Problem 1
Program the conjugate gradient algorithm. Use your program to minimize the function
\[ f(x) := \frac{1}{2} x^T Ax - b^T x \]
where \( A \in \mathbb{R}^{n \times n} \) is the Hilbert matrix with entries \( A_{i,j} = 1/(i + j - 1) \) and \( b = (1, 1, \ldots, 1) \). Use initial point \( x^0 = 0 \). Run the algorithm for dimensions \( n = 5, 10, 15, 20 \). Stop when \( \|\nabla f(x_k)\|_\infty \leq 10^{-6} \). Plot a graph of \( \|\nabla f(x_k)\|_\infty \) versus iteration index \( k \), and a graph of the distance \( \|x_k - x^*\|_2 \) between the iterate \( x_k \) and the optimal solution \( x^* \) versus iteration index \( k \), for each dimension. Interpret the results.

Problem 2
Show that if \( d^0, d^1, \ldots, d^{k-1} \in \mathbb{R}^n \) are linearly independent, and \( f: \mathbb{R}^n \rightarrow \mathbb{R} \) is strongly convex quadratic, then \( h: \mathbb{R}^k \rightarrow \mathbb{R} \) given by \( h(y) := f(x^0 + y_0 d^0 + \cdots + y_{k-1} d^{k-1}) \) is also strongly convex quadratic.

Problem 3
Conjugate gradient methods use directions \( d^0, d^1, \ldots, d^{n-1} \in \mathbb{R}^n \) (for most iterations) generated as follows:
\[ d^0 = -\nabla f(x^0) \]
\[ d^{k+1} = -\nabla f(x^{k+1}) + \beta_{k+1} d^k \]
The Fletcher-Reeves method chooses
\[ \beta_{k+1}^{FR} := \frac{\nabla f(x^{k+1})^T \nabla f(x^{k+1})}{\nabla f(x^{k+1})^T \nabla f(x^k)} \]
The Polak-Ribiére method chooses
\[ \beta_{k+1}^{PR} := \frac{\nabla f(x^{k+1})^T (\nabla f(x^{k+1}) - \nabla f(x^k))}{\nabla f(x^{k+1})^T \nabla f(x^k)} \]
The Hestenes-Stiefel method chooses
\[ \beta_{k+1}^{HS} := \frac{\nabla f(x^{k+1})^T (\nabla f(x^{k+1}) - \nabla f(x^k))}{(\nabla f(x^{k+1}) - \nabla f(x^k))^T d^k} \]
Show that when \( f \) is a quadratic function, and exact line search is done, then the three methods are the same.
Problem 4
Consider a symmetric positive definite matrix $Q \in \mathbb{R}^{n \times n}$, and the associated norm $\|x\|_Q := \sqrt{x^T Q x}$. Consider $Q$-conjugate directions $d_0, d_1, \ldots, d_{n-1} \in \mathbb{R}^n$ generated from linearly independent vectors $p_0, p_1, \ldots, p_{n-1} \in \mathbb{R}^n$. Show that, for each $k = 1, \ldots, n-1$, $d_k = p_k - \hat{p}_k$, where $\hat{p}_k$ is the projection of $p_k$ onto the subspace spanned by $p_0, \ldots, p_{k-1}$ (or the subspace spanned by $d_0, \ldots, d_{k-1}$) with respect to the $\| \cdot \|_Q$-norm, that is,

$$
\hat{p}_k = \arg\min \{\|p_k - p\|_Q : p \in [p_0, \ldots, p_{k-1}] \}
$$

That is, $d_k$ is the part of $p_k$ that remains after we subtract the projection of $p_k$ onto the subspace spanned by $p_0, \ldots, p_{k-1}$. 