# Information on the Final Exam

1. The exam is in class, room 1302 (current classroom) from 5:10 to 7 pm on Thursday, December 16
2. The exam is closed-book. You must put away and out of reach and sight electronic devices such as phones and tablets.
3. You are allowed to bring a "cheat-sheet", which is one piece of US standard sized paper (8.5in × 11in) with anything written on it – front and back.
4. Explain all answers with a few words or sentences. Points may be deducted for correct answers that aren't explained. This applies to true/false and multiple choice questions as well as full answer questions.
5. Cross off anything you think is wrong. Points may be deducted for wrong answers even if the right answer also appears.
6. You will get 25% credit for a blank answer. You can lose those points if you give a wrong answer. Cross off anything you think is wrong.
7. Write all answers in the provided exam books. It is desirable but not required to answer the questions in order. Please write clearly and as neatly as possible under exam conditions.
8. The actual exam will have fewer questions than are given below. These are practice questions, not a practice exam.

# Practice questions, true/false

1. Suppose $A$ is an $n \times m$ matrix with $n < m$. Suppose $B = A^t A$, which is $m \times m$ but still has rank $n < m$ in exact arithmetic. If you find the SVD of $B$ using the SVD function in the `linalg` part of `numpy` in Python, then at most $n$ of the computed singular values will be different from zero.
2. If you're solving a differential equation system and the solution has distinct Lyapunov exponents, then the condition number of the problem (the problem of computing $x(t)$ from $x(0)$) grows exponentially with $t$.
3. If you're trying to minimize a function $V(x)$, and if $\nabla V(x) \neq 0$, then $-\nabla V(x)$ is a descent direction.
4. If you're trying to minimize a function $V(x)$, and if $\nabla V(x) \neq 0$, and $H$ is a symmetric matrix, then $-H \nabla V(x)$ is a descent direction.

5. If you're trying to minimize a function $V(x)$, and if $\nabla V(x) \neq 0$, and $H$ is a symmetric and positive definite matrix, then $-H\nabla V(x)$ is a descent direction.

## Practice questions, multiple choice

1. We saw that including an ill conditioned sub-problem in a solution algorithm for a well conditioned problem can lead to inaccurate results. [This was finding the eigenvalues and eigenvectors of the generator in the matrix exponential problem.] Why is this possible?

   (a) Because the ill conditioned sub-problem changes the condition number of the overall problem?

   (b) Because rounding errors in the ill-conditioned sub-problem can lead to an inaccurate solution of the sub-problem.

   (c) Because you have to adjust the overall problem input to accommodate different algorithms.

   (d) Because the ill-conditioned sub-problem was solved in single precision rather than double precision.

2. Consider two pieces of Python code for computing the matrix product $C = AB$, with $A$ and $B$ pre-defined $n \times n$ matrices.

```
------------- method 1 ------------------

      C = np.zeros([n,n])
      for i in range(n):
         for j in range(n):
            for k in range(n)
               C[i,j] += A[i,k}*B[k,j]

------------- method 2 ------------------

       C = A@B
```

   Which of the following is not a reason for method 2 to run faster than method 1 when $n$ is large?

   (a) The algorithm in method 2 uses fewer multiplications and additions than in method 1

   (b) The fact that Python is interpreted means that the lines of code in method 1 must be interpreted many times.

   (c) The matrix multiply code in method 2 uses cache memory more effectively

3. Which of the following is a reason to use Monte Carlo sampling rather than using quadrature on a product mesh? We want to compute

$$A = \int \cdots \int f(x_1, \cdots, x_d)\, dx_1 \, \cdots \, dx_d$$

The product mesh has points $x_k = (k_1\Delta x, k_2\Delta x, \cdots, k_d\Delta x)$, where $k_1, k_2, \cdots$ are integers and $x_k$ is in the domain of integration. It is called a "product mesh" because it is a sort of product of $d$ one-dimensional meshes, which consist of the one component mesh points $x_{i,j} = j\Delta x$, where $x_i$ is component $i$ of $x \in \mathbb{R}^d$.

(a) It takes less work to evaluate $f$ at a random point than at fixed grid points.

(b) It is hard to calculate the locations of product mesh points.

(c) There are too many points in the product mesh.

(d) Monte Carlo methods have a high order of accuracy.

# Practice questions, full answer

1. In the IEEE floating point standard, a 32 bit single precision floating point number has 8 exponent bits. Your answers to most of the following questions may be off by one or maybe even 2.

   (a) What is the largest integer $n$ so that $10^n$ is represented exactly in this format?

   (b) What is the largest integer $n$ so that $10^n$ is represented to within single precision floating point roundoff error?

   (c) What is the smallest $n$ so that $10^n$ is represented exactly? This $n$ may be negative.

   (d) What is the smallest $n$ so that $10^n$ is represented to within single precision floating point round error?

   (e) What is the smallest $n$ so that $10^n$ may be represented to within 10%? This has to do with de-normalized numbers.

2. Consider a matrix $\begin{pmatrix} a & b \\ b & c \end{pmatrix}$ that is close to $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Show that finding the eigenvalue of $A$ near 2 is a well conditioned problem but finding the eigenvalue near zero is ill conditioned. [This involves eigenvalue perturbation theory and the definition of condition number.]

