BROWN UNIVERSITY
DATA 1010
PRACTICE MIDTERM I
INSTRUCTOR: SAMUEL S. WATSON

Name:

You will have three hours to complete this exam. The exam consists of 12 written questions. No calculators or other materials are allowed, except the Julia-Python-R reference sheet.

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 **[JULIA]**

Write a Julia function called `appeartwice` which accepts two arguments: a vector `x` and a number `a`. The function should return `true` if `x` has two or more entries which are equal to `a` and `false` otherwise.

```
@assert appeartwice([1,4,1,0,2,1],1) == true
@assert appeartwice([-3,2,-5,7,1],-5) == false
@assert appeartwice([-3,2,-5,7,1],11) == false
```

Make your code as close to correct as you can, but minor syntax errors will be disregarded in the grading.

Solution

Here's one approach using a loop:

```
function appeartwice(x,a)
    ctr = 0
    for entry in x
        if entry == a
            ctr += 1
            if ctr ≥ 2
                return true
            end
        end
    end
    false
end
```

Problem 2 **[LINALG]**

Let us say that a vector in a list of vectors is *redundant* if it can be deleted from the list without changing the span of the list. Show that if a list of nonzero vectors is linearly dependent, then the number of redundant vectors is at least two.

Solution

A linearly dependent list of vectors satisfies an equation of the form

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}, \tag{2.1}$$

where one or more of the c_i's are nonzero. If exactly one of the weights were zero, then the corresponding vector would have to be the zero vector. So for at linearly dependent list of nonzero vectors, there must be at least two weights in (2.1) which are nonzero.

Furthermore, any vector $\mathbf{v}_i$ whose weight in (2.1) is nonzero may be dropped from the list without changing its span, because we can solve (2.1) for $\mathbf{v}_i$ and thereby express it in terms of the other vectors. So any linear combination of all the vectors can be expressed as a linear combination not involving $\mathbf{v}_i$.

Problem 3 **[MATALG]**

(a) Suppose that A is an $m \times n$ matrix. Explain why a vector $\mathbf{x}$ is orthogonal to the span of the columns of A if it is in the null space of the transpose of A.

(b) Suppose that A is a 10×5 matrix and that $\mathbf{b}$ is a vector which is in the span of the columns of A. Explain why the equation $A\mathbf{x} = \mathbf{b}$ cannot be solved by left-multiplying by A^{-1} to obtain $\mathbf{x} = A^{-1}\mathbf{b}$.

(c) Suppose A is an $n \times n$ matrix and that $\mathbf{b}$ is a vector in $\mathbb{R}^n$. Solve the matrix equation $A\mathbf{x} + \mathbf{b} = \mathbf{x}$ for $\mathbf{x}$ (you may assume matrix invertibility wherever convenient).

Solution

(a) By definition of matrix multiplication, a vector is in the null space of a matrix if its dot product with every row of the matrix is equal to zero. Therefore, the null space of A' is equal to the orthogonal complement of the span of the columns of A.

(b) No non-square matrix has an inverse, so it does not make sense to consider multiplying by the inverse of A.

(c) We subtract $\mathbf{b} + \mathbf{x}$ from both sides to obtain

$$(A - I)\mathbf{x} = -\mathbf{b}.$$

Therefore, $\mathbf{x} = -(A - I)^{-1}\mathbf{b}$.

Problem 4 **[EIGEN]**

Recall that eigenvectors corresponding to different eigenvalues are linearly independent (in other words, if $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are eigenvectors with eigenvalues $\lambda_1, \ldots, \lambda_n$, and if no pair of the λ_i's are equal, then $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a linearly independent list). Using this fact, explain why an $n \times n$ matrix has at most finitely many eigenvalues.

Solution

If an $n \times n$ matrix has k eigenvalues, then it has k linearly independent eigenvectors. Since there can be at most n linearly independent eigenvectors in $\mathbb{R}^n$, we have $k \leq n$. Therefore, an $n \times n$ matrix has at most n eigenvalues.

