MATH-329 Nonlinear optimization Homework 3: Constrained optimization

Alix Pelletier 346750
Vlad Burca 344876
Ismail Bouhaj 326480

11/2023

Part 1 : Projections to cones and stopping criteria in constrained optimization.

Question 1

Since Q is non-empty, Let $x_0 \in Q$. Then by definition of the minimizer, we have:

$$\min_{x \in Q} \frac{1}{2}\|x - z\|^2 \leq \frac{1}{2}\|x_0 - z\|^2$$

Inspired by this, let us define the space $Q' \subseteq Q$ as the intersection of Q with the closed ball $\bar{B}(\|x_0 - z\|, z)$ of center z and radius $\|x_0 - z\|$:

$$Q' = Q \cap \bar{B}(\|x_0 - z\|, z)$$

We have $Q' \neq \emptyset$ since $x_0 \in Q'$. Moreover, Q' is closed since it is the intersection of two closed sets. Finally, Q' is bounded since it is contained in the closed ball $\bar{B}(\|x_0 - z\|, z)$. Therefore, Q' is compact. By Weierstrass, the function $f(x) = \frac{1}{2}\|x - z\|^2$ attains its minimum on Q'. So the set:

$$\text{Proj}_{Q'}(z) = \left\{ x \in Q' : \frac{1}{2}\|x - z\|^2 = \min_{y \in Q'} \frac{1}{2}\|y - z\|^2 \right\}$$

is well-defined and non-empty.

We want to show that $\text{Proj}_{Q'}(z) = \text{Proj}_Q(z)$. Let $x \in Q \setminus Q'$. Then:

$$\|x - z\| > \|x_0 - z\| \implies$$

$$\frac{1}{2}\|x - z\|^2 > \frac{1}{2}\|x_0 - z\|^2 \geq \min_{y \in Q'} \frac{1}{2}\|y - z\|^2$$

So the minimizer of $\frac{1}{2}\|y - z\|^2$ on Q' is also the minimizer of $\frac{1}{2}\|y - z\|^2$ on Q. Therefore, $\text{Proj}_{Q'}(z) = \text{Proj}_Q(z)$.

Question 2

Let $\mathcal{E} := \mathbb{R}$; $S := \{-1, 1\}$; $z := 0$. Then clearly S is closed (union of 2 singletons). Moreover, we have:

$$\frac{1}{2}\|z - s\|^2 = \frac{1}{2}\|s\|^2 = \frac{1}{2} \text{ for all } s \in S \implies$$

$$\text{Proj}_S(z) = \{1, -1\}$$

Question 3

$\mathbf{Proj}_C(z) = \{0\} \implies z \in C^\circ$**:**

Assume $\text{Proj}_C(z) = \{0\}$. This implies the point in C closest to z is the origin. So for any $x \in C$:

$$\begin{aligned} \frac{1}{2}\|x - z\|^2 &\geq \frac{1}{2}\|0 - z\|^2 \\ \|x\|^2 - 2\langle x, z\rangle + \|z\|^2 &\geq \|z\|^2 \\ \|x\|^2 - 2\langle x, z\rangle &\geq 0 \end{aligned}$$

Now if $\langle x, z\rangle > 0$, for some x then we have:

$$\lambda := \frac{\langle x, z\rangle}{\|z\|^2} > 0 \implies \lambda x \in C$$

But then, plugging λx in the inequality above, we get:

$$\lambda^2\|x\|^2 - 2\lambda\langle x, z\rangle \geq 0 \iff$$

$$\left(\frac{\langle x, z\rangle}{\|z\|^2}\right)^2 \|x\|^2 - 2\frac{\langle x, z\rangle}{\|z\|^2}\langle x, z\rangle \geq 0 \iff$$

$$-\frac{\langle x, z\rangle^2}{\|x\|^2} \geq 0$$

Which is clearly a contradiction. Therefore, we must have $\langle x, z\rangle \leq 0$ for all $x \in C$. This implies $z \in C^\circ$, by definition.