3. Explain an adaptive procedure that tries to make the most accurate estimate possible of $f'(x)$ using a second order centered difference approximation in the presence of roundoff error. The process should be to reduce

$h$ by factors of two until the results are inconsistent with asymptotic error analysis. Is it possible to use Richardson extrapolation to make an estimate substantially more accurate than the centered difference with the smallest reliable centered difference? Asked another way: Does Richardson extrapolation improve the accuracy obtainable by second order finite differences in the presence of roundoff?

4. A code that is supposed to be 3-rd order accurate ran with $\Delta x = 10$ and $\Delta x = 5$ with results $A(10) = 328.5$ and $A(5) = 320.5$. About what $\Delta x$ should give an error less than 1?

5. In the code below we will change the value of $n$ keeping the rest the same. Assuming $n$ is not too large an integer, what values might be printed: $n-1$, $n-1$, $n$, $n+1$, $n+2$?

```
import numpy as np
L = np.pi      # Length of an interval, the mathematical pi = 3.14...
n = 100        # or 200 or 5000, but not too large
dx = L/n       # interval size
x  = 0.        # start of the interval
k  = 0         # k counts points in the interval
while ( x < L):  # step through the interval
   x += dx       # uniform grid cells
   k += 1        # count grid points
print("final count is " + str(k))
```

6. We want to find a minimum norm solution to an under-determined linear least squares problem
$$\min_{Ax=b} \|x\|_2 \ .$$
The dimensions are $x \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$ with $m < n$. Suppose $A$ is an $m \times n$ matrix with full rank.

   (a) Explain how to solve the problem in $O(n^2)$ operations using the information contained in the SVD $A = U\Sigma V^t$.

   (b) Explain how to solve the problem using the information contained in the QR factorization of $A^t$, which is $A^t = QR$, with $Q$ being $n \times n$ orthogonal and $R$ being $n \times m$ upper triangular.

7. Suppose you have the following data: $f(0) = a$, $f(h) = b$, and $f'(0) = c$. What is the best estimate of $f(2h)$ using this data? That means, what combination of data gives $f(2h)$ with the highest order of accuracy?

8. Suppose $A$ is *upper Hessenberg*, which means that $a_{jk}$ (the $(j,k)$ entry of $A$) is zero below the first sub-diagonal: $a_{jk} = 0$ if $j > k+1$. Thus, $a_{3,1} = 0$, $a_{4,2} = 0$, etc. Show that "most" such matrices have an $LU$ factorization where $L$ is bi-diagonal. What is the number of operations needed to compute $L$ and $U$? *Hint.* Write the equations for $LU = A$ and show they can be solved unless some bad ones are zero.

9. Consider the two stage Runge Kutta method for solving $\dot{x} = f(x)$ with time step $\Delta t$ and approximate solution $x_k \approx x(t_k)$, with $t_k = k\Delta t$. First, $\xi_1 = \Delta t f(x_k)$, then $\xi_2 = \Delta t f(x_k + \xi_1)$, then $x_{k+1} = x_k + \frac{1}{2}(\xi_1 + \xi_2)$. [By tradition, we write $\xi_1$ for what should be written $\xi_{1,k}$, etc.] Show that this has local truncation error consistent with a second order accurate method. That means that (in the notation of the book, starting on page 171) $\widehat{S}(x, \Delta t) - S(x, \Delta t) = O(\Delta t^3)$.

10. Consider an iteration of the form

$$x_{k+1} = \lambda x_k + (1 - \lambda)\frac{a}{x_k} .$$

Suppose that $a > 0$ and $0 < \lambda < 1$.

(a) Show that $x_* = \sqrt{a}$ is a fixed point of the iteration.

(b) Show that this fixed point is locally linearly stable, so $x_k \to x_*$ as $k \to \infty$ if $x_0 - x_*$ is small enough.

(c) Show that the iteration is locally quadratically convergent if and only if $\lambda = \frac{1}{2}$.

(d) (not part of the test, but if you're curious, the sequence converges to $x_*$ from any positive starting point if $\lambda - \frac{1}{2}$. You can see this by drawing the Newton's method picture, and the $\lambda = \frac{1}{2}$ iteration is equivalent to Newton's method for $f(x) = 0$ with $f(x) = x^2 - a$.

11. Suppose $M$ is a symmetric positive definite matrix and consider the *pre-conditioned* gradient descent algorithm

$$x_{k+1} = x_k - sM\nabla V(x_k) .$$

Here, $x \in \mathbb{R}^d$ and $V(x)$ is a smooth convex function with a non-degenerate local minimum at $x = 0$, $s > 0$ is a fixed "learning rate".

(a) Show that there is an affine transformation $y = Ax$ so that this is gradient descent in the $y$ variable.

(b) Suppose $V = \frac{1}{2}x^t Hx$ (that is, $V$ is quadratic) with $H$ symmetric and positive definite. Find a matrix related to $H$ and $M$ whose condition number controls the convergence rate. [This is a hard question for an exam, but you have the background to figure it out.]

12. A sequence is given by $f_k = x^k$ for $k = 0, \cdots, n-1$. What is the discrete Fourier transform (dft) of this sequence?

13. Suppose $x_k$ is a periodic sequence with $x_{k+n} = x_k$ for all $k$. Suppose that the mean value of $x_k$ is zero, which is

$$\sum_{k=0}^{n-1} x_k = 0 .$$

We want to find a sequence $y_k$ so that, for all $k$,

$$x_k = y_{k+1} - 2y_k + y_{k-1} \ .$$

Find an algorithm using the FFT to do this using on the order of $n \log(n)$ operations. Is the mean zero condition necessary for a solution to exist?