Homework 3

Noé Hollande, Yujun Kim

May 16, 2022

Part 1

Question 1:

From Weierstrass' theorem, we know that if f(x) is continuous on a nonempty feasible set S that is closed and bounded, then f(x) has a global minimum on S, first, $f(x) = 1/2||x-z||^2$ is continuous on Q, and Q is closed now it's not bounded, but suppose we take a random feasible x in Q for a first approach, and we get f(x)= d, we now that the minimum we looking for is $\leq d$, therefore let's consider the set $Q\bigcup C$ where C is the sphere of center z and radius d, this is a closed on bounded set where f(x) is still feasible (since it contains our first arbitrary point), therefore we can apply Weierstrass and argue our function has a global minimum in $Q\bigcap C$, therefore $Proj_Q(z)$ is non-empty.

Question:2

Let's take for instance $z = 0$ and Q the ring defined as $x \in Q : 1 \leq ||x|| \leq 2$, Q is indeed a closed and bounded set but $Proj_Q(z)$ has an infinite number of elements.

Question 3:

Suppose that $Proj_Q(z) = \{0\}$ i.e, the minimum is realised at $x^* = 0$, then we use the necessary condition for optimality implying that $<\nabla f(0), v> \geq 0$ for all $v \in C^{\circ}$. By a computation we made several times already $\nabla f(x) = x - z$ so that $<z, v> \leq 0$ for all $v \in C^{\circ}$ which gives us the first implication.
Now suppose $z \in C^{\circ}$, therefore $<z, v> \leq 0$ for all $v \in C^{\circ}$. Let's assume ab absurdo that $x = Proj_C(z) \neq \{0\}$ so that
$||x-z||^2 < ||z||^2 \Leftrightarrow <x-z, x-z> < <z, z> \Leftrightarrow <z, z> - 2<z, x> + <x, x> < <z, z>$.
And therefore : $0 \geq <z, x> > <v, v> > 0 ⇜$ therefore we get the desired implication.

Question 4:

From the course we know that x^* is a stationnary point if and only if $-\nabla f(x^*) \in (T_{x^*}C)^{\circ}$, but by the precedent question this is also true if and only if $Proj_{T_{x^*}C}(-\nabla f(x^*)) = 0$.

Question 5:

a) If $v \in Proj_C(z)$ then $<\nabla(1/2||z-v||^2), w> = <z-v, w> \geq 0$ for any $w \in T_vC$. In particular, v and -v are tangent directions, we can see it by setting a curve $c(t) = vt, t \in [0,1]$ such that $c'(0) = v$ and another curve d such that $d(t) = (1-t)v, t \in [0,1]$ such that $d'(0) = -v$. Now, both $<z-v, v> \geq 0$ and $<z-v, -v> \geq 0$, thus, by showing this double inequality, we get $<z-v, v> = 0$.

b) Let v_1 and $v_2 \in Proj_C(z)$ then $<v_i, v_i - v> = 0$ so that $<v_i, v_i> = <v_i, z>$ for $i = 1, 2$.
Since $||v_1 - z||^2 = ||v_2 - z||^2$,
we have $||v_1||^2 - 2<v_1, z> + ||z||^2 = ||v_2||^2 - 2<v_2, z> + ||z||^2$.
Which gives us $<v_1, z> - 2<v_1, z> = <v_2, z> - 2<v_2, z>$
and $<v_1, v_1> = <v_1, z> = <v_2, z> = <v_2, v_2>$.

Question 6:

Let $S = \{x \in R^2 | x_1 \leq 0\}$

$$\text{We have } T_xS = \begin{cases} R^2 \text{ if } x_1 < 0 \\ \text{S if } x_1 = 0 \end{cases}$$

Let f(x) : $R^2 \to R : f(x) = -x_1$

We therefore have $-\nabla f(x) = (1, 0)$

$$\text{Hence, } Proj_{T_xS}(-\nabla f(x) = \begin{cases} (1,0) \text{ if } x_1 < 0 \\ (0,0) \text{ if } x_1 = 0 \end{cases}$$

And therefore q(x) isn't continuous at x = 0, note that this exemple generalizes to R^n.

Question 7:

We compute $T_xS = \{v \in E : <2x, x-v> = 0\} = \{v \in E : <x, v> = 1\}$.
Furthermore, we compute the projection of $-\nabla f(x)$ on the line T_xS.
This gives us :

$Proj_{T_xC}(-\nabla f(x)) = -\nabla f(x) + < -\nabla f(x), x > / <x, x> * x + x = -\nabla f(x) + < -\nabla f(x), x > x + x$ which must be contionous since ∇f is and therefore the norm of this contionous function is still continuous.

