Whenever calculations are needed to solve a problem, those calculations must be submitted as part of the homework assignment.

**Exercise 3.1.** Given a twice-continuously differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, assume that the point $\bar{x}$ is not a stationary point of $f$. Show that any descent direction $p$ at $\bar{x}$ is the unique minimizer of a quadratic function whose Hessian is

$$B = I - \frac{1}{p^T p} pp^T - \frac{1}{\bar{g}^T p} \bar{g} \bar{g}^T,$$

where $\bar{g}$ is the gradient of $f$ at $\bar{x}$.

**Exercise 3.2.** Consider a quasi-Newton method in which the search direction $p_k$ is defined by

$$p_k = -M_k g_k,$$

where $g_k$ is the gradient of $f$ at $x_k$ and the symmetric matrix $M_k$ represents a quasi-Newton approximation to the inverse Hessian of a nonlinear function $f(x)$. Assume that the updated matrix satisfies $M_{k+1} y_k = s_k$, where $y_k = g_{k+1} - g_k$ and $s_k = x_{k+1} - x_k$, and that $M_{k+1}$ is updated using the DFP formula,

$$M_{k+1} = M_k - \frac{1}{y_k^T M_k y_k} M_k y_k y_k^T M_k + \frac{1}{s_k^T y_k} s_k s_k^T.$$

Show that $M_{k+1}$ is singular if $M_k$ is singular.

**Exercise 3.3.** Write code that implements a specialized BFGS quasi-Newton method for minimizing a quadratic with positive definite Hessian $H$, where $x_{k+1} = x_k + \alpha_k p_k$, with two options for computing the step $\alpha_k$:

(i) Version 1: $\alpha_k$ is taken as the exact step to the minimizer along $p_k$, which is given for a quadratic by

$$\alpha_k = -\frac{g_k^T p_k}{p_k^T H p_k};$$

(ii) Version 2: $\alpha_k$ is chosen to satisfy the Armijo condition with $\eta_s = 0.001$, using a backtracking line search with $\gamma_c = \frac{1}{2}$, starting each line search with the unit step.

To keep things simple:
Perform the BFGS updates on a sequence of approximations \( \{B_k\} \) to the Hessian itself (not approximations to the inverse Hessian). If \( y_k^T s_k \leq 0 \), skip the update at iteration \( k \).

Solve \( B_k p_k = -g_k \) for \( p_k \) from scratch at each iteration. When \( y_k^T s_k > 0 \), generate \( B_{k+1} \) explicitly by adding the associated rank-two change to \( B_k \).

Start with \( B_0 = I \) and stop when \( \|g_k\| \) is less than \( \text{ftol} \) or else after \( \text{maxit} \) iterations. Print \( x^* \) and \( f(x^*) \). At each iteration, print \( k, x_k, f_k, \|g_k\|, \alpha_k, B_k \), and a message if the update was skipped.

Consider the four-variable quadratic function

\[
q(x) = c^T x + \frac{1}{2} x^T H x,
\]

with

\[
c = \begin{pmatrix} 2 \\ -1 \\ 2 \\ -1 \end{pmatrix} \quad \text{and} \quad H = \text{diag}(5, 1, 10^{-2}, 10^{-4}).
\]

(a) Use Version 1, with \( \text{ftol} = 1.0 \times 10^{-8} \) and \( \text{maxit} = 10 \), starting at \( x_0 = (-1, 0, 1, 1)^T \). How many iterations are executed before the program terminates? What is \( \|g_k\| \) at the last iteration? Does \( B_k \) converge to the exact Hessian?

(b) Use Version 1, with \( \text{ftol} = 1.0 \times 10^{-8} \) and \( \text{maxit} = 10 \), starting at \( x_0 = (-0.4, 0, 1, 1)^T \). How many iterations are executed before the program terminates? Does \( B_k \) converge to the exact Hessian? If not, please explain how this can happen, given Theorem 13.3 (in Notes 5).

