Sample questions

1. Compare and contrast the line search and trust region methods of globalizing a quasi-Newton method. Your discussion should touch on the following points:
   (a) The cost of taking a step (that is, of solving the line search subproblem and trust region subproblem), and the complexity of the algorithms.
   (b) Dealing with nonconvexity.
   (c) Convergence theorems.

2. Derive the BFGS update used for approximating the Hessian of the objective function in an unconstrained minimization problem. Explain the rationale for the steps in the derivation.

3. (a) State carefully and prove the first-order necessary condition for \( f : \mathbb{R}^n \to \mathbb{R} \) to have a local minimum at \( x = x^* \).
   (b) Give an example to show that the first-order condition is only necessary, not sufficient.
   (c) State carefully and prove the second-order necessary condition for \( f : \mathbb{R}^n \to \mathbb{R} \) to have a local minimum at \( x = x^* \).
   (d) Give an example to show that the second-order condition is only necessary, not sufficient.

4. (a) Let \( \{x_k\} \) be a sequence in \( \mathbb{R}^n \), and suppose
   \[
   x_k \to x^* \in \mathbb{R}^n \quad \text{as} \quad k \to \infty.
   \]
   Define ``\( x_k \to x^* \) q-quadratically.''

(b) Let \( f : \mathbb{R}^n \to \mathbb{R} \) be given. Newton's method for minimizing \( f \) is locally q-quadratically convergent. State carefully and prove this theorem.

5. Suppose \( f : \mathbb{R}^n \to \mathbb{R} \) is convex.
   (a) State and prove a theorem indicating that the usual first-order necessary condition is, in
this case, a sufficient condition.

(b) Prove that every local minimum of \( f \) is, in fact, a global minimum.

6. Consider the equality-constrained problem

\[
\min_{x \in \mathbb{R}^n} f(x) \\
\text{s.t.} \\
g(x) = 0,
\]  

(2)

where \( f : \mathbb{R}^n \to \mathbb{R} \) and \( g : \mathbb{R}^n \to \mathbb{R}^m \) are smooth functions.

(a) Explain how to apply the quadratic penalty method to (2). How does one obtain an estimate of the Lagrange multiplier?

(b) Explain how to apply the augmented Lagrangian method to (2). How does one obtain an estimate of the Lagrange multiplier?

7. Recall that the \( \ell_1 \) norm of \( x \in \mathbb{R}^n \) is defined by

\[
||x||_1 = \sum_{i=1}^{n} |x_i|.
\]

Derive the formula for the corresponding induced matrix norm of \( A \in \mathbb{R}^{n \times n} \), and prove that it is correct.

8. (a) What is the condition number of a matrix \( A \in \mathbb{R}^{n \times n} \)?

(b) Explain how this condition number is related to the problem of computing the solution \( x \) to \( Ax = b \), where \( b \) is regarded as the data of the problem, and \( A \) is regarded as being known exactly.

9. Consider the least-squares problem \( Ax = b \), where \( A \in \mathbb{R}^{m \times n}, \ m > n \), and \( b \in \mathbb{R}^{m} \) are given, and \( x \) is to be determined.

(a) Assume that \( A \) has full rank. Explain how to solve the least-squares problem using:

i. the normal equations;

ii.
the QR factorization of $A$;
iii.  the SVD of $A$.
In each case, your explanation must include a justification that the algorithm leads to the solution of the least-squares problem (e.g. explain why the solution of the normal equations is the solution of the least-squares problem).

(b)  Discuss the advantages and disadvantages of each of the above methods. Which is the method of choice for the full rank least-squares problem?

10.  
(a)  Give a simple example to show that Gaussian elimination without partial pivoting is unstable in finite-precision arithmetic. (Hint: The example can be as small as $2 \times 2$.)

(b)  Using the concept of backward error analysis, explain the conditions under which Gaussian elimination with partial pivoting can be unstable in finite-precision arithmetic. (Note: This question does not ask you to perform a backward error analysis. Rather, you can quote standard results in your explanation.)

(c)  Give an example to show that Gaussian elimination with partial pivoting can be unstable.

11.  
(a)  Suppose that $A \in \mathbb{R}^{n \times n}$ has eigenvalues

$$
\lambda_n > \lambda_{n-1} \geq \lambda_{n-2} \geq \cdots \geq \lambda_1.
$$

Explain how to perform the power method, and under what conditions it converges to an eigenvalue.

(b)  Explain the idea of simultaneous iteration.

(c)  Explain the QR algorithm and its relationship to simultaneous iteration.

12.  
Suppose that $A \in \mathbb{R}^{n \times n}$ is invertible, $B$ is an estimate of $A^{-1}$, and $AB = I + E$. Show that the relative error in $B$ is bounded by $\|E\|$ (using an arbitrary matrix norm).

13.  
Show that if $A$ is symmetric positive definite and banded, say $a_{ij} = 0$ for $|i-j| > p$, then the Cholesky factor $B$ of $A$ satisfies $b_{ij} = 0$ for $j > i$ or $j < i-p$.

14.  
Suppose that Gaussian elimination (without partial pivoting) is applied to a symmetric positive definite matrix $A$. Write
\[
A^{(k)} = E_{k-1}E_{k-2} \cdots E_1 A,
\]

where each \( E_j \) is an elementary (lower triangular) matrix (left-multiplication by \( E_j \) accomplishes the \( j \)th step of Gaussian elimination). None of the \( E_j \)s is a permutation matrix, that is, no row interchanges are performed. The purpose of this exercise is to prove that this is possible (i.e. that Gaussian elimination can be applied without row interchanges) and to prove the following inequality:

\[
\max_{i,j,k=1,\ldots,n} |a_{ij}^{(k)}| \leq \max_{i,j=1,\ldots,n} |a_{ij}|.
\]

Do this by proving the following three lemmas:

(a) Let \( B \) be a symmetric positive definite matrix. Then \( b_{ii} > 0 \) for \( i = 1, \ldots, n \) and the largest entry of \( B \) (in magnitude) occurs on the diagonal.

(b) Let \( A \) be a symmetric positive definite matrix, and suppose one step of Gaussian elimination is applied to \( A \) to obtain

\[
A^{(2)} = \begin{bmatrix}
a_{11} & a^T \\
0 & \hat{A}^{(2)}
\end{bmatrix}.
\]

Then \( \hat{A}^{(2)} \) is also symmetric positive definite.

(c) Using the notation of the previous lemma,

\[
\tilde{a}_{ii}^{(2)} \leq a_{i+1,i+1}, \quad i = 1, \ldots, n - 1.
\]

Now complete the proof by induction. (Note that this result both proves that no partial pivoting is required for a symmetric positive definite matrix, and also that Gaussian elimination is perfectly stable when applied to such a matrix.)

15. Let \( A \in \mathbb{R}^{m \times n} \) have SVD \( A = USV^T \), where \( U \in \mathbb{R}^{m \times m} \) and \( V \in \mathbb{R}^{n \times n} \) are orthogonal and \( S \) is diagonal (\( S_{ij} = 0 \) if \( i \neq j \)). Define \( S^{(k)} \) by

\[
S^{(k)}_{ij} = \begin{cases} 
0, & i = j \geq k, \\
S_{ij}, & \text{otherwise},
\end{cases}
\]
and define $A^{(k)}$ by

$$A^{(k)} = U S^{(k)} V^T.$$ 

What is $\|A - A^{(k)}\|_2$, where $\| \cdot \|_2$ is the matrix norm induced by the Euclidean ($\ell^2$) vector norm? Prove your answer.