Assignment IV

MEMAD

September 6, 2025

Abstract

The goals are to revise some relevant aspects of the quasi-Newton, in particular, BFGS and Gauss-Newton methods, as well as linear and nonlinear least-squares. All the problems are worth 1 point, except the ninth. Please upload your solutions in one compressed file to Classroom before September 16th.

Problem 1

Let's define

$$s_k = x_{k+1} - x_k = \alpha_k p_k, \qquad y_k = \nabla f_{k+1} - \nabla f_k.$$

Show that if α_k and p_k satisfy the Wolfe conditions, then the following inequality (curvature condition) holds:

$$s_k^T y_k > 0. \tag{1}$$

Problem 2

Consider the secant equation

$$B_{k+1} s_k = y_k, \tag{2}$$

where it is assumed that $B_k > 0$ and $B_k = B_k^T$. Remember that if (1) holds, then (2) always has a solution. In fact, that system has an infinite number of solutions since the degrees of freedom are the entries of B_k. How many degrees of freedom does (2) have? How many imposed conditions (equations) does (2) have?

Problem 3

Compute the Frobenius norm of the following matrices:

(a)

$$A = \begin{pmatrix} 16 & -8 & -4 \\ -8 & 29 & 12 \\ -4 & 12 & 41 \end{pmatrix}.$$

(b)

$$A = \begin{pmatrix} 9 & 0 & -8 \\ 6 & -5 & -2 \\ -9 & 3 & 3 \end{pmatrix}.$$

(c)

$$A = \begin{pmatrix} -9 & 0 & -8 \\ 6 & -5 & -2 \\ -9 & 3 & 3 \end{pmatrix}.$$

Problem 4

Consider the Himmelblau's function given by:

$$f(x, y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2. \tag{3}$$

Consider $x_k = (1, 1)$.

(a) Compute p_k as the steepest descent decrease direction.

(b) Compute an α_k which satisfies the Wolfe conditions for x_k and p_k.

(c) Compute the average Hessian of (3).

Important

For problems 5 and 6 consider the following functions $f : \mathbb{R}^n \to \mathbb{R}$:

- Translated sphere's function:

$$f(\mathbf{x}) = \sum_{i=1}^{n} (x_i - c_i)^2, \quad \text{for a given (fixed) } \mathbf{c} \in \mathbb{R}^n. \tag{4}$$

  You can let it be $\mathbf{c} = (1, 1, 1 \cdots, 1)$ for instance.

- Rosenbrock's function:

$$f(\mathbf{x}) = \sum_{i=1}^{n-1} \left[100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2 \right]. \tag{5}$$

- Perm's function n, β:

$$f(\mathbf{x}) = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} (j^i + \beta) \left(\left(\frac{x_j}{j} \right)^i - 1 \right) \right)^2, \quad \text{for a given (fixed) } \beta \in \mathbb{R}. \tag{6}$$

  You can let β be 1 for instance.

Also, as the started point consider $\mathbf{x}_0 = (0.5, 0.5, \cdots, 0.5)$. Furthermore, you can assume $n = 5$.

Problem 5

Make a Python script implementing the BFGS algorithm for functions (4)-(6). Use either Wolfe or strong Wolfe conditions to compute suitable length steps. Use finite differences to approximate the analytical gradients of the functions. Show plots of the iterations versus the function value to show how the latter decreases as the former increases.

Problem 6

Consider the function (5). Solve the associated optimization problem using the Newton, SD and BFGS methods. For all of those algorithms, you can use numerical or analytical gradient. Moreover, for both SD and BFGS consider the Wolfe conditions or the strong Wolfe conditions. Show the function values for the same number of iterations using a table or a figure. The idea is to show the comparison of linear, superlinear and quadratic convergence rates.

Problem 7

Consider the "Linear" dataset $D = \{X, y\}$ provided with this assignment. Consider a model with the following form

$$h(X;\theta) = \sum_{j=0}^{N} \theta_j x_j^i. \tag{7}$$

Find the optimal parameters of (7) considering $N = 1$ using the normal equations. Make a plot of the model and data. What's the error you get with the parameters you've found?

Problem 8

Consider the "Polynomial" dataset $D = \{X, y\}$ provided with this assignment. Consider a model of the form (7). Using the normal equations, find the optimal parameters of the model by trying several values of N. Make some plots of the model and data to support your estimates of N. What's the error you get with the parameters you've found for each N value you tried? What is the optimal value of N according to your analysis?

Problem 9 (3 points)

Consider the "Lennard-Jones energy levels" dataset provided with this assignment. Given N particles, $N \geq 2$, the Lennard Jones potential energy is given by

$$E_N = 4\sum_{i<j}^{N} \left[\left(\frac{1}{r_{ij}}\right)^{12} - \left(\frac{1}{r_{ij}}\right)^{6} \right], \tag{8}$$

where r_{ij} denotes the Euclidean distance between particles X_i and X_j. For several applications, it is of great interest to find the particular spatial configuration of N particles that minimizes the energy (8). Then:

a) Pick a particular value for N greater than 2.

b) Write down the expression for the residual vector $r(X)$.

c) Using the provided data, implement the Gauss-Newton method to solve the nonlinear least squares problem

d) Make a plot of the values of (8) along iterations and compare your final estimation with those reported in Table 1 of [1].

e) Make a 3D plot of the final particle configuration (i.e., the points X).

f) What was the largest value for N that you could solve for?

References

[1] D.J. Wales, J.P.K. Doye, **Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms**, Abstract published in AdVance ACS Abstracts, June 15, 1997.