European Journal of Applied Mathematics (2025), **36**, pp. 292–315
doi:10.1017/S095679252400007X

PAPER

Consensus-based optimisation with truncated noise

Massimo Fornasier[1,2,3], Peter Richtárik[4,5,6], Konstantin Riedl[1,2] and Lukang Sun[4,5]

[1]Technical University of Munich, School of Computation, Information and Technology, Department of Mathematics, Munich, Germany
[2]Munich Center for Machine Learning, Munich, Germany
[3]Munich Data Science Institute, Garching, Germany
[4]King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
[5]KAUST AI Initiative, Thuwal, Saudi Arabia
[6]SDAIA-KAUST Center of Excellence in Data Science and Artificial Intelligence, Thuwal, Saudi Arabia
Corresponding author: Massimo Fornasier; Email: massimo.fornasier@cit.tum.de

Received: 25 October 2023; **Revised:** 31 January 2024; **Accepted:** 01 February 2024; **First published online:** 5 April 2024

Keywords: Global optimisation; derivative-free optimisation; non-smoothness; non-convexity; metaheuristics; consensus-based optimisation; truncated noise

2020 Mathematics Subject Classification: 65K10 (Primary); 90C26, 90C56, 35Q90, 35Q84 (Secondary)

Abstract

Consensus-based optimisation (CBO) is a versatile multi-particle metaheuristic optimisation method suitable for performing non-convex and non-smooth global optimisations in high dimensions. It has proven effective in various applications while at the same time being amenable to a theoretical convergence analysis. In this paper, we explore a variant of CBO, which incorporates truncated noise in order to enhance the well-behavedness of the statistics of the law of the dynamics. By introducing this additional truncation in the noise term of the CBO dynamics, we achieve that, in contrast to the original version, higher moments of the law of the particle system can be effectively bounded. As a result, our proposed variant exhibits enhanced convergence performance, allowing in particular for wider flexibility in choosing the noise parameter of the method as we confirm experimentally. By analysing the time evolution of the Wasserstein-2 distance between the empirical measure of the interacting particle system and the global minimiser of the objective function, we rigorously prove convergence in expectation of the proposed CBO variant requiring only minimal assumptions on the objective function and on the initialisation. Numerical evidences demonstrate the benefit of truncating the noise in CBO.

1. Introduction

The search for a global minimiser v^* of a potentially non-convex and non-smooth cost function

$$f : \mathbb{R}^d \to \mathbb{R}$$

holds significant importance in a variety of applications throughout applied mathematics, science and technology, engineering, and machine learning. Historically, a class of methods known as metaheuristics [1, 2] has been developed to address this inherently challenging and, in general, NP-hard problem. Examples of such include evolutionary programming [3], genetic algorithms [4], particle swarm optimisation (PSO) [5], simulated annealing [6] and many others. These methods work by combining local improvement procedures and global strategies by orchestrating deterministic and stochastic advances, with the aim of creating a method capable of robustly and efficiently finding the globally minimising argument v^* of f. However, despite their empirical success and widespread adoption in practice, most metaheuristics lack a solid mathematical foundation that could guarantee their robust convergence to global minimisers under reasonable assumptions.

Motivated by the urge to devise algorithms which converge provably, a novel class of metaheuristics, so-called consensus-based optimisation (CBO), originally proposed by the authors of [7], has recently emerged in the literature. Due to the inherent simplicity in the design of CBO, this class of optimisation algorithms lends itself to a rigorous theoretical analysis, as demonstrated in particular in the works [8–14]. However, this recent line of research does not just offer a promising avenue for establishing a thorough mathematical framework for understanding the numerically observed successes of CBO methods [9, 11, 15–17], but beyond that allows to explain the effective use of conceptually similar and widespread methods such as PSO as well as at first glance completely different optimisation algorithms such as stochastic gradient descent (SGD). While the first connection is to be expected and by now made fairly rigorous [18–20] due to CBO indisputably taking PSO as inspiration, the second observation is somewhat surprising, as it builds a bridge between derivative-free metaheuristics and gradient-based learning algorithms. Despite CBO solely relying on evaluations of the objective function, recent work [21] reveals an intrinsic SGD-like behaviour of CBO itself by interpreting it as a certain stochastic relaxation of gradient descent, which provably overcomes energy barriers of non-convex function. These perspectives, and, in particular, the already well-investigated convergence behaviour of standard CBO, encourage the exploration of improvements to the method in order to allow overcoming the limitations of traditional metaheuristics mentioned at the start. For recent surveys on CBO, we refer to [22, 23].

While the original CBO model [7] has been adapted to solve constrained optimisations [24–26], optimisations on manifolds [16, 27–30], multi-objective optimisation problems [31–33], saddle point problems [34] or the task of sampling [35], as well as has been extended to make use of memory mechanisms [17, 36, 37], gradient information [17, 38], momentum [39], jump-diffusion processes [40] or localisation kernels for polarisation [41], we focus in this work on a variation of the original model, which incorporates a truncation in the noise term of the dynamics. More formally, given a time horizon $T>0$, a time discretisation $t_0=0<\Delta t<\cdots<K\Delta t=t_K=T$ of $[0,T]$, and user-specified parameters $\alpha,\lambda,\sigma>0$ as well as $v_b, R>0$, we consider the interacting particle system

$$V^i_{k+1,\Delta t}-V^i_{k,\Delta t}=-\Delta t\lambda\big(V^i_{k,\Delta t}-\mathcal{P}_{v_b,R}\big(v_\alpha\big(\widehat{\rho}^N_{k,\Delta t}\big)\big)\big)+\sigma\big(\big\|V^i_{k,\Delta t}-v_\alpha\big(\widehat{\rho}^N_{k,\Delta t}\big)\big\|_2\wedge M\big)\,B^i_{k,\Delta t}, \tag{1}$$

$$V^i_0\sim\rho_0 \quad \text{for all } i=1,\ldots,N, \tag{2}$$

where $((B^i_{k,\Delta t})_{k=0,\ldots,K-1})_{i=1,\ldots,N}$ are independent, identically distributed Gaussian random vectors in $\mathbb{R}^d$ with zero mean and covariance matrix $\Delta t\mathrm{Id}_d$. Equation (1) originates from a simple Euler–Maruyama time discretisation [42, 43] of the system of stochastic differential equations (SDEs), expressed in Itô's form as

$$dV^i_t=-\lambda\big(V^i_t-\mathcal{P}_{v_b,R}\big(v_\alpha\big(\widehat{\rho}^N_t\big)\big)\big)\,dt+\sigma\big(\big\|V^i_t-v_\alpha\big(\widehat{\rho}^N_t\big)\big\|_2\wedge M\big)\,dB^i_t \tag{3}$$

$$V^i_0\sim\rho_0 \quad \text{for all } i=1,\ldots,N, \tag{4}$$

where $((B^i_t)_{t\geq0})_{i=1,\ldots,N}$ are now independent standard Brownian motions in $\mathbb{R}^d$. The empirical measure of the particles at time t is denoted by $\widehat{\rho}^N_t:=\frac{1}{N}\sum_{i=1}^N\delta_{V^i_t}$. Moreover, $\mathcal{P}_{v_b,R}$ is the projection onto $B_R(v_b)$ defined as

$$\mathcal{P}_{v_b,R}(v)\;:=\begin{cases} v, & \text{if } \|v-v_b\|_2\leq R,\\[1ex] v_b+R\dfrac{v-v_b}{\|v-v_b\|_2}, & \text{if } \|v-v_b\|_2>R. \end{cases} \tag{5}$$

As a crucial assumption in this paper, the map $\mathcal{P}_{v_b,R}$ depends on R and v_b in such way that $v^*\in B_R(v_b)$. Setting the parameters can be feasible under specific circumstances, as exemplified by the regularised optimisation problem $f(v):=\mathrm{Loss}\,(v)+\Lambda\|v\|_2$, wherein $v^*\in B_{\mathrm{Loss}\,(0)/\Lambda}(0)$. In the absence of prior knowledge regarding v_b and R, a practical approach is to choose $v_b=0$ and assign a sufficiently large value to R. The first terms in (1) and (3), respectively, impose a deterministic drift of each particle

https://doi.org/10.1017/S095679252400007X Published online by Cambridge University Press

towards the possibly projected momentaneous consensus point $v_\alpha(\widehat{\rho}_t^N)$, which is a weighted average of the particles' positions and computed according to

$$v_\alpha(\widehat{\rho}_t^N) := \int v \frac{\omega_\alpha(v)}{\|\omega_\alpha\|_{L_1(\widehat{\rho}_t^N)}}\, d\widehat{\rho}_t^N(v). \tag{6}$$

The weights $\omega_\alpha(v) := \exp(-\alpha f(v))$ are motivated by the well-known Laplace principle [44], which states for any absolutely continuous probability distribution ϱ on $\mathbb{R}^d$ that

$$\lim_{\alpha\to\infty}\left(-\frac{1}{\alpha}\log\left(\int \omega_\alpha(v)\, d\varrho(v)\right)\right) = \inf_{v\in\operatorname{supp}(\varrho)} f(v) \tag{7}$$

and thus justifies that $v_\alpha(\widehat{\rho}_t^N)$ serves as a suitable proxy for the global minimiser v^* given the currently available information of the particles $(V_t^i)_{i=1,\dots,N}$. The second terms in (1) and (3), respectively, encode the diffusion or exploration mechanism of the algorithm, where, in contrast to standard CBO, we truncate the noise by some fixed constant $M > 0$.

We conclude and re-iterate that both the introduction of the projection $\mathcal{P}_{v_b,R}\big(v_\alpha(\widehat{\rho}_t^N)\big)$ of the consensus point and the employment of truncation of the noise variance $\big(\big\|V_t^i - v_\alpha(\widehat{\rho}_t^N)\big\|_2 \wedge M\big)$ are main innovations to the original CBO method. We shall explain and justify these modifications in the following paragraph.

Despite these technical improvements, the approach to analyse the convergence behaviour of the implementable scheme (1) follows a similar route already explored in [8–11]. In particular, the convergence behaviour of the method to the global minimiser v^* of the objective f is investigated on the level of the mean-field limit [10, 45] of the system (3). More precisely, we study the macroscopic behaviour of the agent density $\rho \in \mathcal{C}([0,T], \mathcal{P}(\mathbb{R}^d))$, where $\rho_t = \mathrm{Law}(\overline{V}_t)$ with

$$d\overline{V}_t = -\lambda\big(\overline{V}_t - \mathcal{P}_{v_b,R}(v_\alpha(\rho_t))\big)\, dt + \sigma\big(\big\|\overline{V}_t - v_\alpha(\rho_t)\big\|_2 \wedge M\big)\, dB_t \tag{8}$$

and initial data $\overline{V}_0 \sim \rho_0$. Afterwards, by establishing a quantitative estimate on the mean-field approximation, that is, the proximity of the mean-field system (8) to the interacting particle system (3) and combining the two results, we obtain a convergence result for the CBO algorithm (1) with truncated noise.

1.1. Motivation for using truncated noise

In what follows, we provide a heuristic explanation of the theoretical benefits of employing a truncation in the noise of CBO as in (1), (3) and (8). Let us therefore first recall that the standard variant of CBO [7] can be retrieved from the model considered in this paper by setting $v_b = 0$, $R = \infty$ and $M = \infty$. For instance, in place of the mean-field dynamics (8), we would have

$$d\overline{V}_t^{\mathrm{CBO}} = -\lambda\Big(\overline{V}_t^{\mathrm{CBO}} - v_\alpha\big(\rho_t^{\mathrm{CBO}}\big)\Big)\, dt + \sigma\Big\|\overline{V}_t^{\mathrm{CBO}} - v_\alpha\big(\rho_t^{\mathrm{CBO}}\big)\Big\|_2\, dB_t.$$

Attributed to the Laplace principle (7), it holds $v_\alpha(\rho_t^{\mathrm{CBO}}) \approx v^*$ for α sufficiently large, that is, as $\alpha \to \infty$, the former dynamics converges to

$$d\overline{Y}_t^{\mathrm{CBO}} = -\lambda\Big(\overline{Y}_t^{\mathrm{CBO}} - v^*\Big)\, dt + \sigma\Big\|\overline{Y}_t^{\mathrm{CBO}} - v^*\Big\|_2\, dB_t. \tag{9}$$

First, observe that here the first term imposes a direct drift to the global minimiser v^* and thereby induces a contracting behaviour, which is on the other hand counteracted by the diffusion term, which contributes a stochastic exploration around this point. In particular, with $\overline{Y}_t^{\mathrm{CBO}}$ approaching v^*, the exploration vanishes so that $\overline{Y}_t^{\mathrm{CBO}}$ converges eventually deterministically to v^*. Conversely, as long as $\overline{Y}_t^{\mathrm{CBO}}$ is far away from v^*, the order of the random exploration is strong. By Itô's formula, we have

$$\frac{d}{dt}\mathbb{E}\Big[\Big\|\overline{Y}_t^{\mathrm{CBO}} - v^*\Big\|_2^p\Big] = p\Big(-\lambda + \frac{\sigma^2}{2}(p+d-2)\Big)\,\mathbb{E}\Big[\Big\|\overline{Y}_t^{\mathrm{CBO}} - v^*\Big\|_2^p\Big]$$

https://doi.org/10.1017/S095679252400007X Published online by Cambridge University Press

and thus

$$\mathbb{E}\left[\left\|\overline{Y}_t^{\mathrm{CBO}} - v^*\right\|_2^p\right] = \exp\left(p\left(-\lambda + \frac{\sigma^2}{2}(p+d-2)\right)t\right)\mathbb{E}\left[\left\|\overline{Y}_0^{\mathrm{CBO}} - v^*\right\|_2^p\right] \tag{10}$$

for any $p \geq 1$. Denoting with μ_t^{CBO} the law of $\overline{Y}_t^{\mathrm{CBO}}$, this means that, given any $\lambda, \sigma > 0$, there is some threshold exponent $p^* = p^*(\lambda, \sigma, d)$, such that

$$\begin{aligned}\lim_{t\to\infty} W_p\left(\mu_t^{\mathrm{CBO}}, \delta_{v^*}\right) &= \lim_{t\to\infty}\left(\mathbb{E}\left[\left\|\overline{Y}_t^{\mathrm{CBO}} - v^*\right\|_2^p\right]\right)^{1/p} \\ &= \lim_{t\to\infty}\exp\left(\left(-\lambda + \frac{\sigma^2}{2}(p+d-2)\right)t\right)\left(\mathbb{E}\left[\left\|\overline{Y}_0^{\mathrm{CBO}} - v^*\right\|_2^p\right]\right)^{1/p} \\ &= 0\end{aligned}$$

for $p < p^*$, while for $p > p^*$ it holds

$$\begin{aligned}\lim_{t\to\infty} W_p\left(\mu_t^{\mathrm{CBO}}, \delta_{v^*}\right) &= \lim_{t\to\infty}\left(\mathbb{E}\left[\left\|\overline{Y}_t^{\mathrm{CBO}} - v^*\right\|_2^p\right]\right)^{1/p} \\ &= \lim_{t\to\infty}\exp\left(\left(-\lambda + \frac{\sigma^2}{2}(p+d-2)\right)t\right)\left(\mathbb{E}\left[\left\|\overline{Y}_0^{\mathrm{CBO}} - v^*\right\|_2^p\right]\right)^{1/p} \\ &= \infty.\end{aligned}$$

Recalling that the distribution of a random variable Y has heavy tails if and only if the moment generating function $M_Y(s) := \mathbb{E}\left[\exp(sY)\right] = \mathbb{E}\left[\sum_{p=0}^{\infty}(sY)^p/p!\right]$ is infinite for all $s > 0$, these computations suggest that the distribution of μ_t^{CBO} exhibits characteristics of heavy tails as $t \to \infty$, thereby increasing the likelihood of encountering outliers in a sample drawn from μ_t^{CBO} for large t.