$z \in C^\circ \implies \mathbf{Proj}_C(z) = \{0\}$**:**

Assume $z \in C^\circ$. We want to show that for all $x \in C \setminus \{0\}$ we have:

$$\frac{1}{2}\|x - z\|^2 > \frac{1}{2}\|0 - z\|^2$$

By the above calculations this is equivalent to showing that for all $x \in C \setminus \{0\}$ we have:

$$\|x\|^2 - 2\langle x, z\rangle > 0$$

But this is true since $z \in C^\circ$, so $\langle x, z\rangle \leq 0$ for all $x \in C$ and $x \neq 0$. Therefore, we have:

$$\text{Proj}_C(z) = \{0\}$$

Question 4

We know from class that $x^* \in S$ is a stationary point of f if and only if $-\nabla f(x^*) \in (T_{x^*}S)^\circ$. By Question 3, this is equivalent to:

$$\text{Proj}_{T_{x^*}S}(-\nabla f(x^*)) = \{0\}$$

Question 5

(a) $v \in \mathbf{Proj}_C(z) \implies \langle v, z-v \rangle = 0$**:**

Let:

$$\begin{aligned} g : \mathcal{E} \to & \quad \mathbb{R} \\ x \mapsto & \frac{1}{2}\|x-z\|^2 \end{aligned}$$

Then g is differentiable and for$h \in \mathbb{R}$ and $v \in \mathcal{E}$ we have:

$$\begin{aligned} g(x+hv) = \frac{1}{2}\|x+hv-z\|^2 &= \frac{1}{2}\|x-z\|^2 + h\langle v, x-z\rangle + \frac{h^2}{2}\|v\|^2 \\ &= g(x) + h\langle v, x-z\rangle + O(h^2) \implies \\ \nabla g(x) = x - z & \end{aligned}$$

If $v \in \mathrm{Proj}_C(z)$ then v is a stationary point of g (as v is a global minimum of g). Therefore, we have:

$$\begin{aligned} \langle \nabla g(v), w \rangle \geq 0, &\quad \forall\ w \in T_vC \\ \langle v - z, w \rangle \geq 0, &\quad \forall\ w \in T_vC \end{aligned}$$

It is clear that if we show that $v,\ -v \in T_vC$ we are done. But this is true since $v \in C$ and C is a cone.

Let $\left((1-\frac{1}{n})v\right)_{n\in\mathbb{N}^*} \subseteq C$, $\left((1+\frac{1}{n})v\right)_{n\in\mathbb{N}^*} \subseteq C$ then:

$$\lim_{n\to\infty}\left((1-\frac{1}{n})v\right) = v, \quad \lim_{n\to\infty}\left((1+\frac{1}{n})v\right) = v$$

and:

$$\begin{aligned} \lim_{n\to\infty}\left(\frac{(1-\frac{1}{n})v - v}{\frac{1}{n}}\right) &= \lim_{n\to\infty}\left(\frac{-\frac{1}{n}v}{\frac{1}{n}}\right) = -v \\ \lim_{n\to\infty}\left(\frac{(1+\frac{1}{n})v - v}{\frac{1}{n}}\right) &= \lim_{n\to\infty}\left(\frac{\frac{1}{n}v}{\frac{1}{n}}\right) = v \end{aligned}$$

Therefore, by the definition of the tangent cone $v,\ -v \in T_vC$. So we have:

$$\begin{aligned} \langle v - z, v \rangle &\geq 0 \\ \langle v - z, -v \rangle &\geq 0 \implies \\ -\langle v - z, v \rangle &\geq 0 \end{aligned}$$

So we have:

$$\langle v, z - v \rangle = 0$$

(b) $v_1,\ v_2 \in \mathbf{Proj}_C(z) \implies \|v_1\| = \|v_2\|$**:**

By part (a), we have:

$$\begin{aligned} \langle v_1, z - v_1 \rangle = 0 &\implies \|v_1\|^2 = \langle v_1, v_1 \rangle = \langle v_1, z \rangle \\ \langle v_2, z - v_2 \rangle = 0 &\implies \|v_2\|^2 = \langle v_2, v_2 \rangle = \langle v_2, z \rangle \end{aligned}$$

