1. Pedro Talk

The standard gradient descent dynamics is given by

$$dx(t) = -\nabla U(x(t))dt + \sigma dW(t), \quad x(0) = x_0.$$

In this note we will try to study the modified dynamics

$$\begin{aligned} &dX_\varepsilon(t) = -e^{-\gamma\left(U(X_\varepsilon(t)) - \min_{s\leq t} U(X_\varepsilon(s))\right)}\nabla U(X_\varepsilon(t))dt + \varepsilon dW(t), \\ &X_\varepsilon(0) = x_0 \end{aligned} \tag{1}$$

2. Idea of Analysis

In the analysis of (1), let us assume that x_0 is a local minimum and a stationary point; that is, $\nabla U(x_0) = 0$. Define the deterministic flow induced by the vector field ∇U starting at an arbitrary point $x \in \mathbb{R}^d$:

$$\frac{d}{dt}S^t x = -\nabla U(S^t x), \quad S^0 x = x. \tag{2}$$

Then, consider the proxy process to (1) given by

$$d\hat{X}_\varepsilon(t) = -e^{-\gamma\left(U(\hat{X}_\varepsilon(t)) - \min_{s<t-\tau_i} U(S^s \hat{X}_\varepsilon(\tau_i))\right)}\nabla U(\hat{X}_\varepsilon(t))dt + \varepsilon dW(t), \tag{3}$$

on the interval $\tau_i < t < \tau_{i+1}$, where the times τ_i are the innovation times defined via

$$\tau_i = \inf\left\{t \geq \tau_{i-1} : U(\hat{X}_\varepsilon(t)) < U(\hat{X}_\varepsilon(\tau_{i-1})\right\},$$

with $\tau_0 = 0$ and the same initial condition as X_ε, $\hat{X}_\varepsilon(0) = x_0$. Let us now make some observations.

The time τ_1 is bounded below by the exit time of $\hat{X}_\varepsilon$ from the basin of attraction, $\mathcal{B}(x_0)$, of x_0; that is, $\tau_1 > \sigma(x_0) = \inf\left\{t : \hat{X}_\varepsilon(t) \in \partial\mathcal{B}(x_0)\right\}$, since for every $y \in \mathcal{B}(x_0)$, $\nabla U(y) \neq 0$. Now, from FW theory, it is well known that $\sigma(x_0)$ is exponentially distributed with mean $\varepsilon^{-2}\left(\min_{y\in\partial\mathcal{B}(x_0)} V(y) - V(x_0)\right)$, where V is the quasi-potential of (3). Since, $\hat{X}_\varepsilon$ does not find a new minimum of the function U before time $\sigma(x_0)$, and since $\min_{t>0} U(S^t x_0) = U(x_0)$ we observe that the drift in equation (3) is given by $b(x) = -\nabla\left(1 - e^{-\gamma(U(y)-U(x_0))}\right)$ the quasi-potential is given by

$$V(y) = 2 - 2e^{-\gamma(U(y)-U(x_0))}.$$

As a consequence,

$$\begin{aligned} \mathbf{E}\sigma(x_0) &= 2\varepsilon^{-2}\left(1 - e^{-\gamma\left(\min_{y\in\partial\mathcal{B}(x_0)} U(y) - U(x_0)\right)}\right) \\ &\approx 2\frac{\gamma}{\varepsilon^2}\left(\min_{y\in\partial\mathcal{B}(x_0)} U(y) - U(x_0)\right) + \mathcal{O}\left(\frac{\gamma}{\varepsilon^2}\right). \end{aligned}$$

Claim 1. *By choosing $\gamma = \varepsilon^2$, as $\varepsilon \to 0$, the exit happens in almost constant time.*