Bob Ghosh, 42039157, j0k0b

Assignment 3, CPSC 406, February 7

Late Days:

Taken: 1, during the weekend.
Submitted: Saturday, Feb-08.

Colloborations:

Shikhar Nandi, 51931153: Help with Questions 1,2 and 3.

Libraries

In [200]:
```
using Plots
using LinearAlgebra
using Printf
```

Question 1:

a:

i

$f(x_1, x_2) = 2x_2^3 - 6x_2^2 + 3x_1^2 x_2.$
$\nabla f(x_1) = 6x_1 x_2$
$\nabla f(x_2) = 6x_2^2 - 12x_2 + 3x_1^2$

$$\therefore Let \;\; \nabla f(x) = \begin{bmatrix} 6x_1 x_2 \\ 6x_2^2 - 12x_2 + 3x_1^2 \end{bmatrix} = 0.$$
When $\nabla f(x_1) = 6x_1 x_2 = 0 => x_1 = 0$ And/Or $x_2 = 0.$

Case 1: $x_1 = 0$
$6x_2^2 = 12x_2 => x_2 = 2$

Case 2: $x_2 = 0$
$3x_1^2 = 0 => x_1 = 0$

Case 3: $x_1 = 0 \cap x_2 = 0$
Default Case

Hence, the two stationary points are: ***(0,0) and (0,2).***

ii

From above, we can deduce that:
$$\nabla^2 f(x_1) = \begin{bmatrix} 6x_2 & 6x_1 \\ 6x_1 & 12x_2 - 12 \end{bmatrix}$$

Discriminant of $\nabla^2 f(x_1 : 0, x_2 : 0) = \begin{bmatrix} 0 & 0 \\ 0 & -12 \end{bmatrix}$ [1] is 0.

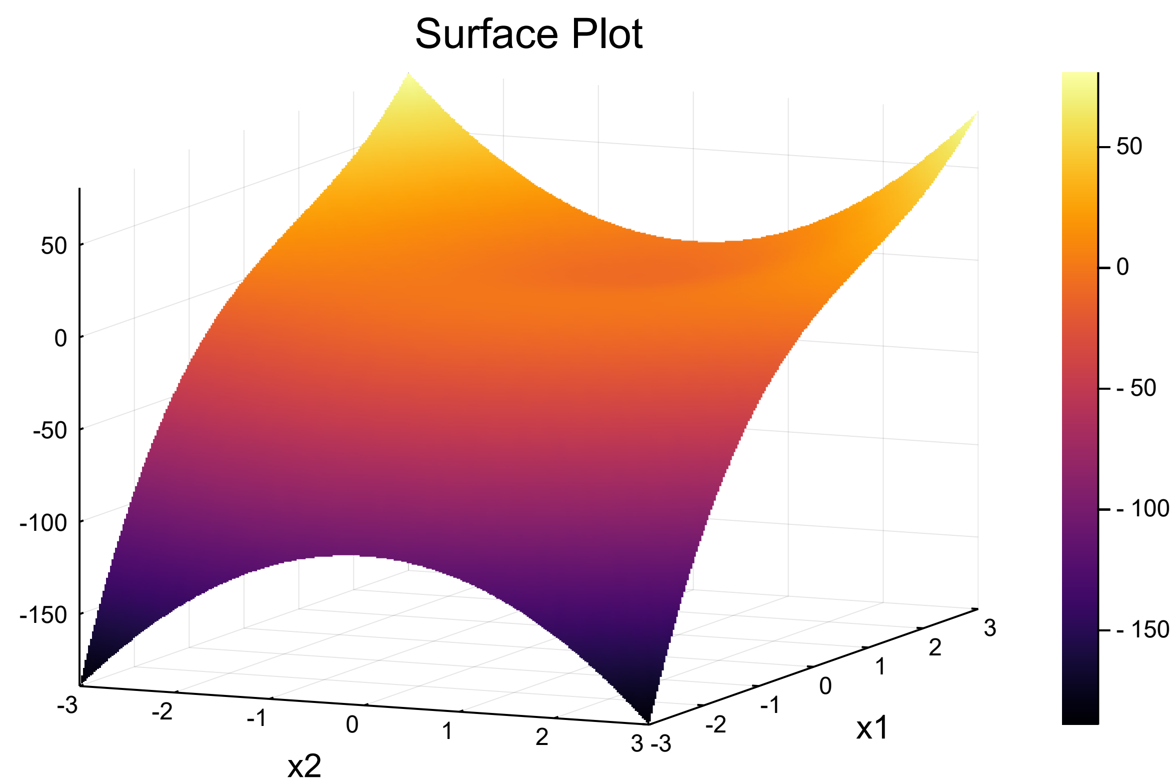
Discriminant of $\nabla^2 f(x_1 : 0, x_2 : 2) = \begin{bmatrix} 12 & 0 \\ 0 & 12 \end{bmatrix}$ [2] is 144.

