UCLA — Electrical Engineering Dept.
EE133A: Applied Numerical Computing
Final Exam
Wednesday, December 7, 2016

This exam has 6 questions, for a total of 100 points.

Closed book. No calculator. Answer the questions in the spaces provided on the question sheets. If you run out of room for an answer, continue on the back of the page.
Please, write your name and ID on the top of each loose sheet!

Name and ID: ______________________________

Name of person on your left: ______________________________

Name of person on your right: ______________________________

Question	Points	Score
1	8	
2	22	
3	26	
4	10	
5	14	
6	20	
Total:	100	

1. Consider the minimization problem

$$\begin{array}{ll} \text{minimize} & x^{\mathsf{T}} M x \\ \text{subject to} & x_1 = x_n = 1 \end{array}$$

where M is an $n \times n$ positive definite matrix, and x an $n \times 1$ vector.

   (a) (4 points) Prove that this problem is a constrained least squares problem, and write it in the form

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|^2 \\ \text{subject to} & Cx = d. \end{array}$$

   Specify what A, b, C, and d are.

   (b) (4 points) Write the expression for the solution to the minimization problem, $\hat{x}$.

2. Suppose we observe some signal $y \in \mathbb{R}^n$. We assume that it takes the form

$$y = T(a)x = \frac{1}{n} W^{\mathsf{H}} \mathrm{diag}(Wa) Wx$$

where $W \in \mathbb{R}^{n \times n}$ is the discrete Fourier transform (DFT) matrix. The inverse of W is given by $W^{-1} = (1/n)W^{\mathsf{H}}$. Both y and a are known vectors; x is an unknown vector. $T(a)$ is a matrix with Toeplitz structure.

(a) (9 points) Find $T(a)^{-1}$. Does this matrix have Toeplitz structure? Provide evidence with your statement.

(b) (9 points) Describe an efficient method for calculating x. State every step, and the number of flops necessary.

(c) (4 points) What signal processing operation have we performed to calculate x?

3. Consider the computation of $x \in \mathbb{R}^n$ in the following problem:

$$x = (A + A^{-1})\, b,$$

where A is a non-singular $n \times n$ matrix, and b is a given $n \times 1$ vector.

(a) (8 points) Describe an efficient algorithm for computing x. State each step in your method, and provide a flop count.

(b) (5 points) Suppose $A > 0$ (i.e., A is positive definite). Prove that $A^{-1} > 0$ and $(A+A^{-1}) > 0$.

(c) (5 points) If $A > 0$, describe a *more efficient* algorithm for computing x. State each step in your method, and provide a flop count.

(d) (8 points) Suppose

$$A = \begin{bmatrix} B & 1 \\ u^{\mathsf{T}} & 1 \end{bmatrix},$$

where $B \in \mathbb{R}^{(n-1)\times(n-1)}$ is a positive definite matrix with **known** Cholesky factorization $B = R^{\mathsf{T}} R$, and $u \in \mathbb{R}^{n-1}$. Describe an *even more efficient* algorithm for computing x. State each step in your method, and provide a flop count.

4. Consider the following function

$$\gamma = \min_x \left\{ \frac{x^\mathsf{T} A x}{x^\mathsf{T} B x} \right\},$$

where A and B are both $n \times n$ positive definite real matrices.

(a) (5 points) Transform the problem into the equivalent formulation:

$$\gamma = \min_y \left\{ \frac{y^\mathsf{T} C y}{\|y\|^2} \right\}.$$

Define y and C.

(b) (5 points) Prove that the previous formulation is equivalent to

$$\tilde{\gamma} = \min_y \left\{ y^\mathsf{T} C y, \quad \text{subject to } \|y\| = 1 \right\}.$$

5. Consider the problem

$$\begin{bmatrix} 0 & A^\top & I \\ A & 0 & 0 \\ I & 0 & D \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} b \\ c \\ d \end{bmatrix},$$

where $A, D \in \mathbb{R}^{n \times n}$ are non-singular matrices.

(a) (8 points) Describe an efficient algorithm for calculating x, y, z. State every step, and the number of flops necessary.

(b) (6 points) Suppose we are given (for no cost) the factorization

$$A = U \Lambda U^\top$$

where U is an orthogonal matrix, and Λ is a diagonal matrix. Describe a faster method to compute x, y, z. State every step, and the number of flops necessary.

6. Consider the problem

$$\text{minimize} \qquad f(x) = \sum_{i=1}^{n} k_i e^{x_i} + \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2,$$

where the $k_i > 0$ are known constants.

(a) (5 points) Compute the gradient $\nabla f(x)$ and Hessian $\nabla^2 f(x)$. Note that $\nabla^2 f(x) > 0$ for all x.

(b) (5 points) Does this problem have a solution? Is it unique? Provide evidence with your statement.

(c) (5 points) Provide an efficient method for computing the Newton step

$$v = -\nabla^2 f(x)^{-1} \nabla f(x).$$

(d) (5 points) What is the cost in flops of each Newton step with this method, for large n?

