Effective *fluctuating* continuum models for stochastic gradient descent

Benjamin Gess
TU Berlin & MPI MiS Leipzig

Conference "Flows on Measure Spaces and Applications in Machine Learning", MFO, 2026

Max-Planck-Institut für Mathematik in den Naturwissenschaften
Funded by the European Union
European Research Council
Established by the European Commission
Technische Universität Berlin

Focus: Effective *fluctuating* continuum models for stochastic gradient descent with small learning rate, or on overparameterized networks.

Agenda

I Recap: Stochastic Gradient Descent (SGD) and its diffusion approximation

II Stochastic flow approximation for SGD with small learning rate

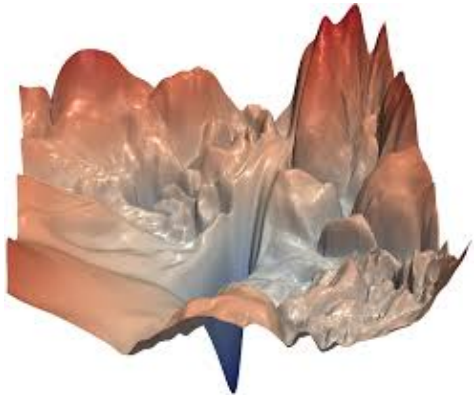
joint work with S. Kassing and V. Konarovskyi, 2024, https://arxiv.org/abs/2302.07125

III Stochastic PDEs / stochastic Wasserstein gradient flow

joint work with R. Gvalani and V. Konarovskyi, 2025, https://arxiv.org/abs/2207.05705.

Aim: Minimize a function

$$R(z) := \mathbb{E}_\nu[\tilde{R}(z, \xi)].$$



$(\Xi, \mathcal{G}, \nu)$ is a measure space with finite measure ν, $L_2((\Xi, \nu); \mathbb{R})$ is separable

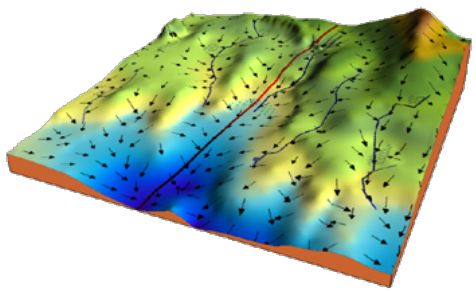
$$\tilde{R} : \mathbb{R}^d \times \Xi \to \mathbb{R} \text{ is differentiable in the first component}$$

E.g. empirical risk

$$R(z) = \mathbb{E}_\nu[(N(z, \xi) - f(\xi))^2] = \frac{1}{m} \sum_{i=1}^{m} (N(z, \xi_i) - f(\xi_i))^2 = \frac{1}{m} \sum_{i=1}^{m} \tilde{R}(z, \xi_i).$$

Stochastic gradient descent:

$$Z_{n+1}(x) = Z_n(x) - \eta \nabla \tilde{R}(Z_n(x), \xi_n), \quad n \in \mathbb{N}_0,$$



where $Z_0(x) = x$ is the starting value, $\eta > 0$ is the step-size/learning rate, $(\xi_n)_{n \in \mathbb{N}}$ is an i.i.d. sequence with $\mathcal{L}(\xi_1) = \nu$

SGD

$$Z_{n+1}(x) = Z_n(x) - \eta \nabla \tilde{R}(Z_n(x), \xi_n), \quad n \in \mathbb{N}_0,$$

may be interpreted as a perturbed Euler step of length η of the ODE

$$\dot{Y}_t = -\nabla R(Y_t).$$

$\rightsquigarrow$ $t_n = n\eta$ "numerical time" of the process

Theorem

Let R be sufficiently regular. Then, there exists a constant $C > 0$ such that, for every $g \in C_b^2(\mathbb{R}^d)$,

$$\sup_{x \in \mathbb{R}^d} \sup_{n : n\eta \leq T} |\mathbb{E}g(Z_n(x)) - \mathbb{E}g(Y_{n\eta}(x))| \leq C\eta.$$

