
On a similar note,

$$\begin{aligned}\mathbb{E}\left[2q(A_n^\lambda)^{q-1}B_n^\lambda|V_n^{i,\lambda,c}\right] = 2q(A_n^\lambda)^{q-1}\mathbb{E}\left[B_n^\lambda|Z_n^\lambda\right] &= 4q\lambda\left(d^\theta/N + d^x\right)(A_n^\lambda)^{q-1} \\ &\le 4q\lambda\left(d^\theta/N + d^x\right)\left(r_{q-1}^\lambda|V_n^{i,\lambda,c}|^{2(q-1)} + w_{q-1}^\lambda\right) \\ &\le 4q\left(2\lambda d\right)\left(r_{q-1}^\lambda|V_n^{i,\lambda,c}|^{2(q-1)} + w_{q-1}^\lambda\right). \end{aligned} \tag{47}$$

The 3rd term on (45) can be further expanded to

$$\begin{aligned}\sum_{k=2}^{q}\binom{q}{k}|A_n^\lambda|^{q-k}|B_n^\lambda|^k &= \sum_{\ell=0}^{q-2}\binom{q}{\ell+2}|A_n^\lambda|^{q-2-k}|B_n^\lambda|^{k+2} \\ &= \frac{q}{\ell+2}\frac{q-1}{\ell+1}\sum_{\ell=0}^{q-2}\binom{q}{\ell}|A_n^\lambda|^{q-2-k}|B_n^\lambda|^k|B_n^\lambda|^2 \\ &\le q(q-1)\left(|A_n^\lambda| + |B_n^\lambda|\right)^{q-2}|B_n^\lambda|^2 \\ &\le q(q-1)2^{q-3}|A_n^\lambda|^{q-2}|B_n^\lambda|^2 + q(q-1)2^{q-3}|B_n^\lambda|^q. \\ &= D + F \end{aligned} \tag{48}$$

Taking the expectation of the above terms yields

$$\mathbb{E}[D|V_n^{i,\lambda,c}] = q(q-1)2^{q-3}|A_n^\lambda|^{q-2}\mathbb{E}\left[|B_n^\lambda|^2|V_n^{i,\lambda,c}\right],$$

where

$$\begin{aligned}
\mathbb{E}\left[|B_n^\lambda|^2|V_n^{i,\lambda,c}\right] &= \mathbb{E}\left[\left|2\langle\Delta_n^\theta, G_n^{\lambda,\theta}\rangle + 2\langle\Delta_n^{\lambda,x,i}, G_n^{\lambda,x,i}\rangle + |G_n^{\lambda,\theta}|^2 + |G_n^{\lambda,x,i}|^2\right|^2 |V_n^{i,\lambda,c}\right] \\
&\leq 4\left(\mathbb{E}\left[4|\Delta_n^{\lambda,\theta}|^2|G_n^{\lambda,\theta}|^2|V_n^{i,\lambda,c}\right] + \mathbb{E}\left[4|\Delta_n^{\lambda,x,i}|^2|G_n^{\lambda,x,i}|^2|V_n^{i,\lambda,c}\right]\right. \\
&\left.+\ \mathbb{E}\left[|G_n^{\lambda,x,i}|^4|V_n^{i,\lambda,c}\right] + \mathbb{E}\left[|G_n^{\lambda,x,i}|^4|V_n^{i,\lambda,c}\right]\right) \\
&\leq 4\left(4(2\lambda d^\theta/N)|\Delta_n^{\lambda,\theta}|^2 + 4(2\lambda d^x)|\Delta_n^{\lambda,x,i}|^2 + 12(2\lambda d^\theta/N)^2 + 12(2\lambda d^x)^2\right) \\
&\leq 16(2\lambda d)\left(|\Delta_n^{\lambda,\theta}|^2 + |\Delta_n^{\lambda,x,i}|^2\right) + 96(2\lambda d)^2 \\
&\leq 16(2\lambda d)A_n^\lambda + 96(2\lambda d)^2.
\end{aligned}$$