Question 8:

a) Suppose that $Dh(x)$ has row rank p (We know it's guaranteed by LICQ conditions). We know by the theorem 8.14 that the cone of linearized feasible directions is T_xC. We therefore have $T_xC = \{x : Dh(x) = 0\} = ker Dh(x)$.

b) T_xC is closed so there's a solution by question 1, by the LICQ conditions, $(\nabla h_1 x, ..., \nabla h_p(x))$ is a basis of $ImDh(x)$, we can get a basis of $KerDh(x)$ by taking the orthogonal of the first basis. Let B denote the matrix composed of the linearly independant vectors that compose our basis of $ImDh(x)$. We know from the linear models course that $B(B^TB)^{-1}B^T(z)$ is the unique projection of z to $kerDh(x)$.

c) By the precedent points, we have that the matrix $B(B^TB)^{-1}B^T(z)$ is continuous since the inversion is and since it's only composed of linear applications, furthermore, the function $-\nabla f(x)$ is continuous as well by hypothesis.

Part 2

Question 1

Note that the mapping φ that $x \mapsto < w, x >$ is a continuous function(because it is linear) and S is a compact set. By extreme value theorem, there exist a point $x_w^8 \in S$ that minimize $< w, x >$ for $\forall x \in S$

Question 2

However, such point of minimization may be not unique as following example: If $w = 0 \in \mathbb{R}^1$ and $S = [0, 1] \subseteq \mathbb{R}^1$, then S is convex and compact while every point in S is an optimal point for $min_{x \in S} < w, x >$.

Question 3

We know that $s(x_k), x_k \in S$, and so $x_{k+1} = (1 - \eta_k)x_k + \eta_k s(k) \in S \ \forall \eta_k \in [0, 1]$. However, if $\eta_k \notin [0, 1]$, then we can't guarantee that $x_{k+1} \in S$ so we cannot further proceed the algorithm.

Question 4

We prove the following lemma:
Lemma If $f : S \longrightarrow \mathbb{R}$ is *L-Lipschitz* continuous for some convex set $S \subseteq \mathbb{R}^n$, then $f(y) - f(x) \leq \nabla f(x)(y - x) + \frac{L}{2}||y - x||^2 \ \forall x, y \in S$.

(proof) Let $G(t) = f(c(t)), c(t) = x + t(y - x)$.

$$\begin{aligned}
f(y) - f(x) - \nabla f(x)(y-x) &\leq |f(y) - f(x) - \nabla f(x)(y-x)| \\
&= |G(1) - G(0) - \nabla f(x)(y-x)| \\
&= |\int_0^1 G'(t) - \nabla f(x)(y-x)dt| \\
&= |\int_0^1 \nabla f(c(t))(y-x) - \nabla f(x)(y-x)dt| \\
&= |\int_0^1 (\nabla f(c(t)) - \nabla f(x))(y-x)dt| \\
&= \int_0^1 |(\nabla f(c(t)) - \nabla f(x))(y-x)|dt \\
&\leq \int_0^1 L|(c(t)-x)(y-x)|dt \\
&= \int_0^1 Lt||y-x||^2 dt \\
&= \frac{L}{2}||y-x||^2
\end{aligned}$$

Thus, the lemma is proved. □

Now, we use this lemma to solve question 4.

(B1) This is direct from the lemma, where we put x_{k+1}, x_k for y, x respectively.

(B2) $x_{k+1} - x_k = (1 - \eta_k)x_k + \eta_k s(x_k) - x_k = \eta_k(s(x_k) - x_k)$

In particular, we have $||x_{k+1} - x_k|| \leq \eta_k ||s(x_k) - x_k|| \leq \eta_k d_s$. Putting these to equations, we get **(B2)** immediate.

(B3) $\nabla f(x_k)s(x_k) \leq \nabla f(x_k)x^*$ by the definition of $s(x_k)$. **(B3)** is immediate.

(B4) $\nabla f(x_k)(x^* - x_k) \leq f(x^*) - f(x_k)$ holds as f in convex function. **(B4)** is immediate.

Question 5

By the lemma from question 4,

$$f(x_1) - f(x^*) \leq \frac{L}{2}||x_1 - x^*||^2 \leq \frac{L}{2}d_s^2$$

Question 6

First, note that $\frac{2k+2}{(k+2)^2} \leq \frac{2}{k+3}\ \forall k \in \mathbb{Z}_+$. This is immediate from $(2k+2)(k+3) = 2k^2 + 8k + 6 \leq 2k^2 + 8k + 8 = 2(k+2)^2$.