But

- Information about (dynamic) fluctuations are lost
- Only first order

[Li, Tai, E; JMLR 2019] Rewrite SGD as

$$\begin{aligned} Z_{n+1}(x) &= Z_n(x) - \eta \nabla \tilde{R}(Z_n(x), \xi_n), \quad n \in \mathbb{N}_0 \\ &= Z_n(x) - \eta \nabla R(Z_n(x)) + \eta \Big(\nabla R(Z_n(x)) - \nabla \tilde{R}(Z_n(x), \xi_n) \Big) \end{aligned}$$

Let

$$\Sigma(z) = \mathbb{E}(\underbrace{\nabla \tilde{R}(z, \xi_1) - \nabla R(z)}_{G(z, \xi_1)}) \otimes (\nabla \tilde{R}(z, \xi_1) - \nabla R(z)) = \mathbb{E} G(z, \xi_1) \otimes G(z, \xi_1).$$

Stochastic modified equation:

$$dY_t(x) = -\nabla R(Y_t(x))dt \; - \frac{\eta}{4} \nabla |\nabla R(Y_t(x))|^2 dt + \sqrt{\eta}\, \Sigma^{1/2}(Y_t(x)) dW_t$$

Note: Modified drift.

Theorem (Li, Tai, E 2019)

Let $\tilde{R}$ and $\Sigma^{1/2}$ be regular enough, i.p. $\Sigma^{1/2} \in C_b^4$. Then, for every $g \in C_b^4(\mathbb{R}^d)$ and $T > 0$, there exists a constant $C > 0$ such that

$$\sup_{x \in \mathbb{R}^d} \sup_{n : n\eta \leq T} |\mathbb{E}[g(Z_n(x))] - \mathbb{E}[g(Y_{n\eta}(x))]| \leq C\eta^2.$$

Proof: Expansion of the generators and regularity of the semigroups.

Limitations:

- Requires regularity of diffusion coefficients $\Sigma^{1/2}$ with

$$\Sigma(z) = \mathbb{E}(\nabla \tilde{R}(z, \xi_1) - \nabla R(z)) \otimes (\nabla \tilde{R}(z, \xi_1) - \nabla R(z))$$

in

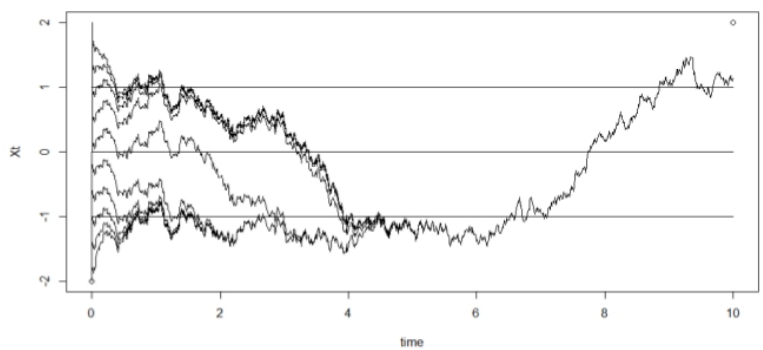
$$dY_t(x) = -\nabla R(Y_t(x))dt - \frac{\eta}{4}\nabla |\nabla R(Y_t(x))|^2 dt + \sqrt{\eta}\,\Sigma(Y_t(x))^{1/2} dW_t.$$

But since Σ can degenerate, only Lipschitz continuity holds. Even if the risks $\tilde{R}$ are smooth.

- Dynamical systems approach to the dynamics of SGD, e.g. [Sato, Tsutsui, Fujiwara; 2022], [Wu, Ma, E; 2018]: Use concepts of attractors, Lyapunov exponents, stochastic synchronization. E.g. asymptotic global stability,

$$d(Z_n(x), Z_n(y)) = |Z_n(x) - Z_n(y)| \to 0$$

for $n \to \infty$, in probability.



Multi-point motions are not matched by SME! Only have $\mathcal{L}(Z_n(x)) \approx \mathcal{L}(Y_{n\eta}(x))$.