On the contrary, for CBO with truncated noise (8), we get, thanks once again to the Laplace principle as $\alpha \to \infty$, that (8) converges to

$$d\overline{Y}_t = -\lambda\left(\overline{Y}_t - v^*\right)dt + \sigma\left(\left\|\overline{Y}_t - v^*\right\|_2 \wedge M\right)dB_t, \tag{11}$$

for which we can compute

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\left[\left\|\overline{Y}_t - v^*\right\|_2^p\right] &\leq -p\lambda\mathbb{E}\left[\left\|\overline{Y}_t - v^*\right\|_2^p\right] + p\frac{\sigma^2}{2}M^2(p+d-2)\,\mathbb{E}\left[\left\|\overline{Y}_t - v^*\right\|_2^{p-2}\right] \\ &\leq -\lambda\mathbb{E}\left[\left\|\overline{Y}_t - v^*\right\|_2^p\right] + \lambda\frac{\sigma^p M^p(d+p-2)^{\frac{p}{2}}}{\lambda^{\frac{p}{2}}},\end{aligned}$$

for any $p \geq 2$. Notice, that to obtain the second inequality we used Young's inequality[1] as well as Jensen's inequality. By means of Grönwall's inequality, we then have

$$\mathbb{E}\left[\left\|\overline{Y}_t - v^*\right\|_2^p\right] \leq \exp(-\lambda t)\,\mathbb{E}\left[\left\|\overline{Y}_0 - v^*\right\|_2^p\right] + \frac{\sigma^p M^p(d+p-2)^{\frac{p}{2}}}{\lambda^{\frac{p}{2}}} \tag{12}$$

and therefore, denoting with μ_t the law of $\overline{Y}_t$,

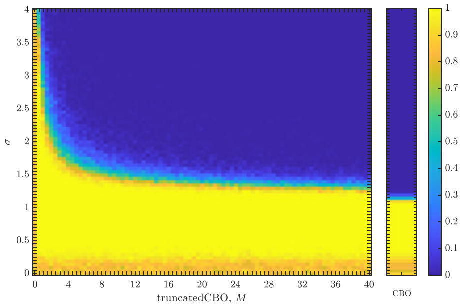
$$\lim_{t\to\infty} W_p(\mu_t, \delta_{v^*}) \leq \frac{\sigma M\sqrt{d+p-2}}{\lambda^{\frac{1}{2}}} < \infty$$

for any $p \geq 2$.

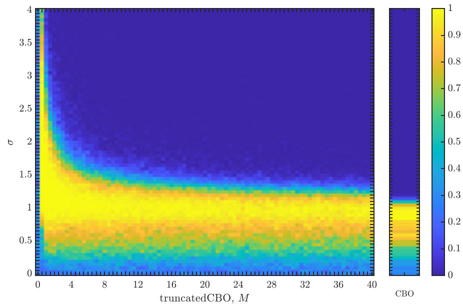
In conclusion, we observe from Equation (10) that the standard CBO dynamics as described in Equation (9) diverges in the setting $\sigma^2 d > 2\lambda$ when considering the Wasserstein-2 distance W_2. Contrarily, according to Equation (12), the CBO dynamics with truncated noise as presented in Equation (11) converges with exponential rate towards a neighbourhood of v^*, with radius $\sigma M\sqrt{d}/\sqrt{\lambda}$.

[1] Choose $a = \lambda^{\frac{p-2}{p}}\mathbb{E}\left[\left\|\overline{Y}_t - v^*\right\|_2^{p-2}\right]$ and $b = \frac{\sigma^2 M^2(d+p-2)}{\lambda^{(p-2)/p}}$, and recall that $ab \leq \frac{p-2}{p}a^{\frac{p}{p-2}} + \frac{2}{p}b^{\frac{p}{2}}$.

https://doi.org/10.1017/S095679252400007X Published online by Cambridge University Press



(a) Phase diagram of success probabilities of isotropic CBO with and without truncated noise at the example of the Ackley function $f(v) = -20\exp\left(-0.2/\sqrt{d}\,\|v\|_2\right) - \exp\left(1/d\sum_{k=1}^{d}\cos(2\pi v_k)\right)$ with $d = 4$



(b) Phase diagram of success probabilities of isotropic CBO with and without truncated noise at the example of the Rastrigin function $f(v) = \sum_{k=1}^{d} v_k^2 + 2.5(1 - \cos(2\pi v_k))$ with $d = 4$

Figure 1. *A comparison of the success probabilities of isotropic CBO with (left phase diagrams) and without (right separate columns) truncated noise for different values of the truncation parameter M and the noise level σ. (Note that standard CBO as investigated in [7, 8, 10] is retrieved when choosing $M = \infty$, $R = \infty$ and $v_b = 0$ in (1)). In both settings (**a**) and (**b**), the depicted success probabilities are averaged over 100 runs and the implemented scheme is given by an Euler–Maruyama discretisation of equation (3) with time horizon $T = 50$, discrete time step size $\Delta t = 0.01$, $R = \infty$, $v_b = 0$, $\alpha = 10^5$ and $\lambda = 1$. We use $N = 100$ particles, which are initialised according to $\rho_0 = \mathcal{N}((1, \dots, 1), 2000)$. In both figures, we plot the success probability of standard CBO (right separate column) and the CBO variant with truncated noise (left phase transition diagram) for different values of the truncation parameter M and the noise level σ, when optimising the Ackley ((**a**)) and Rastrigin ((**b**)) function, respectively. We observe that truncating the noise term (by decreasing M) consistently allows for a wider flexibility when choosing the noise level σ and thus increasing the likelihood of successfully locating the global minimiser.*

This implies that for a relatively small value of M, the CBO dynamics with truncated noise exhibits greater robustness in relation to the parameter $\sigma^2 d/\lambda$. This effect is confirmed numerically in Figure 1.

Remark 1 (Sub-Gaussianity of truncated CBO). *An application of Itô's formula allows to show that, for some $\kappa > 0$, $\mathbb{E}\left[\exp\left(\left\|\overline{Y}_t - v^*\right\|_2^2/\kappa^2\right)\right] < \infty$, provided $\mathbb{E}\left[\exp\left(\left\|\overline{Y}_0 - v^*\right\|_2^2/\kappa^2\right)\right] < \infty$. Thus, by incorporating a truncation in the noise term of the CBO dynamics, we ensure that the resulting distribution μ_t exhibits sub-Gaussian behaviour and therefore we enhance the regularity and well-behavedness of the statistics of μ_t. As a consequence, more reliable and stable results when analysing the properties and characteristics of the dynamics are to be expected.*

1.2. Contributions

In view of the aforementioned enhanced regularity and well-behavedness of the statistics of CBO with truncated noise compared to standard CBO [7] together with the numerically observed improved performance as depicted in Figure 1, a rigorous convergence analysis of the implementable CBO algorithm with truncated noise as given in (1) is of theoretical interest. In this work, we provide theoretical guarantees of global convergence of (1) to the global minimiser v^* for possibly non-convex and non-smooth objective functions f. The approach to analyse the convergence behaviour of the implementable scheme (1) follows a similar route as initiated and explored by the authors of [8–11]. In particular, we first investigate the mean-field behaviour (8) of the system (3). Then, by establishing a quantitative estimate on the mean-field approximation, that is, the proximity of the mean-field system (8) to the interacting particle system (3), we obtain a convergence result for the CBO algorithm (1) with truncated noise. Our proving

https://doi.org/10.1017/S095679252400007X Published online by Cambridge University Press

technique nevertheless differs in crucial parts from the one in [10, 11] as, on the one side, we do take advantage of the truncations, and, on the other side, we require additional technical effort to exploit and deal with the enhanced flexibility of the truncated model. Specifically, the central novelty can be identified in the proof of sub-Gaussianity of the process, see Lemma 8.

1.3. Organisation

In Section 2, we present and discuss our main theoretical contribution about the global convergence of CBO with truncated noise in probability and expectation. Section 3 collects the necessary proof details for this result. In Section 4, we numerically demonstrate the benefits of using truncated noise, before we provide a conclusion of the paper in Section 5. For the sake of reproducible research, in the GitHub repository https://github.com/KonstantinRiedl/CBOGlobalConvergenceAnalysis, we provide the Matlab code implementing CBO with truncated noise.

1.4. Notation

We use $\|\cdot\|_2$ to denote the Euclidean norm on $\mathbb{R}^d$. Euclidean balls are denoted as $B_r(u) := \{v \in \mathbb{R}^d : \|v-u\|_2 \leq r\}$. For the space of continuous functions $f: X \to Y$, we write $\mathcal{C}(X, Y)$, with $X \subset \mathbb{R}^n$ and a suitable topological space Y. For an open set $X \subset \mathbb{R}^n$ and for $Y = \mathbb{R}^m$, the spaces $\mathcal{C}^k_c(X, Y)$ and $\mathcal{C}^k_b(X, Y)$ contain functions $f \in \mathcal{C}(X, Y)$ that are k-times continuously differentiable and have compact support or are bounded, respectively. We omit Y in the real-valued case. All stochastic processes are considered on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$. The main objects of study are laws of such processes, $\rho \in \mathcal{C}([0, T], \mathcal{P}(\mathbb{R}^d))$, where the set $\mathcal{P}(\mathbb{R}^d)$ contains all Borel probability measures over $\mathbb{R}^d$. With $\rho_t \in \mathcal{P}(\mathbb{R}^d)$, we refer to a snapshot of such law at time t. Measures $\varrho \in \mathcal{P}(\mathbb{R}^d)$ with finite p-th moment $\int \|v\|_2^p \, d\varrho(v)$ are collected in $\mathcal{P}_p(\mathbb{R}^d)$. For any $1 \leq p < \infty$, W_p denotes the Wasserstein-p distance between two Borel probability measures $\varrho_1, \varrho_2 \in \mathcal{P}_p(\mathbb{R}^d)$, see, for example, [46]. $\mathbb{E}[\cdot]$ denotes the expectation.

2. Global convergence of CBO with truncated noise

We now present the main theoretical result of this work about the global convergence of CBO with truncated noise for objective functions that satisfy the following conditions.

Definition 2 (Assumptions). *Throughout we are interested in functions $f \in \mathcal{C}(\mathbb{R}^d)$, for which*

A1 *there exist $v^* \in \mathbb{R}^d$ such that $f(v^*) = \inf_{v \in \mathbb{R}^d} f(v) =: \underline{f}$ and $\underline{\alpha}, L_u > 0$ such that*

$$\sup_{v \in \mathbb{R}^d} \left\| v e^{-\alpha(f(v) - \underline{f})} \right\|_2 =: L_u < \infty \tag{13}$$

for any $\alpha \geq \underline{\alpha}$ and any $v \in \mathbb{R}^d$,

A2 *there exist $f_\infty, R_0, \nu, L_\nu > 0$ such that*

$$\|v - v^*\|_2 \leq \frac{1}{L_\nu} \big(f(v) - \underline{f}\big)^\nu \quad \text{for all } v \in B_{R_0}(v^*), \tag{14}$$

$$f_\infty < f(v) - \underline{f} \quad \text{for all } v \in \big(B_{R_0}(v^*)\big)^c, \tag{15}$$

A3 *there exist $L_\gamma > 0, \gamma \in [0, 1]$ such that*

$$|f(v) - f(w)| \leq L_\gamma \big(\|v - v^*\|_2^\gamma + \|w - v^*\|_2^\gamma\big) \, \|v - w\|_2 \quad \text{for all } v, w \in \mathbb{R}^d, \tag{16}$$

$$f(v) - \underline{f} \leq L_\gamma \big(1 + \|v - v^*\|_2^{1+\gamma}\big) \quad \text{for all } v \in \mathbb{R}^d. \tag{17}$$

A few comments are in order: Condition A1 establishes the existence of a minimiser v^* and requires a certain growth of the function f. Condition A2 ensures that the value of the function f at a point v

https://doi.org/10.1017/S095679252400007X Published online by Cambridge University Press

can locally be an indicator of the distance between v and the minimiser v^*. This error-bound condition was first introduced in [10] under the name inverse continuity condition. It in particular guarantees the uniqueness of the global minimiser v^*. Condition A3 sets controllable bounds on the local Lipschitz constant of f and on the growth of f, which is required to be at most quadratic. A similar requirement appears also in [8, 10], but there also a quadratic lower bound was imposed.

2.1. Main result

We can now state the main result of the paper. Its proof is deferred to Section 3.

Theorem 3. *Let $f \in \mathcal{C}(\mathbb{R}^d)$ satisfy A1, A2 and A3. Moreover, let $\rho_0 \in \mathcal{P}_4(\mathbb{R}^d)$ with $v^* \in \operatorname{supp}(\rho_0)$. Let $V^i_{0,\Delta t}$ be sampled i.i.d. from ρ_0 and denote by $((V^i_{k,\Delta t})_{k=1,\dots,K})_{i=1,\dots,N}$ the iterations generated by the numerical scheme (1). Fix any $\epsilon \in \left(0, W_2^2(\rho_0, \delta_{v^*})\right)$, define the time horizon*

$$T^* := \frac{1}{\lambda}\log\left(\frac{2W_2^2(\rho_0, \delta_{v^*})}{\epsilon}\right)$$

and let $K \in \mathbb{N}$ and Δt satisfy $K\Delta t = T^*$. Moreover, let $R \in \left(\|v_b - v^*\|_2 + \sqrt{\epsilon/2}, \infty\right)$, $M \in (0, \infty)$ and $\lambda, \sigma > 0$ be such that $\lambda \geq 2\sigma^2 d$ or $\sigma^2 M^2 d = \mathcal{O}(\epsilon)$. Then, by choosing α sufficiently large and $N \geq \left(16\alpha L_\gamma \sigma^2 M^2\right)/\lambda$, it holds

$$\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^N V^i_{K,\Delta t} - v^*\right\|_2^2\right] \lesssim C_{\mathrm{NA}}(\Delta t)^{2m} + \frac{C_{\mathrm{MFA}}}{N} + \epsilon \tag{18}$$

up to a generic constant. Here, C_{NA} depends linearly on the dimension d and the number of particles N and exponentially on the time horizon T^*, m is the order of accuracy of the numerical scheme (for the Euler–Maruyama scheme $m = 1/2$), and $C_{\mathrm{MFA}} = C_{\mathrm{MFA}}(\lambda, \sigma, d, \alpha, L_\nu, \nu, L_\gamma, L_u, T^*, R, v_b, v^*, M)$.

