Nonlinear Optimization I, Fall 2021
Homework 5
Due before lecture on 12/2
Please only in file formats .pdf, .zip, standard image formats (.jpg, .png, etc.).

Your submitted solutions to homeworks should be entirely your own work. Do not copy solutions from other students or any online source. You are allowed to discuss homework problems at a high-level with other students, but should carry out the execution of any thoughts/directions discussed independently, on your own. Feel free to cite any result presented in class without proof.

You can write solutions by hand or type them up (the LaTeX code for this pdf is on blackboard).

Q1. Consider running a Modified Newton's Method which computes its search direction at each iteration as

$$p_k = \operatorname{argmin}\left\{g_k^T p + \frac{1}{2}p^T B_k p\right\} = -B_k^{-1} g_k$$

where $g_k = \nabla f(x_k)$ and B_k is some positive definite matrix selected at each iteration. Assume that every B_k has eigenvalues at least $\lambda_{min} > 0$ and at most $\lambda_{max} < \infty$. In this question, you will analyze this method's convergence when using a constant stepsize $x_{k+1} = x_k + \alpha p_k$.

(a) For any function f with L-Lipschitz gradient, show that for all x_k, the following relation holds

$$f(x_{k+1}) \leq f(x_k) + \alpha g_k^T p_k + \frac{L\alpha^2}{2}\|p_k\|_2^2$$

and deduce that

$$f(x_{k+1}) \leq f(x_k) - \left(\frac{\alpha}{\lambda_{max}} - \frac{L\alpha^2}{2\lambda_{min}^2}\right)\|g_k\|^2.$$

(b) Consider the quadratic $\left(\frac{\alpha}{\lambda_{max}} - \frac{L\alpha^2}{2\lambda_{min}^2}\right)\|g_k\|^2$ in terms of α above. What value of α maximizes this amount? Is this maximum quantity positive or negative? Is the maximizing value of α positive of negative?

(c) Using this maximizing value of α as a constant stepsize choice, derive a convergence guarantee for this method of the form

$$\min_{i \leq k} \|\nabla f(x_i)\| \leq \frac{M}{\sqrt{k+1}}$$

for some constant M depending on λ_{min}, λ_{max}, L, and $f(x_0) - \min f$.

(d) What can you say about your algorithm and guarantee under the choice of $B_k = LI$?

Q2. Consider running a Quasi-Newton Method where our stepsize is selected to ensure $y_k^T s_k > 0$.

(a) Supposing B_k is positive definite, show the BFGS update produces a positive definite B_{k+1}.

(b) Now that you have guaranteed the update from BFGS is invertible (by showing it is positive definite above), calculate its inverse using the Woodbury matrix identity (a mild generalization of the Sherman-Morrison formula). How does this relate to the DFP update?

Q3. Consider the two-dimensional Rosenbrock function (a common example used to show the steepest descent method slowly converges):

$$\min_{x \in \mathbb{R}^2} (1 - x_1)^2 + 100(x_2 - x_1^2)^2.$$

Note this problem globally minimizes at $(1, 1)$.

(a) Implement and run gradient descent on this problem initialized at $(0, 0)$ for 100 iterations using an exact linesearch[1]. Print out the current point x at each iteration.

(b) Implement and run Newton's Method on this problem for 100 iterations initialized at $(0, 0)$. Print out the current point x at each iteration.

(c) Implement and run the BFGS Quasi-Newton Method for 100 iterations initialized at $(0, 0)$ using an exact linesearch. Print out the current point x at each iteration.

General Guidelines for Programming HW Problems: You can do programming assignments in any programming language you feel comfortable with (python, matlab, java, c/c++, haskell, etc). Programming questions will ask for you to solve a particular problem or describe particular settings to run an algorithm under. You must submit both your code and the requested output/plots from running your code. Grading will focus primarily on the quality of these outputs rather than of your code.

[1] That is, given a direction to search p_k, select α_k minimizing $f(x_k - \alpha_k p_k)$. Here this corresponds to minimizing a single variable, degree four polynomial.