Homework 10. Due Wednesday, Nov. 6.

Reference: [NW] J. Nocedal and S. Wright, "Numerical Optimization", Second Edition.

1. **(5 pts)** Suppose that a smooth function $f(x)$ is approximated by a quadratic model in the neighborhood of a current iterate x:

$$m(p) = f(x) + \nabla f(x)^\top p + \frac{1}{2} p^\top B p,$$

where B is a symmetric positive definite matrix. Show that then the direction p found by setting the gradient of $m(p)$ to zero is a descent direction for $f(x)$, i.e.,

$$\cos\theta := -\frac{\nabla f(x)^\top p}{\|\nabla f(x)\| \|p\|} > 0.$$

Also, bound $\cos\theta$ away from zero in terms of the condition number of B, i.e., $\kappa(B) = \|B\| \|B^{-1}\|$.

2. **(5 pts)** Let $f(x)$, $x \in \mathbb{R}^n$, be a smooth arbitrary function. The BFGS method is a quasi-Newton method with the Hessian approximate built recursively by

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}, \quad \text{where } s_k := x_{k+1} - x_k\,, \;\; y_k := \nabla f_{k+1} - \nabla f_k\,.$$

Let x_0 be the starting point and let the initial approximation for the Hessian is the identity matrix.

   (a) Let p_k be a descent direction. Show that Wolfe's condition 2,

$$\nabla f_{k+1}^\top p_k \ge c_2 \nabla f_k^\top p_k, \quad c2 \in (0,1)$$

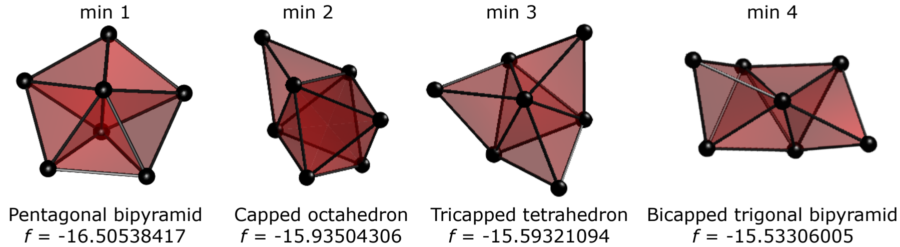
   implies that $y_k^\top s_k > 0$.

   (b) Let B_k be symmetric positive definite (SPD). Prove that then B_{k+1} is also SPD, i.e., for any $z \in \mathbb{R}^n \backslash \{0\}$, $z^\top B_{k+1} z > 0$. You can use the previous item of this problem and the Cauchy-Schwarz inequality for the B_k-inner product $(u, v)_{B_k} := v^\top B_k u$.

3. **(5 pts)** The goal of this problem is to code, test, and compare various optimization techniques on the problem of finding local minima of the potential energy function of the cluster of 7 atoms interacting according to the Lennard-Jones pair potential (for brevity, this cluster is denoted by LJ$_7$):

$$f = 4 \sum_{i=2}^{7} \sum_{j=1}^{i} \left(r_{ij}^{-12} - r_{ij}^{-6} \right), \quad r_{ij} := \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}. \quad (1)$$

It is known that LJ_7 has four local energy minima:



Add the BFGS search directions to the provided Matlab or Python codes. It is recommended to reset the matrix B_k in the BFGS method to the identity of every mth step. Try $m = 5$ and $m = 20$.

Compare the performance of the three algorithms, the steepest descent, Newton's (already encoded), and BFGS in terms of the number of iterations required to achieve convergence and by plotting the graph of f and $\|\nabla f\|$ against the iteration number for each test case. Do it for each of the four initial conditions approximating the four local minima and ten random initial conditions.

4. **(5 pts)** (Approx. Problem 3.1 from [NW])

   (a) Compute the gradient and the Hessian of the Rosenbrock function

   $$f(x, y) = 100(y - x^2)^2 + (1 - x)^2. \tag{2}$$

   Show that $(1, 1)$ is the only local minimizer, and that the Hessian is positive definite at it.

   (b) Program the steepest descent, Newton's, and BFGS algorithms using the backtracking line search. Use them to minimize the Rosenbrock function (2). First start with the initial guess $(1.2, 1.2)$ and then with the more difficult one $(-1.2, 1)$. Set the initial step length $\alpha_0 = 1$ and plot the step length α_k versus k for each of the methods.

   Plot the level sets of the Rosenbrock function using the command `contour` and plot the iterations for each method over it.

   Plot $\|(x_k, y_k) - (x^*, y^*)\|$ versus k in the logarithmic scale along the y-axis for each method. Do you observe a superlinear convergence? Compare the performance of the methods.