Remark 4. *In the statement of Theorem 3, the parameters R and v_b play a crucial role. We already mentioned how they can be chosen in an example after Equation (5). The role of these parameters is bolstered in particular in the proof of Theorem 3, where it is demonstrated that, by selecting a sufficiently large α depending on R and v_b, the dynamics (8) can be set equal to*

$$d\overline{V}_t = -\lambda\left(\overline{V}_t - \mathcal{P}_{v^*,\delta}(v_\alpha(\rho_t))\right)dt + \sigma\left(\left\|\overline{V}_t - v_\alpha(\rho_t)\right\|_2 \wedge M\right)dB_t,$$

where δ represents a small value. For the dynamics (3), we can analogously establish its equivalence to

$$dV_t^i = -\lambda\left(V_t^i - \mathcal{P}_{v^*,\delta}\left(v_\alpha\left(\widehat{\rho}_t^N\right)\right)\right)dt + \sigma\left(\left\|V_t^i - v_\alpha\left(\widehat{\rho}_t^N\right)\right\|_2 \wedge M\right)dB_t^i, \quad i = 1, \dots, N,$$

with high probability, contingent upon the selection of sufficiently large values for both α and N.

Remark 5. *The convergence result in the form of Theorem 3 obtained in this work differs from the one presented in [10, Theorem 14] in the sense that we obtain convergence is in expectation, while in [10] convergence with high probability is established. This distinction arises from the truncation of the noise term employed in our algorithm.*

3. Proof details for section 2

3.1. Well-posedness of equations (1) and (3)

With the projection map $\mathcal{P}_{v_b,R}$ being 1-Lipschitz, existence and uniqueness of strong solutions to the SDEs (1) and (3) are assured by essentially analogous proofs as in [8, Theorems 2.1, 3.1 and 3.2]. The details shall be omitted. Let us remark, however, that due to the presence of the truncation and the projection map, we do not require the function f to be bounded from above or exhibit quadratic growth outside a ball, as required in [8, Theorems 2.1, 3.1 and 3.2].

https://doi.org/10.1017/S095679252400007X Published online by Cambridge University Press

3.2. Proof details for theorem 3

Remark 6. *Since adding some constant offset to f does not affect the dynamics of Equations (3) and (8), we will assume $\underline{f} = 0$ in the proofs for simplicity but without loss of generality.*

Let us first provide a sketch of the proof of Theorem 3. For the approximation error (18), we have the error decomposition

$$\begin{aligned}\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^{N} V^i_{K,\Delta t} - v^*\right\|_2^2\right] &\lesssim \underbrace{\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\left(V^i_{K,\Delta t} - V^i_{T^*}\right)\right\|_2^2\right]}_{I} + \underbrace{\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\left(V^i_{T^*} - \overline{V}^i_{T^*}\right)\right\|_2^2\right]}_{II} \\ &\quad + \underbrace{\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\overline{V}^i_{T^*} - v^*\right\|_2^2\right]}_{III},\end{aligned} \tag{19}$$

where $\left((\overline{V}^i_t)_{t\geq 0}\right)_{i=1,\dots,N}$ denote N independent copies of the mean-field process $\left(\overline{V}_t\right)_{t\geq 0}$ satisfying Equation (8).

In what follows, we investigate each of the three term separately. Term I can be bounded by $C_{\mathrm{NA}}(\Delta t)^{2m}$ using classical results on the convergence of numerical schemes for SDEs, as mentioned for instance in [43]. The second and third terms, respectively, are analysed in separate subsections, providing detailed explanations and bounds for each of the two terms II and III.

Before doing so, let us provide a concise guide for reading the proofs. As the proofs are quite technical, we start for reader's convenience by presenting the main building blocks of the result first and collect the more technical steps in subsequent lemmas. This arrangement should hopefully allow to grasp the structure of the proof more easily and to dig deeper into the details along with the reading.

3.2.1. Upper bound for the second term in (19)

For Term II of the error decomposition (19), we have the following upper bound.

Proposition 7. *Let $f \in \mathcal{C}(\mathbb{R}^d)$ satisfy A1, A2 and A3. Moreover, let R and M be finite such that $R \geq \|v_b - v^*\|_2$ and let $N \geq (16\alpha L_\gamma \sigma^2 M^2)/\lambda$. Then, we have*

$$\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\left(V^i_{T^*} - \overline{V}^i_{T^*}\right)\right\|_2^2\right] \leq \frac{C_{\mathrm{MFA}}}{N}, \tag{20}$$

where $C_{\mathrm{MFA}} = C_{\mathrm{MFA}}(\lambda, \sigma, d, \alpha, L_\nu, \nu, L_\gamma, L_u, T^*, R, v_b, v^*, M)$.

Proof. By a synchronous coupling, we have

$$\begin{aligned} d\overline{V}^i_t &= -\lambda\left(\overline{V}^i_t - \mathcal{P}_{v_b,R}(v_\alpha(\rho_t))\right) dt + \sigma\left(\left\|\overline{V}^i_t - v_\alpha(\rho_t)\right\|_2 \wedge M\right) dB^i_t, \\ dV^i_t &= -\lambda\left(V^i_t - \mathcal{P}_{v_b,R}\left(v_\alpha\left(\widehat{\rho}^N_t\right)\right)\right) dt + \sigma\left(\left\|\overline{V}^i_t - v_\alpha\left(\widehat{\rho}^N_t\right)\right\|_2 \wedge M\right) dB^i_t, \end{aligned}$$

with coinciding Brownian motions. Moreover, recall that $\mathrm{Law}\left(\overline{V}^i_t\right) = \rho_t$ and $\widehat{\rho}^N_t = 1/N\sum_{i=1}^N \delta_{V^i_t}$. By Itô's formula, we then have

$$\begin{aligned} d\left\|\overline{V}^i_t - V^i_t\right\|_2^2 &= \Big(-2\lambda\Big\langle \overline{V}^i_t - V^i_t, \left(\overline{V}^i_t - V^i_t\right) - \left(\mathcal{P}_{v_b,R}(v_\alpha(\rho_t)) - \mathcal{P}_{v_b,R}\left(v_\alpha\left(\widehat{\rho}^N_t\right)\right)\right)\Big\rangle \\ &\quad + \sigma^2 d\left(\left\|\overline{V}^i_t - v_\alpha(\rho_t)\right\|_2 \wedge M - \left\|V^i_t - v_\alpha\left(\widehat{\rho}^N_t\right)\right\|_2 \wedge M\right)^2\Big) dt \\ &\quad + 2\sigma\left(\left\|\overline{V}^i_t - v_\alpha(\rho_t)\right\|_2 \wedge M - \left\|V^i_t - v_\alpha\left(\widehat{\rho}^N_t\right)\right\|_2 \wedge M\right)\left(\overline{V}^i_t - V^i_t\right)^\top dB^i_t, \end{aligned} \tag{21}$$

https://doi.org/10.1017/S095679252400007X Published online by Cambridge University Press

and after taking the expectation on both sides

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\left[\left\|\overline{V}_t^i - V_t^i\right\|_2^2\right] &= -2\lambda\mathbb{E}\left[\left\langle \overline{V}_t^i - V_t^i, \left(\overline{V}_t^i - V_t^i\right) - \left(\mathcal{P}_{v_b,R}(v_\alpha(\rho_t)) - \mathcal{P}_{v_b,R}\left(v_\alpha\left(\widehat{\rho}_t^N\right)\right)\right)\right\rangle\right] \\ &\quad + \sigma^2 d\mathbb{E}\left[\left(\left\|\overline{V}_t^i - v_\alpha(\rho_t)\right\|_2 \wedge M - \left\|V_t^i - v_\alpha\left(\widehat{\rho}_t^N\right)\right\|_2 \wedge M\right)^2\right] \\ &\leq -2\lambda\mathbb{E}\left[\left\|\overline{V}_t^i - V_t^i\right\|_2^2\right] + \sigma^2 d\mathbb{E}\left[\left\|\left(\overline{V}_t^i - V_t^i\right) - \left(v_\alpha(\rho_t) - v_\alpha\left(\widehat{\rho}_t^N\right)\right)\right\|_2^2\right] \\ &\quad + 2\lambda\mathbb{E}\left[\left\|\overline{V}_t^i - V_t^i\right\|_2 \left\|\mathcal{P}_{v_b,R}(v_\alpha(\rho_t)) - \mathcal{P}_{v_b,R}\left(v_\alpha\left(\widehat{\rho}_t^N\right)\right)\right\|_2\right] \\ &\leq -2\lambda\mathbb{E}\left[\left\|\overline{V}_t^i - V_t^i\right\|_2^2\right] + 2\lambda\mathbb{E}\left[\left\|\overline{V}_t^i - V_t^i\right\|_2 \left\|v_\alpha(\rho_t) - v_\alpha\left(\widehat{\rho}_t^N\right)\right\|_2\right] \\ &\quad + \sigma^2 d\mathbb{E}\left[\left\|\left(\overline{V}_t^i - V_t^i\right) - \left(v_\alpha(\rho_t) - v_\alpha\left(\widehat{\rho}_t^N\right)\right)\right\|_2^2\right].\end{aligned} \tag{22}$$

Here, let us remark that the last (stochastic) term in (21) disappears after taking the expectation. This is due to $\mathbb{E}\left[\left\|\overline{V}_t^i - V_t^i\right\|_2^2\right] < \infty$, which can be derived from Lemma 8 after noticing that Lemma 8 also holds for processes V_t^i. Since by Young's inequality, it holds

$$2\lambda\mathbb{E}\left[\left\|\overline{V}_t^i - V_t^i\right\|_2 \left\|v_\alpha(\rho_t) - v_\alpha\left(\widehat{\rho}_t^N\right)\right\|_2\right] \leq \lambda\left(\frac{\mathbb{E}\left[\left\|\overline{V}_t^i - V_t^i\right\|_2^2\right]}{2} + 2\mathbb{E}\left[\left\|v_\alpha(\rho_t) - v_\alpha\left(\widehat{\rho}_t^N\right)\right\|_2^2\right]\right),$$

and

$$\mathbb{E}\left[\left\|\left(\overline{V}_t^i - V_t^i\right) - \left(v_\alpha(\rho_t) - v_\alpha\left(\widehat{\rho}_t^N\right)\right)\right\|_2^2\right] \leq 2\mathbb{E}\left[\left\|\overline{V}_t^i - V_t^i\right\|_2^2 + \left\|v_\alpha(\rho_t) - v_\alpha\left(\widehat{\rho}_t^N\right)\right\|_2^2\right],$$

we obtain

$$\frac{d}{dt}\mathbb{E}\left[\left\|\overline{V}_t^i - V_t^i\right\|_2^2\right] \leq \left(-\frac{3\lambda}{2} + 2\sigma^2 d\right)\mathbb{E}\left[\left\|\overline{V}_t^i - V_t^i\right\|_2^2\right] + 2\left(\lambda + \sigma^2 d\right)\mathbb{E}\left[\left\|v_\alpha(\rho_t) - v_\alpha\left(\widehat{\rho}_t^N\right)\right\|_2^2\right] \tag{23}$$

after inserting the former two inequalities into Equation (22). For the term $\mathbb{E}\left[\left\|v_\alpha(\rho_t) - v_\alpha\left(\widehat{\rho}_t^N\right)\right\|_2^2\right]$, we can decompose

$$\mathbb{E}\left[\left\|v_\alpha(\rho_t) - v_\alpha\left(\widehat{\rho}_t^N\right)\right\|_2^2\right] \leq 2\mathbb{E}\left[\left\|v_\alpha(\rho_t) - v_\alpha\left(\bar{\rho}_t^N\right)\right\|_2^2\right] + 2\mathbb{E}\left[\left\|v_\alpha\left(\bar{\rho}_t^N\right) - v_\alpha\left(\widehat{\rho}_t^N\right)\right\|_2^2\right], \tag{24}$$

where we denote

$$\bar{\rho}_t^N = \frac{1}{N}\sum_{i=1}^N \delta_{\overline{V}_t^i}.$$

For the first term in Equation (24), by Lemma 11, we have

$$\mathbb{E}\left[\left\|v_\alpha(\rho_t) - v_\alpha\left(\bar{\rho}_t^N\right)\right\|_2^2\right] \leq C_0\frac{1}{N}$$

for some constant C_0 depending on $\lambda, \sigma, d, \alpha, L_\gamma, L_u, T^*, R, v_b, v^*$ and M. For the second term in Equation (24), by combining [8, Lemma 3.2] and Lemma 8, we obtain

$$\mathbb{E}\left[\left\|v_\alpha\left(\bar{\rho}_t^N\right) - v_\alpha\left(\widehat{\rho}_t^N\right)\right\|_2^2\right] \leq C_1\frac{1}{N}\sum_{i=1}^N \mathbb{E}\left[\left\|\overline{V}_t^i - V_t^i\right\|_2^2\right],$$

https://doi.org/10.1017/S095679252400007X Published online by Cambridge University Press

for some constant C_1 depending on $\lambda, \sigma, d, \alpha, L_u, R$ and M. Combining these estimates, we conclude

$$\frac{d}{dt}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\left\|\overline{V}_t^i - V_t^i\right\|_2^2\right] \le \left(-\frac{3\lambda}{2} + 2\sigma^2 d + 4C_1\left(\lambda + \sigma^2 d\right)\right)\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\left\|\overline{V}_t^i - V_t^i\right\|_2^2\right] + 4\left(\lambda + \sigma^2 d\right) C_0 \frac{1}{N}.$$

After an application of Grönwall's inequality and noting that $\overline{V}_0^i = V_0^i$ for all $i = 1, \ldots, N$, we have

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\left\|\overline{V}_t^i - V_t^i\right\|_2^2\right] \le 4\left(\lambda + \sigma^2 d\right)\frac{C_0}{N} t e^{\left(-\frac{3\lambda}{2} + 2\sigma^2 d + 4C_1\left(\lambda + \sigma^2 d\right)\right)t}. \tag{25}$$

for any $t \in [0, T^*]$. Finally, by Jensen's inequality and letting $t = T^*$, we have

$$\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\left(V_{T^*}^i - \overline{V}_{T^*}^i\right)\right\|_2^2\right] \le \frac{C_{\mathrm{MFA}}}{N}, \tag{26}$$

where the constant C_{MFA} depends on $\lambda, \sigma, d, \alpha, L_u, L_\gamma, T^*, R, v_b, v^*$ and M. □

In the next lemma, we show that the distribution of $\overline{V}_t$ is sub-Gaussian.

