Brown University
DATA 1010
Fall 2018: Midterm I
Samuel S. Watson

Name:

You will have three hours to complete this exam. The exam consists of 12 written questions and one separate computational problem. You will hand in your answers to the first 12 questions and then get out your laptop to submit a solution to the last question electronically.

For the written part of the exam, no calculators or other materials are allowed, except the Julia-Python-R reference sheet. For the computational part of the exam, you may use any internet technologies which do not involve active communication with another person.

You are responsible for explaining your answer to **every** *question. Your explanations do not have to be any longer than necessary to convince the reader that your answer is correct.*

I verify that I have read the instructions and will abide by the rules of the exam: ____________________________

Problem 1 **[SETFUN]**

Suppose that Ω is a set and that A, B, and C are subsets of Ω. Which of the following is necessarily equal to

$$(A \cap B \cap C)^{\mathsf{c}}?$$

(a) $A \cup B \cup C$
(b) $A \cap B \cap C$
(c) $A^{\mathsf{c}} \cup B^{\mathsf{c}} \cup C^{\mathsf{c}}$
(d) $A^{\mathsf{c}} \cap B^{\mathsf{c}} \cap C^{\mathsf{c}}$

Solution

The complement of an intersection of sets is the union of their complements. To see this, note that the elements of $(A \cap B \cap C)^{\mathsf{c}}$ are elements $\omega \in \Omega$ with the property that it is not the case that ω is in A and B and C. Meanwhile, membership in $A^{\mathsf{c}} \cup B^{\mathsf{c}} \cup C^{\mathsf{c}}$ entails non-membership in at least one of the sets A, B, and C. Since "ω is not in all the sets" is equivalent to "there's at least one set that ω is not in", we see that $(A \cap B \cap C)^{\mathsf{c}}$ and $A^{\mathsf{c}} \cup B^{\mathsf{c}} \cup C^{\mathsf{c}}$ contain the same elements.

Final answer:

(c)

Problem 2 **[LINALG]**

(i) (Multiple choice) Suppose that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent and $\{\mathbf{v}_4, \mathbf{v}_5, \mathbf{v}_6\}$ is linearly dependent. Then

(a) $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5, \mathbf{v}_6\}$ is linearly independent
(b) $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5, \mathbf{v}_6\}$ is linearly dependent
(c) It is impossible to determine whether $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5, \mathbf{v}_6\}$ is linearly independent based on the information given.

(ii) (Multiple choice) Suppose that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent and $\{\mathbf{v}_4, \mathbf{v}_5, \mathbf{v}_6\}$ is linearly independent. Then

(a) $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5, \mathbf{v}_6\}$ is linearly independent
(b) $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5, \mathbf{v}_6\}$ is linearly dependent
(c) It is impossible to determine whether $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5, \mathbf{v}_6\}$ is linearly independent based on the information given.

Solution

(i). The answer is (b). If there is a linear dependence relation among $\mathbf{v}_4$, $\mathbf{v}_5$, and $\mathbf{v}_6$, then there is a linear dependence relation among $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_6$: we use weights of zero for $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ and the same weights for $\mathbf{v}_4$, $\mathbf{v}_5$, and $\mathbf{v}_6$ that demonstrate a linear dependence relation among those three vectors.

(ii) The answer is (c). It is possible that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5, \mathbf{v}_6\}$ is linearly independent. Indeed, any linearly independent list of six vectors has the property that the first three vectors form an independent list and the last three vectors also form an independent list.

On the other hand, it is also possible that the combined list is linearly dependent. For example, suppose that $\mathbf{v}_4 = \mathbf{v}_1$, $\mathbf{v}_5 = \mathbf{v}_2$, and $\mathbf{v}_6 = \mathbf{v}_3$.

Problem 3 **[MATALG]**

(a) Suppose that A and B are 4×4 matrices with the property that $\mathbf{a}'\mathbf{b} = 7$ whenever $\mathbf{a}'$ is a row of A and $\mathbf{b}$ is a column of B. Find the matrix product AB.