Problem 5 **[OPT]**

Suppose $\mathbf{b} \in \mathbb{R}^n$. For each $\lambda \in \mathbb{R}$, consider the problem of finding the value of $\mathbf{x} \in \mathbb{R}^n$ which minimizes the expression

$$|\mathbf{x} - \mathbf{b}| + \lambda |\mathbf{x}|^2$$

Discuss, qualitatively, the behavior of the solution of this optimization problem as λ ranges over the interval $(0, \infty)$. (Note: do not try to differentiate; approach this one qualitatively from start to finish.)

Solution

When λ is very small, the minimizer will be close to $\mathbf{b}$, since that is the value of $\mathbf{x}$ which minimizes the first term. When λ is very large, the minimizer will be close to the zero vector, since the second term would be very large unless $\mathbf{x}$ is small.

Problem 6 **[MATDIFF]**

Suppose that A is an $n \times n$ matrix and that $\lambda \in \mathbb{R}$. Differentiate $A\mathbf{x} + \lambda\mathbf{x}$ with respect to $\mathbf{x}$. Show that the resulting matrix has nonzero determinant for almost all real values of λ (decide on a meaning for "almost all" and state it in your answer).

Solution

The derivative of $A\mathbf{x}$ with respect to $\mathbf{x}$ is A, while the derivative of $\lambda\mathbf{x} = \lambda I\mathbf{x}$ is λI.

The determinant of $A + \lambda I$ is zero for at most n values of λ, since $A + \lambda I$ is invertible unless $-\lambda$ is an eigenvalue of A (and we established in an earlier problem that A has at most n eigenvalues).

Problem 7 **[MACHARITH]**

Which of the following operations results in a number which is exactly equal to `11.0` when evaluated in `Float64` arithmetic?

1. `11.0 + 0.5^30`
2. `11.0 + 0.5^51`
3. `sum([0.125 for i=1:88])`
4. `100.0 + 0.5^48 - 100.0 + 11.0`

Solution

The tick spacing between 8 and 16 is 8 times the tick spacing between 1 and 2. Therefore, the difference between 11 and the next representable `Float64` is $2^{-52}2^3 = 2^{-49}$. Since 0.5^{30} is much larger than this number, `11.0 + 0.5^30` is not equal to `11.0` in `Float64` arithmetic.

By the same token, `11.0 + 0.5^51` is equal to `11.0` in `Float64` arithmetic. It's only 25% of the way from 11.0 to the next representable number.

`sum([0.125 for i=1:88])` is exactly equal to 11, since $0.125 = 1/8$ and all its multiples up to 11.0 are exactly representable.

`100.0 + 0.5^48` is equal to `100.0` because the tick spacing between 64 and 128 is $2^6 2^{-52} = 2^{-46}$. Therefore, subtracting `100.0` from this number gives `0.0`, and adding `11.0` gives `11.0`.

Problem 8 **[NUMERROR]**

Your friend observes that they were able to calculate $A\mathbf{x}$ with error significantly less than $\kappa(A)\epsilon_{\text{mach}}$ (where A is an $m \times n$ matrix and $\mathbf{x}$ is a vector in $\mathbb{R}^n$). Without knowing further details regarding the A and $\mathbf{x}$ values your friend is using, give two reasons why this might have been the case.

Solution

One possibility is that the entries of the matrix and the vector were such that many or all of the operations involved in calculating $A\mathbf{x}$ could be performed exactly. The condition number of a problem measures how it magnifies relative errors, but a relative error of exactly zero still yields a relative error of zero.

The other issue is that the condition number of a matrix is defined to be the *worst-case scenario* relative error magnification factor. This world-case scenario only occurs if $\mathbf{x}$ is in the direction of the column of V corresponding to the least singular value (where $U\Sigma V'$ is the SVD of A), and if the error is in the direction of the column of V corresponding to the greatest singular value. If $\mathbf{x}$ is pointing in a different direction, then we wouldn't necessary expect a relative error to be as large as $\kappa(A)\epsilon_{\text{mach}}$.

Problem 9 **[PRNG]**

Suppose $X_0 \in [0,1]$, and for $n \geq 1$ we define $X_n = \text{mod}(\pi + X_{n-1}, 1)$, where $\text{mod}(x, 1)$ is the difference between x and the greatest integer which is less than or equal to x (so $\text{mod}(5.62, 1) = 0.62$, for example).