Agenda

I Recap: Stochastic Gradient Descent and its diffusion approximation

II **Stochastic flow approximation for SGD with small learning rate**
joint work with S. Kassing and V. Konarovskyi, 2024, https://arxiv.org/abs/2302.07125

III Stochastic PDEs / stochastic Wasserstein gradient flow
joint work with R. Gvalani and V. Konarovskyi, 2025, https://arxiv.org/abs/2207.05705.

Stochastic gradient descent:

$$Z_{n+1}(x) = Z_n(x) - \eta \nabla R(Z_n(x)) + \eta \Big(\nabla R(Z_n(x)) - \nabla \tilde{R}(Z_n(x), \xi_n) \Big).$$

Replace stochastic modified equation

$$dY_t = -\nabla R(Y_t)dt - \frac{\eta}{4} \nabla |\nabla R(Y_t)|^2 dt + \sqrt{\eta} \Sigma^{\frac{1}{2}}(Y_t) dW, \qquad \text{(SME)}$$

by *Stochastic modified flow*

$$dX_t = -\nabla R(X_t)dt - \frac{\eta}{4} \nabla |\nabla R(X_t)|^2 dt + \sqrt{\eta} \int_{\Xi} \overbrace{G(X_t(x), \xi)}^{= \nabla R(X_t(x)) - \nabla \tilde{R}(X_t(x), \xi)} W(d\xi, dt), \qquad \text{(SMF)}$$

$$= -\nabla R(X_t)dt - \frac{\eta}{4} \nabla |\nabla R(X_t)|^2 dt + \sqrt{\eta} \sum_{i=1}^{\infty} \langle G(X_t, \cdot), e_i \rangle_{\nu} dW^{(i)}$$

where

- W is a cylindrical Wiener process on the space $L^2(\Xi, \nu)$
- $(e_i)_{i \in \mathbb{N}}$ is an orthonormal basis of $L^2(\Xi, \nu)$
- $(W_t^{(1)})_{t \geq 0}, (W_t^{(2)})_{t \geq 0}, \ldots$ are independent Brownian motions on $\mathbb{R}$.

Multi point motion

Stochastic gradient descent: $Z_{n+1}(x) = Z_n(x) - \eta\Big(\nabla R(Z_n(x)) + \nabla\tilde{R}(Z_n(x),\xi_n) - \nabla R(Z_n(x))\Big)$.

$$\begin{aligned}
\operatorname{cov}(Z_1(x), Z_1(y)) &= \mathbb{E}(\eta(\nabla R(x) - \nabla\tilde{R}(x,\xi_1)), \eta(\nabla R(y) - \nabla\tilde{R}(y,\xi_1))) \\
&= \eta^2\mathbb{E}((\underbrace{\nabla R(x) - \nabla\tilde{R}(x,\xi_1)}_{:=G(x,\xi_1)}), (\nabla R(y) - \nabla\tilde{R}(y,\xi_1))) \\
&= \eta^2\mathbb{E}_\nu(G(x,\xi_1), G(y,\xi_1)) \\
&= \eta^2\langle(G(x,\cdot), G(y,\cdot)\rangle_\nu.
\end{aligned}$$

Stochastic modified flows: $dX_t(x) = ... + \sqrt{\eta}\sum_{i=1}^{\infty}\langle G(X_t,\cdot), e_i\rangle_\nu dW^{(i)}$.Note:

$$\begin{aligned}
\operatorname{cov}(X_{t=\eta}(x), X_{t=\eta}(y)) &= \eta\,\mathbb{E}\int_0^{t=\eta}(\sum_{i=1}^{\infty}\langle G(X_s(x),\cdot), e_i\rangle_\nu dW^{(i)}, \sum_{j=1}^{\infty}\langle G(X_s(y),\cdot), e_i\rangle_\nu dW^{(j)}) \\
&= \eta\sum_{i=1}^{\infty}\mathbb{E}\int_0^{t=\eta}(\langle G(X_s(x),\cdot), e_i\rangle_\nu, \langle G(X_s(y),\cdot), e_i\rangle_\nu)ds \\
&= \eta\,\mathbb{E}\int_0^{t=\eta}\langle G(X_s(x),\cdot), G(X_s(y),\cdot)\rangle_\nu)ds \\
&\approx \eta^2\,\langle G(x,\cdot), G(y,\cdot)\rangle_\nu = \eta^2\operatorname{cov}(Z_1(x), Z_1(y)).
\end{aligned}$$

