HW6 - Jacob Shkrob

Jacob Shkrob

December 14, 2022

1 Exercise 71

In this exercise, we consider the overdamped Langevin scheme with special precondtioning that involves the Hessian, focussing on the Rosenbrock density $\pi(x)$ given by

$$\pi(x) \propto \exp\left(-\frac{100(y-x^2)^2-(1-x)^2}{20}\right).$$

This density is known to have an extremely curved contours, where the mass is concentrated on this portion. Moreoever, after computing the Hessian, one sees that the 2nd order partial derivative of $-\log\pi(x)$ is equal to $-60x^2+20y-\frac{1}{10}$, which blows up extremely quickly the larger x and y becomes (quadratically in x and linearly in y). This makes optimization using the Hessian much worse, and in particular, overdamped preconditioned Langevin that relies on the Hessian D^2V, where V is $-\log(\pi(x))$. We compare ability of two different sampling mechanisms, an overdamped Langevin with Hessian D^2V preconditioning and a simplified Langevin scheme, to sample from the Rosenbrock density. The simple overdampled Langevin scheme listed in (5.8) is describred by the iteration

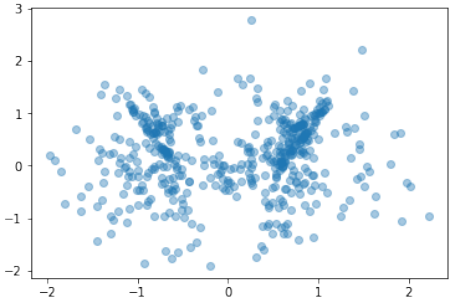
$$X_h^{k+1} = (I - hSM^{-1})X_h^k + \sqrt{2hS}\xi^{k+1},$$

where $S = I$ in our experiment. The affine invariant scheme is described by the iteration

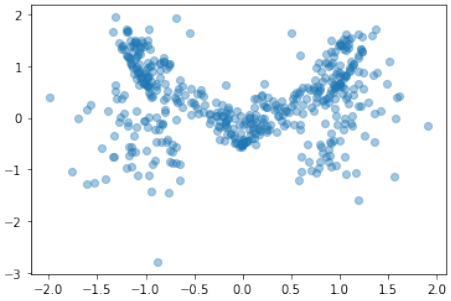
$$X_h^{k+1} = X_h^k - hH^{-1}(X_h^k)\nabla^\top V(H_h^k) + \sqrt{2hH^{-1}(X_h^k)}\xi^{k+1},$$

where $H = D^2(-\log(\pi(x)))$. After computing the gradient and Hessian of the Rosenbrock function, we plotted $N = 100$ iterations of the sampling schemes below, using the formula for the square root of Hessian, R, when necessary. Below is the original Rosebrock density and three different scatterplots of the overdamped Langevin samples. The proposal density in all tests were Gaussians with mean 0 and covariance matrix $0.5\mathbf{I}$. Even with metropolization, the affine-invariant scheme did not outperform the original scheme due to the optimzation difficulties when computing the Hessian. At smaller steps $h = 0.01$, the algorithm did start to gain the shape of Rosenbrock, but the simple Langevin scheme worked quite well (however, there are these strange clusters outside the Rosenbrock

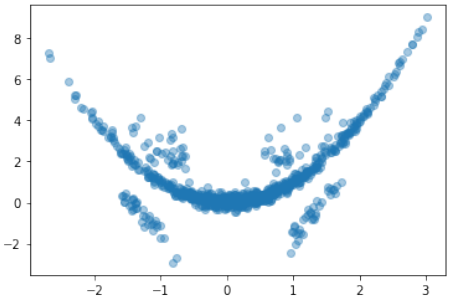
density that I could not seem to explain, possibily due to their appearance being computational errors).



(a) Overdamped Langevin, $h = 0.1$



(b) Overdamped Langevin, $h = 0.01$



(c) Simplified Langevin with no conditioning, $h = 0.1$

Figure 1: Scatterplots of Rosenbrock density using overdamped Langevin scheme at various step size parameters

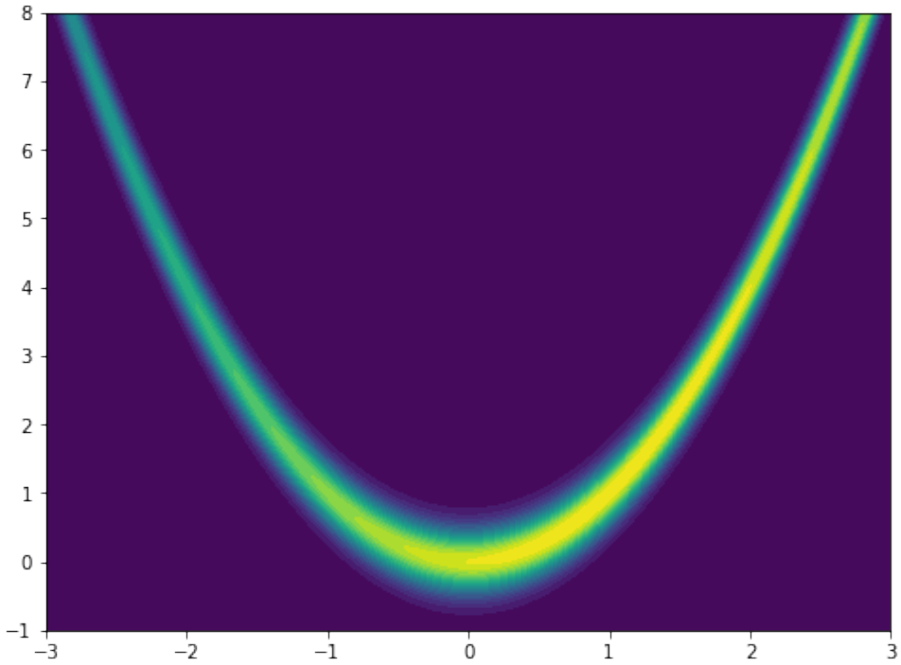
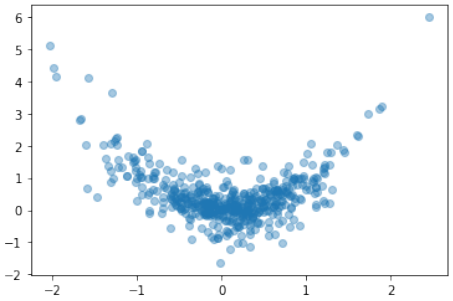


