STAT 946 Fall 2025 Scribes: Jonathan Borenstein
November 12

# Lecture 17: Infinite-Width SDE Limit for Covariance Dynamics

## Setup

We consider the covariance recursion for shaped activations:

$$h_{\ell+1}^{\alpha} = \sqrt{\frac{c}{n}} W_\ell \varphi_S(h_\ell^\alpha), \qquad c^{-1} = \mathbb{E}\,\varphi_S(w)^2, \quad w \sim N(0,1).$$

### Shaped Activations

**ReLU shaping:**

$$\varphi_S(x) = S_+ \max(x,0) + S_- \min(x,0), \qquad S_\pm = 1 + \frac{c_\pm}{n^p}.$$

**Smooth shaping:**

$$\varphi_S(x) = S\varphi\left(\frac{x}{S}\right), \qquad S = an^p, \quad \varphi(0) = 0,\ \varphi'(0) = 1.$$

## Correlation Kernel Expansion

Recall for ReLU:

$$\bar{J}_1(\rho^{\alpha\beta}) = \frac{\sqrt{1-\rho^2} + \rho\ \arccos(-\rho)}{2\pi},$$

with jointly Gaussian input:

$$\begin{pmatrix} w^\alpha \\ w^\beta \end{pmatrix} \sim N\left(0, \begin{bmatrix} 1 & \rho^{\alpha\beta} \\ \rho^{\alpha\beta} & 1 \end{bmatrix}\right).$$

The shaped kernel satisfies

$$cK_1(\rho^{\alpha\beta}) = \mathbb{E}\, c\, \varphi_S(w^\alpha)\varphi_S(w^\beta) = \rho^{\alpha\beta} + \frac{\nu(\rho^{\alpha\beta})}{n^{2p}} + O(n^{-3p}),$$

with

$$\nu(\rho) = \frac{(c_+ - c_-)^2}{2\pi}\big[\sqrt{1-\rho^2} + \rho\ \arccos(\rho)\big].$$

Choosing $p = \frac{1}{2}$ ensures drift $O(n^{-1})$, noise $O(n^{-1/2})$, matching Euler–Maruyama scaling.

# Covariance Recursion

Define

$$\Phi_{\ell+1}^{\alpha\beta} = \frac{c}{n}\,\langle \varphi_{\ell+1}^{\alpha}, \varphi_{\ell+1}^{\beta}\rangle, \qquad \varphi_S(ax) = a\varphi_S(x).$$

## Drift Term

$$\begin{aligned}\mathbb{E}\Big[\Phi_{\ell+1}^{\alpha\beta} \mid \mathcal{F}_\ell\Big] &= \mathbb{E}\left[c\, \varphi_S(h_{\ell+1,i}^{\alpha})\varphi_S(h_{\ell+1,i}^{\beta}) \mid \mathcal{F}_\ell\right] \\ &= (\Phi_\ell^{\alpha\alpha}\Phi_\ell^{\beta\beta})^{1/2}\, cK_1(\rho_\ell^{\alpha\beta}).\end{aligned}$$

Using the expansion,

$$\mathbb{E}\Big[\Phi_{\ell+1}^{\alpha\beta} \mid \mathcal{F}_\ell\Big] = \Phi_\ell^{\alpha\beta} + \frac{(\Phi_\ell^{\alpha\alpha}\Phi_\ell^{\beta\beta})^{1/2}\,\nu(\rho_\ell^{\alpha\beta})}{n} + O(n^{-3/2}).$$

## Noise Term

$$\begin{aligned}\Phi_{\ell+1}^{\alpha\beta} - \mathbb{E}\Big[\Phi_{\ell+1}^{\alpha\beta} \mid \mathcal{F}_\ell\Big] &= \frac{1}{\sqrt{n}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[c\varphi_S(h_{\ell+1,i}^{\alpha})\varphi_S(h_{\ell+1,i}^{\beta}) - cK_1(\rho_\ell^{\alpha\beta})\Big](\Phi_\ell^{\alpha\alpha}\Phi_\ell^{\beta\beta})^{1/2} \\ &= \frac{1}{\sqrt{n}}\,\xi_\ell^{\alpha\beta}.\end{aligned}$$

Thus

$$\Phi_{\ell+1} = \Phi_\ell + \frac{b(\Phi_\ell)}{n} + \frac{1}{\sqrt{n}}\xi_\ell.$$

## Noise Covariance and Wick Expansion

$$\mathrm{Cov}(\xi_\ell^{\alpha\beta}, \xi_\ell^{\gamma\delta}) = \mathbb{E}(w^\alpha w^\beta w^\gamma w^\delta) - \rho^{\alpha\beta}\rho^{\gamma\delta}.$$

Wick/Isserlis:

$$\begin{aligned}\mathbb{E}w^\alpha w^\beta w^\gamma w^\delta &= \mathbb{E}w^\alpha w^\beta\, \mathbb{E}w^\gamma w^\delta + \mathbb{E}w^\alpha w^\gamma\, \mathbb{E}w^\beta w^\delta + \mathbb{E}w^\alpha w^\delta\, \mathbb{E}w^\beta w^\gamma \\ &= \rho^{\alpha\beta}\rho^{\gamma\delta} + \rho^{\alpha\gamma}\rho^{\beta\delta} + \rho^{\alpha\delta}\rho^{\beta\gamma}.\end{aligned}$$

