**DATA 1010**
**Problem Set 3**
**Due 28 September 2018 at 11 PM**

## Problem 1

Calculate, by hand, the gradient and Hessian of the function shown below. Show that the values returned by the `ForwardDiff` package are correct.

```
using ForwardDiff
f(x,y) = x^2 + y^2 - 2y
f(v::Vector) = f(v...) # equivalent to f(v[1],v[2])
x = [1.5,-3.25]
ForwardDiff.gradient(f,x)
ForwardDiff.hessian(f,x)
```

## Solution

The gradient is $[2x, 2y-2]$. Therefore, the gradient at $[1.5, -3.25]$ is $[-3.0, -8.5]$. This is indeed the value returned by `ForwardDiff.gradient`.

The Hessian is

$$\begin{bmatrix} \partial_x^2 f & \partial_{xy} f \\ \partial_{xy} f & \partial_y^2 f \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

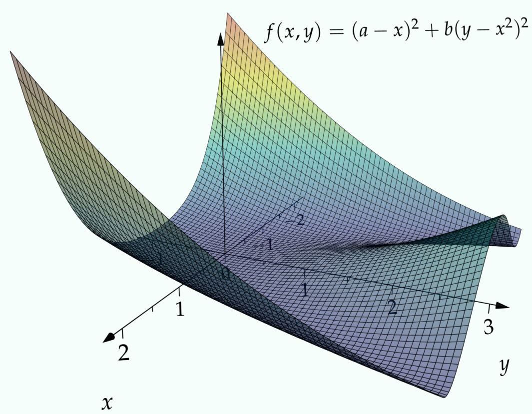
This is the matrix returned by `ForwardDiff.hessian`.

## Problem 2

The *Rosenbrock* function is defined by

$$f(x,y) = (a-x)^2 + b(y-x^2)^2,$$

where $a$ and $b$ are positive constants.




(a) Find the point $(x,y)$ where $f$ has a global minimum.

(b) Show that $f$ does not have a local minimum anywhere else.

(c) Implement (from scratch) the gradient descent algorithm for this function starting from $(0,0)$, with $a = 1$ and $b = 100$, and try some different learning rates. Show that finding the minimum is challenging.

(d) Show that the Julia package `Optim` can nevertheless handle this function just fine. Hint: the code to do this can be found on the GitHub landing page for the Optim.jl package (it's 3 lines).

Notes on (c): you can calculate the gradient of $f$ by hand if you want, and you might find that you need to use a small learning rate. Also, you probably want to use a `for` loop with a controlled number of iterations, since `while` loops might fail to terminate.

## Solution

(a) Since $f$ evaluates to 0 at $(a, a^2)$ and to a strictly positive number at any other point (because the expression for $f$ is a sum of squares), the global minimum occurs at $(a, a^2)$.

(b) The gradient of $f$ is
$$[2(a-x)+2b(y-x^2)(-2x), 2b(y-x^2)].$$
Setting this equal to $[0,0]$, we find that $y = x^2$ (from the second components) and then that $a - x = 0$ (from the first components). Therefore, the only point where $\nabla f = 0$ is $(a, a^2)$. Since $f$ is differentiable everywhere, it has zero gradient at any local extremum. Therefore, the only local minimum is at the point where $f$ has a global minimum.

(c) Here's an implementation of gradient descent for this function.

```
gradf(x,y,a=1,b=100) = [-2*(a-x) - 4b*x*(y-x^2), 2*b*(y-x^2)]
function descend(gradf,p0,ϵ,n)
    p = p0
    for i=1:n
        p = p - ϵ*gradf(p...)
    end
    p
end
descend(gradf,[0,0],0.001,1000000)
```

A small learning rate is required to avoid ending up with `NaN`'s, and quite a few iterations are required because the convergence is slow. However, we do eventually end up near the minimum at $(1, 1)$.

(d) Running

```
using Optim
rosenbrock(x) =  (1.0 - x[1])^2 + 100.0 * (x[2] - x[1]^2)^2
result = optimize(rosenbrock, zeros(2), BFGS())
```

gives a minimizer of $[0.9999999926033423, 0.9999999852005353]$, which is very close to $[1, 1]$.