with $d = \max\{d^\theta/N, d^x\}$. plugging in this result for the term D we get

$$\begin{aligned}
\mathbb{E}[D|V_n^{i,\lambda,c}] &= q(q-1)2^{q-3}|A_n^\lambda|^{q-2}\left(16(2\lambda d)A_n^\lambda + 96(2\lambda d)^2\right) \\
&= 16(2\lambda d)q(q-1)2^{q-3}\left(r_{q-1}^\lambda|V_n^\lambda|^{2(q-1)} + w_{q-1}^\lambda\right) \\
&+ 96(2\lambda d)^2 q(q-1)2^{q-3}\left(r_{q-2}^\lambda|V_n^\lambda|^{2(q-2)} + w_{q-2}^\lambda\right).
\end{aligned}$$

Finally,

$$\begin{aligned}
\mathbb{E}[F|V_n^{i,\lambda,c}] &= q(q-1)2^{q-3}|B_n^\lambda|^q \\
&= q(q-1)2^{q-3}\mathbb{E}\left[\left|2\langle\Delta_n^\theta, G_n^{\lambda,\theta}\rangle + 2\langle\Delta_n^{\lambda,x,i}, G_n^{\lambda,x,i}\rangle + |G_n^{\lambda,\theta}|^2 + |G_n^{\lambda,x,i}|^2\right|^q |V_n^{i,\lambda,c}\right] \\
&\leq q(q-1)2^{q-3}4^{q-1}\left(\mathbb{E}\left[2^q|\Delta_n^{\lambda,\theta}|^q|G_n^{\lambda,\theta}|^q|V_n^{i,\lambda,c}\right] + \mathbb{E}\left[2^q|\Delta_n^{\lambda,x,i}|^q|G_n^{\lambda,x,i}|^q|V_n^{i,\lambda,c}\right]\right. \\
&\left.+\mathbb{E}\left[|G_n^{\lambda,\theta}|^{2q}|V_n^{i,\lambda,c}\right] + \mathbb{E}\left[|G_n^{\lambda,x,i}|^{2q}|V_n^{i,\lambda,c}\right]\right) \\
&\leq q(q-1)2^{3q-5}\left(2^q(2\lambda d^\theta/N)^{q/2}q!!\left(|\Delta_n^{\lambda,\theta}|^2\right)^{q/2} + 2^q q!!(2\lambda d^x)^{q/2}\left(|\Delta_n^{\lambda,x,i}|^2\right)^{q/2}\right. \\
&\left.+(2q)!!(2\lambda d^\theta/N)^q + (2q)!!(2\lambda d^x)^q\right) \\
&\leq (q(q-1))^{q+1}2^{4q-5}\left((2\lambda d)^{q/2}\left(|\Delta_n^{\lambda,\theta}|^2 + |\Delta_n^{\lambda,x,i}|^2\right)^{q/2} + 2(2\lambda d)^q\right) \\
&\leq (q(q-1))^{q+1}2^{4q-5}\left((2\lambda d)^{q/2}\left|A_n^\lambda\right|^{q/2} + 2(2\lambda d)^q\right) \\
&\leq (q(q-1))^{q+1}2^{4q-5}(2\lambda d)^{q/2}\left(r_{q/2}^\lambda|V_n^\lambda|^q + w_{q/2}^\lambda\right) \\
&+ (q(q-1))^{q+1}2^{4q-5}2(2\lambda d)^q.
\end{aligned}$$