Regularity and multi-point motion

Stochastic gradient descent:

$$Z_{n+1}(x) = Z_n(x) - \eta \nabla R(Z_n(x)) + \eta \Big(\nabla R(Z_n(x)) - \nabla \tilde{R}(Z_n(x), \xi_n) \Big).$$

Stochastic modified flows:

$$dX_t(x) = -\nabla R(X_t(x))dt - \frac{\eta}{4} \nabla |\nabla R(X_t(x))|^2 dt + \sqrt{\eta} \int_{\Xi} \underbrace{G(X_t(x), \xi)}_{= \nabla R(X_t(x)) - \nabla \tilde{R}(X_t(x), \xi)} W(d\xi, dt)$$

Theorem (G., Kassing, Konarovskyi '24)

Assume that $\|\tilde{R}(\cdot, \xi)\|_{C_b^{\mathbf{6}}} \in L^2(\Xi, \nu)$. *Then, for every* $\Phi \in C_b^4(\mathcal{P}_2(\mathbb{R}^d))$ *and* $T > 0$, *there exists a constant* $C > 0$ *such that*

$$\sup_{\mu_{\mathbf{0}} \in \mathcal{P}_{\mathbf{2}}(\mathbb{R}^d)} \sup_{n : n\eta \leq T} |\mathbb{E}[\Phi(\mu_0 \circ (Z_n)^{-1})] - \mathbb{E}[\Phi(\mu_0 \circ (X_{n\eta})^{-1})]| \leq C\eta^2.$$

In particular, for all $g \in C_b^4(\mathbb{R}^{md})$,

$$\sup_{x^{\mathbf{1}}, \dots, x^m \in \mathbb{R}^d} \sup_{n : n\eta \leq T} |\mathbb{E}[g(Z_n(x^1), \dots, Z_n(x^m))] - \mathbb{E}[g(X_{n\eta}(x^1), \dots, X_{n\eta}(x^m))]| \leq C\eta^2.$$

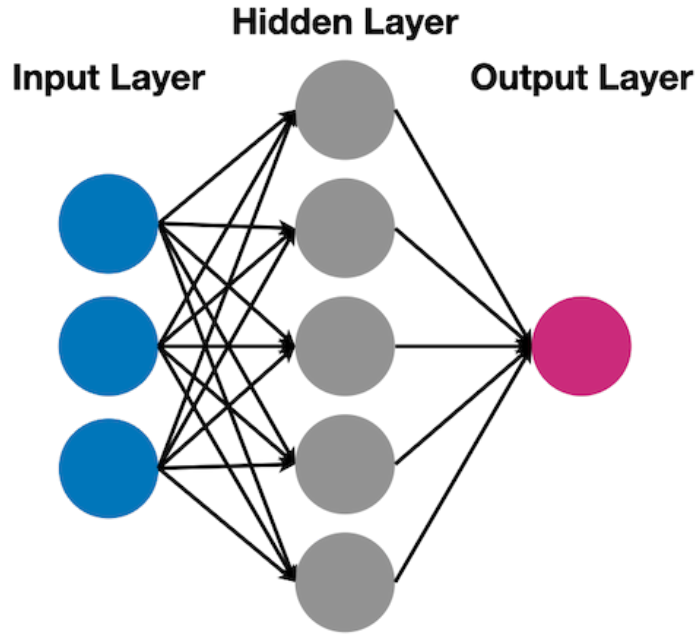
Agenda

I Recap: Stochastic Gradient Descent and its diffusion approximation

II Stochastic flow approximation for SGD with small learning rate
joint work with S. Kassing and V. Konarovskyi, 2024, https://arxiv.org/abs/2302.07125

III **Stochastic PDEs / stochastic Wasserstein gradient flow**
joint work with R. Gvalani and V. Konarovskyi, 2025, https://arxiv.org/abs/2207.05705.

