STA2311 (Fall 2023) - Practice Problems for Class 4 (Stochastic Optimization)

1. Use the simulated annealing algorithm to maximize the function

$$g(x_1, x_2) = [16x_1(1-x_1)x_2(1-x_2)\sin(9\pi x_1)\sin(9\pi x_2)]^2.$$

Draw the contour of the associated distribution function and show the trajectory of the "path": $(x_{1t}, x_{2t})_{1\leq t\leq M}$

2. Consider the sample of size $n = 10,000$ in the file `banana.txt` which we model using the distribution

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{\theta}) = & \frac{1}{(2\pi)^2\theta_1\theta_2} \\ & \times \exp\left(-\frac{1}{2}\left[\frac{y_1^2}{\theta_1^2}+\frac{y_2^2}{\theta_2^2}\right]\right)\exp\left(-\frac{1}{2}\left\{\left[\frac{y_3-\theta_3(y_1^2-\theta_1)}{\theta_1}\right]^2+\left[\frac{y_4-\theta_4(y_2^2-\theta_2)}{\theta_2}\right]^2\right\}\right), \end{aligned}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_4)$ with $\theta_1, \theta_2 > 0$ and $\theta_3, \theta_4 \in \mathbb{R}$, and $\boldsymbol{y} = (y_1, \ldots, y_4)$.

   (a) Use the gradient descent (GD), and stochastic gradient descent (SGD) to find the MLE for $\boldsymbol{\theta}$.

   (b) Compare the computation costs for GD and SGD.

3. Show that if, in the proof of convergence for SGD (done in class), we choose $h_t = \frac{1}{\sqrt{1+t}}$, then

$$\mathbb{E}\left[\|\nabla f(\theta_\tau)\|^2\right] < O\left(T^{-1/2}\right).$$

Remember that $f(x) = O(g(x))$ means $|f(x)|$ is bounded above by a constant multiple of $g(x)$.