Hence the cancellation:

$$\mathrm{Cov}(\xi_\ell^{\alpha\beta}, \xi_\ell^{\gamma\delta}) = \rho^{\alpha\gamma}\rho^{\beta\delta} + \rho^{\alpha\delta}\rho^{\beta\gamma}.$$

Reintroducing the $\Phi$-prefactors:

$$\mathrm{Cov}(\xi_\ell^{\alpha\beta}, \xi_\ell^{\gamma\delta}) = \Phi_\ell^{\alpha\gamma}\Phi_\ell^{\beta\delta} + \Phi_\ell^{\alpha\delta}\Phi_\ell^{\beta\gamma} + O(n^{-1/2}).$$

## Final Covariance Recursion

$$\Phi_{\ell+1}^{\alpha\beta} = \Phi_\ell^{\alpha\beta} + \frac{1}{n}(\Phi_\ell^{\alpha\alpha}\Phi_\ell^{\beta\beta})^{1/2}\,\nu(\rho_\ell^{\alpha\beta}) + \frac{1}{\sqrt{n}}\big(\Sigma(\Phi_\ell)^{1/2}\xi_\ell\big)^{\alpha\beta} + O(n^{-3/2}).$$

## Infinite-Width Limit (Li–Nica–Roy, 2022)

**Theorem 1** (Li–Nica–Roy (2022))**.** *Let $d, n \to \infty$ with $\frac{d}{n} \to \bar{\tau} > 0$. Then*

$$\Phi_\tau^{(n)} := \Phi_{\lfloor \tau n \rfloor} \Longrightarrow \Phi_\tau$$

*where $\Phi_\tau$ satisfies*

$$d\Phi_\tau = b(\Phi_\tau)\,d\tau + \Sigma(\Phi_\tau)^{1/2}\,dB_\tau, \qquad \Phi_0^{\alpha\beta} = \frac{1}{n_0}\langle x^\alpha, x^\beta\rangle.$$

### Smooth Activation Drift

For $\varphi_S(x) = S\varphi(x/S)$ with $S = an^{1/2}$,

$$b(\Phi)^{\alpha\beta} = \frac{\varphi''(0)^2}{4a^2}\big(\Phi^{\alpha\alpha}\Phi^{\beta\beta} + \Phi^{\alpha\beta}(2\Phi^{\alpha\beta} - 3)\big) + \frac{\varphi'''(0)}{2a^2}\Phi^{\alpha\beta}(\Phi^{\alpha\alpha} + \Phi^{\beta\beta} - 2).$$

## Scalar Case and Blow-Up

For $m = 1$:

$$b(\Phi) = b\,\Phi(\Phi - 1), \qquad \Sigma(\Phi) = 2\Phi^2,$$

with

$$b = \frac{3}{4}\varphi''(0)^2 + \varphi'''(0).$$

Example blow-up:

$$\dot{X}_t = X_t^2 \quad \Rightarrow \quad X_t = \frac{1}{1-t}.$$

Example logistic drift:

$$\dot{X}_t = bX_t(X_t - 1) \quad \Rightarrow \quad X_t = \frac{x_0}{x_0 + (1 - x_0)e^{bt}}.$$

If $b < 0$: mean-reverting.

If $b > 0$ and $x_0 > 1$: blow-up at

$$t = \frac{1}{b}\log\left(\frac{x_0}{x_0 - 1}\right).$$

**Proposition 1.** $\Phi$ *avoids finite-time blow-up a.s. iff*

$$b = \frac{3}{4}\varphi''(0)^2 + \varphi'''(0) < 0.$$

*For odd increasing $\varphi$ (sigmoid, tanh), $\varphi''(0) = 0$ and $\varphi'''(0) < 0$, hence stability.*

## ReLU Coefficient Interpretation

$$\Phi_{\ell+1}^{\alpha\beta} = (\Phi_\ell^{\alpha\alpha}\Phi_\ell^{\beta\beta})^{1/2} K_1(\rho_\ell^{\alpha\beta}) + \frac{1}{\sqrt{n}}\xi_\ell^{\alpha\beta}.$$

Thus

$$\Phi_{\ell+1} \approx f \circ f \circ f \circ \cdots (\Phi_0).$$

Shaping:

$$\varphi_S(x) = x + O(n^{-1/2}) \quad \Rightarrow \quad f(\Phi) = \Phi + \frac{b(\Phi)}{n} + O(n^{-3/2}).$$

## ResNet Interpretation

$$h_{\ell+1} = h_\ell + \frac{1}{\sqrt{d}}\frac{1}{\sqrt{n}} W_\ell \varphi(h_\ell),$$

Shaped:

$$h_{\ell+1} = \sqrt{\frac{c}{n}}\, W_\ell \left( h_\ell + \frac{\psi(h_\ell)}{\sqrt{n}} \right),$$

with

$$\frac{\psi(x)}{\sqrt{n}} = \frac{c_+}{\sqrt{n}} \max(x, 0) + \frac{c_-}{\sqrt{n}} \min(x, 0).$$

### Deterministic Limit

For ReLU:

$$\partial_\tau \Phi_\tau = b(\Phi_\tau).$$

same drift as SDE, but deterministic.