Lemma 8. *Let R and M be finite with $R \ge \|v_b - v^*\|_2$. For any $\kappa > 0$, let N satisfy $N \ge \left(4\sigma^2 M^2\right)/\left(\lambda\kappa^2\right)$. Then, provided that $\mathbb{E}\left[\exp\left(\sum_{i=1}^N \|\overline{V}_0^i - v^*\|_2^2/\left(N\kappa^2\right)\right)\right] < \infty$, it holds*

$$C_\kappa := \sup_{t\in[0,T^*]}\mathbb{E}\left[\exp\left(\frac{\sum_{i=1}^N \|\overline{V}_t^i - v^*\|_2^2}{N\kappa^2}\right)\right] < \infty, \tag{27}$$

where C_κ depends on $\kappa, \lambda, \sigma, d, R, M$ and T^*, and where

$$d\overline{V}_t^i = -\lambda\left(\overline{V}_t^i - \mathcal{P}_{v_b,R}\left(v_\alpha(\rho_t)\right)\right)dt + \sigma\left(\left\|\overline{V}_t^i - v_\alpha(\rho_t)\right\|_2 \wedge M\right)dB_t^i$$

for $i = 1, \ldots, N$ with B_t^i being independent to each other and $\mathrm{Law}\left(\overline{V}_t^i\right) = \rho_t$.

Proof. To apply Itô's formula, we need to truncate the function $\exp\left(\|v\|_2^2/\kappa^2\right)$ from above. For this, define for $W > 0$ the function

$$G_W(x) := \begin{cases} x, & x \in [0, W-1], \\ \dfrac{1}{16}(x + 1 - W)^4 - \frac{1}{4}(x + 1 - W)^3 + x, & x \in [W-1, W+1], \\ W, & x \in [W+1, \infty), \end{cases}$$

It is easy to verify that G_W is a $\mathcal{C}^2$ approximation of the function $x \wedge W$ satisfying $G_W \in \mathcal{C}^2(\mathbb{R}^+)$, $G_W(x) \le x \wedge W$, $G'_W \in [0, 1]$ and $G''_W \le 0$.

Since $G_{W,N,\kappa}(t) := \exp\left(G_W\left(\sum_{i=1}^N \|\overline{V}_t^i - v^*\|_2^2/N\right)/\kappa^2\right)$ is upper-bounded, we can apply Itô's formula to it. We abbreviate $G'_W := G'_W\left(\sum_{i=1}^N \|\overline{V}_t^i - v^*\|_2^2/N\right)$ and $G''_W := G''_W\left(\sum_{i=1}^N \|\overline{V}_t^i\|_2^2/N\right)$ in what follows. With the notation $Y_t := \left(\left(\overline{V}_t^1\right)^\top, \cdots, \left(\overline{V}_t^N\right)^\top\right)^\top$, the Nd-dimensional process Y_t satisfies $dY_t = -\lambda\left(Y_t - \overline{\mathcal{P}_{v_b,R}(\rho_t)}\right)dt + \mathcal{M}dB_t$, where $\overline{\mathcal{P}_{v_b,R}(\rho_t)} = \left(\mathcal{P}_{v_b,R}(\rho_t)^\top, \ldots, \mathcal{P}_{v_b,R}(\rho_t)^\top\right)^\top$, $\mathcal{M} = \mathrm{diag}(\mathcal{M}_1, \ldots, \mathcal{M}_N)$ with

$\mathcal{M}_i = \sigma \left\| \overline{V}_t^i - v_\alpha(\rho_t) \right\|_2 \wedge M \mathrm{I}_d$ and B_t the Nd-dimensional Brownian motion. We then have $G_{W,N,\kappa}(t) = \exp\big(G_W\big(\|Y_t\|_2^2/N\big)/\kappa^2\big)$ and

$$
\begin{aligned}
dG_{W,N,\kappa}(t) &= \sum_{i=1}^N \nabla_{Y_t} G_{W,N,\kappa}(t) dY_t + \frac{1}{2} \operatorname{tr}\big(\mathcal{M} \nabla^2_{Y_t,Y_t} G_{W,N,\kappa}(t) \mathcal{M}\big)\, dt \\
&= G_{W,N,\kappa}(t) \frac{G'_W}{\kappa^2} \sum_{i=1}^N \left(2 \frac{\overline{V}_t^i - v^*}{N} \right)^\top d\overline{V}_t^i \\
&\quad + \frac{1}{2} G_{W,N,\kappa}(t) \sum_{i=1}^N \left(G'_W \frac{2d}{N\kappa^2} + G''_W \frac{4\left\|\overline{V}_t^i - v^*\right\|_2^2}{N^2\kappa^2} \right. \\
&\quad \left. + \left(G'_W\right)^2 \frac{4\left\|\overline{V}_t^i - v^*\right\|_2^2}{N^2\kappa^4} \right) \left(\sigma \left\| \overline{V}_t^i - v_\alpha\left(\rho_t\right) \right\|_2 \wedge M \right)^2 dt.
\end{aligned}
\tag{28}
$$

The first term on the right-hand side of (28) can be expanded as

$$
\begin{aligned}
G_{W,N,\kappa}(t) \frac{G'_W}{\kappa^2} \sum_{i=1}^N \left(2 \frac{\overline{V}_t^i - v^*}{N} \right)^\top d\overline{V}_t^i &= G_{W,N,\kappa}(t) G'_W \sum_{i=1}^N \left(2 \frac{\overline{V}_t^i - v^*}{N\kappa^2} \right)^\top d\overline{V}_t^i \\
= G_{W,N,\kappa}(t) G'_W \sum_{i=1}^N \left(2 \frac{\overline{V}_t^i - v^*}{N\kappa^2} \right)^\top & \left(-\lambda \left(\bar{V}_t^i - v^* + v^* - \mathcal{P}_{v_b,R}(\rho_t) \right) dt + \sigma \left(\left\| \overline{V}_t^i - v_\alpha(\rho_t) \right\|_2 \wedge M \right) dB_t^i \right) \\
= G_{W,N,\kappa}(t) G'_W \Bigg\{ \frac{-2\lambda}{N\kappa^2} \sum_{i=1}^N \left\| \overline{V}_t^i - v^* \right\|_2^2 dt &- \frac{2\lambda}{N\kappa^2} \sum_{i=1}^N \left\langle \overline{V}_t^i - v^*, v^* - \mathcal{P}_{v_b,R}(v_\alpha(\rho_t)) \right\rangle dt \\
&+ 2\sigma \sum_{i=1}^N \left(\left\| \overline{V}_t^i - v_\alpha(\rho_t) \right\|_2 \wedge M \right) \left(\frac{(\overline{V}_t^i - v^*)}{N\kappa^2} \right)^\top dB_t^i \Bigg\}.
\end{aligned}
\tag{29}
$$

Notice additionally that

$$
\left\langle \overline{V}_t^i - v^*, v^* - \mathcal{P}_{v_b,R}(v_\alpha(\rho_t)) \right\rangle \leq \left\| \overline{V}_t^i - v^* \right\|_2 \left\| v^* - \mathcal{P}_{v_b,R}(v_\alpha(\rho_t)) \right\|_2 \leq 2R \left\| \overline{V}_t^i - v^* \right\|_2 \tag{30}
$$

as v^* and $\mathcal{P}_{v_b,R}(v_\alpha(\rho_t))$ belong to the same ball $B_R(v_b)$ around v_b of radius R. Similarly, we can expand the coefficient of the second term. According to the properties $G'_W \in [0, 1]$ and $G''_W \leq 0$, we can bound it from above yielding

$$
\begin{aligned}
\frac{1}{2} G_{W,N,\kappa}(t) \sum_{i=1}^N & \left(G'_W \frac{2d}{N\kappa^2} + G''_W \frac{4\left\|\overline{V}_t^i - v^*\right\|_2^2}{N^2\kappa^2} + \left(G'_W\right)^2 \frac{4\left\|\overline{V}_t^i - v^*\right\|_2^2}{N^2\kappa^4} \right) \left(\sigma \left\| \overline{V}_t^i - v_\alpha\left(\rho_t\right) \right\|_2 \wedge M \right)^2 \\
&\leq G_{W,N,\kappa}(t) G'_W \frac{\sigma^2 M^2 d}{\kappa^2} + G_{W,N,\kappa}(t) \left(G'_W\right)^2 \frac{2\sigma^2 M^2}{N^2\kappa^4} \sum_{i=1}^N \left\| \overline{V}_t^i - v^* \right\|_2^2 \\
&\leq G_{W,N,\kappa}(t) G'_W \frac{\sigma^2 M^2 d}{\kappa^2} + G_{W,N,\kappa}(t) G'_W \frac{2\sigma^2 M^2}{N^2\kappa^4} \sum_{i=1}^N \left\| \overline{V}_t^i - v^* \right\|_2^2.
\end{aligned}
\tag{31}
$$

By taking expectations in (28) and combining it with (29), (30) and (31), we obtain

$$
\begin{aligned}
\frac{d}{dt} \mathbb{E}\big[G_{W,N,\kappa}(t)\big] \leq \mathbb{E}\Bigg[& G_{W,N,\kappa}(t) G'_W \left(\frac{-2\lambda}{N\kappa^2} \sum_{i=1}^N \left\| \overline{V}_t^i - v^* \right\|_2^2 + \frac{4R\lambda}{N\kappa^2} \sum_{i=1}^N \left\| \overline{V}_t^i - v^* \right\|_2 \right. \\
&\left. + \ G_{W,N,\kappa}(t) G'_W \frac{\sigma^2 M^2 d}{\kappa^2} + G_{W,N,\kappa}(t) G'_W \frac{2\sigma^2 M^2}{N^2\kappa^4} \sum_{i=1}^N \left\| \overline{V}_t^i - v^* \right\|_2^2 \right) \Bigg].
\end{aligned}
$$

https://doi.org/10.1017/S095679252400007X Published online by Cambridge University Press

Rearranging the former yields

$$\frac{d}{dt}\mathbb{E}\big[G_{W,N,\kappa}(t)\big] \le \mathbb{E}\Bigg[G_{W,N,\kappa}(t)G'_W\Bigg(\Bigg(\Bigg(\frac{4\lambda R}{N\kappa^2}\sum_{i=1}^N \big\|\overline{V}_t^i - v^*\big\|_2\Bigg) + \frac{\sigma^2M^2d}{\kappa^2}\Bigg) - \Bigg(\frac{2\lambda}{N\kappa^2} - \frac{2\sigma^2M^2}{N^2\kappa^4}\Bigg)\sum_{i=1}^N \big\|\overline{V}_t^i - v^*\big\|_2^2\Bigg)\Bigg]. \tag{32}$$

Since by Young's inequality, it holds $4R\big\|\overline{V}_t^i - v^*\big\|_2 \le 4R^2 + \big\|\overline{V}_t^i - v^*\big\|_2^2$, we can continue Estimate (32) by

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\big[G_{W,N,\kappa}(t)\big] &\le \mathbb{E}\Bigg[G_{W,N,\kappa}(t)G'_W\Bigg(\frac{\sigma^2M^2d + 4\lambda R^2}{\kappa^2} - \Bigg(\frac{\lambda}{N\kappa^2} - \frac{2\sigma^2M^2}{N^2\kappa^4}\Bigg)\sum_{i=1}^N \big\|\overline{V}_t^i - v^*\big\|_2^2\Bigg)\Bigg]\\ &\le \mathbb{E}\Bigg[G_{W,N,\kappa}(t)G'_W\Bigg(-A\sum_{i=1}^N \big\|\overline{V}_t^i - v^*\big\|_2^2 + B\Bigg)\Bigg]\end{aligned} \tag{33}$$

with $A := \frac{\lambda}{N\kappa^2} - \frac{2\sigma^2M^2}{N^2\kappa^4}$ and $B := \frac{\sigma^2M^2d+4\lambda R^2}{\kappa^2}$. Now, if $\sum_{i=1}^N \big\|\overline{V}_t^i - v^*\big\|_2^2 \ge (B-1)/A$, we have

$$G_{W,N,\kappa}(t)G'_W\Bigg(-A\sum_{i=1}^N \big\|\overline{V}_t^i - v^*\big\|_2^2 + B\Bigg) \le 0,$$

while, if $\sum_{i=1}^N \big\|\overline{V}_t^i - v^*\big\|_2^2 \le (B-1)/A$, we have

$$G_{W,N,\kappa}(t)G'_W\Bigg(-A\sum_{i=1}^N \big\|\overline{V}_t^i - v^*\big\|_2^2 + B\Bigg) \le Be^{\frac{B-1}{N\kappa^2A}}.$$

Thus, the latter inequality always holds true and consequently we have with (33)

$$\frac{d}{dt}\mathbb{E}\big[G_{W,N,\kappa}(t)\big] \le Be^{\frac{B-1}{N\kappa^2A}},$$

which gives after integration

$$\begin{aligned}\mathbb{E}\big[G_{W,N,\kappa}(t)\big] &\le \mathbb{E}\big[G_{W,N,\kappa}(0)\big] + Be^{\frac{B-1}{N\kappa^2A}}t\\ &\le \mathbb{E}\Bigg[\exp\Bigg(\frac{\sum_{i=1}^N \big\|\overline{V}_0^i - v^*\big\|_2^2}{N\kappa^2}\Bigg)\Bigg] + Be^{\frac{B-1}{N\kappa^2A}}t.\end{aligned}$$

Letting $W \to \infty$, we eventually obtain

$$\mathbb{E}\Bigg[\exp\Bigg(\frac{\sum_{i=1}^N \big\|\overline{V}_t^i - v^*\big\|_2^2}{N\kappa^2}\Bigg)\Bigg] \le \mathbb{E}\Bigg[\exp\Bigg(\frac{\sum_{i=1}^N \big\|\overline{V}_0^i - v^*\big\|_2^2}{N\kappa^2}\Bigg)\Bigg] + Be^{\frac{B-1}{N\kappa^2A}}t < \infty, \tag{34}$$

provided that $\mathbb{E}\Big[\exp\Big(\sum_{i=1}^N \big\|\overline{V}_0^i - v^*\big\|_2^2/N\kappa^2\Big)\Big] < \infty$.

