Bob Ghosh, 42039157, j0k0b

Assignment 4, CPSC 406, March 16.

Libraries

In [1]:
```
using LinearAlgebra, ForwardDiff, Optim
```

Question 1

Proof by Induction:

Let **n = 1**; then $S = S_1$. So, if S_1 is a convex set, then by definition, S is also a convex set.

Now, let us suppose **n = 2**, then $S = S_1 \cap S_2$. Furthermore, let us assume there exists two points a and b, such that $a, b \in S_1 \cap S_2$, that is a, b are the two common points within S_1 and S_2.

If we consider a line segment ab, connecting the points a and b, then $ab \in S_1$ and $ab \in S_2$, by the definition of convexity. Thus $ab \in S_1 \cap S_2$, and hence S is a convex set. [1]

Lastly, let us consider that S is a convex set for **n=k** intersecting sets. We now need to prove that S is a convex set for k+1 intersecting sets.

From above, let us set $S_{k^*} = S_1 \cap S_2 \cap \ldots S_k$

For n = k+1, $S_{k^*+1} = S_1 \cap S_2 \cap \ldots S_k \cap S_{k+1}$

=> $S_{k^*+1} = S_{k^*} \cap S_{k+1}$

Given our assumption, we know that S_{k^*} is a convex set. And from [1], we see that intersection of two convex sets is a convex set. So, now we can claim that S_{k^*+1} is a convex set.

Hence, we have proved by induction that the intersection of convex set is itself a convex set.

Question 2

We are given $S = \{(x,t) : ||x||_2 \leq t\}$. Therefore, we need to pick two points x_1 and x_2 such that $||x_1||_2 \leq t$ and $||x_2||_2 \leq t$.

=> $\theta||x_1||_2 \leq \theta t$ [1]

and $(1-\theta)||x_2||_2 \leq (1-\theta)t$ [2]

for some $\theta \in [0, 1]$.

On adding [1] and [2], we get ->
$\theta||x_1||_2 + (1-\theta)||x_2||_2 \leq \theta t + (1-\theta)t$

Which becomes: $\theta||x_1||_2 + (1-\theta)||x_2||_2 \leq t$, thus by the definition of convexity, S is a convex set.

Question 3

a. Nonnegative Constraint.

Proof by Contradiction

So, in this case, we are assuming (2): $x^* = proj_c(x^* - \gamma \nabla f(x^*))$ for any constant $\gamma > 0$. Furthermore, projection itself is an optimization problem: $proj_c(z) = \underset{x}{argmin}\ ||x - z||_2^2$ subject to $x \in C$.

For $C = \{x : x_i \geq 0\}$, as a point of contradiction, let us assume $\nabla f(x^*) < 0$. In (2) we say that $\gamma > 0$ for $\epsilon > 0$, therefore, $x^* = \underset{x \geq 0}{argmin}\ ||x - x^* - \epsilon||_2^2$. Thus optimally, we would find that $x = x^* + \epsilon$. Thus, there is a contradiction, as each term would not be non-negative. Therefore, $\nabla f(x^*)$ has to be ≥ 0.

b. Normal Cone.

(3) states that $\nabla f(x^*)^T(y - x^*) \geq 0, \forall y \in C$. We will be using (2) to show equivalency.

From (2), $x^* = \underset{x}{argmin}\ ||x - x^* + \gamma \nabla f(x^*)||_2^2$

$$= \underset{x}{argmin}\ (x^T - (x^*)^T + \gamma \nabla f(x^*)^T(x - x^* + \gamma \nabla f(x^*)))$$

$$= \underset{x}{argmin}\ (x^T x - 2x^T x^* + 2\gamma x^T \nabla f(x^*) + (x^*)^T x^* - 2\gamma \nabla f(x^*)^T x^*)$$

$$= \underset{x}{argmin}\ (||x - x^*||_2^2 + 2\gamma \nabla f(x^*)^T(x - x^*))$$

From this we know that $||x - x^*||_2^2 \geq 0$, $f(x^*) \geq 0$ and $x - x^* \geq 0$. So, all the terms in the above equation are positive.

Now we can use $\nabla f(x^*)^T(y - x^*) < 0$ to prove that $x^* \neq \underset{x}{argmin}\ (||x - x^*||_2^2 + 2\gamma\nabla f(x^*)^T(x - x^*))$, by using γ big enough and making the term $argmin < 0$. Hence, that shows the equivalency between (3) and (2).