Since both are minimizers of $\frac{1}{2}\|x-z\|^2$, we have:

$$\begin{aligned} \frac{1}{2}\|v_1 - z\|^2 = \frac{1}{2}\|v_2 - z\|^2 &\implies \\ \|v_1\|^2 - 2\langle v_1, z\rangle + \|z\|^2 = \|v_2\|^2 - 2\langle v_2, z\rangle + \|z\|^2 &\implies \\ \|v_1\|^2 = \|v_2\|^2 \implies \|v_1\| = \|v_2\| & \end{aligned}$$

Question 6

We present an example where the function $q(x) = \|\text{Proj}_{T_xS}(-\nabla f(x))\|$ is discontinuous on the set S. Consider the following:

Function f and Set S:

- Function f: Define $f : \mathbb{R} \to \mathbb{R}$ as $f(x) = -x^2$. Its gradient is $\nabla f(x) = f'(x) = -2x$.
- Set S: Define $S := [0, 1] \subseteq \mathbb{R}$.

First, let's compute T_xS for all $x \in S$. We have:

- $x \in (1, 0)$. Then x is in the interior of S and $T_xS = \mathbb{R}$, by example 7.10. from the lecture notes.
- $x = 0$. Then 0 is on the boundary of S and $T_0S = [0, +\infty)$, by example 7.11. from the lecture notes (we can see 0 as the boundary for the half-space $x \geq 0$ and locally around 0 the two spaces are the same. We conclude by the fact that the tangent cone is a local property, as any sequence $x_k \to 0^+$ will be contained in $(0, 1)$, for k big enough).
- $x = 1$. Then 1 is on the boundary of S and $T_1S = (-\infty, 0]$ (same reasoning as above, but with the half space $x \leq 1$).

Now, let's compute $\text{Proj}_{T_xS}(-\nabla f(x))$ for all $x \in S$. We have:

- $x \in (0, 1)$. Then $\text{Proj}_{T_xS}(-\nabla f(x)) = \{-\nabla f(x)\}$, since $T_xS = \mathbb{R}$. So:
$$q(x) = \|2x\| = 2x$$
- $x = 0$. Then $\text{Proj}_{T_0S}(-\nabla f(0)) = \{0\}$, since $T_0S = [0, +\infty)$ and $-f'(0) = 0$. So:
$$q(0) = \|0\| = 0$$
- $x = 1$. Then $\text{Proj}_{T_1S}(-\nabla f(1)) = \{0\}$, since $T_1S = (-\infty, 0]$ and $-f'(1) = 2$. So:
$$q(1) = \|0\| = 0$$

So:
$$q(x) = \begin{cases} 2x & \text{if } x \in [0, 1) \\ 0 & \text{if } x = 1 \end{cases}$$

Clearly, q is not continuous at $x = 1$.

Question 7

By example 7.14. from the lecture notes, we have:
$$T_xS = \{v \in \mathbb{R} : \langle v, x \rangle = 0\}, \quad \forall\, x \in S$$

For all $x = (x_1, x_2) \in S$, let $x^\perp = (x_2, -x_1) \in S$, then it is clear that:
$$\langle x, x^\perp \rangle = 0 \implies x^\perp \in T_xS$$

Moreover, by a simple argument over the dimensionality of T_xS and $\text{span}(x^\perp)$, we have:
$$T_xS = \text{span}(x^\perp)$$

Since T_xS is a sub-vector space of $\mathbb{R}^2$, of dimension 1 ($\{x^\perp\}$ is an orthogonal basis), we have:

$$\mathrm{Proj}_{T_xS}(-\nabla f(x)) = \mathrm{Proj}_{\mathrm{span}(x^\perp)}(-\nabla f(x)) = \left\{ \frac{\langle -\nabla f(x), x^\perp \rangle}{\|x^\perp\|^2} x^\perp \right\} = \{-\langle \nabla f(x), x^\perp \rangle x^\perp\}$$

So:

$$q(x) = \| - \langle \nabla f(x), x^\perp \rangle x^\perp \| = \| - \langle \nabla f(x), (x_2, -x_1) \rangle (x_2, -x_1) \| = |\langle \nabla f(x), (x_2, -x_1) \rangle|$$

Which is clearly a continuous map as it is a composition of continuous maps (∇f is continuous by assumption).

