S31020, Winter 2024, Practices exam.

STUDENT NAME (PLEASE PRINT) ______________________________
Direction: Please solve all problems and show your work while observing space limitations where required. This exam has 3 problems (each worth 25 points) and 5 discussion questions (each worth 5 points).

Problem 1 (25 points)

The last ten iterates of an iterative minimization method result in the following values for (a) the norms of the gradient function and (b) the ratios of the successive logs of the norms of those gradients when applied to the Rosenbrock function for $n = 6$

$\|\nabla f(x_k)\|$	$\log(\|\nabla f(x_k)\|) / \log(\|\nabla f(x_{k-1})\|)$
3.36E-01	1.94E+00
2.01E-01	1.47E+00
9.86E-02	1.45E+00
4.29E-02	1.36E+00
1.35E-02	1.37E+00
2.75E-03	1.37E+00
2.76E-04	1.39E+00
9.97E-06	1.41E+00
7.72E-08	1.42E+00
5.43E-11	1.44E+00

The algorithm is

$$x_{k+1} = x_k - \left(\nabla^2_{xx} f(x_k) + \|\nabla_x f(x_k)\|^{\frac{1}{p}} I_n \right)^{-1} \nabla f(x_k)$$

where p is one of $1, 2, \infty$ and I_n is the identity matrix of dimension n. You can assume that the sequence x_k converges to x^*, a local minimum of f(x) that satisfies the sufficient second-order condition. Recall, the Rosenbrock function is a multivariate polynomial.

A) Which one of the three is it? Explain why? (no need to prove anything here, though some simple calculations may be necessary)
B) Prove the statement(s) that you used to infer the answer to part A.

S31020, Winter 2024, Practices exam.

Work Space (Page Intentionally Blank).

S31020, Winter 2024, Practices exam.

Problem 2 (25 points).

Consider a line search method for solving the unconstrained optimization problem $\min_x f(x)$ where $f(x): \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable.

A. State the principles of the two standard requirements that ensure global convergence in line search methods, under the assumption of boundedness of the level set $\mathcal{L} = \left\{ x \mid f(x) \le f(x_0) \right\}$. You can use a graph or two to describe your point and to explain why they are necessary (that is, what could happen if they are not satisfied).

B. (hard) Explain how these principles are satisfied by the Armijo (backtracking) line search, and prove your statements under standard assumptions about the search direction (note that the Wolfe conditions are sufficient but not necessary to obtain these two objectives, only the second objective needs a proof in this case). Here you can assume that the search direction satisfied $H_k p_k = -\nabla f(x_k)$ where H_k is a positive definite matrix bounded above and below uniformly with k.

Algorithm: (Armijo-backtracking line search)
Choose $\bar{\alpha} > 0, \rho \in (0,1), c \in (0,1)$; Set $\alpha \leftarrow 1$
repeat until $\mathrm{f}(\mathrm{x}_k + \alpha p_k) \le f(x_k) + c\alpha \nabla f(x_k)^T p_k$
$\qquad \alpha \to \rho\alpha$
end(**repeat**)
Terminate with $\alpha_k = \alpha$

S31020, Winter 2024, Practices exam.

Work Space (Page Intentionally Blank).

S31020, Winter 2024, Practices exam.

Problem 3 (25 points)

Consider a trust-region approach to solve the unconstrained optimization problem $\min_x f(x)$ where $f(x):\mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable. The trust-region subproblem is

$$(T) \quad \min \frac{1}{2} p^T H_k p + \nabla f(x_k)^T p, \text{ s.t. } \|p\| \leq \Delta$$

1. Describe the trust-region adjustment mechanism, sketch its pseudocode (I am not looking for the exact version of the book, but just the part corresponding to the adjustment). Explain why it neither collapses to 0 nor does it get in the way of the possible superlinear or quadratic convergence.
2. Assume that the function $f(x) = \sum_{i=1}^{N} f_i(x)$ that is, it is separable, where $N \gg n$ (for example N can be 10000 and $n = 6$ in some nonlinear regression or maximum likelihood problems) and that $H_k = \nabla_{xx}^2 f(x_k)$. Propose an eigenvalue-decomposition-based approach to solve problem (T) exactly (you can consider only the "easy" case; only a description of how the quantities involved are computed is necessary, and not a pseudocode). If you were to implement it would this result in a major loss in performance compared to a Cholesky based approach to solve (T) (that is, the approach presented in your textbook)?