If $N \ge (4\sigma^2M^2)/(\lambda\kappa^2)$, we have

$$\frac{B-1}{N\kappa^2A} \le \frac{B}{N\kappa^2A} = \frac{N\big(\sigma^2M^2d + 4\lambda R^2\big)}{\lambda N\kappa^2 - 2\sigma^2M^2} \le C(\kappa, \lambda, \sigma, M, R, d).$$

Thus, C_κ is upper-bounded and independent of N. □

Remark 9. *The sub-Gaussianity of $\overline{V}_t$ follows from Lemma 8 by noticing that the statement can be applied in the setting $N = 1$ when choosing κ sufficiently large.*

Remark 10. *In Lemma 8, as the number of particles N increases, the condition for κ to ensure $C_\kappa < \infty$ becomes more relaxed. Specifically, the value of κ can be as small as one needs as N increases. This phenomenon can be easily understood by considering the limit as N approaches infinity. In this case,*

https://doi.org/10.1017/S095679252400007X Published online by Cambridge University Press

C_κ tends to $\sup_{t\in[0,T^*]}\exp\left(\mathbb{E}\left[\left\|\overline{V}_t - v^*\right\|_2^2\right]/\kappa^2\right)$. Therefore, as one shows an upper bound on the second moment of $\overline{V}_t$, it becomes evident that C_κ remains finite as N tends to infinity.

With the help of Lemma 8, we can now prove the following lemma.

Lemma 11. *Let $f\in\mathcal{C}(\mathbb{R}^d)$ satisfy A1 and A3. Then, for any $t\in[0,T^*]$, M and R with $R\geq\|v_b - v^*\|_2$ finite, and N satisfying $N\geq(16\alpha L_\gamma\sigma^2M^2)/\lambda$, we have*

$$\mathbb{E}\left[\left\|v_\alpha(\rho_t) - v_\alpha\left(\bar\rho_t^N\right)\right\|_2^2\right]\leq\frac{C_0}{N}, \tag{35}$$

where $C_0 := C_0(\lambda,\sigma,d,\alpha,L_\gamma,L_u,T^*,R,v_b,v^*,M)$.

Proof. Without the loss of generality, we assume $v^*=0$ and recall that we assumed $\underline{f}=0$ in the proofs as of Remark 6. We have

$$\begin{aligned}
\mathbb{E}\left[\left\|v_\alpha(\rho_t) - v_\alpha\left(\bar\rho_t^N\right)\right\|_2^2\right] &= \mathbb{E}\left[\left\|\frac{\frac{1}{N}\sum_{i=1}^N\overline{V}_t^i e^{-\alpha f(\overline{V}_t^i)}}{\frac{1}{N}\sum_{i=1}^N e^{-\alpha f(\overline{V}_t^i)}} - \frac{\int_{\mathbb{R}^d} v e^{-\alpha f(v)}d\rho_t(v)}{\int_{\mathbb{R}^d} e^{-\alpha f(v)}d\rho_t(v)}\right\|_2^2\right]\\
&\leq 2\mathbb{E}\left[\left\|\frac{1}{\frac{1}{N}\sum_{i=1}^N e^{-\alpha f(\overline{V}_t^i)}}\left(\frac{1}{N}\sum_{i=1}^N\overline{V}_t^i e^{-\alpha f(\overline{V}_t^i)} - \int_{\mathbb{R}^d} v e^{-\alpha f(v)}d\rho_t(v)\right)\right\|_2^2\right]\\
&\quad + 2\mathbb{E}\left[\left\|\frac{v_\alpha(\rho_t)}{\frac{1}{N}\sum_{i=1}^N e^{-\alpha f(\overline{V}_t^i)}}\left(\frac{1}{N}\sum_{i=1}^N e^{-\alpha f(\overline{V}_t^i)} - \int_{\mathbb{R}^d} e^{-\alpha f(v)}d\rho_t(v)\right)\right\|_2^2\right]\\
&\leq 2\mathbb{E}\left[\left\|e^{\alpha\frac{1}{N}\sum_{i=1}^N f(\overline{V}_t^i)}\left(\frac{1}{N}\sum_{i=1}^N\overline{V}_t^i e^{-\alpha f(\overline{V}_t^i)} - \int_{\mathbb{R}^d} v e^{-\alpha f(v)}d\rho_t(v)\right)\right\|_2^2\right]\\
&\quad + 2\left\|v_\alpha(\rho_t)\right\|_2^2\mathbb{E}\left[\left\|e^{\alpha\frac{1}{N}\sum_{i=1}^N f(\overline{V}_t^i)}\left(\frac{1}{N}\sum_{i=1}^N e^{-\alpha f(\overline{V}_t^i)} - \int_{\mathbb{R}^d} e^{-\alpha f(v)}d\rho_t(v)\right)\right\|_2^2\right]\\
&\leq 2T_1T_2 + 2\left\|v_\alpha(\rho_t)\right\|_2^2 T_1T_3,
\end{aligned} \tag{36}$$

where we defined

$$\begin{aligned}
T_1 &:= \left(\mathbb{E}\left[e^{4\alpha\frac{1}{N}\sum_{i=1}^N f(\overline{V}_t^i)}\right]\right)^{\frac{1}{2}},\\
T_2 &:= \left(\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^N\overline{V}_t^i e^{-\alpha f(\overline{V}_t^i)} - \int_{\mathbb{R}^d} v e^{-\alpha f(v)}d\rho_t(v)\right\|_2^4\right]\right)^{\frac{1}{2}},\\
T_3 &:= \left(\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^N e^{-\alpha f(\overline{V}_t^i)} - \int_{\mathbb{R}^d} e^{-\alpha f(v)}d\rho_t(v)\right\|_2^4\right]\right)^{\frac{1}{2}}.
\end{aligned}$$

In the following, we upper-bound the terms T_1, T_2 and T_3 separately. First, recall that by Lemma 8 we have for $t\in[0,T^*]$ that

$$\mathbb{E}\left[\exp\left(\frac{\sum_{i=1}^N\left\|\overline{V}_t^i\right\|_2^2}{N\kappa^2}\right)\right]\leq C_\kappa<\infty, \tag{37}$$

https://doi.org/10.1017/S095679252400007X Published online by Cambridge University Press

where C_κ only depends on $\kappa, \lambda, \sigma, d, R, M$ and T^*. With this,

$$\begin{aligned}
T_1^2 = \mathbb{E}\left[\exp\left(4\alpha\frac{1}{N}\sum_{i=1}^N f(\overline{V}_t^i)\right)\right] &\le \mathbb{E}\left[\exp\left(4\alpha\frac{1}{N}\sum_{i=1}^N L_\gamma\left(1+\|\overline{V}_t^i\|_2^{1+\gamma}\right)\right)\right]\\
&\le e^{4\alpha L_\gamma}\mathbb{E}\left[\exp\left(4\alpha L_\gamma\frac{1}{N}\sum_{i=1}^N \|\overline{V}_t^i\|_2^{1+\gamma}\right)\right]\\
&\le e^{8\alpha L_\gamma}\mathbb{E}\left[\exp\left(4\alpha L_\gamma\frac{1}{N}\sum_{i=1}^N \|\overline{V}_t^i\|_2^{2}\right)\right]\\
&= e^{8\alpha L_\gamma}\mathbb{E}\left[\exp\left(\frac{1}{\kappa^2}\frac{1}{N}\sum_{i=1}^N \|\overline{V}_t^i\|_2^{2}\right)\right]\\
&\le e^{8\alpha L_\gamma}C_\kappa\big|_{\kappa=\frac{1}{2\sqrt{\alpha L_\gamma}}},
\end{aligned}$$

where we set $\kappa^2 = 1/(4\alpha L_\gamma)$ in the next-to-last step and where N should satisfy $N \ge (16\alpha L_\gamma\sigma^2 M^2)/\lambda$.

Second, we have

$$\begin{aligned}
\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^N \overline{V}_t^i e^{-\alpha f(\overline{V}_t^i)} - \int_{\mathbb{R}^d} v e^{-\alpha f(v)}\,d\rho_t(v)\right\|_2^4\right] &= \frac{1}{N^4}\mathbb{E}\left[\sum_{i_1,i_2,i_3,i_4\in\{1,\ldots,N\}}\left\langle \overline{Z}_t^{i_1},\overline{Z}_t^{i_2}\right\rangle\left\langle \overline{Z}_t^{i_3},\overline{Z}_t^{i_4}\right\rangle\right]\\
&\le \frac{4!L_u^4}{N^2},
\end{aligned}$$

where $\left(\overline{Z}_t^i := \overline{V}_t^i e^{-\alpha f(\overline{V}_t^i)} - \int_{\mathbb{R}^d} v e^{-\alpha f(v)}\,d\rho_t(v)\right)_{i=1,\ldots,N}$ are i.i.d. and have zero mean. Thus,

$$T_2 = \left(\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^N \overline{V}_t^i e^{-\alpha f(\overline{V}_t^i)} - \int_{\mathbb{R}^d} v e^{-\alpha f(v)}\,d\rho_t(v)\right\|_2^4\right]\right)^{\frac{1}{2}} \le \frac{5L_u^2}{N}.$$

Similarly, we can derive

$$T_3 = \left(\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^N e^{-\alpha f(\overline{V}_t^i)} - \int_{\mathbb{R}^d} e^{-\alpha f(v)}\,d\rho_t(v)\right\|_2^4\right]\right)^{\frac{1}{2}} \le \frac{5}{N}.$$

Collecting the bounds for the terms T_1, T_2 and T_3 and inserting them in (36), we obtain

$$\mathbb{E}\left[\|v_\alpha(\rho_t) - v_\alpha(\bar\rho_t)\|_2^2\right] \le 10e^{6\alpha L_\gamma} C_\kappa^{\frac{1}{2}}\big|_{\kappa=\frac{1}{2\sqrt{\alpha L_\gamma}}}\left(L_u^2 + \sup_{t\in[0,T^*]}\|v_\alpha(\rho_t)\|_2^2\right)\frac{1}{N}. \tag{38}$$

Since by Lemmas 14, 16 and 17, we know that $\|v_\alpha(\rho_t)\|_2$ can be uniformly bounded by a constant depending on $\alpha, \lambda, \sigma, d, R, v_b, v^*, M, L_\nu$ and ν (see in particular Equation (48) that combines the aforementioned lemmas), we can conclude (38) with

$$\mathbb{E}\left[\|v_\alpha(\rho_t) - v_\alpha(\bar\rho_t)\|_2^2\right] \le \frac{C_0}{N} \tag{39}$$

for some constant C_0 depends on $\lambda, \sigma, d, \alpha, L_\nu, \nu, L_\gamma, L_u, T^*, R, v_b, v^*$ and M. □

https://doi.org/10.1017/S095679252400007X Published online by Cambridge University Press

3.2.2. Upper bound for the third term in (19)

In this section, we bound Term *III* of the error decomposition (19). Before stating the main result of this section, Proposition 15, we first need to provide two auxiliary lemmas, Lemma 12 and Lemma 14.

Lemma 12. *Let* $R, M \in (0, \infty)$. *Then, it holds*

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\left[\left\|\overline{V}_t - v^*\right\|_2^2\right] \leq &-\lambda\mathbb{E}\left[\left\|\overline{V}_t - v^*\right\|_2^2\right] \\ &+\lambda\left(\left\|\mathcal{P}_{v_b,R}(v_\alpha(\rho_t)) - v^*\right\|_2^2 + \left\|v_\alpha(\rho_t) - \mathcal{P}_{v_b,R}(v_\alpha(\rho_t))\right\|_2^2\right) \\ &+\sigma^2 M^2 d.\end{aligned} \tag{40}$$

If further $\lambda \geq 2\sigma^2 d$, *we have*

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\left[\left\|\overline{V}_t - v^*\right\|_2^2\right] \leq &-\lambda\mathbb{E}\left[\left\|\overline{V}_t - v^*\right\|_2^2\right] \\ &+\lambda\left(\left\|\mathcal{P}_{v_b,R}(v_\alpha(\rho_t)) - v^*\right\|_2^2 + \left\|v_\alpha(\rho_t) - \mathcal{P}_{v_b,R}(v_\alpha(\rho_t))\right\|_2^2\right).\end{aligned} \tag{41}$$

Proof. By Itô's formula, we have

$$\begin{aligned}d\left\|\overline{V}_t - v^*\right\|_2^2 &= 2\left(\overline{V}_t - v^*\right)^\top d\overline{V}_t + \sigma^2 d\left(\left\|\overline{V}_t - v_\alpha(\rho_t)\right\|_2^2 \wedge M^2\right) dt \\ &= -2\lambda\left\langle\overline{V}_t - v^*, \overline{V}_t - \mathcal{P}_{v_b,R}(v_\alpha(\rho_t))\right\rangle dt + 2\sigma\left(\left\|\overline{V}_t - v_\alpha(\rho_t)\right\|_2 \wedge M\right)\left(\overline{V}_t - v^*\right)^\top dB_t \\ &\quad + \sigma^2 d\left(\left\|\overline{V}_t - v_\alpha(\rho_t)\right\|_2^2 \wedge M^2\right) dt \\ &= -\lambda\left[\left\|\overline{V}_t - v^*\right\|_2^2 + \left\|\overline{V}_t - \mathcal{P}_{v_b,R}(v_\alpha(\rho_t))\right\|_2^2 - \left\|\mathcal{P}_{v_b,R}(v_\alpha(\rho_t)) - v^*\right\|_2^2\right] dt \\ &\quad + 2\sigma\left(\left\|\overline{V}_t - v_\alpha(\rho_t)\right\|_2 \wedge M\right)\left(\overline{V}_t - v^*\right)^\top dB_t + \sigma^2 d\left(\left\|\overline{V}_t - v_\alpha(\rho_t)\right\|_2^2 \wedge M^2\right) dt,\end{aligned}$$

which, after taking the expectation on both sides, yields

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\left[\left\|\overline{V}_t - v^*\right\|_2^2\right] = &-\lambda\mathbb{E}\left[\left\|\overline{V}_t - v^*\right\|_2^2\right] \\ &+\lambda\left\|\mathcal{P}_{v_b,R}(v_\alpha(\rho_t)) - v^*\right\|_2^2 - \lambda\mathbb{E}\left[\left\|\overline{V}_t - \mathcal{P}_{v_b,R}(v_\alpha(\rho_t))\right\|_2^2\right] \\ &+\sigma^2 d\mathbb{E}\left[\left\|\overline{V}_t - v_\alpha(\rho_t)\right\|_2^2 \wedge M^2\right].\end{aligned} \tag{42}$$