Question 4

a

b

In [2]:
```
## Projected Gradient Descent
A = [1 1 1; 2 1 0];
b = [1;0.5];
alpha = 0.01; # step-size
x_temp = zeros(size(A)[2],1);
x_0 = zeros(size(A)[2],1);

function gradient(x, A, b)
    f(x)= 100.0 * (x[2] - x[1]^2)^2 + (1-x[1])^2 + 100.0 * (x[3]-x[2]^2)^2+(1-x[
    return ForwardDiff.gradient(f,x)
end

function gradientDescent(A, b, x_temp, alpha, n)
    m = length(b)
    for iter in 1:100
        g = gradient(x_temp, A, b)
        x_before = A\b
        x_temp = x_temp-alpha * g
        fun(x_temp) = norm((x_before - x_temp), 2)
        res = Optim.optimize(fun, x_temp)
        x_temp = Optim.minimizer(res)
    return x_temp
    end
end

Theta = gradientDescent(A, b, x_temp, alpha, 1000)
```

Out[2]:
```
3×1 Array{Float64,2}:
 0.08333333250256361
 0.3333333327835978
 0.5833333399671755
```

In [3]:
```julia
## Reduced Gradient Method
N = 3;
f3(x)= 100.0 * (x[2] - x[1]^2)^2 + (1-x[1])^2 + 100.0 * (x[3]-x[2]^2)^2+(1-x[2])^
g(x) = ForwardDiff.gradient(f3, x);
H(x) = ForwardDiff.hessian(f3, x);
x = [1,1,1];
g(x);
cholesky(H(x));
a1 = ones(1,3) ; b1 = 1;
a2 = [2 1 0]; b2 = 0.5;
A = vcat(a1,a2); b = vcat(b1,b2);
Q, R = qr(A'); Z = reshape(Q[:,3],3,1);
xbar = A \ b
```

Out[3]:
```
3-element Array{Float64,1}:
 0.08333333333333334
 0.3333333333333333
 0.5833333333333336
```

In [4]:
```julia
norm(Theta - xbar, 2)
```

Out[4]:
```
6.708222277663692e-9
```

The solutions are not strictly equal, although they are somewhat similar, depending on the degree of accuracy. Perhaps, more iterations with different step size could lead to a better, or even same result.

Question 5

a. Entropy

We are given that $f(x) = -\sum_{i=1}^{n} x_i log(x_i)$.

From this we can derive $\nabla f(x)$ as $\nabla f(x) = -(log(x_i) + 1)$. And from this, we can find the Hessian to be $H = \begin{bmatrix} -1/x_1 & 0 & \dots & 0 \\ 0 & -1/x_2 & \dots & 0 \\ \vdots & \ddots & \dots & \dots \\ 0 & 0 & \dots & -1/x_n \end{bmatrix}$

$\therefore H = (-1/x_i)I_n$

We also know that $x \in [0, 1]$, so H is a negative definite matrix. And hence, f(x) is strictly concave when $x \neq y$.

b. Log-Sum-Exp

We are given $f(x) = \log(\sum_{i=1}^{m} e^{a_i^T x})$. For simplicity, let us define $u_i = e^{a_i^T x}$ and $w_i = e^{a_i^T y}$. [1]

Because of the nature of convexity, we can claim that
$f(\theta x + (1 - \theta)y) = \log(\sum_{i=1}^{n} e^{\theta a_i^T x + (1-\theta) a_i^T y})$ [2] for some $\theta \in [0, 1]$.

From [1] and [2], we can simplify the equation as $\log(\sum_{i=1}^{n} e^{\theta a_i^T x + (1-\theta) a_i^T y}) = \log(\sum_{i=1}^{n} u_i^{\theta} w_i^{1-\theta})$.

Using the properties of inequality we can state that:
$\log(\sum_{i=1}^{n} u_i^{\theta} w_i^{1-\theta}) \leq \log((\sum_{i=1}^{n} u_i)^{\theta}(\sum_{i=1}^{n} w_i)^{1-\theta})$

$\leq \theta \log \sum_{i=1}^{n} u_i + (1 - \theta) \log \sum_{i=1}^{n} w_i$

This brings us to $f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta) f(y)$, which proves the convexity.