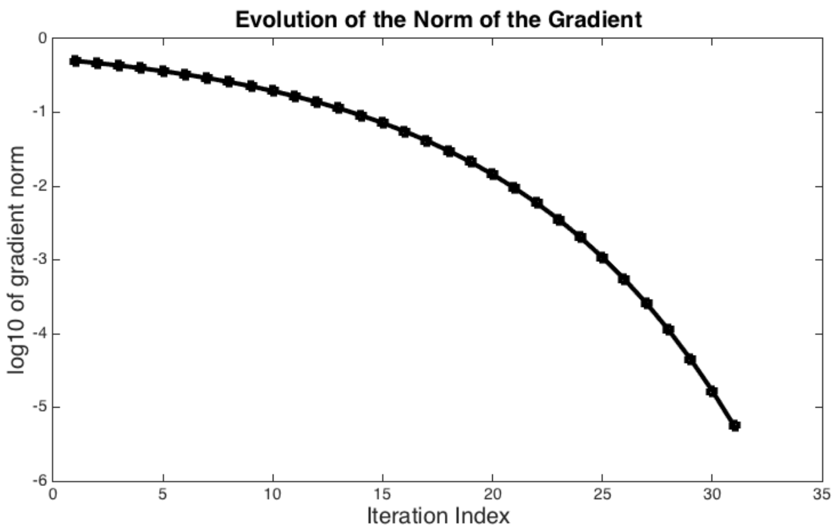
Work Space (Page Intentionally Blank).

S31020, Winter 2024, Practices exam.

Discussion questions.
Note that the answer is limited in size (no more than 4-8 lines and use a readable font please). The answer must still include all important elements (which are not many)

Question 1. (5 points)

You are minimizing a function without constraint and you are plotting the following diagnostic:

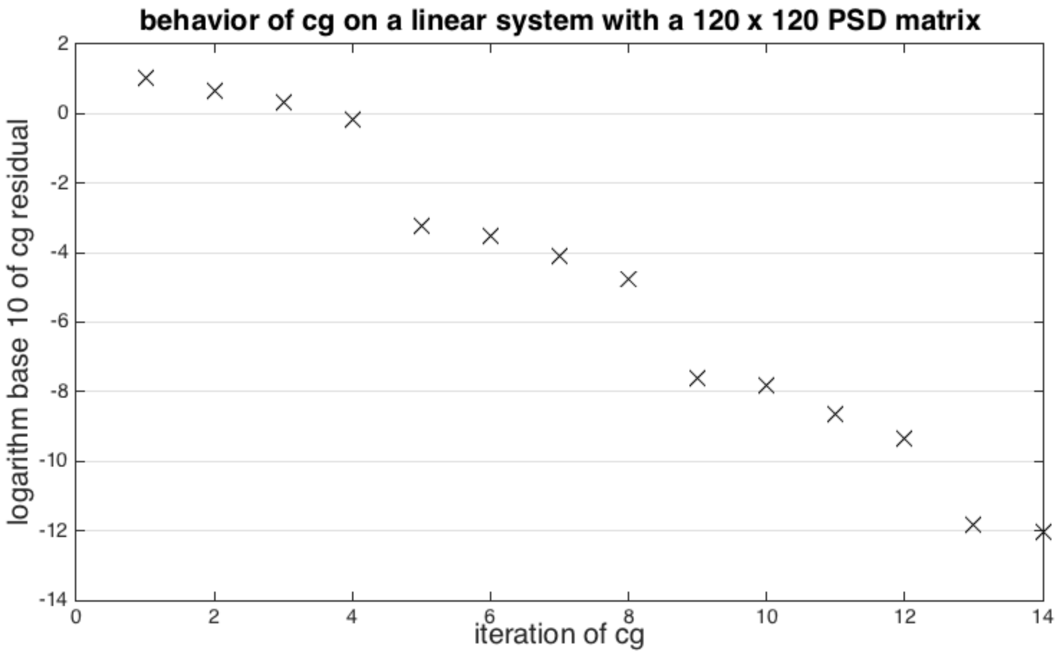


What can you say is the rate of convergence of the method? If the problem would satisfy the standard assumptions, and you would have to choose between this being the steepest descent, quasi-Newton or Newton method, which one seems more likely to have produced this output?

Question 2. (5 points) You are using the central difference approximation $f'(x) \approx \dfrac{f(x+h)-f(x-h)}{2h}$ to estimate the derivative of a function. Explain what is roughly the optimal value of the perturbation parameter h and why.

S31020, Winter 2024, Practices exam.

Question 3 (5 points). You are applying the un-preconditioned conjugated gradient to a system $Ax=b$ where $A \in {\sim}^{120\times120}$ is a positive definite matrix. You are observing the following behavior of the residual.



What can you say about the matrix A (and more precisely, its spectrum)?

Question 4 (5 points). In certain circumstances, the Gauss-Newton method may be superlinearily convergent when applied to nonlinear least squares. TRUE --- FALSE (Explain in no more than 4 lines)

Question 5 (5 points). To achieve superlinear convergence for an inexact Newton method (such as Steihaug's method) it is sufficient to choose the inner loop tolerance $\acute{U}_k$ to be smaller than a sufficiently small number for all k. TRUE---- FALSE (Explain in no more than 4 lines)