For the term $\mathbb{E}\left[\left\|\overline{V}_t - \mathcal{P}_{v_b,R}(v_\alpha(\rho_t))\right\|_2^2\right]$, we notice that

$$\begin{aligned}\mathbb{E}\left[\left\|\overline{V}_t - \mathcal{P}_{v_b,R}\left(v_\alpha(\rho_t)\right)\right\|_2^2\right] &= \mathbb{E}\left[\left\|\overline{V}_t - v_\alpha(\rho_t)\right\|_2^2\right] + \mathbb{E}\left[\left\|v_\alpha(\rho_t) - \mathcal{P}_{v_b,R}\left(v_\alpha(\rho_t)\right)\right\|_2^2\right] \\ &\quad + 2\mathbb{E}\left[\left\langle\overline{V}_t - v_\alpha(\rho_t), v_\alpha(\rho_t) - \mathcal{P}_{v_b,R}\left(v_\alpha(\rho_t)\right)\right\rangle\right] \\ &\geq \mathbb{E}\left[\left\|\overline{V}_t - v_\alpha(\rho_t)\right\|_2^2\right] + \mathbb{E}\left[\left\|v_\alpha(\rho_t) - \mathcal{P}_{v_b,R}\left(v_\alpha(\rho_t)\right)\right\|_2^2\right] \\ &\quad - \left(\frac{1}{2}\mathbb{E}\left[\left\|\overline{V}_t - v_\alpha(\rho_t)\right\|_2^2\right] + 2\mathbb{E}\left[\left\|v_\alpha(\rho_t) - \mathcal{P}_{v_b,R}\left(v_\alpha(\rho_t)\right)\right\|_2^2\right]\right) \\ &= \frac{1}{2}\mathbb{E}\left[\left\|\overline{V}_t - v_\alpha(\rho_t)\right\|_2^2\right] - \mathbb{E}\left[\left\|v_\alpha(\rho_t) - \mathcal{P}_{v_b,R}\left(v_\alpha(\rho_t)\right)\right\|_2^2\right],\end{aligned}$$

which, inserted into Equation (42), allows to derive

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\left[\left\|\overline{V}_t - v^*\right\|_2^2\right] \leq &-\lambda\mathbb{E}\left[\left\|\overline{V}_t - v^*\right\|_2^2\right] \\ &+\lambda\left(\left\|\mathcal{P}_{v_b,R}\left(v_\alpha(\rho_t)\right) - v^*\right\|_2^2 + \left\|v_\alpha(\rho_t) - \mathcal{P}_{v_b,R}\left(v_\alpha(\rho_t)\right)\right\|_2^2\right) \\ &-\frac{1}{2}\lambda\mathbb{E}\left[\left\|\overline{V}_t - v_\alpha(\rho_t)\right\|_2^2\right] + \sigma^2 d\left(\left\|\overline{V}_t - v_\alpha(\rho_t)\right\|_2^2 \wedge M^2\right).\end{aligned}$$

https://doi.org/10.1017/S095679252400007X Published online by Cambridge University Press

From this, we get for any λ and σ that

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\Big[\big\|\overline{V}_t - v^*\big\|_2^2\Big] \leq &-\lambda\mathbb{E}\Big[\big\|\overline{V}_t - v^*\big\|_2^2\Big] \\ &+\lambda\left(\big\|\mathcal{P}_{v_b,R}(v_\alpha(\rho_t)) - v^*\big\|_2^2 + \big\|v_\alpha(\rho_t) - \mathcal{P}_{v_b,R}(v_\alpha(\rho_t))\big\|_2^2\right) \\ &+\sigma^2M^2d\end{aligned} \tag{43}$$

as well as

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\Big[\big\|\overline{V}_t - v^*\big\|_2^2\Big] \leq &-\lambda\mathbb{E}\Big[\big\|\overline{V}_t - v^*\big\|_2^2\Big] \\ &+\lambda\left(\big\|\mathcal{P}_{v_b,R}(v_\alpha(\rho_t)) - v^*\big\|_2^2 + \big\|v_\alpha(\rho_t) - \mathcal{P}_{v_b,R}(v_\alpha(\rho_t))\big\|_2^2\right) \\ &+\left(-\frac{1}{2}\lambda + \sigma^2 d\right)\mathbb{E}\Big[\big\|\overline{V}_t - v_\alpha(\rho_t)\big\|_2^2\Big].\end{aligned} \tag{44}$$

If $\lambda \geq 2\sigma^2 d$, by Equation (44), we get

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\Big[\big\|\overline{V}_t - v^*\big\|_2^2\Big] \leq &-\lambda\mathbb{E}\Big[\big\|\overline{V}_t - v^*\big\|_2^2\Big] \\ &+\lambda\left(\big\|\mathcal{P}_{v_b,R}(v_\alpha(\rho_t)) - v^*\big\|_2^2 + \big\|v_\alpha(\rho_t) - \mathcal{P}_{v_b,R}(v_\alpha(\rho_t))\big\|_2^2\right).\end{aligned} \tag{45}$$

□

Remark 13. *When $R = M = \infty$, we can show*

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\Big[\big\|\overline{V}_t - v^*\big\|_2^2\Big] = &-\lambda\mathbb{E}\Big[\big\|\overline{V}_t - v^*\big\|_2^2\Big] \\ &+\lambda\,\|v_\alpha(\rho_t) - v^*\|_2^2 - \left(\lambda - \sigma^2 d\right)\mathbb{E}\Big[\big\|\overline{V}_t - v_\alpha(\rho_t)\big\|_2^2\Big].\end{aligned}$$

If further $\lambda \geq \sigma^2 d$, we have

$$\frac{d}{dt}\mathbb{E}\Big[\big\|\overline{V}_t - v^*\big\|_2^2\Big] \leq -\lambda\mathbb{E}\Big[\big\|\overline{V}_t - v^*\big\|_2^2\Big] + \lambda\,\|v_\alpha(\rho_t) - v^*\|_2^2.$$

This differs from [10, Lemma 18].

The next result is a quantitative version of the Laplace principle as established in [10, Proposition 21].

Lemma 14. *For any $r > 0$, define $f_r := \sup_{v\in B_r(v^*)} f(v)$. Then, under the inverse continuity condition A2, for any $r \in (0, R_0]$ and $q > 0$ such that $q + f_r \leq f_\infty$, it holds*

$$\|v_\alpha(\rho) - v^*\|_2 \leq \frac{(q+f_r)^\nu}{L_\nu} + \frac{\exp(-\alpha q)}{\rho(B_r(v^*))}\int \|v - v^*\|_2\, d\rho(v) \tag{46}$$

With the above preparation, we can now upper bound Term *III*. We have by Jensen's inequality

$$III = \mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^N \overline{V}^i_{T^*} - v^*\right\|_2^2\right] \leq \frac{1}{N}\sum_{i=1}^N \mathbb{E}\Big[\big\|\overline{V}^i_{T^*} - v^*\big\|_2^2\Big], \tag{47}$$

that is, it is enough to upper-bound $\mathbb{E}\Big[\big\|\overline{V}_{T^*} - v^*\big\|_2^2\Big]$, which is the content of the next statement.

Proposition 15. *Let $f \in \mathcal{C}(\mathbb{R}^d)$ satisfy A1, A2 and A3. Moreover, let $\rho_0 \in \mathcal{P}_4(\mathbb{R}^d)$ with $v^* \in \operatorname{supp}(\rho_0)$. Fix any $\epsilon \in (0, W_2^2(\rho_0, \delta_{v^*}))$ and define the time horizon*

$$T^* := \frac{1}{\lambda}\log\left(\frac{2W_2^2(\rho_0, \delta_{v^*})}{\epsilon}\right).$$

https://doi.org/10.1017/S095679252400007X Published online by Cambridge University Press

Moreover, let $R \in (\|v_b - v^*\|_2 + \sqrt{\epsilon/2}, \infty)$, $M \in (0, \infty)$ *and* $\lambda, \sigma > 0$ *be such that* $\lambda \geq 2\sigma^2 d$ *or* $\sigma^2 M^2 d = \mathcal{O}(\epsilon)$. *Then, we can choose* α *sufficiently large, depending on* $\lambda, \sigma, d, T^*, R, v_b, M, \epsilon$ *and properties of* f, *such that* $\mathbb{E}\left[\left\|\overline{V}_{T^*} - v^*\right\|_2^2\right] = \mathcal{O}(\epsilon)$.

Proof. We only prove the case $\lambda \geq 2\sigma^2 d$ in detail. The case $\sigma^2 M^2 d = \mathcal{O}(\epsilon)$ follows similarly.

According to Lemmas 14 and 17, we have

$$\begin{aligned} \|v_\alpha(\rho_t) - v^*\|_2 &\leq \frac{(q + f_r)^\nu}{L_\nu} + \frac{\exp(-\alpha q)}{\rho_t\left(B_r\left(v^*\right)\right)} \mathbb{E}\left[\left\|\overline{V}_t - v^*\right\|_2\right] \\ &\leq \frac{(q + f_r)^\nu}{L_\nu} + \exp(-\alpha q) C_2 C_3, \end{aligned} \tag{48}$$

where $C_2 := (\exp q' T^*)/C_4 < \infty$, q' and C_4 are from Lemma 17, and where, as of Lemma 16, $C_3 := \sup_{[0,T^*]} \mathbb{E}\left[\left\|\overline{V}_t - v^*\right\|_2\right] < \infty$. In what follows, let us deal with the two terms on the right-hand side of (48). For the term $(q + f_r)^\nu / L_\nu$, let $q = f_r$. Then by A2 and A3, we can choose proper r, such that $2(L_\nu r)^{1/\nu} \leq 2f_r \leq f_\infty$. Further by A3, we have

$$\frac{(q + f_r)^\nu}{L_\nu} = \frac{(2f_r)^\nu}{L_\nu} \leq \frac{(2L_\gamma)^\nu r^{(1+\gamma)\nu}}{L_\nu},$$

so if

$$r < r_0 := \min\left\{\left(\frac{\epsilon}{8}\right)^{\frac{1}{2(1+\gamma)\nu}} \left(\frac{L_\nu}{(2L_\gamma)^\nu}\right)^{\frac{1}{(1+\gamma)\nu}}, \sqrt{\frac{\epsilon}{2}}\right\},$$

we can bound

$$\frac{(q + f_r)^\nu}{L_\nu} = \frac{(2f_r)^\nu}{L_\nu} \leq \frac{\sqrt{\epsilon}}{2\sqrt{2}}.$$

For term $\exp(-\alpha q) C_2 C_3$, we can choose α large enough such that

$$\exp(-\alpha q) C_2 C_3 \leq \frac{\sqrt{\epsilon}}{2\sqrt{2}}.$$

With these choices of r and α and by integrating them into Equation (48), we obtain

$$\|v_\alpha(\rho_t) - v^*\|_2^2 < \frac{\epsilon}{2},$$

for all $t \in [0, T^*]$, and thus

$$\|v_\alpha(\rho_t) - v_b\|_2 \leq \|v_\alpha(\rho_t) - v^*\|_2 + \|v^* - v_b\|_2 \leq \sqrt{\frac{\epsilon}{2}} + \|v^* - v_b\|_2 \leq R.$$

Consequently, by Lemma 12, we have

$$\begin{aligned} \frac{d}{dt} \mathbb{E}\left[\left\|\overline{V}_t - v^*\right\|_2^2\right] &\leq -\lambda \left(\mathbb{E}\left[\left\|\overline{V}_t - v^*\right\|_2^2\right] - \|v_\alpha(\rho_t) - v^*\|_2^2\right) \\ &\leq -\lambda \left(\mathbb{E}\left[\left\|\overline{V}_t - v^*\right\|_2^2\right] - \frac{\epsilon}{2}\right), \end{aligned}$$

since now $\mathcal{P}_{v_b,R}(v_\alpha(\rho_t)) = v_\alpha(\rho_t)$. Finally by Grönwall's inequality, $\mathbb{E}\left[\left\|\overline{V}_{T^*} - v^*\right\|_2^2\right] \leq \epsilon$. □

Lemma 16. *Let* $\|v_b - v^*\|_2 < R < \infty$ *and* $0 < M < \infty$. *Then, it holds*

$$\sup_{t \in [0,T^*]} \mathbb{E}\left[\left\|\overline{V}_t - v^*\right\|_2\right] \leq \sqrt{\max\left\{\mathbb{E}\left[\left\|\overline{V}_0 - v^*\right\|_2^2\right], \lambda R^2 + \sigma^2 M^2 d\right\}}. \tag{49}$$

https://doi.org/10.1017/S095679252400007X Published online by Cambridge University Press

Proof. By Equation (42), we have

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\left[\left\|\overline{V}_t - v^*\right\|_2^2\right] &\le -\lambda\mathbb{E}\left[\left\|\overline{V}_t - v^*\right\|_2^2\right] \\ &\quad + \lambda\left\|\mathcal{P}_{v_b,R}\left(v_\alpha(\rho_t)\right) - v^*\right\|_2^2 - \lambda\mathbb{E}\left[\left\|\overline{V}_t - \mathcal{P}_{v_b,R}\left(v_\alpha(\rho_t)\right)\right\|_2^2\right] \\ &\quad + \sigma^2 d\mathbb{E}\left[\left\|\overline{V}_t - v_\alpha(\rho_t)\right\|_2^2 \wedge M^2\right] \\ &\le -\lambda\mathbb{E}\left[\left\|\overline{V}_t - v^*\right\|_2^2\right] + \lambda R^2 + \sigma^2 M^2 d,\end{aligned}$$

yielding

$$\mathbb{E}\left[\left\|\overline{V}_t - v^*\right\|_2^2\right] \le \max\left\{\mathbb{E}\left[\left\|\overline{V}_0 - v^*\right\|_2^2\right], \lambda R^2 + \sigma^2 M^2 d\right\},$$

after an application of Grönwall's inequality for any $t \ge 0$. □

Lemma 17. *For any $M \in (0, \infty)$, $\tau \ge 1$, $r > 0$ and $R \in (\|v_b - v^*\|_2 + r, \infty)$, it holds*

$$\rho_t\left(B_r\left(v^*\right)\right) \ge C_4 \exp(-q' t) > 0,$$

where

$$C_4 := \int_{B_r(v^*)} 1 + (\tau - 1)\left\|\frac{v - v^*}{r}\right\|_2^\tau - \tau\left\|\frac{v - v^*}{r}\right\|_2^{\tau-1} d\rho_0(v)$$

and where q' depends on $\tau, \lambda, \sigma, d, r, R, v_b$ and M.

