HW for 2019-05-24

(due: 2019-05-31)

1: Stochastic gradient descent Consider optimizing the objective

$$\phi(c) = \frac{1}{2N} \sum_{i=1}^{N} (c_1 + c_2 x_i^2 + c_3 x_i^4 - cos(x_i))^2$$

where x_i is a uniform mesh from $[-4, 4]$ For $N = 100$ points, implement a stochastic gradient descent iteration for 2×10^5 steps with a fixed step size of 10^{-4} and a gradient estimate based on random samples of $B = 20$ points, starting at an initial guess of $c = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$. Plot $\phi(\hat{c}) - \phi(c^*)$ on a semilog plot. What do you observe?

2: Least $2p$-norm regression Consider minimizing

$$\phi(x) = \frac{1}{2p} \|r\|_{2p}^{2p} = \frac{1}{2p} \sum_{i=1}^{m} r_i^{2p} \quad \text{where } r = b - Ax$$

Rewrite this problem in forms of a nonlinear least squares problem, i.e.

$$\phi(x) = \frac{1}{2} \|f(x)\|^2.$$

What is the form of $f(x)$ and the Jacobian $J(x)$? Argue that a Gauss-Newton step solves a *weighted* least squares problem

$$\min \|D^k (Ap^k - r^k)\|^2$$

where D^k is a diagonal weight matrix, $r^k = b - Ax^k$ is the residual at step k, and p^k is the Gauss-Newton step.