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:

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

is well-defined and non-empty.

We want to show that $\mathrm{Proj}_{Q'}(z) = \mathrm{Proj}_Q(z)$. 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, $\mathrm{Proj}_{Q'}(z) = \mathrm{Proj}_Q(z)$.

Question 2

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

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

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

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:

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

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:

$$\langle \nabla g(v), w\rangle \geq 0, \quad \forall\ w \in T_vC$$
$$\langle v - z, w\rangle \geq 0, \quad \forall\ w \in T_vC$$

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:

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

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

$$\langle v - z, v\rangle \geq 0$$
$$\langle v - z, -v\rangle \geq 0 \implies$$
$$-\langle v - z, v\rangle \geq 0$$

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:

$$\langle v_1, z - v_1\rangle = 0 \implies \|v_1\|^2 = \langle v_1, v_1\rangle = \langle v_1, z\rangle$$
$$\langle v_2, z - v_2\rangle = 0 \implies \|v_2\|^2 = \langle v_2, v_2\rangle = \langle v_2, z\rangle$$

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

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

Question 6

We present an example where the function $q(x) = \|\mathrm{Proj}_{T_xS}(-\nabla f(x))\|$ 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 $\mathrm{Proj}_{T_xS}(-\nabla f(x))$ for all $x \in S$. We have:

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

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

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

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:

$$UH(x)V^T = D$$

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

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

Then, since U and V are orthogonal, we have:

$$v \in T_xS \iff H(x)v = 0 \iff UH(x)V^TVv = 0 \iff DVv = 0 \iff Vv \in \ker D$$

and, as V is orthogonal, we have:

$$\frac{1}{2}\|z - v\|^2 = \frac{1}{2}\|Vz - Vv\|^2$$

So, we have:

$$\text{Proj}_{T_xS}(z) = \text{Proj}_{\ker D}(Vz)$$

Part 2 : A Frank-Wolfe algorithm

Question 1

Question 2

To demonstrate that the minimization problem may have multiple solutions, consider the following example:

- **Set** S: Let S be a line segment in $\mathbb{R}^2$ defined as $S = \{(1, y) : 0 \leq y \leq 1\}$.
- **Linear Function**: Consider a linear function $\langle w, x\rangle$ with $w = (0, 0)$. For any $x \in S$, we have $\langle w, x\rangle = 0$.

In this case, every point in S minimizes the function $\langle w, x\rangle$ as the value is zero for all $x \in S$. Hence, the problem has multiple solutions, with every point in the set S being a solution.

Question 3

Why is the Restriction $0 \leq \eta_k \leq 1$ Important?

1. **Feasibility**: The feasible set S is assumed to be convex. By convexity, for any $x, y \in S$ and $\lambda \in [0, 1]$, the convex combination $(1 - \lambda)x + \lambda y$ is also in S. In the algorithm, both x_k and $s(x_k)$ are in S, so for η_k in $[0, 1]$, the updated point $(1 - \eta_k)x_k + \eta_k s(x_k)$ remains within S.
2. **Convergence**: The step size η_k controls the magnitude of the move towards the direction of minimization. Values of η_k outside the interval $[0, 1]$ can lead to overshooting or even divergence. Specifically, $\eta_k > 1$ may cause the algorithm to take excessively large steps, while negative values of η_k would reverse the direction of the update, both hindering convergence.
3. **Controlled Progress**: The interval $[0, 1]$ allows for dynamic adjustment of η_k to control the algorithm's progress. Smaller values of η_k can be used for cautious steps near the optimal solution, enhancing stability and precision.
4. **Balance Between Exploration and Exploitation**: η_k balances exploration of the feasible set S and exploitation towards the minimizer of the linearized function. $\eta_k = 0$ implies no movement (pure exploitation), while $\eta_k = 1$ means moving entirely towards the new direction (pure exploration). Intermediate values facilitate a balanced approach.

In conclusion, the restriction $0 \leq \eta_k \leq 1$ in the Frank-Wolfe algorithm is essential for ensuring feasibility, convergence, controlled progress, and a balanced approach between exploration and exploitation.

Question 4

We analyze four key inequalities (B1) to (B4) arising in the Frank-Wolfe algorithm 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 inequality stems from the first-order Taylor expansion and the Lipschitz continuity of ∇f, bounding the error of the linear approximation.