Shallow ReLU networks



Shallow neural network with

- input dimension D
- number of hidden neurons M
- activation function $\sigma : \mathbb{R} \to \mathbb{R}$

is given by

$$N^M(\overbrace{x}^{\text{parameters}}, \overbrace{\xi}^{\text{argument}}) = \frac{1}{M}\sum_{i=1}^{M} x_{(D+2)i+D+2}\,\sigma\Big(\sum_{j=1}^{D} x_{(D+2)i+j}\,\xi_j + x_{(D+2)i+D+1}\Big)$$

$$= \frac{1}{M}\sum_{i=1}^{M} \Psi(\overbrace{z_i}^{\text{parameters}}, \xi)$$

$$= \int \Psi(z,\xi)\mu^M(dz) = \langle \Psi(\cdot,\xi), \mu^M \rangle$$

where $\mu^M = \frac{1}{M}\sum_{j=1}^{M} \delta_{z_j}$ is the empirical measure of the parameters.

The square loss is given by

$$\tilde{R}(z,\xi) = \frac{1}{2}|f(\xi) - N^M(z,\xi)|^2.$$

Going back to the ODE approximation

$$(\dot{Z}_t^M)_j = -V((Z_t^M)_j, \mu_t^M)$$

we get that $\mu^M = \frac{1}{M}\sum_{j=1}^M \delta_{Z_j}$ satisfies

$$d\mu_t^M = \nabla\left(V(\cdot,\mu_t^M)\mu_t^M\right)dt$$

Draw starting values Z_0^j, $j \in [M]$, i.i.d. from a distribution μ_0. Then, expect

$$\lim_{M\to\infty} \mu^M \to \mu$$

with

$$d\mu_t = \nabla\left(V(\cdot,\mu_t)\mu_t\right)dt.$$

Theorem (Convergence to deterministic PDE; Mei, Montanari, Nguyen '18; Chizat, Bach '18, Rotskoff, Vanden-Eijnden '18)

If $Z_k(0) \sim \mu_0$ *– i.i.d., then*

$$d(\mu_t^M, \mu_t) = O(\frac{1}{\sqrt{M}})$$

with

$$d\mu_t = -\nabla \left(\mu_t V(\cdot, \mu_t)\right) dt$$

or in Lagragian perspective

$$\frac{d}{dt} X_t(x) = V(\mu_t, X_t(x))$$
$$X_0(x) = x, \quad \mu_t = \mu_0 \circ (X_t)^{-1}, \quad x \in \mathbb{R}^d, \quad t \geq 0.$$

But:

- No information on the (dynamic) fluctuations is retained
- Limited order of convergence in η:

$$d(\mu_n^M, \mu_{\eta n}) = O(\eta + \frac{1}{\sqrt{M}})$$

Keep fluctuations? [Rotskoff, Vanden-Eijnden, '18, '22], [Chen, Rotskoff, Bruna, Vanden-Eijnden '20] Stochastic modified equation:

$$\begin{aligned} dZ_t(x) &= -\nabla R(Z_t(x))dt + \sqrt{\eta}\,\Sigma^{1/2}(Z_t(x))dW_t \\ &= V(Z_t(x),\mu^M)dt + \sqrt{\eta}\,\tilde{\Sigma}(Z_t(x),\mu^M)dW_t \end{aligned}$$

Get Dean's equation for the empirical density

$$d\mu_t^M = -\nabla\cdot\left(V(\cdot,\mu_t^M)\,\mu_t^M\right)dt + \frac{\sigma}{2}D^2 : \left(A(\cdot,\mu_t^M)\,\mu_t^M\right)dt + \sigma^{\frac{1}{2}}\nabla\cdot\left(d\mathcal{M}_t^M\right)$$

with

$$[\langle\psi,\mathcal{M}^M\rangle]_t = \int_0^t \iint \psi(x)\otimes\psi(y) : \tilde{A}(x,y,\mu_s^M)\,\mu_s^M(dx)\,\mu_s^M(dy)\,ds$$

and

$\tilde{A}(x,y,\mu) = \mathbb{E}_\vartheta[\,G(x,\mu,\xi)\otimes G(y,\mu,\xi)\,]$

$$G(x,\mu,\xi) = \left(f(\xi) - \int \Psi(y,\xi)\,\mu(dy)\right)\nabla_x\Psi(x,\xi) - \mathbb{E}_\vartheta\left[\left(f(\xi) - \int \Psi(y,\xi)\,\mu(dy)\right)\nabla_x\Psi(x,\xi)\right].$$