## Problem 3

Suppose that five dice are rolled simultaneously. The result of the roll is identified as: all different (like 5, 6, 4, 3, 1), one pair (like 2, 3, 4, 2, 5), two pairs (6, 6, 3, 1, 3), three-of-a-kind (1, 4, 3, 4, 4), full house (5, 4, 5, 4, 4), four-of-a-kind (3, 1, 3, 3, 3), or five-of-a-kind (2, 2, 2, 2, 2).

(a) Show that the probability of five-of-a-kind is approximately 0.08%.

(b) Show that the probability of a full house is approximately 3.86%.

## Solution

(a) There are $6^5$ total possibilities, and only six of them are five-of-a-kind. Since each outcome is equally likely, the probability is
$$6/6^5 = 7.716 \times 10^{-4} \approx 0.08\%.$$

(b) We can think of producing a full house by choosing the number to appear 3 times and the number to appear 2 times (4 and 6, let's asy), and then choosing the two positions for the number that appears twice (like the first and third positions, for the outcome 6,4,6,4,4). All together, there are $6 \times 5 \times \binom{5}{3} = 300$ full houses. So the probability is
$$300/6^5 = 0.0385802469136 \approx 3.86\%.$$

## Problem 4

Consider the following random experiment: you flip a coin, and if it turns up heads, you roll a die. If it turns up tails, then you draw a ball from an urn containing balls labeled 1 to 11.

Define a sample space $\Omega$ for this random experiment.

**Solution**

The set of outcomes is

$$\Omega = \{(\mathtt{H},1),(\mathtt{H},2),(\mathtt{H},3),(\mathtt{H},4),(\mathtt{H},5),(\mathtt{H},6),(\mathtt{T},1),(\mathtt{T},2),(\mathtt{T},3),(\mathtt{T},4),(\mathtt{T},5),(\mathtt{T},6),(\mathtt{T},7),(\mathtt{T},8),(\mathtt{T},9),(\mathtt{T},10),(\mathtt{T},11)\}.$$

**Problem 5**

Suppose that $E$ is an event. Using the axioms of a probability measure (Theorem 6.2.1), show that $\mathbb{P}(E^{\mathsf{c}}) = 1 - \mathbb{P}(E)$.

**Solution**

Since $E$ and $E^{\mathsf{c}}$ are disjoint, the additivity property tells us that

$$\mathbb{P}(E \cup E^{\mathsf{c}}) = \mathbb{P}(E) + \mathbb{P}(E^{\mathsf{c}}).$$

Since $E \cup E^{\mathsf{c}} = \Omega$, this implies that

$$1 = \mathbb{P}(E) + \mathbb{P}(E^{\mathsf{c}}).$$

Subtracting $\mathbb{P}(E)$ from both sides yields the desired result.

**Problem 6**

Suppose that $E$, $F$, and $G$ are events. Come up with events $\widetilde{E}, \widetilde{F}$, and $\widetilde{G}$ which are pairwise disjoint and which satisfy

$$E = \widetilde{E}, \quad E \cup F = \widetilde{E} \cup \widetilde{F}, \quad \text{and} \quad E \cup F \cup G = \widetilde{E} \cup \widetilde{F} \cup \widetilde{G}.$$

**Solution**

The idea is to remove from each set any elements that belong to preceding sets: we define $\widetilde{E} = E$, $\widetilde{F} = F \setminus E$, and $\widetilde{G} = G \setminus (F \cup E)$. These sets are pairwise disjoint, since any $\omega \in E \cup F \cup G$ appears in exactly one of the sets $\widetilde{E}, \widetilde{F}, \widetilde{G}$ (namely, the first one it appears in, with respect to the order $E, F, G$).

**Problem 7**

Suppose that $A$ is the event that the high temperature in Providence next Tuesday is at least 65 degrees Fahrenheit, let $B$ be the event that the high temperature in Providence next Tuesday is at least 60 degrees Fahrenheit, and let $C$ be the event that it rains in Providence next Tuesday. Write each of the following events using the sets $A$, $B$, and $C$ and the operations $\cap$, $\cup$, and $^{\mathsf{c}}$.

(a) It will be less than 60 degrees all day and rainy next Tuesday in Providence.

(b) The high temperature in Providence next Tuesday will be at least 60 degrees but not as high as 65 degrees.

(c) In Providence next Tuesday, it will either be dry or warm (where warm is defined to mean "daily high of at least 65 degrees").