(c) Use Version 2, with \( \text{ftol} = 1.0 \times 10^{-8} \) and \( \text{maxit} = 30 \), starting at \( x_0 = (-1, 0, 1, 1)^T \). How many iterations are executed before the program terminates? Please comment on (i) the values of \( \alpha_k \) and (ii) the rate of convergence of \( \{\|g_k\|\} \) to zero during the final few iterations. Does \( B_k \) seem to be converging to the exact Hessian?

(d) Use Version 2, with \( \text{ftol} = 1.0 \times 10^{-9} \) (i.e., \( \text{ftol} \) is smaller than in (c)) and \( \text{maxit} = 30 \), starting at \( x_0 = (-1, 0, 1, 1)^T \). How many iterations are executed before the program terminates? Please comment on (i) the values of \( \alpha_k \) during the final few iterations and (ii) the rate of convergence of \( \{\|g_k\|\} \) to zero during the final few iterations. Describe and explain any differences from the results of part (d).

Exercise 3.4. Let \( B \) be a symmetric positive-definite matrix, and let \( s \) be a nonzero \( n \)-vector. Prove that the matrix

\[
B_+ = B - \frac{1}{s^T B s} B s s^T B
\]

has \( n - 1 \) positive eigenvalues and one zero eigenvalue. Give a nonzero vector \( v \) for which \( B_+ v = 0 \).

Exercise 3.5. Consider the following local quadratic model of a twice-continuously differentiable objective function \( f(x) \) at the point \( x_k \):

\[
q_k(x) = f_k + g_k^T (x - x_k) + \frac{1}{2} (x - x_k)^T B_k (x - x_k),
\]

Let \( B \) be a symmetric positive-definite matrix, and let \( s \) be a nonzero \( n \)-vector. Prove that the matrix

\[
B_+ = B - \frac{1}{s^T B s} B s s^T B
\]

has \( n - 1 \) positive eigenvalues and one zero eigenvalue. Give a nonzero vector \( v \) for which \( B_+ v = 0 \).
where \( f_k = f(x_k), \) \( g_k = g(x_k), \) and \( B_k \) is a symmetric approximation to the exact Hessian \( H(x_k). \) When we move from \( x_k \) to the new point \( x_k + d, \) the change in the quadratic model is

\[
\Delta_k(d) = q_k(x_k + d) - q_k(x_k) = g_k^T d + \frac{1}{2} d^T B_k d.
\]

Assuming that \( d = \alpha p_k, \) where the search direction \( p_k \) satisfies \( B_k p_k = -g_k, \) show that \( \Delta_k(\alpha p_k) < 0 \) if \( B_k \) is positive definite and \( 0 < \alpha < 2. \)

**Exercise 3.6.** When minimizing \( f(x) \) subject to the \( m \) equality constraints \( c(x) = 0, \) where \( m < n \) and \( c(x) = (c_1(x), c_2(x), \ldots, c_m(x))^T, \) let \( x^* \) be a feasible point, let \( J^* \) denote the \( m \times n \) Jacobian of \( c \) at \( x^*, \) and let \( Z^* \) denote a basis for the null space of \( J^*. \) The associated Lagrangian function is defined as \( L(x, \lambda) = f(x) - \lambda^T c(x), \) where \( \lambda \) is an \( m \)-vector of Lagrange multipliers. The Hessian of the Lagrangian function with respect to \( x \) is denoted by \( W(x, \lambda) = H(x) - \sum_{i=1}^m \lambda_i H_i(x), \) where \( H(x) = \nabla^2_{xx} f \) and \( H_i = \nabla^2_{xx} c_i. \)

Three background results are needed for this problem: the first-order KKT conditions (Definition 19.5 in Notes 7), the second-order necessary conditions (Theorem 19.7 in Notes 7), and the sufficient conditions (Theorem 19.8 in Notes 7).

