Questions for the written examination “Optimization for Computer Science”

WS 2018/2019

Mode of the exam: No additional documents are allowed. The duration of the examination is 90 minutes. Total number of points: 100. Grading key: 50.5-62.5 (4), 63-75 (3), 75.5-87.5 (2), 88-100 (1)

1. Explain the definition of an induced matrix norm. Explain the meaning of the $1, 2, \infty$ matrix norms? What are Schatten norms?
2. What is the spectral decomposition theorem of a symmetric $n \times n$ matrix?
3. What are the linear and quadratic approximation theorems?
4. What is the fundamental theorem of calculus? Show how it can be applied to a multivariate function $f(x)$ and to its gradient $f(x)$.
5. What is the definition of local and global minima. Give the first order necessary condition of optimality and prove it. What are stationary points? What are the second order necessary and sufficient conditions of optimality? What are the conditions under which a function has a global minimizer or maximizer?
6. Consider the following optimization problem:

$$\min_x f(x) = ..., \text{ s.t....}$$

   Compute all stationary points and classify them as local/global minima or saddle points. Prove/disprove the existence of a global minimizer. (Hint: For examples of this question consider the examples in the book of Amir Beck (Teach Center), page 26-36)
7. Compute the gradient and the Hessian matrix of the quadratic function

$$\min_x \frac{1}{2} x^T A x + b^T x + c.$$

   Give a characterization of its stationary points based on the definiteness of A.

8. What are linear and non-linear least-squares problems. Show the application of linear least squares to data fitting. Show how to compute the optimal solution.

9. Give the definition of a descent direction. Draw a simple example explaining the properties of a descent direction. Prove that taking a small enough step along the descent direction decreases the objective function.

10. Give the general form of a descent method and show that $d^k = -D^k \nabla f(x^k)$ with D^k symmetric and positive definite is a descent direction. Give three different standard choices for descent directions based on choosing the scaling matrix D^k and discuss their numerical performance.

11. Prove the formula for exact line search based on a descent direction d for quadratic functions of the form

$$\min_x \frac{1}{2} x^T A x + b^T x + c.$$

12. Explain the Armijo step size rule and draw a figure. Prove the sufficient decrease condition and show its relation to the Armijo step size condition.

13. What is the "zig-zag" effect? Prove that the differences of successive iterates are orthogonal to each other when using an exact line search.

14. What is the rate of convergence of the gradient method for quadratic functions when using exact line search. What is the condition number?

15. What is a Lipschitz continuous gradient and show how it is related to the norm of the Hessian matrix? What is the Lipschitz constant of the gradient of a linear least squares problem of the form

$$\min_x \frac{1}{2} \|Ax - b\|^2$$

16. Prove the descent lemma for a differentiable function f with Lipschitz-continuous gradient. Interpret the inequality of the descent lemma in terms of an upper bound to the function f. Use the descent lemma to also prove the sufficient decrease lemma.

17. Give the general form of non-linear least-squares problems. Show how the Gauss-Newton method is obtained from performing a first-order Taylor approximation. What is the Levenberg-Marquadt method?

18. Explain the Kalman filter and show its relation to incremental Gauss-Newton methods. What are its applications?

19. Show that the plain form of Newton's method can be derived from a second order Taylor approximation of the objective function. Show that Newton's method is invariant with respect to affine scalings.

20. Give the theorem of locally quadratic convergence of the pure form of Newton's method and prove it. What can you say about the global convergence of Newton's method?

21. What is Q-conjugacy? Show how the Gram-Schmidt procedure can be used to generate a set of Q-conjugate directions from a set of linearly independent vectors.

22. Write down the conjugate gradients (CG) method. What is its convergence rate and how can it be generalized to non-linear problems.

23. Explain the heavy-ball algorithm and show that it is obtained from a finite differences approximation of the heavy-ball with friction dynamical system. How is it related to the CG method?

24. Give the definition of stationary points for minimizing a differentiable function over a convex set. Give an example showing the necessary optimality condition for minimizing a differentiable function over a convex set. Why does it fail in case the constraint set is non-convex?

25. Explain the projection on a convex set? Write down the projection theorem. Compute the projection of a point $z \in \mathbb{R}^n$ to the constraint set $C = \{x \in \mathbb{R}^n : Ax = 0\}$.

26. Show that stationarity in optimization over a convex set can also be written in terms of the projection operator $\text{proj}_C(x)$.

27. What is the gradient projection method. Give the algorithm and give choices for the step size selection. Draw an example illustrating the iterations of the gradient projection method.

28. What is the scaled gradient projection method? Show how to specialize your algorithm such that it becomes a projected Newton algorithm.

29. Derive the affine scaling method for solving an equality constrained LP of the form
$$\min_x \; c^T x, \quad \text{s.t.} \quad Ax = b, \; x \geq 0.$$
Show how the LP is solved using a scaled gradient projection method. Why can the inequality constraint be skipped? How can the step size be computed?

30. Derive the dual affine scaling method for an inequality constrained LP of the form
$$\max_y \; b^T x, \quad \text{s.t.} \quad A^T y \leq c$$
Derive the scaled gradient projection method and show how the metric is chosen.

31. What is the Lagrange multiplier theorem for equality constrained optimization problems? Draw a simple example and explain why the gradients of the constraint functions need to be linearly independent.