Quiz 5

Convex Optimization, 10-725

Due Friday November 15, 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. Suppose A is $n \times n$, nonsingular, and dense. Which of the following statements is not true?
   ☐ a. Computing a QR decomposition of A costs $O(n^3)$ flops.
   ☐ b. Once we have its QR decomposition, solving a linear system in A costs $O(n^2)$ flops.
   ☐ c. Computing a Cholesky decomposition is only possible if A is positive definite.
   ☐ d. Computing a Choleksy decomposition costs less than $O(n^3)$ flops.

2. Consider solving a dense linear system versus solving a banded linear system, both $n \times n$. (Treat the bandwidth as constant, so that it does not appear in the $O(\cdot)$ notation.) As measured by flops, their computation time is:
   ☐ a. $O(n^2)$ versus $O(n)$;
   ☐ b. $O(n^3)$ versus $O(n)$;
   ☐ c. $O(n^3)$ versus $O(n^2)$;
   ☐ d. $O(n^2)$ versus (it depends).

3. For solving the linear system $Ax = b$, gradient descent and conjugate gradient both converge at the same $O(\log(1/\epsilon))$ rate, but differ in how the constants here depend on the condition number of A.
   ☐ True
   ☐ False

4. The convergence of coordinate descent (meaning exact coordinatewise minimization) is believed to be generally better than that of a first-order method.
   ☐ True
   ☐ False

5. For the lasso problem, one cycle of coordinate descent updates (by which we mean exact coordinatewise minimizations), implemented carefully, costs the same order of flops as one proximal gradient descent update.
   ☐ True
   ☐ False

6. Coordinate descent most generally applies to convex problems that:
   ☐ a. have smooth plus separable criterions;
   ☐ b. have no constraints;
   ☐ c. decompose directly into separate univariate minimizations;
   ☐ d. are not solvable by ADMM.

7. If f is strongly convex, closed, and has a Lipschitz gradient, then dual gradient ascent and primal gradient descent both converge at the same rate.
   ☐ True
   ☐ False

8. The dual decomposition method has great convergence properties but is not parallelizable, while the method of the augmented Lagrangian has poor convergence properties but can be parallelized.

☐ True
☐ False

9. The particular parametrization of a problem that we use, before we apply ADMM (i.e., the way we introduce auxiliary variables to put a problem in “ADMM form”) is not too important, because the convergence behavior of the resulting ADMM algorithm will be more or less the same, regardless of the parametrization.
☐ True
☐ False

10. The augmented Lagrangian parameter ρ in ADMM can be routinely chosen by backtracking search.
☐ True
☐ False

11. The convergence of ADMM is believed to be generally on par with that of a first-order method (if not slightly better).
☐ True
☐ False

12. ADMM cannot be applied to an optimization problem with no constraints.
☐ True
☐ False

13. The Frank-Wolfe method is based on minimizing a:
☐ a. quadratic approximation of the criterion, then projecting onto the constraint set;
☐ b. linear approximation of the criterion, then projecting onto the constraint set;
☐ c. linear approximation of the criterion over the constraint set;
☐ d. quadratic approximation of the criterion over the constraint set.

14. For a constrained, convex problem, the Frank-Wolfe method achieves the same convergence rate as projected gradient descent, but only under much stronger conditions.
☐ True
☐ False

15. Both the Frank-Wolfe method and projected gradient descent can be carried out efficiently with ℓ_q norm constraints, for any $q \geq 1$.
☐ True
☐ False