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

Solution: Taking the gradient of m, we find
$$\nabla m(p) = \nabla f(x) + Bp.$$
Hence, setting this gradient equal to 0 and solving for p yields the direction
$$p = -B^{-1}\nabla f(x)$$
and so
$$\cos\theta := -\frac{\nabla f(x)^\top p}{\|\nabla f(x)\|\|p\|} = \frac{\nabla f(x)^\top B^{-1}\nabla f(x)}{\|\nabla f(x)\|\|p\|}.$$
But now note that since B is symmetric positive definite, so is B^{-1}, and hence
$$\nabla f(x)^\top B^{-1}\nabla f(x) > 0$$
assuming $\nabla f(x) \neq 0$. As norms, we also have that $\|\nabla f(x)\|, \|p\| > 0$ and hence
$$\cos\theta = \frac{\nabla f(x)^\top B^{-1}\nabla f(x)}{\|\nabla f(x)\|\|p\|} > 0$$
as desired.

TODO

2. 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
$$B_{k+1} = B_k - \frac{B_k s_k s_k^\top B_k}{s_k^\top B_k s_k} + \frac{y_k y_k^\top}{y_k^\top s_k}, \text{ where } s_k := x_{k+1} - x_k \text{ and } y_k := \nabla f_{k+1} - \nabla f_k.$$
Let x_0 be the starting point and let the initial approximation for the Hessian be the identity matrix.

   (a) Let p_k be a descent direction. Show that Wolfe's condition 2,
$$\nabla f_{k+1}^\top p_k \geq c_2 \nabla f_k^\top p_k, \quad c_2 \in (0,1)$$
   implies that $y_k^\top s_k > 0$.

Solution: Recalling the definitions of s_k and y_k,

$$s_k := x_{k+1} - x_k \text{ and } y_k := \nabla f_{k+1} - \nabla f_k.$$

In particular, since $x_{k+1} = x_k + \alpha_k p_k$, we can write $s_k = \alpha_k p_k$ for some $\alpha_k > 0$. Therefore, we have that

$$\begin{aligned} y_k^\top s_k = \alpha_k y_k^\top p_k = \alpha_k(\nabla f_{k+1} - \nabla f_k)^\top p_k = \alpha_k(\nabla f_{k+1}^\top - \nabla f_k^\top)p_k \\ = \alpha_k(\nabla f_{k+1}^\top p_k - \nabla f_k^\top p_k). \end{aligned}$$

Further, from Wolfe's condition 2, there exists some $c_2 \in (0,1)$ such that

$$\nabla f_{k+1}^\top p_k \geq c_2 \nabla f_k^\top p_k$$

and hence

$$y_k^\top s_k = \alpha_k(\nabla f_{k+1}^\top p_k - \nabla f_k^\top p_k) \geq \alpha_k(c_2 \nabla f_k^\top p_k - \nabla f_k^\top p_k) = \alpha_k(c_2 - 1)\nabla f_k^\top p_k.$$

Finally, since p_k is a descent direction, $\nabla f_k^\top p_k < 0$ and $(c_2 - 1) < 0$ since $c_2 \in (0,1)$, therefore $(c_2 - 1)\nabla f_k^\top p_k > 0$ and so

$$y_k^\top s_k \geq \alpha_k(c_2 - 1)\nabla f_k^\top p_k > 0$$

as desired.

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

Solution: The Cauchy-Schwarz inequality for the B_k-inner product asserts that

$$v^\top B_k u u^\top B_k v = (u,v)_{B_k}(v,u)_{B_k} = (u,v)_{B_k}^2 \leq (u,u)_{B_k}(v,v)_{B_k} = u^\top B_k u v^\top B_k v.$$

Now let $z \in \mathbb{R}^n \setminus \{0\}$ and observe that by the definition of B_{k+1},

$$z^\top B_{k+1} z = z^\top B_k z - \frac{z^\top B_k s_k s_k^\top B_k z}{s_k^\top B_k s_k} + \frac{z^\top y_k y_k^\top z}{y_k^\top s_k}.$$

