COL726 Major

Due date: 9 March 2021, 4:00pm

Total marks: 35

You are permitted to refer to the textbook, the lecture notes, and online resources during the exam. You are not permitted to collaborate or discuss with other people, other than asking for clarifications on Piazza.

---

1. Give short answers for the following: [6 marks]

   (a) Let $\mathbf{A} = \hat{\mathbf{Q}}\hat{\mathbf{R}}$ be a reduced QR factorization with columns $\mathbf{a}_i, \mathbf{q}_i, \mathbf{r}_i$ respectively. Under what conditions are $\mathbf{q}_i$ and $\mathbf{r}_i$ ill-conditioned with respect to perturbations in the corresponding $\mathbf{a}_i$?

   (b) Briefly explain why the following statement is true or false: Given a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ with singular values $\sigma_1 > \sigma_2 > \dots$, for most vectors $\mathbf{x}^{(0)}$ the iterates of the algorithm

   $$\mathbf{y}^{(k+1)} = \frac{\mathbf{A}\mathbf{x}^{(k)}}{\|\mathbf{A}\mathbf{x}^{(k)}\|},$$
   $$\mathbf{x}^{(k+1)} = \frac{\mathbf{A}^T\mathbf{y}^{(k+1)}}{\|\mathbf{A}^T\mathbf{y}^{(k+1)}\|}$$

   converge linearly to singular vectors $\mathbf{x}^{(k)} \to \pm\mathbf{v}_1, \mathbf{y}^{(k)} \to \pm\mathbf{u}_1$ with rate σ_2/σ_1.

   (c) Given $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$, we could try solving $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ either (a) directly by Newton's method, or (b) by minimizing $\|\mathbf{f}(\mathbf{x})\|_2^2$ with Newton's method for optimization. What are the matrices arising in the linear solves of the two algorithms? Near the solution, do they have similar condition number?

   (d) Given m points $\mathbf{c}_i \in \mathbb{R}^n$ and symmetric positive definite matrices $\mathbf{P}_i \in \mathbb{R}^{n\times n}$, I want to find a point $\mathbf{x} \in \mathbb{R}^n$ lying in the intersection of all m ellipsoids $E_i = \{\mathbf{x} : (\mathbf{x}-\mathbf{c}_i)^T\mathbf{P}_i(\mathbf{x}-\mathbf{c}_i) \leq 1\}$, or else prove that the intersection is empty. State an algorithm for doing so.

2. Consider the conjugate gradient algorithm applied to the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$, where $\mathbf{A} \in \mathbb{R}^{m\times m}$ is symmetric positive definite with all eigenvalues distinct. [5 marks]

   (a) Show that $\mathbf{b}^T\mathbf{x}^{(k)} \geq 0$ at each iteration.
   Hint: Consider what property $\mathbf{x}^{(k)}$ minimizes in the CG algorithm.

   (b) Suppose $\mathbf{b}$ lies in the span of only n eigenvectors of $\mathbf{A}$, where $n < m$. Show that the CG algorithm finds the solution $\mathbf{x}_*$ in $\leq n$ iterations.

3. For a nonsymmetric matrix $\mathbf{A} \in \mathbb{R}^{m \times m}$, suppose you know a Schur factorization $\mathbf{A} = \mathbf{Q}\mathbf{T}\mathbf{Q}^T$, and all diagonal entries $t_{11}, \ldots, t_{mm}$ of $\mathbf{T}$ are distinct. [5 marks]

   (a) Give an efficient algorithm for finding an eigenvector of $\mathbf{A}$ with eigenvalue $\lambda = t_{ii}$.

   (b) Analyze the operation count of your algorithm to leading order (i.e. $\sim$ and not $O(\cdot)$) in terms of i and m.

4. Suppose you have a system of equations of the form

$$\begin{aligned} x_1 &= f_1(x_2), \\ x_2 &= f_2(x_3), \\ &\vdots \\ x_{n-1} &= f_{n-1}(x_n), \\ x_n &= f_n(x_1), \end{aligned}$$

   and let $\mathbf{x}^*$ be a solution. [5 marks]

   (a) Naturally you consider updating each x_i in parallel, i.e. $x_i^{(k+1)} = f_i(x_{i+1}^{(k)})$ for $i < n$ and $x_n^{(k+1)} = f_n(x_1^{(k)})$. Under what conditions is this scheme guaranteed to converge when started near $\mathbf{x}^*$, and what is its convergence rate?

   (b) Give Newton's method for this system of equations. Under what conditions can you expect quadratic convergence?

5. Instead of performing line search, some optimization algorithms choose the next iterate $\mathbf{x}^+$ as the point $\mathbf{y}$ which minimizes $f(\mathbf{y}) + \frac{1}{2t}\|\mathbf{y} - \mathbf{x}\|_2^2$. [5 marks]

   (a) Show that this is always a descent step (i.e. $f(\mathbf{x}^+) \leq f(\mathbf{x})$ for any $t > 0$), and that when f is differentiable it satisfies $\mathbf{x}^+ = \mathbf{x} - t\nabla f(\mathbf{x}^+)$.

   (b) Give a closed form for $\mathbf{x}^+$ when $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} - \sum \log x_i$, which is only defined for $\mathbf{x} \succ \mathbf{0}$. Verify that we always have $\mathbf{x}^+ \succ \mathbf{0}$ even if $\mathbf{x} \not\succ \mathbf{0}$.
   Hint: Is it possible to solve for each x_i^+ independently?

6. Consider the problem of minimizing $f(\mathbf{x})$ subject to $\mathbf{A}\mathbf{x} = \mathbf{b}$, where $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{p \times n}$ has full rank. Let us analyze the three different approaches we know of for solving such problems. [9 marks]

   (a) In the elimination method, we need to find a matrix $\mathbf{F} \in \mathbb{R}^{n \times (n-p)}$ such that $\text{range}(\mathbf{F}) = \text{null}(\mathbf{A})$. Show how it can be computed from the full QR factorization of $\mathbf{A}^T$.

   (b) Suppose the objective is quadratic, $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{q}^T\mathbf{x} + r$ with $\mathbf{P} \succ \mathbf{0}$. Find the dual function $g(\boldsymbol{\nu})$, and explain how to compute $\mathbf{x}^*$ using one iteration of Newton's method applied to g.

(c) To apply Newton's method directly to the constrained problem, we need to solve a linear system with matrix $\mathbf{K} = \begin{bmatrix} \nabla^2 f(\mathbf{x}) & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix}$. Show that if $\nabla^2 f(\mathbf{x})$ is symmetric positive definite and $\mathbf{A}$ is full rank, then $\mathbf{K}$ has a factorization $\mathbf{LDL}^T$ where $\mathbf{D} = \begin{bmatrix} \mathbf{I} & \\ & -\mathbf{I} \end{bmatrix}$. Explain how to compute it using a standard Cholesky factorization routine.

(d) Suppose $f(\mathbf{x})$ is of the form $\sum_{i=1}^{n-1} f_i(x_i, x_{i+1})$ where each $f_i : \mathbb{R}^2 \to \mathbb{R}$ is convex. Show that $\nabla^2 f(\mathbf{x})$ is tridiagonal and symmetric positive semidefinite. Give the big-O operation count of any one of the above three algorithms (considering precomputation and per-iteration costs separately) if this tridiagonal structure is taken into account.