(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, this inequality bounds the change in f in terms of the diameter of S and step size η_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 $s(x_k)$ to any $x^* \in S$, including the optimal point.

(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

Given $x_0 \in S$, let x_1 be produced by the Frank-Wolfe algorithm with $\eta_0 = \frac{2}{0+2} = 1$. We show that $f(x_1) - f(x^*) \leq \frac{L}{2} d_S^2$, where L is the Lipschitz constant of ∇f and d_S is the diameter of S.

Proof

1. The update rule for x_{k+1} in the Frank-Wolfe algorithm is $x_{k+1} = (1-\eta_k)x_k + \eta_k s(x_k)$. For $k = 0$, this becomes $x_1 = s(x_0)$.
2. By the L-Lipschitz continuity of ∇f, we have
$$f(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{L}{2}\|y - x\|^2 \quad \text{for all } x, y \in S.$$
3. Setting $x = x_1$ and $y = x^*$, we get
$$f(x^*) \leq f(x_1) + \nabla f(x_1)^\top (x^* - x_1) + \frac{L}{2}\|x^* - x_1\|^2.$$
4. Rearranging, we obtain
$$f(x_1) - f(x^*) \geq -\nabla f(x_1)^\top (x^* - x_1) - \frac{L}{2}\|x^* - x_1\|^2.$$
5. Since x_1 and x^* are in S and $\|x^* - x_1\|^2 \leq d_S^2$, we have
$$f(x_1) - f(x^*) \leq \frac{L}{2} d_S^2.$$

Thus, after the first iteration with $\eta_0 = 1$, the function value at x_1 is within $\frac{L}{2} d_S^2$ of the optimal value $f(x^*)$.

Question 6

We prove that for the Frank-Wolfe algorithm with step sizes $\eta_k = \frac{2}{k+2}$, the inequality $f(x_k) - f(x^*) \leq \frac{2Ld_S^2}{k+2}$ holds for all $k \geq 1$.

Proof by Induction

Base Case ($k = 1$)

From the previous analysis, we have $f(x_1) - f(x^*) \leq \frac{Ld_S^2}{2}$, which satisfies the inequality for $k = 1$, as $\frac{2Ld_S^2}{1+2} = \frac{Ld_S^2}{2}$.

Inductive Step

Assume the inequality holds for some $k \geq 1$:

$$f(x_k) - f(x^*) \leq \frac{2Ld_S^2}{k+2}$$

We need to show it holds for $k + 1$:

$$f(x_{k+1}) - f(x^*) \leq \frac{2Ld_S^2}{k+3}$$

Using the inequality (B4) from the algorithm:

$$f(x_{k+1}) - f(x_k) \leq \eta_k(f(x^*) - f(x_k)) + \frac{L}{2}\eta_k^2 d_S^2$$

where $\eta_k = \frac{2}{k+2}$. Substituting and rearranging gives:

$$\begin{aligned}
f(x_{k+1}) - f(x^*) &\leq \left(1 - \frac{2}{k+2}\right)\frac{2Ld_S^2}{k+2} + \frac{L}{2}\left(\frac{2}{k+2}\right)^2 d_S^2 \\
&= \frac{kLd_S^2}{(k+2)^2} + \frac{2Ld_S^2}{(k+2)^2} \\
&= \frac{(k+2)Ld_S^2}{(k+2)^2} \\
&= \frac{Ld_S^2}{k+2}
\end{aligned}$$

Using $\frac{2}{k+2} \leq \frac{2}{k+3}$, we get:

$$f(x_{k+1}) - f(x^*) \leq \frac{2Ld_S^2}{k+3}$$

Thus, by induction, the inequality holds for all $k \geq 1$.

Question 7

We show that the simplex $\Delta_n = \{x \in \mathbb{R}^n | x_1, \ldots, x_n \geq 0 \text{ and } x_1 + \cdots + x_n = 1\}$ is convex, compact, and non-empty.

Convexity

A set is convex if for any two points in the set, the line segment between them is also in the set. For $x, y \in \Delta_n$ and $\lambda \in [0, 1]$, consider $z = (1 - \lambda)x + \lambda y$. Since $x_i, y_i \geq 0$, each component $z_i = (1 - \lambda)x_i + \lambda y_i \geq 0$. Also, $\sum_{i=1}^n z_i = (1 - \lambda)\sum_{i=1}^n x_i + \lambda \sum_{i=1}^n y_i = 1$. Hence, $z \in \Delta_n$, proving convexity.