Formulas

- Inner product, norm, angle.
  - Relation between inner product, norms, and angle: $a^T b = \|a\|\,\|b\| \cos \angle(a,b)$.
  - Average value of elements of an n-vector: $\mathrm{avg}(a) = (\mathbf{1}^T a)/n$.
  - Root-mean-square value of an n-vector: $\mathrm{rms}(a) = \|a\|/\sqrt{n}$.
  - Standard deviation of an n-vector: $\mathrm{std}(a) = \mathrm{rms}(\tilde{a})$ where $\tilde{a} = a - \mathrm{avg}(a)\mathbf{1}$.
  - Correlation coefficient of two n-vectors:

$$\rho = \frac{\tilde{a}^T \tilde{b}}{\|\tilde{a}\| \|\tilde{b}\|} \quad \text{where } \tilde{a} = a - \mathrm{avg}(a)\mathbf{1} \text{ and } \tilde{b} = b - \mathrm{avg}(b)\mathbf{1}.$$

- Complexity of basic matrix and vector operations (α is a scalar, x and y are n-vectors, A is an $m \times n$ matrix, B is an $n \times p$ matrix).
  - Inner product $x^T y$: $2n - 1$ flops ($\approx 2n$ flops for large n).
  - Vector addition $x + y$: n flops.
  - Scalar-vector multiplication αx: n flops.
  - Scalar-matrix multiplication αA: mn flops.
  - Matrix-vector multiplication Ax: $m(2n - 1)$ flops ($\approx 2mn$ flops for large n).
  - Matrix-matrix multiplication AB: $mp(2n - 1)$ flops ($\approx 2mpn$ flops for large n).
- Pseudo-inverses.
  - Pseudo-inverse of left invertible matrix A: $A^\dagger = (A^T A)^{-1} A^T$.
  - Pseudo-inverse of right invertible matrix A: $A^\dagger = A^T (AA^T)^{-1}$.
- Complexity of forward or back substitution with triangular $n \times n$ matrix: n^2 flops.
- Complexity of matrix factorizations.
  - QR factorization of $m \times n$ matrix: $2mn^2$ flops.
  - LU factorization of $n \times n$ matrix: $(2/3)n^3$ flops.
  - Cholesky factorization of $n \times n$ matrix: $(1/3)n^3$ flops.
- Least-squares and constrained least squares problems.
  - Solution of least-squares problem 'minimize $\|Ax - b\|^2$' if A has linearly independent columns: $x = (A^T A)^{-1} A^T b$.
  - Least-norm solution of $Cx = d$ if C has linearly independent rows: $x = C^T (CC^T)^{-1} d$.
  - Optimality conditions for 'minimize $\|Ax - b\|^2$ subject to $Cx = d$':

$$\begin{bmatrix} A^T A & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = \begin{bmatrix} A^T b \\ d \end{bmatrix}.$$

Please go to next page...

- Newton's method.
  - Newton iteration for solving a set of nonlinear equations $f(x) = 0$: $x^+ = x - Df(x)^{-1}f(x)$.
  - Newton step at x for minimizing a function g: $v = -\nabla^2 g(x)^{-1}\nabla g(x)$.
  - Condition of sufficient decrease in line search: $g(x + tv) \leq g(x) + \alpha t \nabla g(x)^T v$
- Formulas for gradient and Hessian.
  - $g(x) = x^T P x + q^T x + r$ with P symmetric: $\nabla g(x) = 2Px + q$, $\nabla^2 g(x) = 2P$.
  - $g(x) = f(Ax + b)$:

$$\nabla g(x) = A^T \nabla f(Ax + b), \qquad \nabla^2 g(x) = A^T \nabla^2 f(Ax + b)A.$$

  - $g(x) = \sum_{i=1}^{m} r_i(x)^2$:

$$\begin{aligned} \nabla g(x) &= 2\sum_{i=1}^{m} r_i(x)\nabla r_i(x), \\ \nabla^2 g(x) &= 2\sum_{i=1}^{m} \left(r_i(x)\nabla^2 r_i(x) + \nabla r_i(x)\nabla r_i(x)^T\right). \end{aligned}$$

- Properties of matrix norm $\|A\| = \max_{x \neq 0} \|Ax\|/\|x\|$.
  - $\|\alpha A\| = |\alpha|\|A\|$ for $\alpha \in \mathbf{R}$.
  - $\|A\| \geq 0$ for all A. $\|A\| = 0$ only if $A = 0$.
  - $\|A + B\| \leq \|A\| + \|B\|$.
  - $\|A\| = \|A^T\|$.
  - $\|Ax\| \leq \|A\|\,\|x\|$ if the matrix-vector product exists.
  - $\|AB\| \leq \|A\|\,\|B\|$ if the matrix-matrix product exists.
- Condition number.
  - Definition: $\kappa(A) = \|A\|\,\|A^{-1}\|$.
  - Error bounds for $Ax = b$, $A(x + \Delta x) = b + \Delta b$:

$$\|\Delta x\| \leq \|A^{-1}\|\|\Delta b\|, \qquad \frac{\|\Delta x\|}{\|x\|} \leq \kappa(A)\frac{\|\Delta b\|}{\|b\|}.$$

- IEEE floating point numbers.
  - Floating-point numbers with base 2

$$\pm(.d_1 d_2 \ldots d_n)_2 \cdot 2^e = \pm(d_1 2^{-1} + d_2 2^{-2} + \cdots + d_n 2^{-n}) \cdot 2^e \qquad (d_1 = 1,\ d_i \in \{0, 1\}).$$

  - Machine precision: $\epsilon_M = 2^{-n}$.
  - IEEE double precision arithmetic: $n = 53$, $\epsilon_M \approx 1.1102 \cdot 10^{-16}$.