(b) Suppose that A and B are 4×4 matrices with the property that $\mathbf{a}\mathbf{b}' = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$ whenever $\mathbf{a}$ is a column of A and $\mathbf{b}'$ is a row of B. Find the matrix product AB.

Solution

(a) By the definition of the matrix-matrix product, the entries of AB are the dot products of the rows of A and the columns of B. Therefore, we have

$$AB = \begin{bmatrix} 7 & 7 & 7 & 7 \\ 7 & 7 & 7 & 7 \\ 7 & 7 & 7 & 7 \\ 7 & 7 & 7 & 7 \end{bmatrix}$$

(b) If we divide A into columns and B into rows, then we can block multiply to get

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} \mathbf{b}'_1 \\ \mathbf{b}'_2 \\ \vdots \\ \mathbf{b}'_n \end{bmatrix} = \mathbf{a}_1\mathbf{b}'_1 + \mathbf{a}_2\mathbf{b}'_2 + \cdots + \mathbf{a}_n\mathbf{b}'_n,$$

where $n = 4$. Therefore,

$$AB = \begin{bmatrix} 4 & 4 & 4 & 4 \\ 4 & 4 & 4 & 4 \\ 4 & 4 & 4 & 4 \\ 4 & 4 & 4 & 4 \end{bmatrix}.$$

Problem 4 **[EIGEN]**

The matrix $A = \begin{bmatrix} \frac{11}{25} & \frac{6}{125} & -\frac{8}{125} \\ \frac{6}{125} & \frac{977}{1250} & \frac{182}{625} \\ -\frac{8}{125} & \frac{182}{625} & \frac{1147}{1875} \end{bmatrix}$ diagonalizes as $\begin{bmatrix} 0 & -\frac{4}{5} & -\frac{3}{5} \\ \frac{4}{5} & -\frac{9}{25} & \frac{12}{25} \\ \frac{3}{5} & \frac{12}{25} & -\frac{16}{25} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 0 & \frac{4}{5} & \frac{3}{5} \\ -\frac{4}{5} & -\frac{9}{25} & \frac{12}{25} \\ -\frac{3}{5} & \frac{12}{25} & -\frac{16}{25} \end{bmatrix}$.

Which of the following is closest to A^{100}?

(a) $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ (b) $\begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{16}{25} & \frac{12}{25} \\ 0 & \frac{12}{25} & \frac{9}{25} \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

Solution

We can raise a matrix to a power by diagonalizing it as $V\Lambda V^{-1}$ and raising Λ to that power. So we get

$$\begin{aligned} \lim_{n\to\infty} A^n = V(\lim_{n\to\infty} \Lambda^n)V^{-1} &= V \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} V^{-1} \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{16}{25} & \frac{12}{25} \\ 0 & \frac{12}{25} & \frac{9}{25} \end{bmatrix}. \end{aligned}$$

Final answer:

(b)

Problem 5 **[OPT]**

The code below represents an attempt to implement gradient descent for the function $f(x, y) = (x-1)^4 + x(y-2)^2$.

```
f(x,y) = (x-1)^4 + x*(y-2)^2
df(x,y) = [4(x-1)^3 + (y-2)^2, 2x*(y-2)]
function gradient_descent(f,df,start,ϵ,threshold)
    (x,y) = start
    while norm(df(x,y)) > threshold
        x,y = [x,y] + ϵ*df(x,y)
    end
    x,y
end
gradient_descent(f,df,(1,1),0.1,1e-5)
```

(a) Find the bug. (Note: all of the syntax is correct. The mistake has to do with what is being calculated in the body of the function).

(b) Even after the bug you found in (a) is fixed, the algorithm still doesn't get anywhere close to the global minimum at $(1,2)$ if you start at $(0,0)$. However, it does find the global minimum if you start at $(1/4, 1/4)$ instead. Explain what goes wrong with the $(0,0)$ starting point.