Now, we use induction on k to prove question 6. (i) When $k = 0$, this is question 5. (ii)

Suppose this hold for k. i.e. $f(x_k) - f(x^*) \le \frac{2Ld_s^2}{k+2} = \eta_k L d_s^2$. Then,

$$\begin{aligned} f(x_{k+1}) - f(x^*) &= f(x_{k+1}) - f(x_k) + f(x_k) - f(x^*) \\ &\le \eta_k(f(x^*) - f(x_k)) + \frac{L}{2}\eta_k^2 d_s^2 + f(x_k) - f(x^*) \\ &= (1-\eta_k)(f(x_k) - f(x^*)) + \frac{L}{2}\eta_k^2 d_s^2 \\ &\le (1-\eta_k)\eta_k L d_s^2 + \frac{L}{2}\eta_k^2 d_s^2 \\ &= (1 - \frac{\eta_k}{2})\eta_k L d_s^2 \\ &= \frac{2}{k+2}\frac{k+1}{k+2} L d_s^2 = \frac{2k+2}{(k+2)^2} L d_s^2 \\ &\le \frac{2}{k+3} L d_s^2 = \eta_{k+1} L d_s^2 \end{aligned}$$

Hence, question 6 is justified.

Question 7

(Convexity) Let $x, y \in \Delta^n$. Then, $x_i, y_i \ge 0\ \forall i, \sum x_i = \sum y_i = 1$. Then for any $\lambda \in [0,1]$, $(1-\lambda)x + \lambda y$ has
(1)$(1-\lambda)x_i + \lambda y_i \ge 0\ \forall i$
(2) $\sum (1-\lambda)x_i + \lambda y_i = (1-\lambda)\sum x_i + \lambda \sum y_i = (1-\lambda) + \lambda = 1$ Hence, $\lambda \in [0,1]$, $(1-\lambda)x + \lambda y \in \Delta^n$.

(Compact) Note that $A = \{x \in \mathbb{R}^n | x_i \ge 0\ \forall i\}$ is closed, by the definition of product topology. Also, the mapping $\sigma : \mathbb{R}^n \longrightarrow \mathbb{R}$ by $x \mapsto \sum x_i$ is continuous. Hence, $B = \sigma^{-1}(\{1\})$ is closed. Also, as $||x|| \le \sqrt{n}$ for all $x \in B$. Thus, B is bounded.

By definition, $\Delta^n = A \cap B$. Thus, it is closed and bounded. i.e. it is compact.

Question 8

Let $i_{min} = argmin_{1 \le i \le n} w_i$. Then for $x \in \Delta^n$,

$$\begin{aligned} < w, e_{i_{min}} > = w_{i_{min}} &= w_{i_{min}} \sum x_i \\ &\le \sum w_i x_i =< w, x > \end{aligned}$$

As $e_{i_{min}} \in \Delta^n$, we have minimum value $w_{i_{min}}$ by picking $x = e_{i_{min}} \in \Delta^n$.

Question 9

For picking the optimal x, we only have to find the index of minimal elements among n elements. Hence, if we let $y = w_1$, and update it through increasing i by letting $y = w_i$ if $w_i < y$ gives $O(n)$ algorithm to choose optimal x.

Question 10

This problem has solution as Δ^n is compact and f is continuous function. However, the solution may not be unique. For instance, suppose $A = 0$, the zero matrix. Then, f has same value for every input in Δ^n.

Question 11

$\nabla f(x) = \nabla \frac{1}{2} < Ax - b, Ax - b >= A^T(Ax - b)$ is its gradient.

Question 12

$g(\eta) = f((1-\eta)x + \eta y) = \frac{1}{2}||A((1-\eta)x + \eta y) - b||^2$. Then,

$$\begin{aligned} g'(\eta) &=< \nabla f((1-\eta)x + \eta y), (y - x) > \\ &=< \nabla f(x + \eta(y - x)), (y - x) > \\ &=< A^T(A((x + \eta(y - x)) - b), y - x > \\ &= \eta < A^T A(y - x), y - x > + < A^T(Ax - b), y - x > \\ &= 0 \end{aligned}$$

gives the optimal η:

$$\eta' = -\frac{< A^T(Ax - b), y - x >}{||A(y - x)||^2}$$

However, this η' might not be in $[0, 1]$. Thus, the optimal η is

$$\eta = \begin{cases} \eta' & \text{if } \eta' \in [0, 1] \\ 0 & \text{if } \eta' \notin [0, 1] \text{ and } g(0) \leq g(1) \\ 1 & \text{if } \eta' \notin [0, 1] \text{ and } g(0) > g(1) \end{cases}$$