Combining the above results in view of (48) yields

$$\begin{aligned}
&\mathbb{E}\left[\sum_{k=2}^q \binom{q}{k}|A_n^\lambda|^{q-k}|B_n^\lambda|^k\right] \\
&\leq 16(2\lambda d)q(q-1)2^{q-3}\left(r_{q-1}^\lambda|V_n^\lambda|^{2(q-1)} + w_{q-1}^\lambda\right) \\
&+ 96(2\lambda d)^2 q(q-1)2^{q-3}\left(r_{q-2}^\lambda|V_n^\lambda|^{2(q-2)} + w_{q-2}^\lambda\right) \\
&+ (q(q-1))^{q+1}2^{4q-5}(2\lambda d)^{q/2}\left(r_{q/2}^\lambda|V_n^\lambda|^q + w_{q/2}^\lambda\right) \\
&+ (q(q-1))^{q+1}2^{4q-4}(2\lambda d)^q.
\end{aligned} \tag{49}$$

Substituting (46),(47) and (49) into (45) also yields

$$
\begin{aligned}
\mathbb{E}\left[|V_{n+1}^{i,\lambda,c}|^{2q}|V_n^{i,\lambda,c}\right] &\leq r_q^\lambda|V_n^{i,\lambda,c}|^{2q} + w_q^\lambda \\
&+ 4q\,(2\lambda d)\left(1+4(q-1)2^{q-3}\right)\left(r_{q-1}^\lambda|V_n^{i,\lambda,c}|^{2(q-1)} + w_{q-1}^\lambda\right) \\
&+ 96(2\lambda d)^2 q(q-1)2^{q-3}\left(r_{q-2}^\lambda|V_n^{i,\lambda,c}|^{2(q-2)} + w_{q-2}^\lambda\right) \\
&+ (q(q-1))^{q+1}2^{4q-5}(2\lambda d)^{q/2}\left(r_{q/2}^\lambda|V_n^{i,\lambda,c}|^{q} + w_{q/2}^\lambda\right) \\
&+ (q(q-1))^{q+1}2^{4q-4}(2\lambda d)^{q}.
\end{aligned}
\tag{50}
$$

Consider $|V_n^{i,\lambda,c}| \geq \sqrt{\dfrac{2d}{\mu/4}}\left\{2(q(q-1))^{q+1}2^{4(q-1)}\right\}^{1/2} \geq \sqrt{\dfrac{2d}{\mu/4}}\left\{2(q(q-1))^{q+1}2^{4(q-1)}\right\}^{1/q}$. Then one observes

$$
\begin{aligned}
\mathbb{E}\left[|V_{n+1}^{i,\lambda,c}|^{2q}|V_n^{i,\lambda,c}\right] &\leq r_q^\lambda|V_n^{i,\lambda,c}|^{2q} + w_q^\lambda + \frac{\lambda\mu}{2\cdot 4}|V_n^{i,\lambda,c}|^{2}\left(r_{q-1}^\lambda|V_n^{i,\lambda,c}|^{2(q-1)}\right) \\
&+ \frac{(\lambda\mu)^2}{2\cdot 4^2}|V_n^{i,\lambda,c}|^{4}\left(r_{q-2}^\lambda|V_n^{i,\lambda,c}|^{2(q-2)}\right) \\
&+ \frac{(\lambda\mu)^{q/2}}{2\cdot 4^{q/2}}|V_n^{i,\lambda,c}|^{q}\left(r_{q/2}^\lambda|V_n^{i,\lambda,c}|^{q}\right) + (q(q-1))^{q+1}2^{4(q-1)}(2\lambda d)^q \\
&+ (q(q-1))^{q+1}2^{4(q-1)}\left((2\lambda d)w_{q-1}^\lambda + (2\lambda d)^2 w_{q-2}^\lambda + (2\lambda d)^{q/2}w_{q/2}^\lambda\right) \\
&\leq r_{q/2}^\lambda\left(\left(1-\frac{\lambda\mu}{4}\right)^{q/2} + \frac{1}{2}\left(\frac{\lambda\mu}{4}\right)\left(1-\frac{\lambda\mu}{4}\right)^{q/2-1}\right. \\
&\left.+\frac{1}{2}\left(\frac{\lambda\mu}{4}\right)^{2}\left(1-\frac{\lambda\mu}{4}\right)^{q/2-2} + \frac{1}{2}\left(\frac{\lambda\mu}{4}\right)^{q/2}\right)|V_n^{i,\lambda,c}|^{2q} \\
&+ w_q^\lambda + (q(q-1))^{q+1}2^{4(q-1)}(2\lambda d)^q \\
&+ (q(q-1))^{q+1}2^{4(q-1)}\left((2\lambda d)w_{q-1}^\lambda + (2\lambda d)^2 w_{q-2}^\lambda + (2\lambda d)^{q/2}w_{q/2}^\lambda\right).
\end{aligned}
$$