Solution

(a) The sign in front of the ϵ should be negative, since we should move in the direction *opposite* to the gradient to minimize the function.

(b) The algorithm goes nowhere if it starts at $(0,0)$, because that is already a critical ponit.

Problem 6 **[MATDIFF]**

Suppose that A and C are $m \times n$ matrices and $\mathbf{b}$ and $\mathbf{d}$ are vectors in $\mathbb{R}^m$ such that there is a unique vector $\mathbf{x}$ which minimizes the expression

$$|A\mathbf{x} - \mathbf{b}|^2 + |C\mathbf{x} - \mathbf{d}|^2.$$

Find the minimizing vector $\mathbf{x}$ (in terms of A, $\mathbf{b}$, C, and $\mathbf{d}$). You may assume invertibility wherever convenient.

Solution

We minimize the expression by differentiating and setting the result equal to 0:

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{x}}\left[(A\mathbf{x} - \mathbf{b})'(A\mathbf{x} - \mathbf{b}) + (C\mathbf{x} - \mathbf{d})'(C\mathbf{x} - \mathbf{d})\right] = 0 &\implies \\
(A\mathbf{x} - \mathbf{b})'(A) + (C\mathbf{x} - \mathbf{d})'(C) = 0 &\implies \\
\mathbf{x}'A'A - \mathbf{b}'A + \mathbf{x}'C'C - \mathbf{d}'C = 0 &\implies \\
(A'A + C'C)\mathbf{x} = A'\mathbf{b} + C'\mathbf{d} &\implies \\
\mathbf{x} = (A'A + C'C)^{-1}(A'\mathbf{b} + C'\mathbf{d}).&
\end{aligned}$$

Since the function $\mathbf{x} \mapsto |A\mathbf{x} - \mathbf{b}|^2 + |C\mathbf{x} - \mathbf{d}|^2$ is continuous goes to ∞ as $|\mathbf{x}| \to \infty$, it has a global minimum on $\mathbb{R}^n$. Since it only has a single critical point (assuming that $A'A + C'C$ is invertible), the global minimum must occur there. Therefore, the minimizing value of $\mathbf{x}$ is $(A'A + C'C)^{-1}(A'\mathbf{b} + C'\mathbf{d})$.

Final answer:

$$(A'A + C'C)^{-1}(A'\mathbf{b} + C'\mathbf{d})$$

Problem 7 **[MACHARITH]**

Find the largest possible value of the quotient $\frac{|a-b|}{b}$, where a is a real number between 2^{-1000} and 2^{1000} and b is the `Float64` value to which a is rounded.

Solution

Since b is the nearest `Float64` to a, we know that $|a-b|$ is not larger than half the tick spacing at a. Furthermore, between any two successive powers of 2, the ratio of the `Float64` tick spacing in that interval to the smallest number in that interval is 2^{-52}. Therefore, the ratio of half the tick spacing at a point to the value of that point can be as large as $\boxed{2^{-53}}$ but no larger.

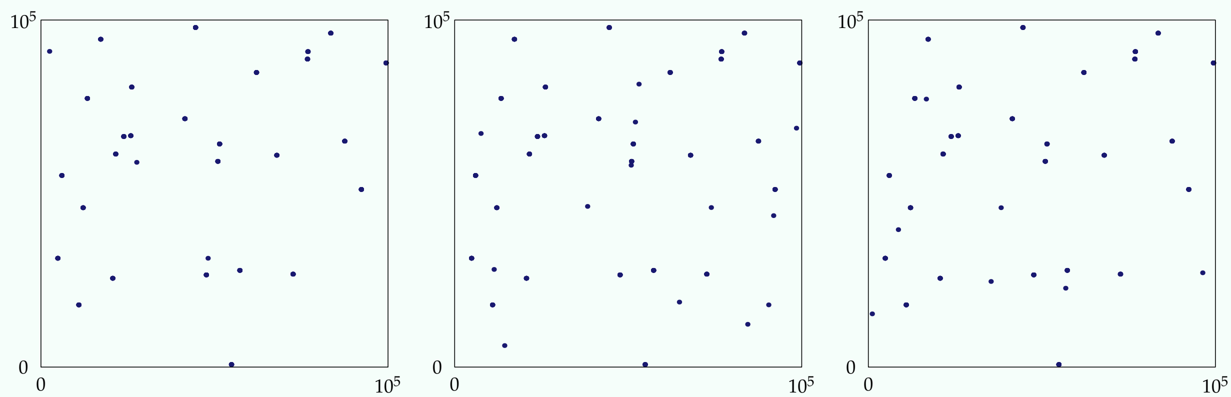
Final answer:

2^{-53}

Problem 8 **[PRNG]**

Consider the following pseudorandom number generator: begin with a four-digit number s, square s, and take second through fifth digits of the result (if the result has fewer than 5 digits, fill in with zeros on the right). For example, if we start with 2431, then the next number would be 9097, since $2431^2 = 5909761$.

If you generate the first 10,000 terms of this sequence starting from the seed 2431, split the terms into blocks of 2, and plot the resulting 5,000 pairs in the plane, you get the first figure shown below. Repeating with the seeds 7532 and 1074, you get the second and third plots.



Using appropriate terminology, discuss one or more major weaknesses of this pseudorandom number generator.

Solution

Each figure clearly contains far fewer than 10,000 dots. Therefore, the period of this PRNG is very short. This is highly undesirable, since the sequences produced by this PRNG have glaring statistical anomalies. For example, there are large intervals which will never contain a sample.

Problem 9 [NUMERROR]

(a) Explain why the last line in the code block below does not return 0, even though the functions f and g are mathematically equivalent for $x \neq 2$.

```
f(x) = (x^2 - 4) / (x-2)
g(x) = x + 2
f(2.002) - g(2.002) # returns -1.6431300764452317e-13
```

(b) Consider the problem of evaluating the (mathematical) function $f(x) = (x^2 - 4)/(x - 2)$ for values of x near 2 (but not equal to 2). This problem is

(i) well-conditioned
(ii) ill-conditioned
(iii) stable
(iv) unstable

(c) The algorithm executed by the Julia function `f` is

(i) well-conditioned
(ii) ill-conditioned
(iii) stable
(iv) unstable

(d) The algorithm executed by the Julia function `g` is

(i) well-conditioned
(ii) ill-conditioned
(iii) stable
(iv) unstable

Solution

(a) The functions `f` and `g` execute different algorithms, since no algebraic simplification is performed prior to evaluating the body of a function. Since `Float64` arithmetic is performed with rounding, there is no reason to expect that `f` and `g` should return exactly equal values.

(b) This function f is equivalent to $f(x) = x + 2$ for $x \neq 2$. The condition number of this function is $\left|\frac{x}{x+2}\right|$, which is not large when x is near 2. Therefore, f is **well-conditioned**.

(c) The algorithm executed by `f` is **unstable**, since two of the steps (subtracting 4 from a number near 4, and subtracting 2 from a number near 2) are ill-conditioned.

(d) The algorithm executed by `g` is **stable**, since it only contains one step (adding 2), and that step is well-conditioned.

Problem 10 **[COUNTING]**

You have 4 mystery novels and 6 textbooks, and you want to arrange 3 of the mystery novels and 4 of the textbooks in order on your bookshelf. If you want the mystery novels to appear all together and the textbooks to appear all together, how many ways are there to do this?

(For example, $T_6T_1T_3T_4M_1M_4M_2$ be an acceptable arrangement, while $T_3T_1M_3M_4M_1T_5T_4$ would not be. Furthermore, $T_6T_1T_3T_4M_1M_2M_3$ and $M_1M_2M_3T_6T_1T_3T_4$ count as different arrangements.)

