Assignment #4

Due Monday 10 October at the start of class

Please read *everything* in Chapter 3 in the textbook (Nocedal & Wright). Do the following Exercises and Problems.

> In fact the homework and Midterm Exam[1] problems will *not* require you to know the following specific material:
> - strong Wolfe conditions (equations (3.7) page 34)
> - Goldstein conditions (page 36)
> - "line search algorithm for the Wolfe conditions" (pages 60–62)

Exercise 3.2

Exercise 3.3

Exercise 3.6

Exercise 3.13

Problem P7. Suppose $A \in \mathbb{R}^{n\times n}$ is symmetric and positive semi-definite with eigenvalues $\lambda_1 \ge \lambda_2 \ge \dots \ge \lambda_n \ge 0$. Show that if $x \in \mathbb{R}^n$ then

$$x^\top A x \ge \lambda_n x^\top x.$$

(*Hint*. You may use the fact that, because A is symmetric, any vector can be expanded in the eigenvectors of A—i.e. the eigenvectors form a basis. You may use the fact that, because A is symmetric, eigenvectors for distinct eigenvalues are orthogonal.)

[1] Happens on October 24.

Problem P8. (*This problem replaces and clarifies Exercise 3.5.*)

In this problem, $\|A\|$ denotes the *matrix 2-norm*.[2] It is defined and discussed in Appendix A.1—see particularly formulas (A.7) and (A.8b)—but this problem restates the definition and basic properties. In this problem we use the Euclidean norm (2-norm) for vectors, so that if $x \in \mathbb{R}^n$ then $\|x\| = \sqrt{x^\top x} = (\sum_i x_i^2)^{1/2}$.

Suppose $A \in \mathbb{R}^{n\times n}$ is a square matrix. We define

$$\|A\| = \max_{\substack{x\in\mathbb{R}^n \\ x\neq 0}} \frac{\|Ax\|}{\|x\|}.$$

It is *not* trivial to compute $\|A\|$, but it is always true that

$$\|A\| = \left(\text{largest eigenvalue of } A^\top A\right)^{1/2}.$$

(*Because $A^\top A$ is positive semi-definite, its eigenvalues are nonnegative.*) If also A is symmetric then

$$\|A\| = \max_{\substack{\lambda \text{ is an} \\ \text{eigenvalue of } A}} |\lambda|.$$

If A itself is symmetric and positive semi-definite then $\|A\| = \max \lambda$.

Now for the exercise itself.

(a) Show that if $A \in \mathbb{R}^{n\times n}$ is any matrix then $\|Ax\| \le \|A\|\|x\|$ for all $x \in \mathbb{R}^n$.

(b) For an invertible[3] matrix A, let

$$\kappa(A) = \operatorname{cond}(A) = \|A\|\|A^{-1}\|.$$

Show that if A is symmetric and positive definite with eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n > 0$ then

$$\kappa(A) = \frac{\lambda_1}{\lambda_n}.$$

Give a geometric interpretation to this ratio.

(c) Suppose we do a quasi-Newton step, namely $p_k = -B_k^{-1}\nabla f(x_k)$, for some B_k which is symmetric and positive-definite. As in (3.12), define

$$\cos\theta_k = \frac{-\nabla f(x_k)^\top p_k}{\|\nabla f(x_k)\|\|p_k\|}.$$

Show that

$$\cos\theta_k \ge \frac{1}{\kappa(B_k)}.$$

(*Hint*. This is the main part of the problem. You will use **P7** and parts **(a)** and **(b)**.)

(d) Show (3.19) and (3.20).

[2] Likewise true everywhere in the textbook unless otherwise stated.

[3] By tradition one defines $\kappa(A) = +\infty$ if A is not invertible.

Problem P9. (*This problem replaces, clarifies, and simplifies Exercise 3.13.*)
The BFGS algorithm is described in section 2.2 on page 24. See especially formulas (2.17) and (2.19). The algorithm:

chose x_0 and B_0 — *B_0 should be positive-definite*

for $k = 0, 1, 2, \dots$

$p_k = -B_k^{-1}\nabla f(x_k)$ — *usual quasi-Newton search vector (3.34)*

$\alpha_k =$ (result from a line search)

$s_k = \alpha_k p_k$ — *the step itself*

$x_{k+1} = x_k + s_k$ — *take step*

if $\|\nabla f(x_{k+1})\| \le \texttt{tol}$ — *absolute tolerance criterion ... minimal*

break

end

$y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ — *right side of this goal: $\nabla^2 f(x_{k+1}) s_k \approx y_k$*

$z_k = B_k s_k$ — *efficient to get this vector first*

$$B_{k+1} = B_k - \frac{z_k z_k^\top}{s_k^\top z_k} + \frac{y_k y_k^\top}{y_k^\top s_k}$$ — *so that "secant equation" $B_{k+1} s_k = y_k$ is true*

end

Implement this algorithm, using $B_0 = I$ and the usual back-tracking line search.[4] Apply to the Rosenbrock function[5] using the two initial iterates x_0 stated in Exercise 3.1. Compare the performance to that of Newton's method; refer to the Assignment #3 solutions for results from Newton.

[4]Online at `bueler.github.io/M661F16/matlab/bt.m`.

[5]Also online at `bueler.github.io/M661F16/matlab/rosenbrock.m`.