Quiz 4

Convex Optimization, 10-725

Due Friday October 25, 2019

Name:

Andrew ID:

Each question is either in true/false format, or multiple choice. For multiple choice, just choose the single best option. In each case, make sure to fill in the box according to the answer you choose (true or false, or the multiple choice option) completely. All questions are worth 1 point.

1. Consider Newton's method on a convex function, under lower and upper bounds on the eigenvalues of the Hessian, and a Lipschitz Hessian. In its region of quadratic convergence, backtracking will always select pure step sizes (equal to 1).
   ☐ True
   ☐ False

2. Which of the following statements comparing Newton's method and gradient descent is not accurate (for convex problems)?
   ☐ a. One Newton iteration is typically more computationally costly than one gradient descent iteration.
   ☐ b. Newton's method has faster local rate of convergence under suitable assumptions.
   ☐ c. Each Newton step can be viewed as an exact minimization of a suitable quadratic approximation, whereas that is not the case for gradient descent.
   ☐ d. In both Newton's method and gradient descent, we can use backtracking to ensure global convergence.

3. Generally speaking, Newton's method can be used for: a. finding the roots of a nonlinear equation, b. minimizing a function. These ideas are completely separate (they don't have any relationship).
   ☐ True
   ☐ False

4. Compared to gradient descent, Newton's method, roughly speaking:
   ☐ a. uses more accurate quadratic approximations, admits more expensive iterations, but requires fewer iterations to converge to high accuracy;
   ☐ b. uses less accurate quadratic approximations, and cheaper iterations, so it requires more iterations to converge to high accuracy;
   ☐ c. uses cubic approximations, and its iterations and convergence are not really comparable to gradient descent;
   ☐ d. only approximates the smooth part of the criterion by a quadratic, and thus applies to a broader class of nonsmooth optimization problems.

5. If we run Newton's method on $f(x)$ starting at x_0 for k iterations, and for the same sequence of step sizes, run Newton's method on $g(y) = f(Ay)$ starting at $y_0 = A^{-1}x_0$ for k iterations, then the achieved sequence of criterion values is the same in both cases.
   ☐ True
   ☐ False

6. Pure Newton's method (with step sizes equal to 1) will always converge on a convex function.
   ☐ True
   ☐ False

7. Newton's method on a convex function, under lower and upper bounds on the eigenvalues of the Hessian, and a Lipschitz Hessian, converges (with suitable step sizes) at the rate:
   ☐ a. $O(\log\log(1/\epsilon))$;
   ☐ b. $O(\log\log(1/\epsilon))$, locally;
   ☐ c. it depends on whether the function is self-concordant;
   ☐ d. none of the above.

8. Let $f^\star$ denote the optimal criterion value of the convex problem

$$\min_x \; f(x) \quad \text{subject to} \quad h_j(x) \le 0, \; j = 1, \dots, m,$$

   and $x^\star(t)$ denote the solution in the barrier problem

$$\min_x \; tf(x) + \phi(x).$$

   Then $f(x^\star(t)) - f^\star \le m/t$.
   ☐ True
   ☐ False

9. Each main iteration of the barrier method performs just one one Newton update.
   ☐ True
   ☐ False

10. The main idea behind the barrier method is add terms to the criterion that:
   ☐ a. smoothly approximate indicator functions of the constraints;
   ☐ b. make the new criterion strongly convex;
   ☐ c. make the new criterion smooth;
   ☐ d. get rid of equality constraints.

11. The barrier method solves the problem:

$$\min_x \; f(x) \quad \text{subject to} \quad h_j(x) \le 0, \; j = 1, \dots, m,$$

   by solving a:
   ☐ a. single problem of the form $\min_x\,(tf(x) + \phi(x))$, where $\phi(x) = -\sum_{j=1}^m \log(-h_j(x))$ and $t > 0$;
   ☐ b. sequence of problems of the form $\min_x\,(t_k f(x) + \phi(x))$, where $\phi(x) = -\sum_{j=1}^m \log(-h_j(x))$ and $t_k \to \infty$;
   ☐ c. sequence of problems of the form $\min_x\,(t_k f(x) + \phi(x))$, where $\phi(x) = -\sum_{j=1}^m \log(-h_j(x))$ and $t_k \to 0$;
   ☐ d. sequence of problems of the form $\min_x\,(t_k f(x) + \phi(x))$, where $\phi(x) = \sum_{j=1}^m \log(-h_j(x))$ and $t_k \to \infty$;
   ☐ e. sequence of problems of the form $\min_x\,(t_k f(x) + \phi(x))$, where $\phi(x) = \sum_{j=1}^m \log(-h_j(x))$ and $t_k \to 0$.

12. For constrained convex minimization, barrier methods approach the solution from the outside of the constraint set.
   ☐ True
   ☐ False

13. Which of the following statements about the barrier method and the primal-dual interior-point method is not true (for convex problems)?
   ☐ a. Both barrier method and primal-dual interior-point method can be interpreted as solving a perturbed version of the KKT conditions.
   ☐ b. Both methods have local $O(\log(1/\epsilon))$ rate of convergence.
   ☐ c. Primal-dual interior-point method is more commonly used in practice because it tends to be more efficient.
   ☐ d. Both methods perform just one Newton update before taking a step along the central path (adjusting the barrier parameter t).

14. The iterates of the primal-dual interior-point method are always primal and dual feasible.
☐ True
☐ False

15. Each main iteration of a primal-dual interior-point method performs just one one Newton update.
☐ True
☐ False

16. Which of the following statements about the barrier method and the primal-dual interior-point method is not true (for convex problems)?
☐ a. Both barrier method and primal-dual interior-point method can be interpreted as solving a perturbed version of the KKT conditions.
☐ b. Both require solving a linear system at the lowest level of iteration.
☐ c. Both methods have local $O(\log(1/\epsilon))$ rate of convergence.
☐ d. Both yield feasible primal and dual iterates at every step.

17. Consider the convex problem:

$$\min_x \ f(x) \ \text{ subject to } \ Ax = b, \ h(x) \leq 0.$$

Which one of the following is not a consideration, when choosing the step size in each main iteration of a primal-dual interior-point algorithm applied to the this problem?
☐ a. Take a full Newton step (step size equal to 1) if possible.
☐ b. Take a step size that ensures $h(x) < 0$.
☐ c. Take a step size that ensures $u > 0$.
☐ d. Take a step size that ensures $Ax = b$.

18. DFP and BFGS differ in that only one of them satisfies the secant equation.
☐ True
☐ False

19. DFP and BFGS differ in that only one of them preserves positive definiteness (of the approximated Hessian, from one iteration to the next).
☐ True
☐ False

20. In the DFP and BFGS updates, each update on the approximation to the Hessian matrix and its inverse are:
☐ a. symmetric rank-one updates for both the Hessian and its inverse;
☐ b. symmetric rank-two updates for both the Hessian and its inverse;
☐ c. a symmetric rank-one update for the Hessian and a rank-two update for its inverse;
☐ d. a symmetric rank-two update for the Hessian and a rank-one update for its inverse.