Question 13

(Step 1) First take any $x_0 \in \Delta^n$. For example, choose $x_0 = e_1$.
For $k = 0, 1, 2, \cdots$,
(Step 2) Given x_k, find index i_k that minimize the component of x_k. Then $s(x_k) = e_{i_k}$. Now, compute η' with formula in question 12 with $y = x_k$, $x = x_k$. Compute $\eta_k = \eta$ as in question 12 as well.
(Step 3) $x_{k+1} = (1 - \eta_k)x_k + \eta_k s(x_k)$. Go to **Step 2**

Question 14

The whole latex code for this question can be found on appendix.

Question 15

The following figure 1 show the Frank Wolfe gap and the function value $f(x)$ respectively.

Question 16

Figure 2 shows the sparsity of the obtained solutions.

Appendix: Codes

Part 2: Frank Wolfe Algorithm

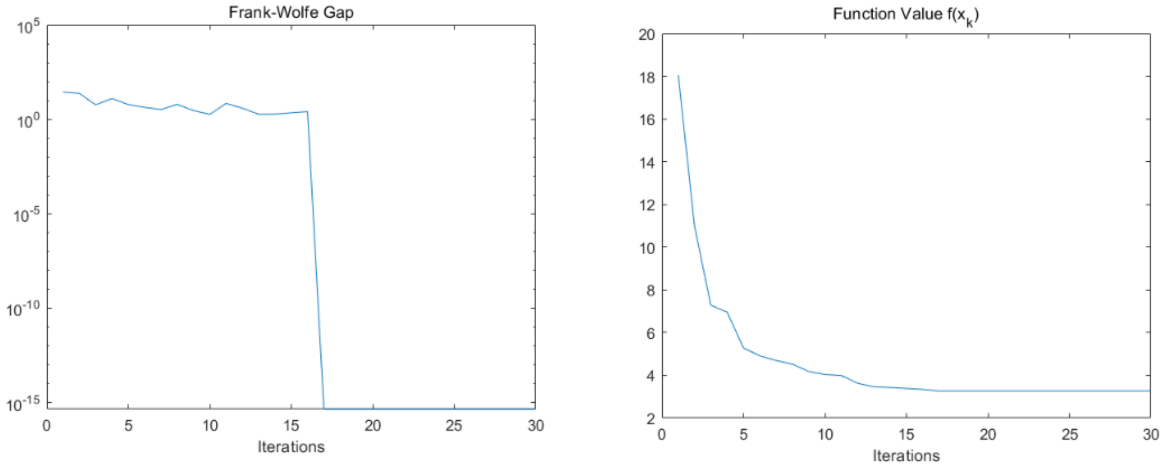

Figure 1: Question 15

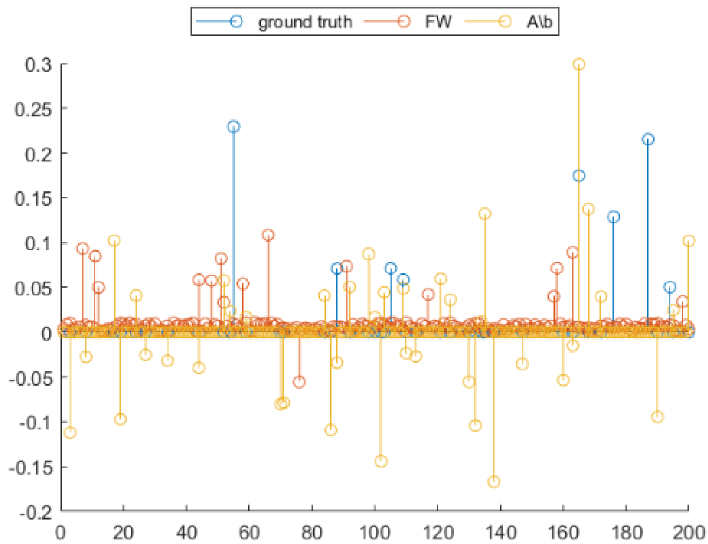

Figure 2: Question 16

```
clc
clear all
load data.mat
x_gt = x;
N = 30;
epsilon = 1e-50
gaps = zeros(N, 1);
fs = zeros(N, 1);

% Functions
e=@(k) [zeros(k-1,1);1;zeros(n-k,1)];
eta_gen=@(x,y) -(A*x-b)'*A*(y-x)/norm(A*(y-x))^2;
f = @(x) 1/2*norm(A*x-b)^2;
gap = @(x,y) (A*x-b)'*A*(x-y);

% Looping
x = rand(n, 1);
x = x/norm(x);
iter = 1;
while iter≤N
    fs(iter) = f(x);
    [w, i] = min(x);
    y = e(i);
    % Check gap
    gaps(iter) = abs(gap(x,y));

    if gaps(iter) < epsilon
        break
    end
    eta0 = eta_gen(x, y);

    % Setting eta
    if 0 ≤ eta0 ≤ 1
        eta = eta0;
    elseif f(x) ≤ f(y)
        eta = 0;
    else
        eta = 1;
    end

    % Next step
    x = (1-eta)*x + eta*y;
    iter = iter + 1;
end

% Plotting
figure
semilogy(gaps)
title('Frank-Wolfe Gap')
xlabel('Iterations')

figure
plot(fs)
title('Function Value f(x_k)')
xlabel('Iterations')

figure
plot_data(x_gt, x, A\b);
```

Part 3

Question 1:

Suppose $f(x^*) = z$ is a an optimal solution for (4), then (x^*, z) is feasible in (5), since $z \geq f_i(x^*)$ for any i (because it's feasible in (4)), we have $f_i(x^*) - z \leq 0$ for all i. Now suppose by contradiction that z is not optimal for (5), i.e., $\exists(x_1, y)$ such that $f_i(x_1) - y \leq 0 \forall i$ and $y < z$ then, we first could conclude that $y = f(x_1) = max f_i(x_1)$ because if it was inferior, (x_1, y) would not be in S, if it was superior, (x_1, y) would not be an optimal solution for (5) since $(x_1, max f_i(x_1)$ would be a better one. But then, $y = max f_i(x_1)$ is a better solution for (4) than z as well, which contradicts the optimality of z. Therefore we conclude that z must be optimal in (5) as well. By reusing the same arguments, we also conclude that an optimal solution in (5) is optimal in (4).