Proof. Recall that the law ρ_t of $\overline{V}_t$ satisfies the Fokker–Planck equation

$$\partial_t \rho_t = \lambda \operatorname{div}\left(\left(v - \mathcal{P}_{v_b,R}\left(v_\alpha(\rho_t)\right)\right)\rho_t\right) + \frac{\sigma^2}{2}\Delta\left(\left(\|v - v_\alpha\left(\rho_t\right)\|^2 \wedge M^2\right)\rho_t\right).$$

Let us first define for $\tau \ge 1$ the test function

$$\phi_r^\tau(v) := \begin{cases} 1 + (\tau - 1)\left\|\frac{v}{r}\right\|_2^\tau - \tau\left\|\frac{v}{r}\right\|_2^{\tau-1}, & \|v\|_2 \le r, \\ 0, & \text{else}, \end{cases} \tag{50}$$

for which it is easy to verify that $\phi_r^\tau \in \mathcal{C}_c^1(\mathbb{R}^d, [0, 1])$. Since $\operatorname{Im}\phi_r^\tau \subset [0, 1]$, we have $\rho_t(B_r(v^*)) \ge \int_{B_r(v^*)} \phi_r^\tau(v - v^*)\, d\rho_t(v)$. To lower bound $\rho_t(B_r(v^*))$, it is thus sufficient to establish a lower bound on $\int_{B_r(v^*)} \phi_r^\tau(v - v^*)\, d\rho_t(v)$. By Green's formula,

$$\begin{aligned}\frac{d}{dt}\int_{B_r(v^*)} \phi_r^\tau(v - v^*)\, d\rho_t(v) &= -\lambda\int_{B_r(v^*)} \left\langle v - \mathcal{P}_{v_b,R}\left(v_\alpha(\rho_t)\right), \nabla\phi_r^\tau(v - v^*)\right\rangle d\rho_t(v) \\ &\quad + \frac{\sigma^2}{2}\int_{B_r(v^*)} \left(\|v - v_\alpha(\rho_t)\|_2^2 \wedge M^2\right)\Delta\phi_r^\tau(v - v^*) d\rho_t(v) \\ &= \tau(\tau - 1)\int_{B_r(v^*)} \frac{\|v - v^*\|_2^{\tau-3}}{r^{\tau-3}}\left(\left(1 - \frac{\|v - v^*\|_2}{r}\right)\left(\lambda\left\langle\frac{v - \mathcal{P}_{v_b,R}\left(v_\alpha(\rho_t)\right)}{r}, \frac{v - v^*}{r}\right\rangle\right.\right. \\ &\quad \left.\left. - \frac{\sigma^2}{2}(d + \tau - 2)\frac{\|v - v_\alpha(\rho_t)\|_2^2 \wedge M^2}{r^2}\right) + \frac{\sigma^2}{2}\frac{\|v - v_\alpha(\rho_t)\|_2^2 \wedge M^2}{r^2}\right) d\rho_t(v).\end{aligned}$$

For simplicity, let us abbreviate

$$\begin{aligned}\Theta := &\left(1 - \frac{\|v - v^*\|_2}{r}\right)\left(\lambda\left\langle\frac{v - \mathcal{P}_{v_b,R}\left(v_\alpha(\rho_t)\right)}{r}, \frac{v - v^*}{r}\right\rangle\right. \\ &\left. - \frac{\sigma^2}{2}(d + \tau - 2)\frac{\|v - v_\alpha(\rho_t)\|_2^2 \wedge M^2}{r^2}\right) + \frac{\sigma^2}{2}\frac{\|v - v_\alpha(\rho_t)\|_2^2 \wedge M^2}{r^2}.\end{aligned}$$

https://doi.org/10.1017/S095679252400007X Published online by Cambridge University Press

We can choose ϵ_1 small enough, depending on τ and d, such that when $\|v-v^*\|_2/r > 1-\epsilon_1$, we have

$$\begin{aligned}\Theta &= \left(1-\frac{\|v-v^*\|_2}{r}\right)\lambda\left\langle\frac{v-\mathcal{P}_{v_b,R}\left(v_\alpha(\rho_t)\right)}{r},\frac{v-v^*}{r}\right\rangle\\ &\quad+\left(\frac{\sigma^2}{2}-\left(1-\frac{\|v-v^*\|_2}{r}\right)\frac{\sigma^2}{2}(d+\tau-2)\right)\frac{\|v-v_\alpha(\rho_t)\|_2^2\wedge M^2}{r^2}\\ &\geq \left(1-\frac{\|v-v^*\|_2}{r}\right)\lambda\left\langle\frac{v-\mathcal{P}_{v_b,R}\left(v_\alpha(\rho_t)\right)}{r},\frac{v-v^*}{r}\right\rangle+\frac{\sigma^2}{3}\frac{\|v-v_\alpha(\rho_t)\|_2^2\wedge M^2}{r^2},\end{aligned}$$

where the last inequality works if $\|v-v^*\|_2/r \geq 1-1/(6(d+\tau-2))$.

If $v_\alpha(\rho_t)\notin B_R(v_b)$, we have $\left|\left\langle v-\mathcal{P}_{v_b,R}(v_\alpha(\rho_t))\,,v-v^*\right\rangle\right|/r^2 \leq C(r,R,v_b)$ and, since $R > \|v_b-v^*\|_2 + r$, $(\|v-v_\alpha(\rho_t)\|_2^2\wedge M^2)/r^2 \geq C(r,M,R,v_b)$, which allows to choose ϵ_2 small enough, depending on $\lambda, r, \sigma, R, v_b$ and M, such that $\Theta > 0$ when $\|v-v^*\|_2/r > 1-\min\{\epsilon_1,\epsilon_2\}$.

If $v_\alpha(\rho_t)\in B_R(v_b)$ and $\|v-v_\alpha(\rho_t)\|_2 \leq M$, we have by Lemma 18

$$\begin{aligned}\Theta &\geq \left(1-\frac{\|v-v^*\|_2}{r}\right)\lambda\left\langle\frac{v-v_\alpha(\rho_t)}{r},\frac{v-v^*}{r}\right\rangle+\frac{\sigma^2}{3}\frac{\|v-v_\alpha(\rho_t)\|_2^2}{r^2}\\ &= \left(\frac{\sigma^2}{3}+\left(1-\frac{\|v-v^*\|_2}{r}\right)\lambda\right)\frac{\|v-v^*\|_2^2}{r^2}+\frac{\sigma^2}{3}\frac{\|v_\alpha(\rho_t)-v^*\|_2^2}{r^2}\\ &\quad-\left(\frac{2\sigma^2}{3}+\left(1-\frac{\|v-v^*\|_2}{r}\right)\lambda\right)\left\langle\frac{v_\alpha(\rho_t)-v^*}{r},\frac{v-v^*}{r}\right\rangle\\ &\geq 0,\end{aligned}$$

when $\|v-v^*\|_2\,/r\in\left[1-2\sigma^2/(3\lambda),1\right]$.

If $v_\alpha(\rho_t)\in B_R(v_b)$ and $\|v-v_\alpha(\rho_t)\|_2 > M$, we have

$$\Theta \geq \left(1-\frac{\|v-v^*\|_2}{r}\right)C(\lambda,r,R,v_b)+\frac{\sigma^2}{3}M^2,$$

that is, we can choose ϵ_3 small enough, depending on $\lambda, r, \sigma, R, v_b$ and M, such that $\Theta \geq 0$ when $\|v-v^*\|_2/r > 1-\min\{\epsilon_1,\epsilon_2,\epsilon_3,2\sigma^2/3\lambda\}$.

Combining the cases from above, we conclude that $\Theta\geq 0$ when $\|v-v^*\|_2/r \geq 1-\min\{\epsilon_1,\epsilon_2,\epsilon_3,2\sigma^2/3\lambda\}$. On the other hand, when $\|v-v^*\|_2/r \leq 1-\min\{\epsilon_1,\epsilon_2,\epsilon_3,2\sigma^2/3\lambda\}$, we have

$$\tau(\tau-1)\frac{\|v-v^*\|_2^{\tau-3}}{r^{\tau-3}}\Theta = \tau(\tau-1)\frac{\|v-v^*\|_2^{\tau-3}}{r^{\tau-3}}\frac{\Theta}{\phi_r^\tau(v)}\phi_r^\tau(v-v^*) \geq -C_5\phi_r^\tau(v-v^*)$$

for some constant C_5 depending on $r, R, M, v_b, \lambda, \sigma, d$ and τ, since $|\Theta|$ is upper-bounded and $\phi_r^\tau(v-v^*) \geq \phi_r^\tau((1-\min\{\epsilon_1,\epsilon_2,\epsilon_3,2\sigma^2/3\lambda\})r) > 0$ for any v satisfies $\|v-v^*\|_2/r \leq 1-\min\{\epsilon_1,\epsilon_2,\epsilon_3,2\sigma^2/3\lambda\}$.

All in all we have

$$\frac{d}{dt}\int_{B_r(v^*)}\phi_r^\tau(v-v^*)\,d\rho_t(v) \geq -q'\int_{B_r(v^*)}\phi_r^\tau(v-v^*)\,d\rho_t(v),$$

where $q' := \max\{C_5, 0\}$. By Grönwall's inequality, we thus have

$$\rho_t(B_r(v^*)) \geq \int_{B_r(v^*)}\phi_r^\tau(v-v^*)\,d\rho_t(v) \geq e^{-q't}\int_{B_r(v^*)}\phi_r^\tau(v-v^*)\,d\rho_0(v),$$

which concludes the proof. □

Lemma 18. *Let $a, b > 0$. Then, we have*

$$(a+b(1-x))x^2+ay^2-(2a+b(1-x))xy \geq 0,$$

for any $x\in[1-2a/b,1]\cap(0,\infty)$ and $y\geq 0$.

https://doi.org/10.1017/S095679252400007X Published online by Cambridge University Press

Proof. For $y = 0$, this is true. For $y > 0$, divide both side by ay^2 and denote $c = b/a$. Then the lemma is equivalent to showing $(1 + c(1 - x))(x/y)^2 - (2 + c(1 - x))x/y + 1 \geq 0$, that is, it is enough to show $\min_{r\geq 0}(1 + c(1 - x))r^2 - (2 + c(1 - x))r + 1 \geq 0$, when $x \in [1 - 2/c, 1]$. We have

$$\arg\min_r (1 + c(1 - x))r^2 - (2 + c(1 - x))r + 1 = \frac{2 + c(1 - x)}{2 + 2c(1 - x)},$$

and thus

$$\begin{aligned} &\min_{r\geq 0}(1 + c(1 - x))r^2 - (2 + c(1 - x))r + 1 \\ &= (1 + c(1 - x))\left(\frac{2 + c(1 - x)}{2 + 2c(1 - x)}\right)^2 - (2 + c(1 - x))\frac{2 + c(1 - x)}{2 + 2c(1 - x)} + 1 \\ &= -\frac{1}{2}\frac{(2 + c(1 - x))^2}{2 + 2c(1 - x)} + 1 \geq 0, \end{aligned}$$

when $x \in [1 - 2/c, 1]$. This finishes the proof. □

4. Numerical experiments

In this section, we numerically demonstrate the benefit of using CBO with truncated noise. For isotropic [7, 8, 10] and anisotropic noise [9, 11], we compare the CBO method with truncation $M = 1$ to standard CBO for several benchmark problems in optimisation, which are summarised in Table 1.

Table 1. *Benchmark test functions*

Name	Objective function f	v^*	$\underline{f}$
Ackley	$-20\exp\left(-0.2\sqrt{\frac{1}{d}\sum_{i=1}^d v_i^2}\right) - \exp\left(\frac{1}{d}\sum_{i=1}^d \cos(2\pi v_i)\right) + 20 + e$	$(0, \ldots, 0)$	0
Griewank	$1 + \sum_{i=1}^d \frac{v_i^2}{4000} - \prod_{i=1}^d \cos\left(\frac{v_i}{i}\right)$	$(0, \ldots, 0)$	0
Rastrigin	$10d + \sum_{i=1}^d \left[v_i^2 - 10\cos(2\pi v_i)\right]$	$(0, \ldots, 0)$	0
Alpine	$10\sum_{i=1}^d \left\| \left(v_i - v_i^*\right)\sin\left(10\left(v_i - v_i^*\right)\right) - 0.1\left(v_i - v_i^*\right)\right\|_2$	$(0, \ldots, 0)$	0
Salomon	$1 - \cos\left(200\pi\sqrt{\sum_{i=1}^d v_i^2}\right) + 10\sqrt{\sum_{i=1}^d v_i^2}$	$(0, \ldots, 0)$	0

In the subsequent tables, we report comparison results for the two methods for the different benchmark functions as well as different numbers of particles N and, potentially, different numbers of steps K. Throughout, we set $v_b = 0$ and $R = \infty$, which is out of convenience. Any sufficiently large but finite choice for R yields identical results.

The success criterion is defined by achieving the condition $\left\|\frac{1}{N}\sum_{i=1}^N V_{K,\Delta t}^i - v^*\right\|_2 \leq 0.1$, which ensures that the algorithm has reached the basin of attraction of the global minimiser. The success rate is averaged over 1000 runs.

4.1. Isotropic case

Let $d = 15$. In the case of isotropic noise, we always set $\lambda = 1$, $\sigma = 0.3$, $\alpha = 10^5$ and step size $\Delta t = 0.02$. The initial positions $(V_0^i)_{i=1,\ldots,N}$ are sampled i.i.d. from $\rho_0 = \mathcal{N}(0, I_d)$. In Table 2, we report results comparing the isotropic CBO method with truncation $M = 1$ and the original isotropic CBO method [7,

https://doi.org/10.1017/S095679252400007X Published online by Cambridge University Press

Table 2. *For the* 15*-dimensional Ackley and Salomon function, the CBO method with truncation* ($M = 1$) *is able to locate the global minimum using only* $N = 300$ *particles. In comparison, even with an larger number of particles (up to* $N = 1200$*), the original CBO method* ($M = +\infty$) *cannot achieve a flawless success rate. In the case of the Griewank function, the original CBO method* ($M = +\infty$) *exhibits a quite low success rate, even when utilising* $N = 1200$ *particles. Contrarily, in the same setting, the CBO method with truncation* ($M = 1$) *achieves a success rate of* 0.791.

Number of steps $K = 200$						
Test function	M	$N = 150$	$N = 300$	$N = 600$	$N = 900$	$N = 1200$
Ackley	1	0.978	0.999	1	1	1
	$+\infty$	0.001	0.056	0.478	0.824	0.935
Griewank	1	0.060	0.188	0.5013	0.671	0.791
	$+\infty$	0	0	0.010	0.013	0.032
Salomon	1	0.970	1	1	1	1
	$+\infty$	0.005	0.068	0.603	0.909	0.979

Table 3. *For the* 15*-dimensional Rastrigin and Alpine function, both algorithms have difficulties in finding the global minimiser. However, the success rates for the CBO method with truncation* ($M = 1$) *are significantly higher compared to those of the original CBO method* ($M = +\infty$.*)*

Number of steps $K = 200$						
Test function	M	$N = 300$	$N = 600$	$N = 900$	$N = 1200$	$N = 1500$
Rastrigin	1	0.180	0.256	0.298	0.322	0.337
	$+\infty$	0	0	0.004	0.004	0.007
Alpine	1	0.029	0.049	0.051	0.070	0.080
	$+\infty$	0	0.001	0.004	0.004	0.004
Number of steps $K = 500$						
Test function	M	$N = 300$	$N = 600$	$N = 900$	$N = 1200$	$N = 1500$
Rastrigin	1	0.213	0.265	0.316	0.326	0.343
	$+\infty$	0.001	0.004	0.005	0.009	0.010
Alpine	1	0.103	0.115	0.147	0.165	0.173
	$+\infty$	0.010	0.015	0.033	0.037	0.040

8, 10] ($M = +\infty$) for the Ackley, Griewank and Salomon function. Each algorithm is run for $K = 200$ steps.