Consider the following problem with two variables and a single nonlinear equality constraint:

\[
\begin{align*}
\text{minimize} \quad f(x) &= (x_1 - 1)^2 + x_2^2 \\
\text{subject to} \quad c(x) &= -x_1 + \frac{x_2^2}{\beta} = 0,
\end{align*}
\]

where \( \beta \) is a positive constant.

(a) In each of the two cases \( \beta = 1 \) and \( \beta = 4, \) explain whether or not \( \bar{x} = (0, 0)^T \) satisfies the (a) first-order KKT conditions, (b) second-order necessary conditions, and (c) sufficient conditions for optimality.

(b) Find the smallest positive value \( \bar{\beta} \) such that for all \( \beta > \bar{\beta}, \) the point \( \bar{x} = (0, 0)^T \) satisfies the sufficient conditions for a strict local solution of (3.1).

**Exercise 3.7.** Assume that \( \rho_k \) and \( \rho_{k+1} \) are positive penalty parameters such that \( \rho_{k+1} > \rho_k > 0. \) Consider the quadratic penalty function

\[
P_Q(x, \rho_k) = f(x) + \frac{1}{2} \rho_k c(x)^T c(x),
\]

where \( f(x) \) and \( c_i(x), i = 1, 2, \ldots, m, \) are smooth nonlinear functions. Assume that \( P_Q(x, \rho_k) \) and \( P_Q(x, \rho_{k+1}) \) have unique unconstrained minimizers \( x_k \) and \( x_{k+1} \) respectively. Show that

(a) \( f(x_{k+1}) \geq f(x_k); \) and

(b) \( ||c(x_{k+1})||^2 \leq ||c(x_k)||^2. \)

**Exercise 3.8.** This problem involves trying (simplified versions of!) two approaches for solving a 2-variable nonlinear equality-constrained optimization problem of minimizing \( f(x) \) subject to \( c(x) = 0, \) where \( c \) is a scalar-valued function.

Approach 1 involves using a pure Newton method to perform unconstrained minimization of the quadratic penalty function \( P_Q(x, \rho) = f(x) + \frac{1}{2} \rho \|c(x)\|^2, \) so that the \( k \)th search direction \( p_k \) for a given value of \( \rho \) satisfies the Newton equations

\[
\nabla^2 P_Q(x_k, \rho) p_k = -\nabla P_Q(x_k, \rho).
\]
At $x^*(\rho)$, an unconstrained minimizer of $P_Q(x, \rho)$, the vector $\lambda(\rho) = -\rho c(x^*(\rho))$ is an estimate of the Lagrange multiplier vector for the original problem (if Lagrange multipliers exist).

Approach 2, sometimes called the Newton-Lagrange method, is applicable when the problem satisfies a constraint qualification (so that Lagrange multipliers exist). The strategy is to use Newton’s method for multidimensional zero-finding to find $(x, \lambda)$ such that the $(n + m)$-dimensional nonlinear function $F(x, \lambda)$, the gradient of the Lagrangian function with respect to $x$ and $\lambda$, is equal to zero:

$$F(x, \lambda) \triangleq \begin{pmatrix} g(x) - J(x)^T \lambda \\ c(x) \end{pmatrix}.$$ 

The associated Newton equations are

$$\begin{pmatrix} W(x_k, \lambda_k) & -J(x_k)^T \\ J(x_k) & 0 \end{pmatrix} \begin{pmatrix} p_k \\ \delta_k \end{pmatrix} = -F(x_k, \lambda_k),$$

where $W(x_k, \lambda_k)$ is the Hessian of the Lagrangian function at $(x_k, \lambda_k)$, $p_k$ is the step in $x$, and $\delta_k$ is the step in $\lambda$. Note that there is no penalty parameter, but that the Lagrange multiplier is treated as an independent variable.