Then taking $v = z$ and $u = s_k$ in the Cauchy-Schwarz inequality, we see that

$$z^\top B_k s_k s_k^\top B_k z \leq s_k^\top B_k s_k z^\top B_k z$$

and hence

$$\begin{aligned} z^\top B_{k+1} z \geq z^\top B_k z - \frac{s_k^\top B_k s_k z^\top B_k z}{s_k^\top B_k s_k} + \frac{z^\top y_k y_k^\top z}{y_k^\top s_k} \\ = z^\top B_k z - z^\top B_k z + \frac{z^\top y_k y_k^\top z}{y_k^\top s_k} = \frac{z^\top y_k y_k^\top z}{y_k^\top s_k} = \frac{(y_k^\top z)^2}{y_k^\top s_k}. \end{aligned}$$

From item (a), we know that $y_k^\top s_k > 0$ and hence since $(y_k^\top z)^2 \geq 0$,

$$z^\top B_{k+1} z \geq \frac{(y_k^\top z)^2}{y_k^\top s_k} \geq 0.$$

It remains to show that one of the two inequalities above must be strict; if z and s_k are linearly independent, then the Cauchy-Schwarz inequality is strict, so $z^\top B_{k+1} z > 0$. Otherwise, we can write $z = cs_k$ with $c \neq 0$ since $z \neq 0$, and so

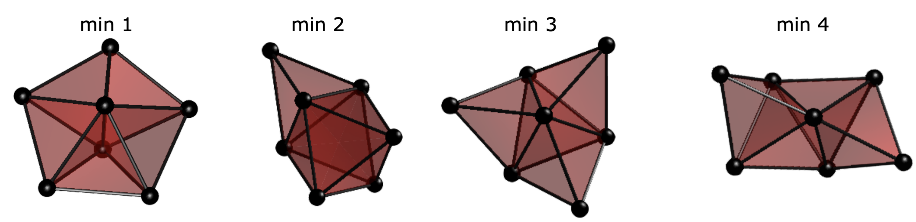
$$z^\top B_{k+1} z \geq \frac{(y_k^\top z)^2}{y_k^\top s_k} = \frac{(y_k^\top c s_k)^2}{y_k^\top s_k} = c^2 y_k^\top s_k > 0.$$

In either case, we see that $z^\top B_{k+1} z > 0$, so $z^\top B_{k+1} z > 0$ for all $z \in \mathbb{R}^n \setminus \{0\}$, meaning B_{k+1} is also symmetric positive definite.

3. 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$_7$):

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

It is known that LJ$_7$ has four local energy minima:



Pentagonal bipyramid f = -16.50538417 | Capped octahedron f = -15.93504306 | Tricapped tetrahedron f = -15.59321094 | Bicapped trigonal bipyramid f = -15.53306005

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 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.

Solution: TODO

4. (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.

Solution: The first-order partial derivatives of the Rosenbrock function are

$$\frac{\partial f}{\partial x} = -400x(y - x^2) - 2(1 - x) \text{ and } \frac{\partial f}{\partial y} = 200(y - x^2),$$

hence its gradient is given by

$$\nabla f(x,y) = \begin{bmatrix} -400x(y - x^2) - 2(1 - x) \\ 200(y - x^2) \end{bmatrix}.$$

The second-order partial derivatives of the Rosenbrock function are

$$\frac{\partial^2 f}{\partial x^2} = -400(y - x^2) + 800x^2 + 2 \text{ and } \frac{\partial^2 f}{\partial x \partial y} = -400x$$

and

$$\frac{\partial^2 f}{\partial y \partial x} = -400x \text{ and } \frac{\partial^2 f}{\partial y^2} = 200$$

so its Hessian is

$$\begin{bmatrix} -400(y - x^2) + 800x^2 + 2 & -400x \\ -400x & 200 \end{bmatrix}.$$

TODO

(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.

Solution: TODO