Using the fact that $\lambda\mu \leq 1$ we get

$$
\begin{aligned}
\mathbb{E}\left[|V_{n+1}^{i,\lambda,c}|^{2q}|V_n^{i,\lambda,c}\right] &\leq \left(1-\frac{\lambda\mu}{2}\right)|V_n^{i,\lambda,c}|^{2q} \\
&+ w_q^\lambda + (q(q-1))^{q+1}2^{4(q-1)}(2\lambda d)^q \\
&+ (q(q-1))^{q+1}2^{4(q-1)}\left((2\lambda d)w_{q-1}^\lambda + (2\lambda d)^2 w_{q-2}^\lambda + (2\lambda d)^{q/2}w_{q/2}^\lambda\right) \\
&\leq \left(1-\frac{\lambda\mu}{2}\right)|V_n^{i,\lambda,c}|^{2q} + (q(q-1))^{q+1}2^{4(q-1)}10^q C^q(d\lambda)^q.
\end{aligned}
$$

Consequently, on $\{|V_n^{i,\lambda,c}| \leq \sqrt{\frac{2d}{\mu/4}}\left\{2(q(q-1))^{q+1}2^{4(q-1)}\right\}^{1/2}\}$, we get

$$\begin{aligned}\mathbb{E}\left[|V_{n+1}^{i,\lambda,c}|^{2q}|V_n^{i,\lambda,c}\right] \leq & \left(1-\frac{\lambda\mu}{2}\right)|V_n^{i,\lambda,c}|^{2q} \\ & + (q(q-1))^{q+1}2^{4(q-1)}10^q C^q (d\lambda)^q \\ & + 3\left(q(q-1)^{q+1}2^{4(q-1)}\right)^q (2d)^q \left(\frac{4}{\mu}\right)^q 2^{q-1}.\end{aligned}$$

Hence the bound in (50) is further improved to

$$\mathbb{E}\left[|V_{n+1}^{i,\lambda,c}|^{2q}\right] \leq \left(1-\frac{\lambda\mu}{2}\right)\mathbb{E}\left[|V_n^{i,\lambda,c}|^{2q}\right] + C(d,q,\lambda,\mu),$$

and the usual bound for recursive sequences follows. $\square$

A.3.3 Proof of Proposition 6

Proof. Consider the usual split on the difference between the interpolation and the scaled dynamics and apply the Itô's formula for $x \to |x|^2$

$$\begin{aligned}|\overline{Z}_t^{\lambda,c} - \mathcal{Z}_{\lambda t}|^2 &= |\overline{\theta}_t^{\lambda,c} - \vartheta_{\lambda t}|^2 + \frac{1}{N}\sum_{i=1}^N |\overline{X}_t^{i,\lambda,c} - \mathcal{X}_{\lambda t}^i|^2 \\ = -2\lambda \int_0^t & \left\langle \frac{1}{N}\sum_{i=1}^N h_{\lambda,c}^\theta(\theta_{\lfloor t\rfloor}^{\lambda,c}, X_{\lfloor t\rfloor}^{i,\lambda,c}) - \frac{1}{N}\sum_{i=1}^N h^\theta(\vartheta_{\lambda t}^\lambda, \mathcal{X}_{\lambda t}^i), \overline{\theta}_t^{i,\lambda,c} - \vartheta_{\lambda t}^i \right\rangle ds \\ -\frac{2\lambda}{N}\sum_{i=1}^N & \int_0^t \left\langle h_{\lambda,c}^x(\theta_{\lfloor t\rfloor}^{\lambda,c}, X_{\lfloor t\rfloor}^{i,\lambda,c}) - h^x(\vartheta_{\lambda t}^\lambda, \mathcal{X}_{\lambda t}^i), \overline{X}_t^{i,\lambda,c} - \mathcal{X}_{\lambda t}^i \right\rangle ds \\ &= -\frac{2\lambda}{N}\sum_{i=1}^N \int_0^t \langle h_{\lambda,c}(V_{\lfloor t\rfloor}^{i,\lambda,c}) - h(\mathcal{V}_{\lambda t}^i), \overline{V}_t^{i,\lambda,c} - \mathcal{V}_{\lambda t}^i\rangle ds.\end{aligned}$$