Since the benchmark functions Rastrigin and Alpine are more challenging, we use more particles N and a larger number of steps K, namely $K = 200$ and $K = 500$. We report the results in Table 3.

4.2. Anisotropic case

Let $d = 20$. In the case of anisotropic noise, we set $\lambda = 1, \sigma = 5, \alpha = 10^5$ and step size $\Delta t = 0.02$. The initial positions of the particles are initialised with $\rho_0 = \mathcal{N}(0, 100 I_d)$. In Table 4, we report results comparing the anisotropic CBO method with truncation $M = 1$ and the original anisotropic CBO method

https://doi.org/10.1017/S095679252400007X Published online by Cambridge University Press

Table 4. *For the 20-dimensional Rastrigin, Ackley and Salomon function, the original anisotropic CBO method ($M = +\infty$) works better than the anisotropic CBO method with truncation ($M = 1$), in particular when the particle number N is small. In the case of the Salomon function, when increasing the number of particle to $N = 900$, the success rates of the original anisotropic CBO method ($M = +\infty$) decreases. In the case of the Griewank function, however, we find that the anisotropic CBO method with truncation ($M = +\infty$) works considerably better than the original anisotropic CBO method ($M = 1$.)*

Number of steps $K = 1000$						
Test function	M	$N = 75$	$N = 150$	$N = 300$	$N = 600$	$N = 900$
Rastrigin	1	0.285	0.928	0.990	1	1
	$+\infty$	0.728	0.952	0.993	1	1
Ackley	1	0.510	0.997	1	1	1
	$+\infty$	0.997	1	1	1	1
Griewank	1	0.097	0.458	0.576	0.625	0.665
	$+\infty$	0.093	0.101	0.157	0.159	0.167
Salomon	1	0.010	0.434	0.925	0.998	1
	$+\infty$	0.622	0.954	0.970	0.934	0.891

Table 5. *For the 15-dimensional Alpine function, the anisotropic CBO method with truncated noise ($M = 1$) works better than the original anisotropic CBO method ($M = +\infty$.)*

Number of steps $K = 200$						
Test function	M	$N = 300$	$N = 600$	$N = 900$	$N = 1200$	$N = 1500$
Alpine	1	0	0.006	0.006	0.008	0.025
	$+\infty$	0.001	0.004	0.008	0.007	0.021
Number of steps $K = 500$						
Test function	M	$N = 300$	$N = 600$	$N = 900$	$N = 1200$	$N = 1500$
Alpine	1	0.130	0.224	0.291	0.336	0.365
	$+\infty$	0.083	0.175	0.250	0.292	0.330
Number of steps $K = 1000$						
Test function	M	$N = 300$	$N = 600$	$N = 900$	$N = 1200$	$N = 1500$
Alpine	1	0.102	0.198	0.293	0.340	0.368
	$+\infty$	0.097	0.179	0.250	0.295	0.331

[9, 11] ($M = +\infty$) for the Rastrigin, Ackley, Griewank and Salomon function. Each algorithm is run for $K = 200$ steps.

Since the benchmark function Alpine is more challenging and none of the algorithms work in the previous setting, we reduce the dimensionality to $d = 15$, choose $\sigma = 1$, use $\rho_0 = \mathcal{N}(0, I_d)$ to initialise, employ more particles and use a larger number of steps K, namely $K = 200$, $K = 500$ and $K = 1000$. We report the results in Table 5.

https://doi.org/10.1017/S095679252400007X Published online by Cambridge University Press

5. Conclusions

In this paper, we establish the convergence to a global minimiser of a potentially non-convex and non-smooth objective function for a variant of CBO which incorporates truncated noise. We observe that truncating the noise in CBO enhances the well-behavedness of the statistics of the law of the dynamics, which enables enhanced convergence performance and allows in particular for a wider flexibility in choosing the noise parameter of the method, as we observe numerically. For rigorously proving the convergence of the implementable algorithm to the global minimiser of the objective, we follow the route devised in [10].

Acknowledgements. MF acknowledges the support of the Munich Center for Machine Learning. PR acknowledges the support of the Extreme Computing Research Center at KAUST. KR acknowledges the support of the Munich Center for Machine Learning and the financial support from the Technical University of Munich – Institute for Ethics in Artificial Intelligence (IEAI). LS acknowledges the support of KAUST Optimization and Machine Learning Lab. LS also thanks the hospitality of the Chair of Applied Numerical Analysis of the Technical University of Munich for discussions that contributed to the finalisation of this work.

Funding statement. This work has been funded by the KAUST Baseline Research Scheme and the German Federal Ministry of Education and Research, and the Bavarian State Ministry for Science and the Arts.

Competing interests. None.

References

[1] Bäck, T., Fogel, D. B. & Michalewicz, Z. (1997). *Handbook of evolutionary computation*, Institute of Physics Publishing, Oxford University Press, Bristol, New York
[2] Blum, C. & Roli, A. (2003) Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Comput. Surv* **35**(3), 268–308.
[3] Fogel, D. B. (2006) *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*, John Wiley & Sons, Ltd, Piscataway, NJ.
[4] Holland, J. H. (1975) *Adaptation in Natural and Artificial Systems. An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, University of Michigan Press, Ann Arbor, Michigan, US.
[5] Kennedy, J. & Eberhart, R. (1995). Particle swarm optimization. (*Proceedings of ICNN'95 - International Conference on Neural Networks*, IEEE, **4**, pages 1942–1948.
[6] Aarts, E. & Korst, J. (1989). Simulated annealing and boltzmann machines. A stochastic approach to combinatorial optimization and neural computing. In: *Wiley-Interscience Series in Discrete Mathematics and Optimization*, John Wiley & Sons, Ltd, Chichester
[7] Pinnau, R., Totzeck, C., Tse, O. & Martin, S. (2017) A consensus-based model for global optimization and its mean-field limit. *Math. Models Methods Appl. Sci* **27**(1), 183–204.
[8] Carrillo, J. A., Choi, Y.-P., Totzeck, C. & Tse, O. (2018) An analytical framework for consensus-based global optimization method. *Math. Models Methods Appl. Sci* **28**(6), 1037–1066.
[9] Carrillo, J. A., Jin, S., Li, L. & Zhu, Y. (2021) A consensus-based global optimization method for high dimensional machine learning problems. *ESAIM Control Optim. Calc. Var* **27**(S5), 22.
[10] Fornasier, M., Klock, T. & Riedl, K. (2021) Consensus-based optimization methods converge globally. arXiv:2103.15130, 2021.
[11] Fornasier, M., Klock, T. & Riedl, K. (2022) Convergence of anisotropic consensus-based optimization in mean-field law. In: Laredo, J. L. Jiménez, Hidalgo, J. I. & Babaagba, K. O. (editors), *Applications of Evolutionary Computation*, Springer International Publishing, Cham, pp. 738–754.
[12] Ha, S.-Y., Jin, S. & Kim, D. (2020) Convergence of a first-order consensus-based global optimization algorithm. *Math. Models Methods Appl. Sci* **30**(12), 2417–2444.
[13] Ha, S.-Y., Jin, S. & Kim, D. (2021) Convergence and error estimates for time-discrete consensus-based optimization algorithms. *Numer. Math* **147**(2), 255–282.
[14] Ko, D., Ha, S.-Y., Jin, S. & Kim, D. (2022) Convergence analysis of the discrete consensus-based optimization algorithm with random batch interactions and heterogeneous noises. *Math. Models Meth. Appl. Sci.* **32**(06), 1071–1107.
[15] Carrillo, J. A., Trillos, N. G., Li, S. & Zhu, Y. (2023) FedCBO: Reaching group consensus in clustered federated learning through consensus-based optimization. arXiv:2305.02894.
[16] Fornasier, M., Huang, H., Pareschi, L. & Sünnen, P. (2021) Consensus-based optimization on the sphere: Convergence to global minimizers and machine learning. *J. Mach. Learn. Res* **22**, 55.
[17] Riedl, K. (2023) Leveraging memory effects and gradient information in consensus-based optimisation: On global convergence in mean-field law. *European J. Appl. Math.*, 1–32.

[18] Cipriani, C., Huang, H. & Qiu, J. (2022) Zero-inertia limit: From particle swarm optimization to consensus-based optimization. *SIAM J. Appl. Math* **54**(3), 3091–3121.
[19] Grassi, S. & Pareschi, L. (2021) From particle swarm optimization to consensus based optimization: Stochastic modeling and mean-field limit. *Math. Models Methods Appl. Sci* **31**(8), 1625–1657.
[20] Huang, H., Qiu, J. & Riedl, K. (2023) On the global convergence of particle swarm optimization methods. *Appl Math Optim* **88**(2), 30.
[21] Riedl, K., Klock, T., Geldhauser, C. & Fornasier, M. (2023) Gradient is all you need? arXiv:2306.09778, 2023.
[22] Grassi, S., Huang, H., Pareschi, L. & Qiu, J. (2023) Mean-field particle swarm optimization. In:*Modeling and Simulation for Collective Dynamics*, World Scientific, Singapore, pp. 127–193.
[23] Totzeck, C. (2021) Trends in consensus-based optimization. In: *Active Particles*, Vol. 3, Springer, 201–226. Advances in Theory, Models, and Applications
[24] Bae, H.-O., Ha, S.-Y., Kang, M., Lim, H., Min, C. & Yoo, J. (2022) A constrained consensus based optimization algorithm and its application to finance. *Appl. Math. Comput* **416**, 126726.
[25] Borghi, G., Herty, M. & Pareschi, L. (2023) Constrained consensus-based optimization. *Siam J Optimiz* **33**(1), 211–236.
[26] Carrillo, J. A., Totzeck, C. & Vaes, U. (2023) Consensus-based optimization and ensemble kalman inversion for global optimization problems with constraints. In: *Modeling and Simulation for Collective Dynamics*, World Scientific, Singapore, pp. 195–230.
[27] Fornasier, M., Huang, H., Pareschi, L. & Sünnen, P. (2020) Consensus-based optimization on hypersurfaces: Well-posedness and mean-field limit. *Math. Models Methods Appl. Sci* **30**(14), 2725–2751.
[28] Fornasier, M., Huang, H., Pareschi, L. & Sünnen, P. (2022) Anisotropic diffusion in consensus-based optimization on the sphere. *SIAM J. Optim* **32**(3), 1984–2012.
[29] Ha, S.-Y., Kang, M. & Kim, D. (2022) Emergent behaviors of high-dimensional Kuramoto models on stiefel manifolds. *Automatica* **136**, 110072.
[30] Kim, J., Kang, M., Kim, D., Ha, S.-Y. & Yang, I. (2020) A stochastic consensus method for nonconvex optimization on the Stiefel manifold. In:*2020 59th IEEE Conference on Decision and Control (CDC)*, IEEE, 1050–1057.
[31] Borghi, G., Herty, M. & Pareschi, L. (2022). A consensus-based algorithm for multi-objective optimization and its mean-field description. In: *2022 IEEE 61st Conference on Decision and Control (CDC)*, IEEE, pp. 4131–4136.
[32] Borghi, G., Herty, M. & Pareschi, L. (2023) An adaptive consensus based method for multi-objective optimization with uniform pareto front approximation. *Appl Math Optim* **88**(2), 1–43.
[33] Klamroth, K., Stiglmayr, M. & Totzeck, C. (2024) Consensus-based optimization for multi-objective problems: A multi-swarm approach. *J. Glob. Optim.*, 1–32. arXiv:2211.15737.
[34] Huang, H., Qiu, J. & Riedl, K. (2022) Consensus-based optimization for saddle point problems. arXiv:2212.12334, 2022.
[35] Carrillo, J. A., Hoffmann, F., Stuart, A. M. & Vaes, U. (2022) Consensus-based sampling. *Stud. Appl. Math* **148**(3), 1069–1140.
[36] Borghi, G., Grassi, S. & Pareschi, L. (2023) Consensus based optimization with memory effects: Random selection and applications. *Cha. Solit. Fract.*, **174**, 113859. arXiv preprint arXiv:2301.13242, 2023.
[37] Totzeck, C. & Wolfram, M.-T. (2020) Consensus-based global optimization with personal best. *Math. Biosci. Eng* **17**(5), 6026–6044.
[38] Schillings, C., Totzeck, C. & Wacker, P. (2023) Ensemble-based gradient inference for particle methods in optimization and sampling. *SIAM/ASA J. Uncert. Quant.* **11**(3), 757–787.
[39] Chen, J., Jin, S. & Lyu, L. (2020) A consensus-based global optimization method with adaptive momentum estimation, arXiv:2012.04827.
[40] Kalise, D., Sharma, A. & Tretyakov, M. V. (2023) Consensus-based optimization via jump-diffusion stochastic differential equations. *Math. Model Meth. Appl. Sci.* **33**(02), 289–339.
[41] Bungert, L., Wacker, P. & Roith, T. (2022). Polarized consensus-based dynamics for optimization and sampling, arXiv:2211.05238, 2022.
[42] Higham, D. J. (2001) An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev* **43**(3), 525–546.
[43] Platen, E. (1999) An introduction to numerical methods for stochastic differential equations. In: *Acta Numerica*, Vol. 8, Cambridge Univ. Press, Cambridge, pp. 197–246.
[44] Dembo, A. & Zeitouni, O. (1998) Large deviations techniques and applications. In: *Applications of Mathematics (New York)*, Vol. 38, second ed. Springer-Verlag, New York
[45] Huang, H. & Qiu, J. (2022) On the mean-field limit for the consensus-based optimization. *Math. Methods Appl. Sci* **45**(12), 7814–7831.
[46] Ambrosio, L., Gigli, N. & Savaré, G. (2008). Gradient flows in metric spaces and in the space of probability measures. In: *Lectures in Mathematics ETH Zürich*. second ed. Birkhäuser Verlag, Basel

Cite this article: Fornasier M., Richtárik P., Riedl K. and Sun L. (2025). Consensus-based optimisation with truncated noise. *European Journal of Applied Mathematics*, **36**, 292–315. https://doi.org/10.1017/S095679252400007X