Difficulties:

- Infinite dimensional martingale problem (nonlocal coefficients, degenerate coercivity)
- Approaching via SPDE: Naive guess leads to taking square root of terms in quadratic variation $\rightarrow$ irregular coefficients

Choose process: stochastic modified flow PDE

$$d\mu_t = -\nabla \cdot (V(\cdot, \mu_t)\mu_t)dt + \sqrt{\eta'}\nabla \cdot \int_\Xi G(\cdot, \mu_t, \xi)\mu_t \, W(d\xi, \circ dt)$$

where $A(x, \mu) = \mathbb{E}_\nu G(x, \mu) \otimes G(x, \mu)$.

Well-posedness results for similar SPDEs: "Nonlinear Fokker-Planck equation with common noise"

$$\partial_t \mu = \partial^2_{i,j}\big(a^{ij}(t, x, \mu)\mu\big) - \partial_i\big(b^i(t, x, \mu)\mu\big) - \partial_i\big(\sigma^{ik}(t, x, \mu)\mu\big)\, dW^k_t.$$

- Deterministic continuity equation with irregular coefficients [DiPerna, Lions, Ambrosio, Trevisan, Crippa...]. There $A = G = 0$.
- Particle representations for a class of nonlinear SPDEs [Kurtz, Xiong '99], with initial condition μ_0 having L_2-density.
- Stochastic nonlinear Fokker-Planck equation [Coghi, G. '19]. Duality method, Backward SPDEs & their regularity theory, Sobolev embeddings.
- ... this work ...
- Stochastic nonlinear Fokker-Planck equation [Bugini, Friz, Stannat, '25]. Duality method, Backward SPDEs & their regularity theory via rough paths.

Theorem (Well-posedness and superposition principle, G., Gvalani , Konarovskyi 25)

Let the coefficients V, G be Lipschitz continuous and smooth enough w.r.t. spatial variable. Then the SMF PDE

$$d\mu_t = -\nabla \cdot (V(\mu_t, \cdot)\mu_t)\, dt - \sqrt{\eta'}\nabla \cdot \int_\Xi G(\mu_t, \cdot, \xi)\mu_t W(d\xi, \circ dt)$$

has a unique solution.Moreover, μ_t is a superposition solution, i.e.,

$$\mu_t = \mu_0 \circ X_t^{-1}(\cdot), \quad t \geq 0,$$

where X solves

$$dX_t(x) = V(\mu_t, X_t(x))dt + \sqrt{\eta'} \int_\Xi G(\mu_t, X_t(x), \xi) W(d\xi, dt),$$
$$X_0(x) = x, \quad \mu_t = \mu_0 \circ (X_t)^{-1}, \quad x \in \mathbb{R}^d, \quad t \geq 0.$$

Aim: Further relax regularity assumptions on the coefficients.

Comments on the proof:

- Show well-posedness of the Lagrangian system

$$dX_t(x) = V(\mu_t, X_t(x))dt + \sqrt{\eta'} \int_\Xi G(\mu_t, X_t(x), \xi) W(d\xi, dt),$$
$$X_0(x) = x, \quad \mu_t = \mu_0 \circ (X_t)^{-1}, \quad x \in \mathbb{R}^d, \quad t \geq 0.$$

- Hence: Have uniqueness iff every solution a superposition solution
- Freeze the measure coefficient

$$d\mu_t = -\nabla \cdot (\overbrace{V(\cdot, \mu_t)}^{=:V(\cdot,t)} \mu_t)dt + \sqrt{\eta'}\nabla \cdot \int_\Xi \overbrace{G(\cdot, \mu_t, \xi)}^{=:G(\cdot,t)} \mu_t \, W(d\xi, \circ dt).$$

