ECE 530 (Fall 2023) **Instructor:** S. Bose

Homework 4

Assigned: Sep 27 Due: Oct 4

Instructions: *You are encouraged to discuss with your peers. You need to, however, turn in your own solutions. Mention the names of your collaborators. Crediting other classmates does not take away any points from you. Any submissions made by one week after the due date will be marked on 50% of the total points. You don't need to submit after that.*

Problem 1. Proving what Newton method is up to. [4+3+3 = 10 points]

Newton method minimizes the local quadratic approximation of a function, if its Hessian at the current iterate is positive definite. We showed that Newton method indeed finds a *local* minimizer of the quadratic approximation. Here, we show that it finds the *global* minimizer through the following steps.

(a) If $\mathbf{Q}$ is any positive definite matrix, then show that

$$\frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x} + \mathbf{c}^\top \mathbf{x} = \frac{1}{2}\left(\mathbf{x} + \mathbf{Q}^{-1}\mathbf{c}\right)^\top \mathbf{Q}\left(\mathbf{x} + \mathbf{Q}^{-1}\mathbf{c}\right) - \frac{1}{2}\mathbf{c}^\top \mathbf{Q}^{-1}\mathbf{c}.$$

(b) Using the result in part (a), argue that the function $\frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x} + \mathbf{c}^\top \mathbf{x}$ is minimized *globally* at $\mathbf{x}^* = -\mathbf{Q}^{-1}\mathbf{c}$.

(c) Recall that for any function $f : \mathbb{R} \to \mathbb{R}$, its local quadratic approximation at $\mathbf{x}^k$ is given by

$$f^q(\mathbf{x}) := f(\mathbf{x}^k) + [\nabla f(\mathbf{x}^k)]^\top (\mathbf{x} - \mathbf{x}^k) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^k)^\top \mathbf{H}(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k).$$

Assume that the Hessian $\mathbf{H}(\mathbf{x}^k)$ is positive definite. Utilize the result in part (b) to conclude that the *global* minimizer of f^q is given by $\mathbf{x}^k - [\mathbf{H}(\mathbf{x}^k)]^{-1}\nabla f(\mathbf{x}^k)$. Notice that the minimizer is indeed $\mathbf{x}^{k+1}$, as defined by the Newton method.

Problem 2. Newton method needs a touch up. [10+5+5+15 = 35 points]

In Newton method, if the Hessian at the current iterate $\mathbf{H}(\mathbf{x}^k)$ is not positive definite, then $-[\mathbf{H}(\mathbf{x}^k)]^{-1}\boldsymbol{\nabla} f(\mathbf{x}^k)$ may not be a descent direction. Then, we modify the Hessian to $\mathbf{H}(\mathbf{x}^k) + \mathbf{D}^k$, where $\mathbf{D}^k$ is a *diagonal* matrix with nonnegative diagonal entries. Let us design $\mathbf{D}^k$ to ensure that $\mathbf{H}(\mathbf{x}^k) + \mathbf{D}^k$ is positive definite. The following corollary of Gershgorin's circle theorem will prove useful.

Theorem 1. *If λ is any eigenvalue of an arbitrary matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, then*

$$|\lambda - \mathbf{A}_{ii}| \le \sum_{j\neq i} |\mathbf{A}_{ij}|, \tag{1}$$

*for **some** $i = 1, \ldots, n$.*

(a) Using (1), find a sufficient condition on the entries of $\mathbf{A}$ such that all eigenvalues of $\mathbf{A}$ are positive.

(b) Use your result in (a) to find a diagonal matrix $\mathbf{D}^k$, such that all eigenvalues of $\mathbf{H}(\mathbf{x}^k) + \mathbf{D}^k$ are nonnegative, and all diagonal entries of $\mathbf{D}^k$ are nonnegative.

(c) If any eigenvalue of $\mathbf{H}(\mathbf{x}^k) + \mathbf{D}^k$ is close to zero but positive, then it is close to being singular, and its inverse is susceptible to noise. Modify your answer in (b) to ensure that all eigenvalues of $\mathbf{H}(\mathbf{x}^k) + \mathbf{D}^k$ are greater than $\frac{1}{2}$.

(d) The file `applyNewtonMethod.m` implements a basic Newton method to the function

$$f(x_1, x_2) := \cos(x_1^2 - 3x_2) + \sin(x_1^2 + x_2^2).$$

starting from $(x_1^0, x_2^0) = (1.2, 0.5)$. The program also draws the contour-plot and the surface-plot of f.

(i) Verify that the Hessian at the starting point is *not* positive definite.
Hint: The code to compute the Hessian is given in the attached file.

(ii) Does the Newton method converge? If yes, does it converge to a local minimizer of f?

(iii) Fill the missing code in `modifyHessian.m` that takes $\mathbf{H}(\mathbf{x}^k)$ as input and gives $\mathbf{H}(\mathbf{x}^k) + \mathbf{D}^k$ as output. Utilize your condition in part (c) to design $\mathbf{D}^k$. Submit your code. Using your code, compute $\mathbf{H}(\mathbf{x}^0) + \mathbf{D}^0$, i.e., the modified Hessian at the starting point.

(iv) Uncomment the relevant lines in `applyNewtonMethod.m` to run the modified Newton method. Report if the algorithm converges to a local minimizer of f.