Problems for Chapter 3: Optimization

Reading

- ML822 Monte Carlo notes: Secs 3.3-3.5 of Chap. 3.
- AM783 Markov Processes notes: Secs 5.1-5.4 of Chap. 5.

Theoretical

Q1. Gradient diffusions. Consider the SDE in $\mathbb{R}^d$ defined by

$$dX_t = -\nabla U(X_t)dt + \sigma dW_t, \tag{1}$$

where $U(x)$ is a smooth function, called the potential, σ is the noise amplitude (real, positive), and W_t is a d-dimensional Brownian motion. Assume that $U(x)$ is convex (U-shaped), and so has a unique minimum at some point x^*. Assume also that X_t is ergodic, so there exists a unique stationary distribution p^*.

(a) Consider the SDE *without* noise by setting $\sigma = 0$. What is the long-time behavior of the corresponding ODE?

(b) We have seen in class that, under some conditions on $U(x)$, the stationary distribution of this system is

$$p^*(x) = c\, e^{-2U(x)/\sigma^2}$$

where c is a normalization constant. Discuss the shape of p^* in relation to x^*. Where does p^* concentrate as $\sigma \to 0$?

Numerical

Q2. Stochastic gradient descent. We seek to find the global minimum of

$$U(x) = \frac{x^4}{2} - 5x^2 + x. \tag{2}$$

This potential has a positive local minimum in addition to its global minimum, which is negative.

(a) Find numerically the positions of the local and global minima of $U(x)$ using any routine or function in Python or Mathematica.

(b) Solve the gradient descent dynamics, defined by the ODE

$$\dot{x}(t) = -U'(x(t)), \tag{3}$$

for various initial conditions. You can use `odeint` in Python or your own Euler discretization scheme. Analyse your results in view of locating the global minimum of $U(x)$.

(c) Solve the stochastic gradient descent dynamics, defined by the SDE

$$dX_t = -U'(X_t)dt + \sigma dW_t, \tag{4}$$

for various initial conditions and noise amplitudes σ using the Euler–Maruyama scheme. Analyse your results and compare them with part (b). Does X_t always reach the global minimum? [Note: Use $T = 10$ and $\sigma = 1.0$, then try $T = 100$ and $\sigma = 2.5$ and only show the latter.]

(d) Repeat part (c), but now decrease the noise in time according to $\sigma_t = \frac{\alpha}{t+1}$. Try $\alpha \approx 5$ and $T \approx 10$ to 50 to see if you can locate the global minimum. [Note: Decreasing σ in time is referred to as *annealing* or *stochastic relaxation*.]

(e) What is the advantage of stochastic gradient descent over deterministic gradient descent?

Q3. Simulated annealing. We want to find the global minimum of

$$V(x) = x^2(2 + (\sin(10x))^2)$$

using simulated annealing.

(a) What is the global minimum of $V(x)$ on $\mathbb{R}$? No calculations required.

(b) Use the Metropolis algorithm with simulated annealing to locate the global minimum of $V(x)$. Use the annealing scheme $\beta_n = 1 + \log n$ and initial value X_1 uniform in $[-10, 10]$. Try different displacement distributions and show the evolution of $V(X_n)$ or X_n as a function of the simulation time n.

Q4. Graph coloring. [No help for this problem.] Every planar graph can be colored using only four colours such that no two adjacent nodes have the same color. Finding such a coloring for a planar graph of m nodes is a difficult optimization problem, since the size of the solution space is 4^m. The cost function for this problem is the number of adjacent nodes with the same coloring.

(a) Draw a connected planar graph with 10 nodes (planar means no edge crossings) and enter its adjacency matrix. [Hint: Use a grid graph, which can be generated easily with the `networkx` package.]

(b) Generate a vector of size 10 that contains random colors coded as 1, 2, 3, and 4. This vector represents the coloring of your graph. Plot the graph with the `networkx` package with its coloring.

(c) Write a function for the cost of the problem. The function receives the adjacency matrix and the color vector and outputs the number of adjacent nodes with the same colors.

(d) Use simulated annealing to find an optimal coloring of the graph. Plot the resulting graph.