- Then transform away the (linear) transport noise.
- Then apply superposition principle on the random PDE: I.e. can write

$$\mu_t = \mu_0 \circ (X_t)^{-1}.$$

How about the rate of convergence in η? Need to include the bias correction / modification of the drift. Lagrangian system becomes

$$
\begin{aligned}
dX_t(x) = &-\nabla R(\mu_t, X_t(x))dt - \frac{\eta'}{4}\nabla|\nabla R(\mu_t, X_t(x))|^2 dt \\
&- \frac{\eta'}{4}\left\langle \underbrace{D}_{\text{Lions' derivative}} |\nabla R(\mu_t, X_t(x))|^2, \mu_t \right\rangle dt \\
&+ \sqrt{\eta'}\int_{\Xi} G(\mu_t, X_t(x), \xi) W(d\xi, dt), \\
X_0(x) = x, \quad &\mu_t = \mu_0 \circ (X_t)^{-1}, \quad x \in \mathbb{R}^d, \quad t \geq 0.
\end{aligned}
$$

Theorem (G., Kassing, Konarovskyi 24)

Let $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$, and $\mu_0^M = \frac{1}{M}\sum_{j=1}^m \delta_{Z_0^j}$, where Z_0^j, $j \in [M]$, are i.i.d. random variables with distribution μ. Assume that $\|\Psi(\cdot,\xi)\|_{C_b^6} + |f(\xi)| \in L^4(\Xi, \nu)$. Then, for every $\Phi \in C_b^4(\mathcal{P}_2(\mathbb{R}^d))$ there exists a constant $C > 0$ independent of η' and M such that

$$
\sup_{n:n\eta' \leq T} \left|\mathbb{E}\Phi(\mu_{n\eta'}) - \mathbb{E}\Phi(\mu_n^M)\right| \leq C(\eta')^2 + C\sqrt{\mathbb{E}\mathcal{W}_2^2(\mu_0, \mu_0^M)} \tag{1}
$$

for all $\eta' > 0$ and $M \in \mathbb{N}$.
Proof: Matching generators, but now for diffusions on the space of measures.

Higher order approximation?

Theorem (Quantified Central Limit Theorem for SMFE, G., Gvalani, Konarovskyi 25)

Let μ_t^η be the solution to the martingale problem. Assume $\lim \eta^{-1} M^{-1}(\eta) < \infty$. Then, $N_t^\eta := \sqrt{\eta}\left(\mu_t^\eta - \mu_t^0\right) \to N_t$ where N_t is a Gaussian process solving

$$dN_t = -\nabla \cdot \left(\tilde{V}(\mu_t^0, \cdot) N_t + \langle D\tilde{V}(\mu_t^0, \cdot)\mu_t^0, N_t \rangle \mu_t^0(dx) \right) dt - \nabla \cdot \int_\Xi G(\mu_t^0, \cdot, \xi)\mu_t^0 W(d\xi, dt).$$

Moreover,

$$\mathcal{W}_2^2(\mathcal{L}(N^\eta), \mathcal{L}(N)) \lesssim \eta.$$

Remark. *[Sirignano, Spiliopoulos, '20] Let $\nu_t^{M, \frac{1}{M}} := \frac{1}{M}\sum_{i=1}^M \delta_{x_i(Mt)}$. For fluctuation field $\tilde{N}_t^{\ \eta} := \sqrt{\eta}(\nu_t^{M, \frac{1}{M}} - \mu_t^0)$ have*

$$\mathbb{E} \sup_{t \in [0,T]} \|\tilde{N}_t^\eta\|_{-J}^2 \leq C \quad \textit{and} \quad \tilde{N}^\eta \to N.$$

We get

$$\mu_t^{\frac{1}{M}} = \mu_t^0 + M^{-1/2} N + O(M^{-1})$$
$$\nu_t^{M, \frac{1}{M}} = \mu_t^0 + M^{-1/2} N + o(M^{-1/2}).$$

Therefore

References

- Gess, Kassing, Konarovskyi; *Stochastic Modified Flows, Mean-Field Limits and Dynamics of Stochastic Gradient Descent*, JMLR 2024
- Gess, Kassing, Rana; *Stochastic Modified Flows for Riemannian Stochastic Gradient Descent*, to appear in SICON 2024+
- Gess, Gvalani, Konarovskyi; *Conservative SPDEs as fluctuating mean field limits of stochastic gradient descent*, PTRF 2025