(d) The daily high temperature in Providence next Tuesday will be higher than 65 degrees and less than 60 degrees.

**Solution**

(a) Since $A^{\mathsf{c}}$ and $C$ must both occur, this event is $A^{\mathsf{c}} \cap C$.

(b) Since $B$ must occur and $A$ must not, this event is $B \setminus A$.

(c) We require either $C^{\mathsf{c}}$ or $A$ to occur. Therefore, this event is $A \cup C^{\mathsf{c}}$.

(d) We require both $A$ and $B^{\mathsf{c}}$, so this event is $A \cap B^{\mathsf{c}}$. In this instance, $A \subset B$, so $A \cap B^{\mathsf{c}} = \varnothing$ (either answer is acceptable).

**Problem 8**

The matrix $A_1 = \begin{bmatrix} -2 & -6 & 7 \\ -2 & -2 & 2 \\ -5 & 5 & 5 \end{bmatrix}$ has exactly one eigenvalue (approximately $-4.532$), while $A_2 = \begin{bmatrix} -6 & -1 \\ 5 & 0 \end{bmatrix}$ has two eigenvalues ($-5$ and $-1$). Find all of the eigenvalues of the matrix

$$A = \begin{bmatrix} -2 & -6 & 7 & 0 & 0 \\ -2 & -2 & 2 & 0 & 0 \\ -5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & -6 & -1 \\ 0 & 0 & 0 & 5 & 0 \end{bmatrix}.$$

**Solution**

We claim that $A$ has three eigenvalues: $\approx -4.532$ and $-5$ and $-1$. We will show that these are eigenvalues, and we will show that these are the only ones.

If $[v_1, v_2, v_3]$ is an eigenvector of $A_1$ with eigenvalue $\lambda$, then $A[v_1, v_2, v_3, 0, 0] = \lambda[v_1, v_2, v_3, 0, 0]$, so $\lambda$ is an eigenvalue of $A$ as well. Similarly, any eigenvalue of $A_2$ is also an eigenvalue of $A$.

Conversely, if $\mathbf{v} = [v_1, v_2, v_3, v_4, v_5]$ is an eigenvector of $A$ with eigenvalue $\lambda$, then block multiplying tells us that $A\mathbf{v} = [A_1[v_1, v_2, v_3], A_2[v_4, v_5]]$, which implies that $[v_1, v_2, v_3]$ is either the zero vector or an eigenvector of $A_1$ with eigenvalue $\lambda$, and similarly for $[v_4, v_5]$ and $A_2$. Therefore, every eigenvalue of $A$ is an eigenvalue of either $A_1$ or $A_2$.

**Problem 9**

Suppose that $A$ is a $5 \times 5$ diagonalizable matrix with eigenvalues $-1, -1, -1, +1$, and $+1$. Show that $A^2 = I$.

**Solution**

Since $A$ is diagonalizable, we have $A = V\Lambda V^{-1}$ for some matrix $V$, where $\Lambda$ is a diagonal matrix with entries $-1, -1, -1, +1$, and $+1$ along the diagonal. Note that $\Lambda^2 = I$, since each diagonal entry square to give 1. Therefore,

$$A^2 = V\Lambda V^{-1}V\Lambda V^{-1} = V\Lambda^2 V^{-1} = VV^{-1} = I.$$

Geometrically, $A$ acts as a reflection along three of the five axes represented by the columns of $V$. Reflecting twice is the same as doing nothing, so $A^2 = I$.

**Problem 10**

Write a Julia function which accepts a two-column array as an argument and returns the number of rows for which the first column contains a number strictly greater than 200 and the second column contains the string `"blue"`.

```
M = [370.512 "red"; 937.542 "blue"; 782.404 "blue"; 697.21 "blue";
     154.13 "blue"; 819.013 "red"; 568.343 "red"; 928.226 "red";
     947.238 "red"; 656.98 "blue"]
myrowcount(M) # should return 4
```

**Solution**

Here's a solution with a loop:

```julia
function myrowcount(M)
    ctr = 0
    for i=1:size(M,1)
        if M[i,1] > 200 && M[i,2] == "blue"
            ctr += 1
        end
    end
    ctr
end
```

Here's a one-line solution using elementwise array operations:

```julia
myrowcount(M) = sum((M[:,1] .> 200) .* (M[:,2] .== "blue"))
```