HW for 2018-06-15

(due: 2018-06-22)

1: Trying SGD Code the basic stochastic gradient algorithm with a (small) fixed step size for the cricket chirp least squares fit from the June 13 lecture. Draw two plots:

- Starting from a zero initial guess, plot the optimality gap (the true smallest squared residual vs. the SGD squared residual) against the number of steps.
- For different values of α, run SGD for 10^5 steps starting from the true solution and show a histogram of the optimality gap. Also plot the mean optimality gap as a function of α; ideally, the two should be proportional.

2: Randomized scaling For $A \in \mathbb{R}^{m \times n}$, consider the iteration

$$x^{k+1} = x^k - \hat{G}^{-1} A^T (Ax^k - b).$$

Here, $\hat{G}$ is an unbiased estimator for the Gram matrix $A^T A$, defined by randomly choosing a subset of at least n indices $\mathcal{I}$:

$$\hat{G} = \frac{m}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} A_{i,:}^T A_{i,:}$$

where $A_{i,:}$ representing row i of A. Define $\rho(\hat{G}) = \|I - \alpha_{\mathrm{opt}} \hat{G}^{-1} A^T A\|$ to be the rate of convergence for gradient descent with an optimal step size. For the cricket data least squares problem, what is $\log \rho(I)$? Compare to a histogram of $\log \rho(\hat{G})$ for a few different sample sizes $|\mathcal{I}|$.

3: Gauss-Newton Code the Gauss-Newton iteration to minimize

$$\phi(x) = \sum_j \exp(r_j^2) - 1, \quad r = Ax - b,$$

and illustrate the behavior for the cricket data.