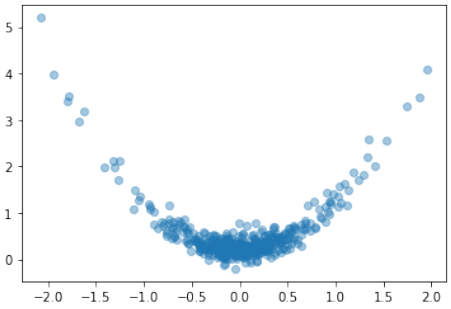
Figure 2: Rosenbrock contour plot

2 Exercise 75

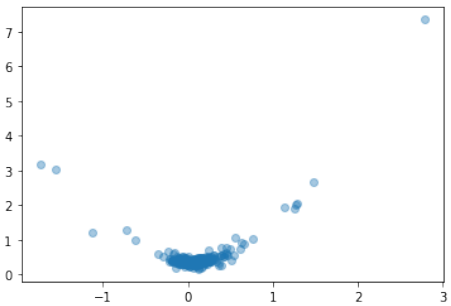
Instead of Langevin, we consider an ensemble-based algorithm using a symmetric proposal distribution $g_\alpha(z)$ which satisfies the symmetry condition $g(1/z) = zg(z)$ where $g(z) \approx 1/\sqrt{z}$ for $z \in [1/\alpha, \alpha]$ and 0 otherwise. The actual algorithm is described in (6.15) and (6.16), and we briefly discuss the result below. The sampling scheme works incredibly fast and is able to capture most of the distribution quite well, as compared to the schemes in Exercise 71. As the parameter α increased, more of the distribution was filled by the ensemble (the right part of the curve was sparse when α was low, but covered when α is high). As L increased, so did the overall cover of the density. We found that only $n = 10$ iterations of ensemble sampling is generally good enough for the Rosenbrock density. This is quite surprising, since the algorithm is quite simple to implement.



(a) Ensemble scheme, $L = 500$, $n = 5$

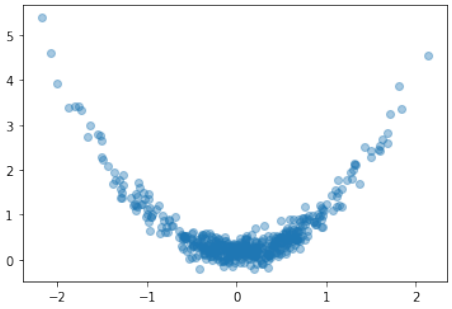


(b) Ensemble scheme, $L = 500$, $n = 10$

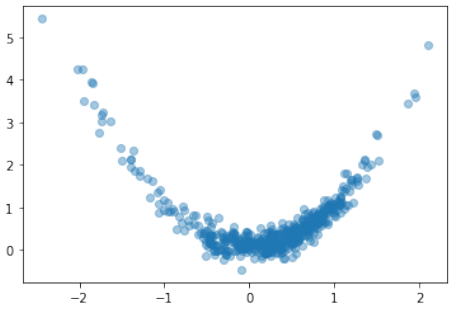


(c) Ensemble scheme, $L = 500$, $n = 20$

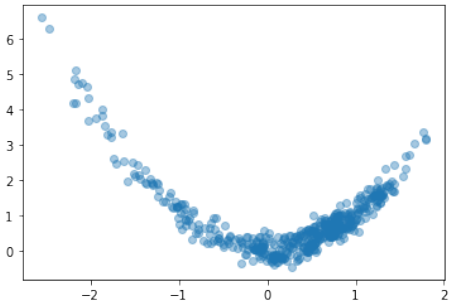
Figure 3: Scatterplots of Rosenbrock density using ensemble scheme at various iteration numbers



(a) Ensemble scheme, $L = 500$, $n = 5$, $\alpha = 2$



(b) Ensemble scheme, $L = 500$, $n = 10$, $\alpha = 10$



(c) Ensemble scheme, $L = 500$, $n = 20$, $\alpha = 30$

Figure 4: Scatterplots of Rosenbrock density using ensemble scheme at various α parameter values