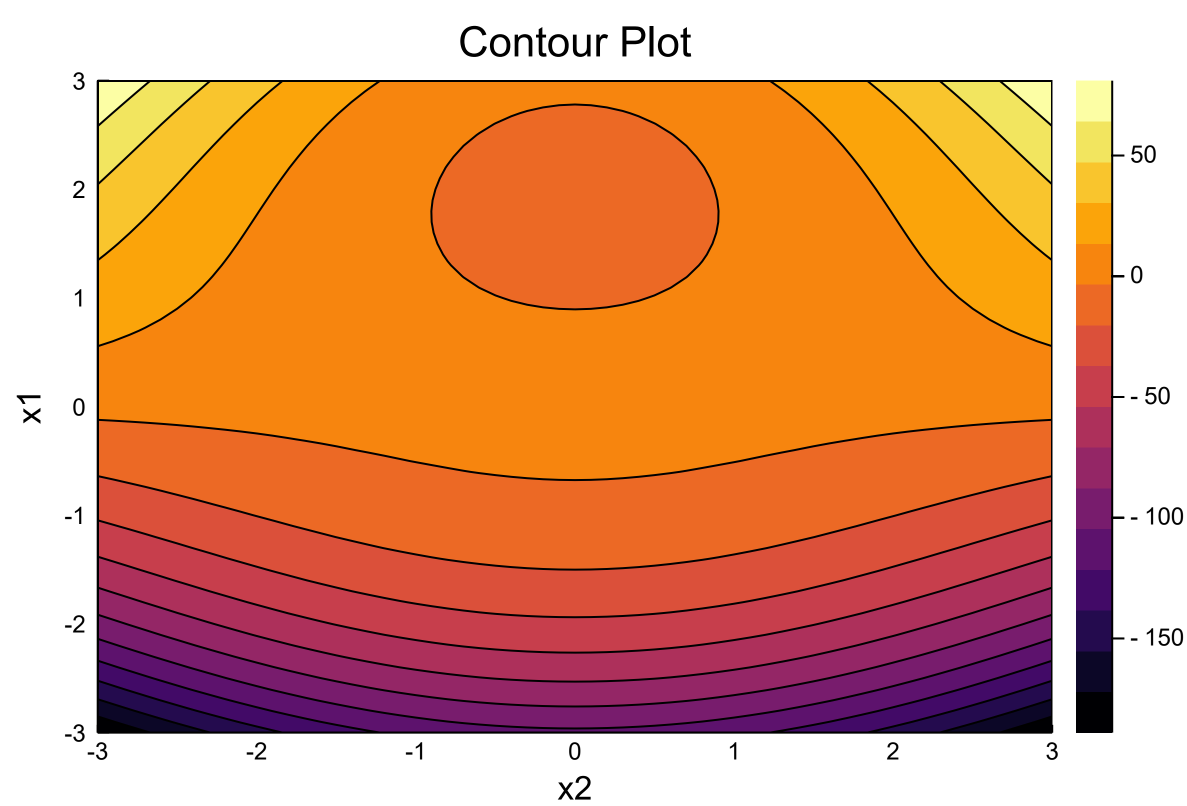
In [1], we can see that the eigenvalues are $\lambda_1 = 0 \; and \; \lambda_2 = -12$. Thus, the Hessian in [1] is Non-Positive. We also find the Discriminant to be 0, hence, this is a **saddle point**.

