（装订线内不要答题）

复旦大学大数据学院
2022年春季学期课程期末考试卷

☒ A 卷　　☐B 卷　　☐C 卷

课程名称：最优化方法
课程代码：DATA130026.01
开课院系：大数据学院 考试形式：闭卷

姓　名：__________ 学　号：_________ 专　业：______________

声明：我已知悉学校对于考试纪律的严肃规定，将秉持诚实守信宗旨，严守考试纪律，不作弊，不剽窃；若有违反学校考试纪律的行为，自愿接受学校严肃处理。

学生（签名）：_____________

年　　月　　日

题　目	1	2	3	4	5	6	总　分
得　分							

1. (20 points) Please answer true or false. (You may use the notation "T" for "true" and "F" for "false".) No explanation is needed. A correct answer is worth 2 points, no answer 0 points, a wrong answer -1 points.

   (1) $f(X) = -\log\det X$ is convex on $\operatorname{dom} f = \mathbf{S}^n_{++}$.

   (2) A set is convex if and only if its intersection with any line is convex.

   (3) Given two sets $S, T \subseteq \mathbb{R}^n$, the set of points closer to one set than another, i.e.,
   $$\{x \mid \operatorname{dist}(x,S) \le \operatorname{dist}(x,T)\}$$
   is convex, where $\operatorname{dist}(x,S) = \inf\{\|x-z\|_2 \mid z \in S\}$.

   (4) Let f be a twice continuously differentiable function defined over $\mathbb{R}^n$. If f is strongly convex and $\nabla^2 f$ is Lipschitz continuous, then Newton's method converges quadratically from any starting point.

   (5) Due to affine invariance, neither the Newton method nor the gradient descent method is affected by the Hessian condition number.

   (6) The subgradient method is a descent method.

   (7) Suppose $\phi : \mathbb{R} \to \mathbb{R}$ is increasing and convex, and $f : \mathbb{R}^n \to \mathbb{R}$ is convex, so $g(x) = \phi(f(x))$ is convex. The Newton direction for g is
   $$-\left(\phi''(f(x))\nabla f(x)\nabla f(x)^T + \phi'(f(x))\nabla^2 f(x)\right)^{-1}\nabla f(x).$$

   (8) Let f be a continuously differentiable function defined over $\mathbb{R}^n$. A point $x \in \mathbb{R}^n$ is a stationary point of f, if and only if it holds $\langle \nabla f(x), d\rangle \ge 0$ for all $d \in \mathbb{R}^n$.

   (9) Given $A \in \mathbf{S}^n$, the semidefinite program
   $$\min \lambda \quad \text{s.t. } \lambda I - A \succeq 0$$
   finds the smallest eigenvalue for A.

   (10) The function $f(x) = -\sqrt{4-x^2}$ with $\operatorname{dom}(f) := \{x : |x| \le 2, x \in \mathbb{R}\}$ is not subdifferentiable at $x = 2$.

2. (15 points)

   (1) (5 points) Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ given by
   $$f(x) = \|x\|_2^{3/2}.$$
   Show that the gradient Lipschitz condition $\|\nabla f(x) - \nabla f(y)\|_2 \le L\|x-y\|_2$ for all x and y is not satisfied for any L.

   (2) (5 points) Derive the subdifferential for the indicate function $I_C(x)$ with
   $$C = \{x \mid \|x\|_\infty \le 1\}.$$

   (3) (5 points) Compute a closed form for the proximal mapping for function $f(x) = -\sum_{i=1}^n \log x_i$.

3. (20 points) Consider the Rosenbrock function
   $$f(x) = 100(x_2 - x_1^2)^2 + (1-x_1)^2.$$

   (1) (10 points) Calculate the gradient $\nabla f(x)$ and Hessian $\nabla^2 f(x)$.

   (2) (10 points) Prove that $x^* = (1,1)^T$ is the only local minimizer and $\nabla^2 f(x^*)$ is positive definite.

4. (20 points) In a Quasi-Newton method, after the $(k+1)$-th iteration, a symmetric positive definite matrix B_{k+1} is sought to satisfy $B_{k+1}(x_{k+1} - x_k) = \nabla f(x_{k+1}) - \nabla f(x_k)$. A necessary condition for such a matrix to exist is
   $$(x_{k+1} - x_k)^T(\nabla f(x_{k+1}) - \nabla f(x_k)) \ge 0. \qquad (*)$$

   (1) (10 points) Prove that for a strong convex function f, $(*)$ always holds if $x_{k+1} \ne x_k$.

（装订线内不要答题）

(2) (10 points) Show that the strong Wolfe curvature condition

$$|\nabla f(x_k + \alpha_k p_k)^T p_k| \le -c_2 \nabla f(x_k)^T p_k$$

with $c_2 \in (0,1)$ and $\alpha_k \in (0,1)$ implies $(*)$ for $x_{k+1} = x_k + \alpha_k p_k$, where $p_k = -B_k^{-1}\nabla f(x_k)$, B_k is positive definite and $\nabla f(x_k) \neq 0$.

5. (25 points) Consider the following linear program,

$$\begin{aligned} \min_x \quad & \sum_{i,j} a_{ij} x_{ij} \\ \text{s.t.} \quad & \sum_{j=1}^n x_{ij} = \alpha_i, \quad i = 1,\dots,m, \\ & \sum_{i=1}^m x_{ij} = \beta_j, \quad j = 1,\dots,n, \\ & x_{ij} \ge 0, \quad i = 1,\dots,m,\ j = 1,\dots,n, \end{aligned}$$

where α_i and β_j are positive scalars, which for feasibility must satisfy

$$\sum_{i=1}^m \alpha_i = \sum_{j=1}^n \beta_j.$$

(a) (10 points) Assign Lagrange multipliers λ_i, μ_j to the equality constraints $\alpha_i - \sum_{j=1}^n x_{ij} = 0$ and $\beta_j - \sum_{i=1}^m x_{ij} = 0$ for $i = 1,\dots,m$, $j = 1,\dots,n$, i.e., define the Lagrange function as

$$L(x;\lambda,\mu) = \sum_{i,j} a_{ij}x_{ij} + \sum_{i=1}^m \lambda_i(\alpha_i - \sum_{j=1}^n x_{ij}) + \sum_{j=1}^n \mu_j(\beta_j - \sum_{i=1}^m x_{ij}).$$

Derive the dual of the above linear program.

(b) (15 points) Show that if x^* is an optimal solution of the primal problem, there is a set of $\{\mu_j^* \mid j = 1,\dots,n\}$ such that if $x_{ij}^* > 0$, then

$$a_{ij} - \mu_j^* = \min_{1\le k\le n}\{a_{ik} - \mu_k^*\}.$$