Question 8

Consider $\mathcal{E} = \mathbb{R}^n$ with a continuously differentiable function $h : \mathbb{R}^n \to \mathbb{R}^p$ and $S = \{x \in \mathbb{R}^n : h(x) = 0\}$, assuming LICQ holds for all x in S.

(a) T_xS?

From the lecture notes, we have that if LICQ holds for $x \in S$ then:

$$T_xS = F_xS = \{v \in \mathcal{E} \ : \ \langle \nabla h_i(x), v \rangle = 0; \ \forall i \in \{1, 2, \dots, p\}\}$$

(b)

Let $H(x) \in \mathbb{R}^{p \times n}$ such that:

$$H(x) = \begin{pmatrix} \nabla h_1(x)^T \\ \nabla h_2(x)^T \\ \vdots \\ \nabla h_p(x)^T \end{pmatrix}$$

Then we can rewrite T_xS as:

$$T_xS = \{v \in \mathcal{E} \ : \ H(x)v = 0\} = \ker H(x)$$

As T_xS is a sub-vector space of $\mathcal{E}$ of dimension $n - p$ (since all the lines of $H(x)$ are linearly independent), we know that the projection of $z \in \mathcal{E}$ exists and is unique.

By SVD, there exist $V \in \mathbb{R}^{n \times n}$ and $U \in \mathbb{R}^{p \times p}$ orthogonal matrices and $D \in \mathbb{R}^{p \times n}$ such that:

$$UH(x)V = D$$

$$D = \begin{pmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & O_{p \times (n-p)} \\ & & & \sigma_p & \end{pmatrix}$$

Where $\sigma_1, \sigma_2, \dots, \sigma_p$ are the singular values of $H(x)$.

Let $V = (v_1, v_2, \dots v_n)$, where $v_i \in \mathcal{E}$. Then we can see that $\{v_{n-p+1}, v_{n-p+2}, \dots, v_n\}$ form an orthonormal basis of $\ker H(x)$. So we have, based on the definition of the projection, for a sub-vector space:

$$\mathrm{Proj}_{\ker H(x)}(z) = \left\{ \sum_{i=n-p+1}^{n} \langle z, v_i \rangle v_i \right\}$$

(c)

We can easily see that q is a continuous function, as it is a composition of continuous functions. Since $H(x)$ is continuous, we expect that the (normalized) basis elements obtained through Gram-Schmidt (so through SVD) to not change too much. And since $H(x)$ is always of full rank p, we expect that $\|\text{Proj}_{\ker H(x)}(z)\|$ to be continuous. Then q is just this composed with ∇f which is continuous by assumption. So q is continuous.

Part 2 : A Frank-Wolfe algorithm

Question 1

we can see that $x \mapsto \langle w, x \rangle$ is a linear function, so it is continuous. Moreover, S is a compact, so by Weierstrass the value:

$$\min_{x \in S} \langle w, x \rangle$$

is attained.

Question 2

Set $S = [-1, 1] \times \{0\}$ and $w = (0, 1)$. Then we have:

$$\langle w, x \rangle = 0 \quad \forall\, x \in S$$

So all of S is a minimizer.

Question 3

As S is convex, we have that the points:

$$(1 - \eta)x_k + \eta s(x_k) \in S$$

Are in S for all $x_k \in S$ and $\eta \in [0, 1]$. It is important to enforce the restriction as otherwise x_{k+1} might not be feasible.

Question 4

We analyze four key inequalities (B1) to (B4) under the assumptions that f is convex and continuously differentiable, and its gradient ∇f is L-Lipschitz continuous.

Inequality Analysis

(B1) $f(x_{k+1}) - f(x_k) \leq \nabla f(x_k)^\top (x_{k+1} - x_k) + \frac{L}{2} \|x_{k+1} - x_k\|^2$

This is due to ∇f is L-Lipschitz continuous, and Theorem 3.2. from the lecture notes.