Compactness

A set is compact if it is closed and bounded. Δ_n is closed as it contains all its limit points. It is bounded because for all $x \in \Delta_n$, $0 \leq x_i \leq 1$ and $\sum_{i=1}^n x_i = 1$. Therefore, Δ_n is compact.

Non-emptiness

Δ_n is non-empty as it contains at least the point $x = (1, 0, \ldots, 0)$, which satisfies $x_i \geq 0$ and $\sum_{i=1}^n x_i = 1$.

In conclusion, the simplex Δ_n is convex, compact, and non-empty.

Question 8

Minimization Problem on the Simplex Δ_n

Given a vector $w \in \mathbb{R}^n$, we consider the problem of minimizing $\langle w, x \rangle$ subject to $x \in \Delta_n$, where Δ_n is the simplex in $\mathbb{R}^n$.

Minimum of the Problem

The problem is formulated as:

$$\text{minimize} \quad \langle w, x \rangle \quad \text{subject to} \quad x \in \Delta_n.$$

Strategy to Attain the Smallest Value

To minimize $\langle w, x \rangle$, we allocate the entire weight to the component of x corresponding to the smallest component of w. Let $i^* = \arg\min_i w_i$. The minimizing vector x is such that $x_{i^*} = 1$ and $x_i = 0$ for all $i \neq i^*$.

Computational Complexity

The computational complexity of finding this solution is $O(n)$, as it requires a linear scan to find the minimum component of the vector w. The minimizing x is then directly obtained from the index of this minimum component.

Question 9

Given the optimization problem $\min_{x \in \Delta_n} f(x)$ with $f(x) = \frac{1}{2}\|Ax - b\|^2$, we analyze whether this problem always has a solution and if the solution is unique.

Existence of a Solution

- **Convexity of $f(x)$**: The function $f(x) = \frac{1}{2}\|Ax - b\|^2$ is convex as it is the composition of a convex function (norm squared) with an affine function.
- **Convexity and Compactness of Δ_n**: The simplex Δ_n is convex and compact.
- **Existence of Solution**: A convex function over a compact convex set attains its minimum. Hence, the problem always has at least one solution.

Uniqueness of the Solution

- **Strict Convexity**: Strict convexity is necessary for uniqueness. The function $f(x)$ is strictly convex if the matrix A has full column rank. However, with $m < n$, A cannot have full column rank.
- **Multiple Solutions**: When A does not have full column rank, there can be multiple minimizers of $f(x)$, due to directions in which A is not injective.

In conclusion, the optimization problem always has a solution, but it does not always have a unique solution due to the potential rank deficiency of A.

Given the optimization problem $\min_{x\in\Delta_n} f(x)$ with $f(x) = \frac{1}{2}\|Ax-b\|^2$, we analyze whether this problem always has a solution and if the solution is unique.

Existence of a Solution

- **Convexity of $f(x)$**: The function $f(x) = \frac{1}{2}\|Ax-b\|^2$ is convex as it is the composition of a convex function (norm squared) with an affine function.
- **Convexity and Compactness of Δ_n**: The simplex Δ_n is convex and compact.
- **Existence of Solution**: A convex function over a compact convex set attains its minimum. Hence, the problem always has at least one solution.

Uniqueness of the Solution

- **Strict Convexity**: Strict convexity is necessary for uniqueness. The function $f(x)$ is strictly convex if the matrix A has full column rank. However, with $m < n$, A cannot have full column rank.
- **Multiple Solutions**: When A does not have full column rank, there can be multiple minimizers of $f(x)$, due to directions in which A is not injective.

In conclusion, the optimization problem always has a solution, but it does not always have a unique solution due to the potential rank deficiency of A.

Question 10

Gradient of the Function for Frank-Wolfe Algorithm

We derive the gradient of the function $f(x) = \frac{1}{2}\|Ax-b\|^2$ for applying the Frank-Wolfe algorithm to the problem $\min_{x\in\Delta_n} f(x)$.