In order to make the forthcoming calculations more readable we define the following quantities:

$$e_t^i = \overline{V}_t^{i,\lambda,c} - \mathcal{V}_{\lambda t}^i \text{ and } e_t = \overline{Z}_t^{\lambda,c} - \mathcal{Z}_{\lambda t},$$

so that it follows: $1/N\sum_{i=1}^N |e_t^i|^2 = |e_t|^2$.

Then by taking the derivative of the above expansion we obtain

$$
\begin{aligned}
\frac{d}{dt}|e_t|^2 &= -\frac{2\lambda}{N}\sum_{i=1}^{N}\langle h_{\lambda,c}(V^{i,\lambda,c}_{\lfloor t\rfloor}) - h(\mathcal{V}^{i}_{\lambda t}), e^i_t\rangle \\
&= -\frac{2\lambda}{N}\sum_{i=1}^{N}\langle h(\overline{V}^{i,\lambda,c}_{t}) - h(\mathcal{V}^{i}_{\lambda t}), e^i_t\rangle \\
&\quad -\frac{2\lambda}{N}\sum_{i=1}^{N}\langle h_{\lambda,c}(V^{i,\lambda,c}_{\lfloor t\rfloor}) - h(V^{i,\lambda,c}_{\lfloor t\rfloor}), e^i_t\rangle \\
&\quad -\frac{2\lambda}{N}\sum_{i=1}^{N}\langle h(V^{i,\lambda,c}_{\lfloor t\rfloor}) - h(\overline{V}^{i,\lambda,c}_{t}), e^i_t\rangle \\
&= k_1(t) + k_2(t) + k_3(t).
\end{aligned} \tag{51}
$$

The first term $k_1(t)$ is controlled through the convexity of the initial potential, the taming error $k_2(t)$ can be controlled via the properties of the taming function and the additional moment bounds that we have established, which are also used to control the discretisation error $k_3(t)$. In particular,

$$
\begin{aligned}
k_1(t) &= -\frac{2\lambda}{N}\sum_{i=1}^{N}\langle h(\overline{V}^{i,\lambda,c}_{t}) - h(\mathcal{V}^{i,\lambda}_{\lambda t}), e^i_t\rangle \\
&\leq -\frac{2\lambda\mu}{N}\sum_{i=1}^{N}|e^i_t|^2 = -2\lambda\mu|e_t|^2.
\end{aligned} \tag{52}
$$

Using the Cauchy-Schwarz inequality and the ϵ-Young inequality for $\epsilon = \mu/2$ yields