(B2) $\leq \eta_k \nabla f(x_k)^\top (s(x_k) - x_k) + \frac{L}{2} \eta_k^2 d_S^2$

Using the update formula $x_{k+1} = (1 - \eta_k)x_k + \eta_k s(x_k)$ and the definition of d_S, the diameter of S:

$$\begin{aligned}
\|x_{k+1} - x_k\|^2 &= \|(1 - \eta_k)x_k + \eta_k s(x_k) - x_k\|^2 \\
&= \|\eta_k (s(x_k) - x_k)\|^2 \\
&= \eta_k^2 \|s(x_k) - x_k\|^2 \\
&\leq \eta_k^2 d_S^2
\end{aligned}$$

and

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

(B3) $\leq \eta_k \nabla f(x_k)^\top (x^* - x_k) + \frac{L}{2} \eta_k^2 d_S^2$

Given that $s(x_k)$ minimizes the linear approximation over S, the inequality follows by comparing $\nabla f(x_k)^T s(x_k)$ to $\nabla f(x_k)^T x^*$.

(B4) $\leq \eta_k(f(x^*) - f(x_k)) + \frac{L}{2}\eta_k^2 d_S^2$

This follows from the convexity of f, which implies $f(x^*) - f(x_k) \geq \nabla f(x_k)^\top (x^* - x_k)$. Substituting this into (B3) yields (B4).

Question 5

We plug $k = 0$ in the inequality (B4) from the previous question:

$$f(x_1) - f(x_0) \leq \eta_0(f(x^*) - f(x_0)) + \frac{L}{2}\eta_0^2 d_S^2$$

As $\eta_0 = 1$ we have :

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

Question 6

We will prove the inequality by induction on k. The base case has been proven in the previous question. Let us assume that the inequality holds for all $i \leq k$. Then we have:

$$\begin{aligned}
f(x_{k+1}) - f(x^*) &= f(x_{k+1}) - f(x_k) + f(x_k) - f(x^*) \\
&\leq \eta_k(f(x^*) - f(x_k)) + \frac{L}{2}\eta_k^2 d_S^2 + f(x_k) - f(x^*) \\
&\leq (1 - \eta_k)(f(x_k) - f(x^*)) + \frac{L4d_S^2}{2(k+2)^2} \\
&\leq \frac{k}{k+2} \cdot \frac{2Ld_S^2}{k+2} + \frac{2Ld_S^2}{(k+2)^2} \\
&= Ld_S^2 \frac{2k+2}{(k+2)^2} \leq Ld_S^2 \frac{2}{k+3}
\end{aligned}$$

This concludes the proof by induction.

Question 7

$$\Delta^n = \left\{ x \in \mathbb{R}^n : \sum_{i=1}^{n} x_i = 1;\; x_i \geq 0;\; \forall i \in \{1, 2, \ldots, n\} \right\} \implies$$

$$\Delta^n = \{x \in \mathbb{R}^n : Ax \leq b\}$$

Where :

$$A = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ -1 & -1 & \ldots & -1 \\ -1 & 0 & \ldots & 0 \\ 0 & -1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & -1 \end{pmatrix} \in \mathbb{R}^{(n+2)\times n}$$

and:

$$b = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{n+2}$$

Then clearly, Δ^n is closed and convex, as it is a convex polytope. Moreover, we have:

$$\|x\|_\infty \leq 1, \forall\, x \in \Delta^n$$

So Δ^n is compact and non-empty since $(1, 0, \ldots, 0) \in \Delta^n$.

Question 8

Let $w = (w_1, w_2, \ldots, w_n)$ and $i \in \mathbb{N}$ such that $w_i = \min(\{w_1, w_2, \ldots, w_n\})$
Then we have:

$$\langle w, x\rangle = \sum_{j=1}^{n} w_j x_j \geq \sum_{j=1}^{n} w_i x_j = w_i \sum_{j=1}^{n} x_j = w_i$$

This is exactly attained for $\bar{x} = (0, \ldots, 1, \ldots, 0)$ where the 1 is at the i-th position. So the linear program has at least one solution.

The computational complexity of finding $\bar{x}$ is $O(n)$ as we need to find the minimum of n numbers.

Question 9

f is a continuous map defined on a compact set, so the min is well-defined and attained by Weierstrass.