In [2], we can see that the eigenvalues are $\lambda_1 = 12 \; and \; \lambda_2 = 12$. Thus, the Hessian in [2] is Positive. Also, the discriminant at this point is > 0, so it's a minima. Since the domain of this function is R^2, therefore this point is a **non-strict local minima**.

iii

```
In [154]: # y = 2*x2^3-6*x2^2+3*(x1^2)*x2;
x1 = range(-3,stop = 3, length = 61);
x2 = range(-3,stop = 3, length = 61);
f(x1,x2) = 2*x2^3-6*x2^2+3*(x1^2)*x2;
surf = surface(x1,x2,f, legend = true,
    title = "Surface Plot", xaxis = "x2", yaxis = "x1");
cont = contour(x1,x2,f, fill = true,
    title = "Contour Plot", xaxis = "x2", yaxis = "x1");
display(plot(surf));
display(plot(cont));
```





From both of the plots we can deduce that at $(0, 0)$ there is no change, and since this is a critical point, we can state that this is a **saddle point**. On the other hand, we can say that at $(0, 2)$ we see a change, and thus we can conclude from the visual depth that this point is a **local minima**.

iv

From everything above, we can state that (0,0) is **not an extrema**, while (0,2) is a **local minima**.

b:

We are given that $f(x) = \frac{1}{2}x^T Ax + b^T x + c$; where $A \in R^{n\times n}$ is PSD ($A \succeq 0$), $b \in R^n$ and $c \in R$.

Firstly, we take the gradient of $f(x)$
$\nabla f(x) = Ax + b^T$; and we will show that $b \in range(A)$.
Now, if we set this to 0, we get $Ax = -b^T$. Let's say we have $\tilde{x}$ that solves the equation, such that $\nabla f(\tilde{x}) = 0$. We already know that A is PSD, $\therefore$ $\nabla^2 f(x) = A$ is also a PSD. From both of the aforementioned facts, we can conclude that $f(x) \geq f(\tilde{x}) > -\infty; \forall x \in R^n$.

Now, for a more formal proof, let us assume that $b \notin range(A)$. We already know that the main function is bounded, and so there must be a $\tilde{x}$ such that $f(x) \geq f(\tilde{x}) > -\infty; \forall x \in R^n$. And since $\tilde{x}$ is a solution, $\therefore$ $\nabla f(\tilde{x}) = 0$ and $\nabla^2 f(\tilde{x})$ is a PSD.
When we compute the $\nabla f(x) = 0$, we get $Ax = -b^T$. But, we already know that there is $\tilde{x}$ exists, and $\nabla f(\tilde{x}) = 0$. Thus we hit a contradiction to our original assumption for this part, that $b \notin A$, which makes it impossible to happen.

Thus, now we can finally conclude that f is bounded below over R^n iff $b \in range(A)$.