Question 2:

The KKT conditions can be stated as follow : There must exist $\lambda_1, ..., \lambda_m$ such that $-\nabla f(x) = \sum_{i=1}^{m} \lambda_i \nabla g_i(x)$ where the $g_i' s$ are the inequality constraints (there are no equality constraints) in this problem. Which applied to the problem gives :
$(0, 1) = \sum_{i=1}^{m} \lambda_i (\nabla f_i(x), -1)$

The complementary conditions are : $\lambda_i(f_i(x) - y) = 0 \forall i.$

Question 3:

Suppose we have $f_1 = x^2$ and $f_2 = x^2$, then, ∇g_1 and ∇g_2 are obviously not independant and they're both active at (0,0), hence the LICQ conditions aren't respected. Note that since $\nabla g_i = (\nabla f_i(x), -1)$ $\forall$ i, the gradient of all the active constraints needs to be equal for them to be linearly dependant. (Because of the second coordinate).

Question 4: Since we know from the course that LICQ implies MFCQ, we just need to focus on the case where the LICQ doesn't hold, in addition, since there are no active constraints in (5), we don't need to focus on the $h_i' s$ part. Therefore, we want to prove that for $i = 1, ..., p$, such that the i-th constraint is active, there exists $x^* \in R^n * R$ such that $\langle \nabla g_i(x), x^* - x \rangle < 0$. But as discussed at the end of Question 3, if the LICQ doesn't hold, then all the gradients of the active constraints are equal, therefore we only need to prove the existence of x^* for a single i-th case : $\langle \nabla g_i(x), x^* - x \rangle < 0 \rightarrow (\nabla f_i(x))$ * (the R^n coordinate of x^* - the R^n coordinate of x) - 1 * (the R coordinate of x^*-the R coordinate of x^*) ¡ 0. We just need to take x^* with the R^n coordinate equal to those of x, and the R coordinate stricly bigger than the one of x to get the desired result, aka, the respect of MFCQ.

Question 5: I'm not sure if that's what asked, but from what we'eve discussed on the previous questions, we can conclude that $(\nabla g_i(x)$ must be equal to (0,1) for all the active constraints, which is coherent with the fact that these point are intuitively stationnary.