The solution is not unique, for example, consider:

$$A = (1, 0, \ldots 0)$$

And $b < 0$, then clearly all the points of the form $(0, x_2, \ldots, x_n)$ are minimizers.

Question 10

From Exercise sheet 4 Exercise 2 we have that:

$$\nabla f(x) = A^T(Ax - b)$$

Question 11

We analyze the line-search function $g(\eta) = f((1-\eta)x + \eta y)$ where $x, y \in \Delta_n$ and $f(x) = \frac{1}{2}\|Ax - b\|^2$ to determine the optimal values of $\eta \in [0, 1]$. Since f is convex, g is convex as well. So we search for solutions of the equation $g'(\eta) = 0$ to find the minimum of g. We have:

$$g'(\eta) = \langle \nabla f((1-\eta)x + \eta y), y - x\rangle = 0$$
$$\langle A^T(A((1-\eta)x + \eta y) - b), y - x\rangle = 0$$
$$\langle A^T(b - Ax), y - x\rangle = \eta \|A(y - x)\|^2 \implies$$

$$\eta = \frac{\langle A^T(b - Ax), y - x\rangle}{\|A(y - x)\|^2}$$

If this value is not in $[0, 1]$ then we set $\eta := 0$ if $\eta < 0$ and $\eta := 1$ if $\eta > 1$.

Question 12

```
function nabla= nablaf(x,A,b)
    nabla= A.'*(A*x-b);
end
```

```
function sx=linearsubproblem(w)
    sx=double(w<=min(w));
    sx=sx./sum(sx);
end
```

```
function [xbar, gaps]= frank_wolfe(A, x0, b)
steps_max=1e5;
tol=1e-3;

xbar=x0;
i=0;
nabla=nablaf(xbar,A,b);
sx=linearsubproblem(nabla);
gap=dot(nabla,xbar-sx);
gaps=[gap];
while(i<steps_max&& gap>tol)
    eta=min( dot(sx-xbar,-nabla)/dot(A*(sx-xbar),A*(sx-xbar)), 1)
        ;
    eta=max(eta,0);
    xbar=(1-eta)*xbar+eta*sx;
    nabla=nablaf(xbar,A,b);
    sx=linearsubproblem(nabla);
    gap=dot(nabla,xbar-sx);
    gaps=[gaps,gap];
    i=i+1;
end
```