(a) Does this sequence have a finite period, and if so, what is the period?

(b) If the values of the sequence are computed iteratively in `Float64` arithmetic rather than real arithmetic, give an upper bound on the period of the resulting sequence.

(c) If we think of the (`Float64`) values $X_0, X_1, X_2, \ldots$ as the output of a pseudorandom number generator, is this PRNG cryptographically secure?

Solution

(a) The sequence is not periodic. If the jth term and the kth term were equal, then $\pi(k-j)$ would differ by an integer. In other words, π would be rational.

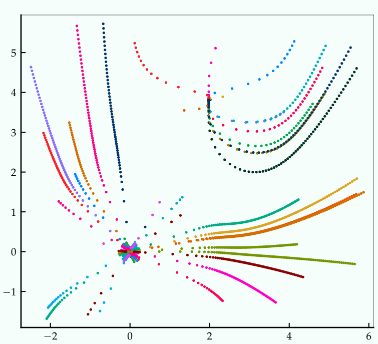
(b) There are 2^{64} `Float64` values, so the sequence must revisit the same value more than once in the first $2^{64}+1$ iterations. Since each term is computed as a function of the term before it, the sequence will repeat from there. Therefore, the repetition block has length no greater than $2^{64}+1$.

(Note: better upper bounds are possible. For example, only a quarter of the `Float64` values are between 0 and 1.)

(c) This PRNG is not cryptographically secure. A malicious agent could easily look at the terms of the sequence and figure out that the sequence is adding π each time and modding out by 1.

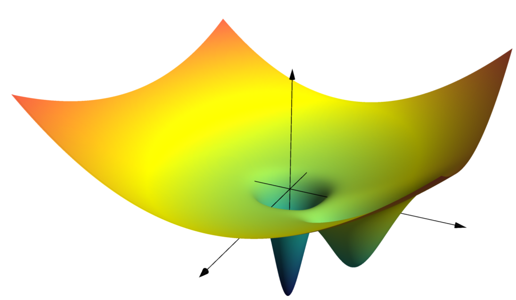
Problem 10 **[NUMOPT]**

Several (plain vanilla) gradient descent trajectories are shown for a function $f : \mathbb{R}^2 \to \mathbb{R}$, starting from various points in the square $[-3,6] \times [-3,6]$. Describe the graph of f. (For example, how many local minima does it have, and where is the graph steepest?)



Solution

The graph appears to have two local minima: one at $(0,0)$ and another at $(2,4)$. The gradient descent iterates proceed gradually toward one of the local minima, *accelerating* when they get close to the actual minimum. This means that the graph slopes gently toward the origin throughout most of the box, but with a fairly sharp bowl-shaped indentation around each local minimum (see the graph on the right). The one at the origin is steeper than the other one, which you can tell since the gradient descent jumps are larger.



Problem 11 **[PROBSPACE]**

Let $(\Omega, \mathbb{P})$ be a probability space. Use the axioms of probability to show that $\mathbb{P}(A \cap B) \leq \mathbb{P}(B)$ for any events A and B.

Solution

Since B can be written as a disjoint union $(A \cap B) \cup (A^{\mathsf{c}} \cap B)$, we have

$$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^{\mathsf{c}} \cap B) \geq \mathbb{P}(A \cap B),$$

since $\mathbb{P}(A^{\mathsf{c}} \cap B) \geq 0$.

Problem 12 **[CONDPROB]**

Three cards are drawn without replacement from a well-shuffled standard deck. Find the conditional probability that the cards are all diamonds given that they are all red cards. (Note: 13 of the cards are diamonds, 26 of the cards are red, and all of the diamonds are red).

Solution

The probability that all of the cards are red is

$$\frac{1}{2} \cdot \frac{25}{51} \cdot \frac{24}{50},$$

and the probability that all of the cards are diamonds is

$$\frac{1}{4} \cdot \frac{12}{51} \cdot \frac{11}{50}.$$

Therefore, the conditional probability is

$$\frac{1}{2} \cdot \frac{12}{25} \cdot \frac{11}{24} = \frac{11}{100}.$$

Final answer:

$$\frac{11}{100}$$