Question 2:

$A \in R^{m \times n}$ and $B \in R^{m \times n}$

Let $A = \begin{bmatrix} A_{1,1} & A_{1,2} & \dots & A_{1,n} \\ A_{2,1} & A_{2,2} & \dots & A_{2,n} \\ \vdots & \vdots & \dots & \vdots \\ A_{m,1} & A_{m,2} & \dots & A_{m,n} \end{bmatrix}$. Let $B = \begin{bmatrix} B_{1,1} & B_{1,2} & \dots & B_{1,n} \\ B_{2,1} & B_{2,2} & \dots & B_{2,n} \\ \vdots & \vdots & \dots & \vdots \\ B_{m,1} & B_{m,2} & \dots & B_{m,n} \end{bmatrix}$

$\therefore A^T = \begin{bmatrix} A_{1,1} & A_{2,1} & \dots & A_{m,1} \\ A_{1,2} & A_{2,2} & \dots & A_{m,2} \\ \vdots & \vdots & \dots & \vdots \\ A_{1,n} & A_{2,n} & \dots & A_{m,n} \end{bmatrix}$

$(A^TB)_{1,1} = A_{1,1}B_{1,1} + A_{2,1} + B_{2,1} + \dots + A_{m,1}B_{m,1}$
$(A^TB)_{2,2} = A_{1,2}B_{1,2} + A_{2,2} + B_{2,2} + \dots + A_{m,2}B_{m,2}$
$\vdots$
$(A^TB)_{n,n} = A_{1,n}B_{1,n} + A_{2,n} + B_{2,n} + \dots + A_{m,n}B_{m,n}$

$\therefore$ We see that $tr(A^TB) = \sum_{i,j} A_{ij}B_{ij}$.

In [202]:
```
m = 100;
n = 50;
A = randn(m,n);
B = randn(m,n);
@time tr(A'*B);

@time function rhs()
    s = 0;
    for i in 1:m
        for j in 1:n
            s += A[i,j]*B[i,j];
        end
    end
    return s;
end;
@printf("LHS = %f \n",tr(A'*B));
@printf("RHS = %f",rhs());
```

```
  0.000042 seconds (8 allocations: 19.828 KiB)
  0.000034 seconds (20 allocations: 1.845 KiB)
LHS = -13.460622
RHS = -13.460622
```

Above, I used Random Matrices of size 100×50 to calculate the time and storage complexities of two sides of the equation. Computationaly, they take around the same time, albeit, RHS is a bit faster and takes less storage space. So, RHS seems to be a superior formula.

Question 3:

a

We are given that A is a diagonal matrix. Now, let us assume that $A \succeq 0$. Stating what's given along with our assumption, we can write x^TAx as $\lambda_1 x_1^2 + \lambda_2 x_2^2 + \dots + \lambda_n x_n^2$ for a vector $\{x_1, x_2, \dots, x_n\}$. A is a diagonal matrix, so the diagonal values are the eigenvalues.

Since, we are assuming $A \succeq 0$, $x^TAx \geq 0$ has to be true.
$\therefore \lambda_1 x_1^2 + \lambda_2 x_2^2 + \dots + \lambda_n x_n^2 \geq 0 => \lambda_i \geq 0, \forall i \in \{1, \dots, n\}$ which confirms that for $A \succeq 0$ iff $A_{ii} \geq 0$. In similar way we can prove for $A \succ 0$.

If we assume that $A \succ 0$, then $x^TAx > 0$ has to be true.
$\therefore \lambda_1 x_1^2 + \lambda_2 x_2^2 + \dots + \lambda_n x_n^2 > 0 => \lambda_i > 0, \forall i \in \{1, \dots, n\}$ which confirms that for $A \succ 0$ iff $A_{ii} > 0$.

b

We are given

$$A = \begin{bmatrix} A^{(1)} & 0 & \dots & 0 \\ 0 & A^{(2)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A^{(l)} \end{bmatrix}$$

Similar to part (a), we can conclude that since A is a diagonal matrix, and if we are given that $A \succeq 0$ or $A \succ 0$, then the eigenvalues of A are non-negative. And, also since A is a diagonal matrix, the determinant of A is the total product of each block $A^{(k)}$, where $k \in \{1, \dots, l\}$. Thus, for $A \succeq 0$ or $A \succ 0$ to be true, $A^{(k)} \succeq 0$ or $A^{(k)} \succ 0$ has to be true, which we could also see is parallel to what we conclude in part (a).