Question 13

```
clear
load data.mat


x0=abs(randn(size(x)));
x0=x0./sum(x0);

[xbar, gaps] =frank_wolfe(A,x0,b);

semilogy(gaps);

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

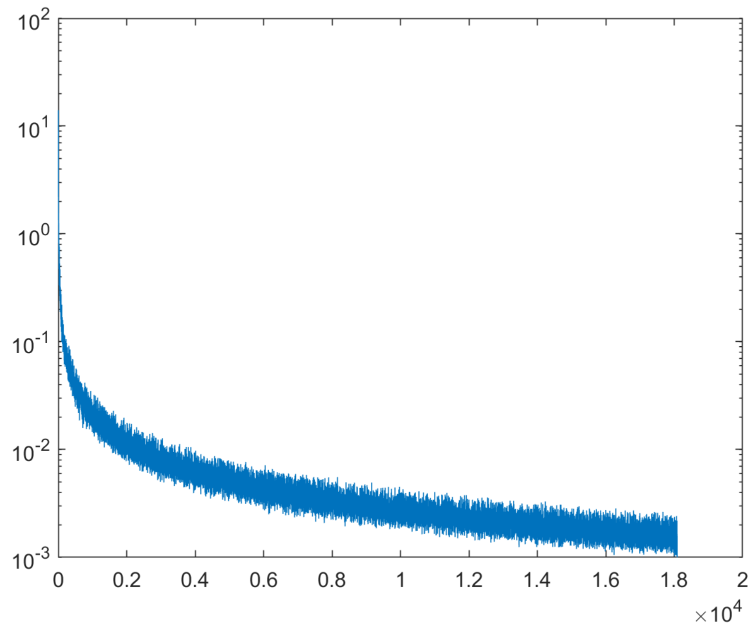

Figure 1: Plot of the Frank-Wolfe gap versus the number of iterations.

Question 14

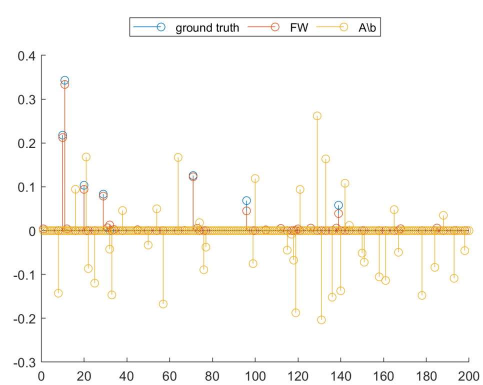

Figure 2: Comparison of the Frank-Wolfe algorithm, the true value and the vector obtained by the Matlab solver.

Part 3 : KKT conditions and constraint qualifications

Question 1

Notice that:

$$
\begin{gathered}
f_i(x) \leq y;\ \forall\ i = 1, \ldots, N \iff \\
\max_{i=1,\ldots,N} f_i(x) \leq y \iff \\
f(x) \leq y \implies \\
\min_x f(x) \leq \min_{(x,y)\in S} y
\end{gathered}
$$

The other inequality is trivial as $(x, f(x)) \in S$. So we can easily conclude that the 2 programs have the same optimal value.

Question 2

Notice that the constraints for S are:

$$g_i(x, y) = f_i(x) - y \leq 0;\ \forall\ i = 1, \ldots, N$$

It is easy to see that:

$$\nabla g_i(x, y) = \begin{pmatrix} \nabla f_i(x) \\ -1 \end{pmatrix}$$

The KKT conditions for $(x, y) \in S$ are: (x, y) is a KKT point if there exists $\lambda \in \mathbb{R}^N$, with $\lambda \geq 0$ such that:

$$-(0, \ldots, 0, 1) = \sum_{i=1}^{N} \lambda_i (\nabla f_i(x)^T, -1)$$

and

$$\lambda_i (f_i(x) - y) = 0;\ \forall\ i = 1, \ldots, N$$

Question 3

Let $n = 1$ and:

$$f_1(x) = x,\ f_2(x) = x^2,\ f_3(x) = x^3$$

Then:

$$\nabla g_1(x, y) = \begin{pmatrix} 1 \\ -1 \end{pmatrix},\ \nabla g_2(x, y) = \begin{pmatrix} 2x \\ -1 \end{pmatrix},\ \nabla g_3(x, y) = \begin{pmatrix} 3x^2 \\ -1 \end{pmatrix}$$

For $(x, y) = (1, 1) \in S$, it is clear that LICQ doesn't hold as $\nabla g_1(1, 1),\ \nabla g_2(1, 1),\ \nabla g_3(1, 1)$ are linearly dependent as a family of 3 vectors in $\mathbb{R}^2$ ($g_i(1, 1) = 0$ for all i).

Question 4

Let $I(x, y) = \{i \in \{1, \ldots, N\} \mid g_i(x, y) = 0\}$. Then for MFCQ to hold, for all $(x, y) \in S$ we need to find a point $(\tilde{x}, \tilde{y}) \in S$ such that:

$$\langle \nabla g_i(x, y), (\tilde{x} - x, \tilde{y} - y) \rangle < 0;\ \forall\ i \in I(x)$$

Substituting ∇g_i, we need to find $(\tilde{x}, \tilde{y}) \in S$ such that:

$$\left\langle \begin{pmatrix} \nabla f_i(x) \\ -1 \end{pmatrix}, (\tilde{x} - x, \tilde{y} - y) \right\rangle < 0;\ \forall\ i \in I(x)$$

But notice that if we set $\tilde{x} := x$ and $\tilde{y} := y + 1$ then:

$$\left\langle \begin{pmatrix} \nabla f_i(x) \\ -1 \end{pmatrix}, (0,1) \right\rangle = \langle -1, 1 \rangle = -1 < 0; \ \forall\, i \in I(x)$$

So MFCQ holds for all $(x, y) \in S$.
(**Note**:$(x, y+1) \in S$ as $y + 1 > y \geq f_i(x)$, for all i)