In this exercise, you are asked to apply Approach 1 for four values of the penalty parameter ($\rho = 1, 10, 100, 1000$), starting at a specified point $x_0$ for the smallest $\rho$. After optimizing the penalty function for the first penalty parameter, the Newton iterations to minimize the penalty function for the next value of $\rho$ should begin at the last iterate $x$ for the previous $\rho$, as in classical penalty function methods. Your program should terminate the Newton iterations for each value of $\rho$ after maxit iterations or when $\|\nabla P_Q\|$ is less than ftol.

If you use the starting points given below, a pure Newton method should work well in both Approaches 1 and 2 (i.e., the step $\alpha$ can safely be taken as 1). Hence you may wish to adapt programs used in earlier homework.

The problem functions are

$$f(x) = x_1^3 - x_1 x_2 \quad \text{and} \quad c(x) = \frac{5}{2} x_1^2 + \frac{1}{4} x_2^2 - \frac{7}{2} = 0.$$  \hspace{1cm} (3.2)$$

The contours of $f$ are shown in the figure, along with the blue boundary of the ellipse where the constraint is satisfied.

(a) Verify numerically that $x^* = (-1, -2)^T$ satisfies the sufficient optimality conditions for problem (3.2). What is the optimal Lagrange multiplier $\lambda^*$ and how did you obtain it?

(b) Show that $\bar{x} = (1, -2)^T$ is a first-order KKT point for this problem. Explain how you know that $\bar{x}$ is not optimal.

(c) Apply Approach 1 with maxit = 15 and ftol = 1.0e-05. Take the initial point as $x_0 = (-1.2, -1.9)^T$ for $\rho = 1$. At the $k$th iteration of Newton’s method for each value of $\rho$, print $k$, $x$, $f$, $c$, and $\|\nabla P_Q\|$, using scientific notation for the latter values and showing at least 6 significant digits. For each value of $\rho$, after your Newton iterations have terminated, print $\|x^*(\rho) - x^*\|$ and $\|\lambda(\rho) - \lambda^*\|$. Comment on how these differences seem to be related to the value of $\rho$. 
(d) Letting $\text{maxit} = 8$ and $\text{ftol} = 1.0\text{e-05}$, apply Approach 2, with $x_0 = (-1.2, -1.9)^T$ and $\lambda_0 = 0$. Comment on the behavior of the iterates. Do they appear to be converging quadratically to $(x^*, \lambda^*)$? Explain.

(e) Letting $\text{maxit} = 12$ and $\text{ftol} = 1.0\text{e-05}$, repeat (c) with $x_0 = (1.2, -1.8)^T$. How do the results differ from those of (c)?

(f) Letting $\text{maxit} = 8$ and $\text{ftol} = 1.0\text{e-05}$, repeat (d) with $x_0 = (1.2, -1.8)^T$ and $\lambda_0 = 0$. How do the results differ from those of (d)? Explain the behavior of the Newton-Lagrange method by considering the figure.

**Exercise 3.9.** Augmented Lagrangian methods rely on the fact that there is a finite penalty parameter $\bar{\rho}$ such that the Hessian of the augmented Lagrangian,

$$\nabla^2 L_A(x, \lambda, \rho) = H(x) - \sum_{i=1}^{m} \lambda_i H_i(x) + \rho J(x)^T J(x),$$

is positive definite for $\rho > \bar{\rho}$.

Consider the eigenvalues of the Hessian of the augmented Lagrangian function for the objective function and constraint of (3.2), evaluated at the optimal $x$ and $\lambda$. Show by computation that, when $\rho = 0$, then $\nabla^2 L_A(x^*, \lambda^*, \rho)$ is not positive definite. Find a value of $\rho$ for which $\nabla^2 L_A(x^*, \lambda^*, \rho)$ is positive definite. If you wish (i.e., this is interesting but optional), use a 1-d zero-finding method (suggestion: bisection) to find $\bar{\rho}$ such that $\nabla^2 L_A(x^*, \lambda^*, \bar{\rho})$ is positive semidefinite and singular.