Solution

There are $\binom{4}{3}$ ways to choose the mystery novels to put on the shelf and $\binom{6}{4}$ ways to choose the textbooks. Once the books are chosen, there are 3! ways to choose the order for the novels, 4! ways to choose the order for the textbooks, and 2 ways to choose whether the novels or textbooks come first. So altogether, we get

$$\left(\frac{4 \cdot 3 \cdot 2}{3!}\right)(3!)\left(\frac{6 \cdot 5 \cdot 4 \cdot 3}{4!}\right)(4!)\,(2) = 17280$$

ways to arrange the books.

Problem 11 **[PROBSPACE]**

Suppose that E and F are events with the property that if E does not occur, then F occurs.

(a) Translate the real-world statement "if E does not occur, then F occurs" into a statement about E and F as subsets of the sample space Ω.

(b) Show that $\mathbb{P}(E) + \mathbb{P}(F) \geq 1$ using the probability space axioms and subadditivity.

Solution

(a) The statement "if E does not occur, then F occurs" means that every outcome in the complement of E is contained in the set F. In other words, $E^{\mathsf{c}} \subset F$.

(b) Since $E^{\mathsf{c}} \subset F$, every element of Ω is either in E or in F. In other words, we have $E \cup F = \Omega$. Therefore,

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(E \cup F) \leq \mathbb{P}(E) + \mathbb{P}(F),$$

by subadditivity.

Problem 12 [JULIA]

Write a Julia function `zigzag` which accepts a vector as an argument and returns `true` if the second entry is greater than the first, the third is *less* than the second, the fourth is greater than the third, and so on.

```
@assert zigzag([1,4,3,7,-1,5,2]) == true
@assert zigzag([1,4,0,1,-2,0,16,3]) == false
@assert zigzag([7,1,7,1]) == false
@assert zigzag([4,4,2,3,0,10]) == false
```

Solution

One way to do it is loop through the array and return `false` whenever a violation of the zigzag condition is discovered.

Julia

```
function zigzag(a)
   for i=2:length(a)
        if iseven(i) && a[i-1] ≥ a[i] || isodd(i) && a[i-1] ≤ a[i]
            return false
        end
   end
   true
end
```

Problem 13 **[JULIA]**

The Babylonian method for approximating $\sqrt{x}$ works as follows. We begin with $t_0 = 1$, and for $n \geq 1$, we define $t_n = \frac{1}{2}(t_{n-1} + x/t_{n-1})$. As $n \to \infty$, t_n converges to $\sqrt{x}$ quite quickly.

(a) Write a Julia function `babylonsqrt` which takes `x` as an argument and computes the 20th iterate of the above sequence (in other words, t_{20}) for the given value of `x`.

```
@assert isapprox(babylonsqrt(5),sqrt(5))
```

(b) Show that if you apply the Babylon square root algorithm with 20 iterations to `[5 1; 0 5]`, you get

$$\begin{bmatrix} 2.2361 & 0.2236 \\ 0.0 & 2.2361 \end{bmatrix}.$$

You will have to alter your function a bit so that it works for matrices. For example, the iteration should start at `I` instead of 1, and x/t should be replaced with `x*inv(t)`.

(c) Interpret the top-right entry 0.2236.

Solution

The version for numbers looks like this:

```
function babylonsqrt(x::Real)
    t = 1
    for i=1:20
        t = 1/2 * (t + x/t)
    end
    t
end
```

And for matrices:

```
function babylonsqrt(x::AbstractArray)
    t = I
    for i=1:20
        t = 1/2 * (t + x*inv(t))
    end
    t
end
```

The top right entry is equal to the derivative function $1/(2\sqrt{x})$ evaluated at $x = 5$. We can see that automatic differentiation manages to compute the derivative of the square root function without having any access to the rule $(\sqrt{x})' = 1/(2\sqrt{x})$.

Additional space

Additional space