The function $f(x)$ is given by:

$$f(x) = \frac{1}{2}\|Ax-b\|^2 = \frac{1}{2}(Ax-b)^\top(Ax-b).$$

Differentiating $f(x)$ with respect to x using matrix calculus, we obtain the gradient of $f(x)$:

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

This gradient represents the direction of the steepest ascent at any point x for the function $f(x)$ and is essential for determining the search direction in each iteration of the Frank-Wolfe algorithm.

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 $\eta \in [0,1]$.

Expression for $g(\eta)$

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

Optimal Value of η

To find the optimal η, we differentiate $g(\eta)$ with respect to η and set the derivative to zero:

$$g'(\eta) = \frac{d}{d\eta}\frac{1}{2}\|A((1-\eta)x + \eta y) - b\|^2$$

$$= (A((1-\eta)x + \eta y) - b)^\top A(y-x)$$

Setting $g'(\eta) = 0$ gives:

$$(A((1-\eta)x + \eta y) - b)^\top A(y-x) = 0$$

Solving this equation for η gives the optimal value.

Closed-Form Formula

A closed-form expression for η depends on the specific structure of A, b, x, and y. Without additional assumptions, the exact solution might be complex or not directly obtainable.

Question 12

Question 13

Question 14

1

(4):

$$\min_{x \in \mathbb{R}^n} f(x), \quad \text{where} \quad f(x) = \max_{i=1,\ldots,N} f_i(x).$$

(5):

$$\min_{(x,y) \in S} y, \quad \text{where} \quad S = \{(x, y) \in \mathbb{R}^n \times \mathbb{R} \,|\, f_i(x) - y \leq 0,\, i = 1, \ldots, N\}.$$

To establish the equivalence in optimal value between problems (4) and (5), we proceed as follows:

For any $x \in \mathbb{R}^n$, define $y_x = \max_{i=1,\ldots,N} f_i(x)$. Clearly, $y_x \geq f_i(x)$ for all i, which implies $(x, y_x) \in S$. This means for every feasible x in problem (4), there is a corresponding feasible (x, y_x) in problem (5) with the same objective value. Conversely, for any feasible (x, y) in problem (5), since $f_i(x) \leq y$ for all i, it follows that $\max_{i=1,\ldots,N} f_i(x) \leq y$. Therefore, the optimal value of x in problem (4) cannot exceed the optimal value of (x, y) in problem (5). This establishes that problems (4) and (5) have the same optimal value.

2

The KKT conditions for problem (5) involve:

- Primal feasibility: $f_i(x) - y \leq 0$ for all i.
- Dual feasibility: $\lambda_i \geq 0$ for all i.
- Complementary slackness: $\lambda_i(f_i(x) - y) = 0$ for all i.
- Stationarity: The gradient of the Lagrangian with respect to x and y is zero.

The Lagrangian for problem (5) is:

$$\mathcal{L}(x, y, \lambda) = y + \sum_{i=1}^{N} \lambda_i (f_i(x) - y).$$

The stationarity condition then yields:

$$\nabla_x \mathcal{L}(x, y, \lambda) = \sum_{i=1}^{N} \lambda_i \nabla f_i(x) = 0,$$

$$\nabla_y \mathcal{L}(x, y, \lambda) = 1 - \sum_{i=1}^{N} \lambda_i = 0.$$

3. Example of Functions $f_1, \ldots, f_N$ Where LICQ does not Hold

Consider the functions $f_1(x) = x^2$ and $f_2(x) = (x-1)^2$ for $x \in \mathbb{R}$ (here $N = 2$). At the point $x = 0.5$, both f_1 and f_2 are active, and their gradients are equal ($\nabla f_1(0.5) = \nabla f_2(0.5)$). Thus, the gradients of the active constraints are not linearly independent, violating the Linear Independence Constraint Qualification (LICQ).

4

The Mangasarian-Fromovitz Constraint Qualification (MFCQ) requires the existence of a direction where all the inequality constraints are strictly decreasing. For problem (5), the constraint is $f_i(x) - y \leq 0$. By increasing y slightly (while keeping x constant), we strictly decrease $f_i(x) - y$ for all i. Therefore, MFCQ is always satisfied for any feasible point in problem (5), as there is always a direction (increasing y) that strictly reduces all the constraints.