$$
\begin{aligned}
k_2(t) &= -\frac{2\lambda}{N}\sum_{i=1}^{N}\langle h_{\lambda,c}(V^{i,\lambda,c}_{\lfloor t\rfloor}) - h(V^{i,\lambda,c}_{\lfloor t\rfloor}), e^i_t\rangle \\
&\leq \frac{2\lambda}{N}\sum_{i=1}^{N}\left(\frac{1}{\mu}|h_{\lambda,c}(V^{i,\lambda,c}_{\lfloor t\rfloor}) - h(V^{i,\lambda,c}_{\lfloor t\rfloor})|^2 + \frac{\mu}{4}|e^i_t|^2\right) \\
&\leq 2\frac{\lambda\mu}{4}|e_t|^2 + \frac{2\lambda}{\mu N}\sum_{i=1}^{N}2\lambda C_2^2(1+|V^{i,\lambda,c}_{\lfloor t\rfloor}|^{4(\ell+1)}) \\
&\leq 2\frac{\lambda\mu}{4}|e_t|^2 + \frac{4\lambda^2}{\mu}C_2^2 + \frac{4\lambda^2}{\mu N}\sum_{i=1}^{N}|V^{i,\lambda,c}_{\lfloor t\rfloor}|^{4(\ell+1)}.
\end{aligned} \tag{53}
$$

Similarly for the last term we get

$$
\begin{aligned}
k_3(t) &= -\frac{2\lambda}{N}\sum_{i=1}^{N}\langle h(V_{\lfloor t\rfloor}^{i,\lambda,c}) - h(\overline{V}_t^{i,\lambda,c}), e_t^i\rangle \\
&\le \frac{2\lambda}{N}\sum_{i=1}^{N}\left(\frac{1}{\mu}|h(V_{\lfloor t\rfloor}^{i,\lambda,c}) - h(\overline{V}_t^{i,\lambda,c})|^2 + \frac{\mu}{4}|e_t^i|^2\right) \\
&\le 2\frac{\lambda\mu}{4}|e_t|^2 + \frac{2\lambda L^2}{\mu N}\sum_{i=1}^{N}\left((1+|V_{\lfloor t\rfloor}^{i,\lambda,c}|^\ell + |\overline{V}_t^{i,\lambda,c}|^\ell)^2|V_{\lfloor t\rfloor}^{i,\lambda,c} - \overline{V}_t^{i,\lambda,c}|^2\right). \qquad (54)
\end{aligned}
$$

By combining the above bounds (52),(53) and (54) in view of (51) and taking the expectation, one obtains

$$
\begin{aligned}
\frac{d}{dt}\mathbb{E}\left[|e_t|^2\right] &\le -\mu\lambda\mathbb{E}\left[|e_t|^2\right] + \frac{4\lambda^2 C_2^2}{\mu} + \frac{4\lambda^2}{\mu}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[|V_{\lfloor t\rfloor}^{i,\lambda,c}|^{4(\ell+1)}\right] \\
&\quad + \frac{2\lambda L^2}{\mu}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}^{1/2}\left[(1+|V_{\lfloor t\rfloor}^{i,\lambda,c}|^\ell + |\overline{V}_t^{i,\lambda,c}|^\ell)^4\right]\mathbb{E}^{1/2}\left[|V_{\lfloor t\rfloor}^{i,\lambda,c} - \overline{V}_t^{i,\lambda,c}|^4\right] \\
&\le -\mu\lambda\mathbb{E}\left[|e_t|^2\right] + \frac{4\lambda^2(C_2^2 + C_3)}{\mu} \\
&\quad + \frac{2\lambda L^2}{\mu}\left(6\max\{1, 2C_4\}\right)\lambda\left(\mu C_5 + 6\sqrt{d\mu}C_6 + 10d + 2\right) \\
&\le -\mu\lambda\mathbb{E}\left[|e_t|^2\right] + \lambda^2 M(L,\mu,d,\ell) \le \lambda\left(-\frac{\mu}{2}\mathbb{E}\left[|e_t|^2\right] + \lambda M\right).
\end{aligned}
$$

by multiplying both sides by the integrating factor $e^{\lambda\mu t/2}$ and rearranging the term one gets

$$
\begin{gathered}
\frac{d}{dt}\left(e^{\lambda\mu t/2}\mathbb{E}\left[|e_t|^2\right]\right) \le \lambda^2 M e^{\lambda\mu t/2} \\
\mathbb{E}\left[|e_t|^2\right] \le \lambda\frac{2M}{\mu}.
\end{gathered}
$$

$\square$