Technical University of Crete

School of Electrical and Computer Engineering

Course : **Optimization**

Exercise 3

Instructor : Athanasios P. Liavas

Students : Toganidis Nikos - 2018030085

Tzimpimpaki Evangelia - 2018030165

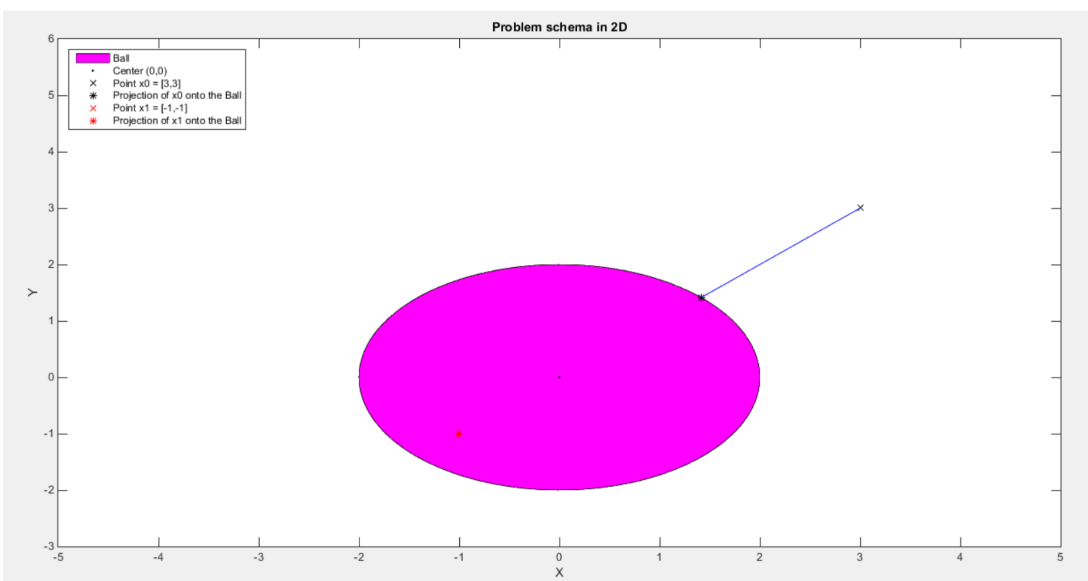
---

(1) Computing the projection of $\mathbf{x_0} \in \mathbb{R}^n$ onto the set $\mathbf{B}(\mathbf{0}, r) := \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_2 \leq r\}$.

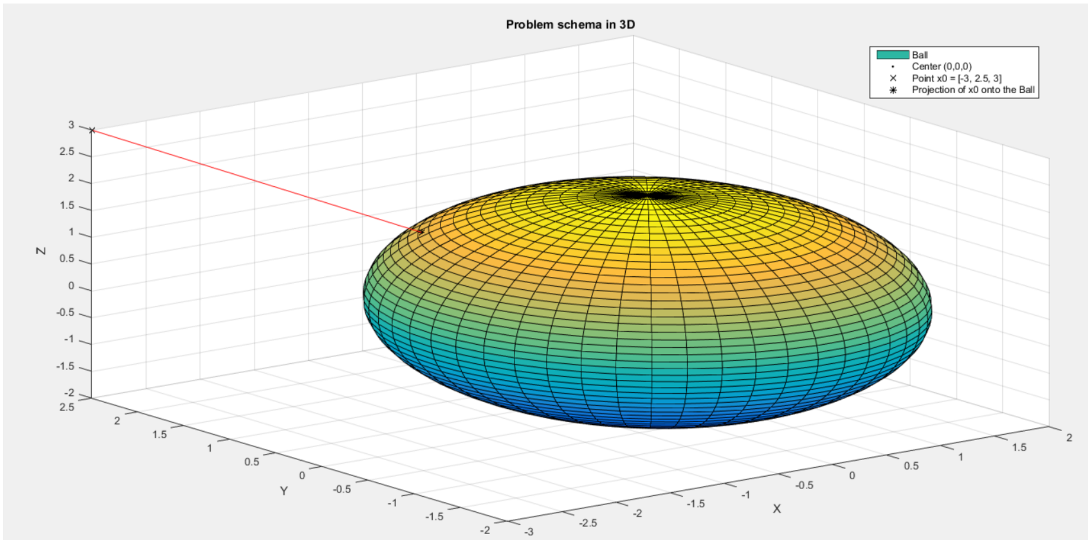
(a) Drawing a scheme of the problem

Set $\mathbf{B}(\mathbf{0}, r) := \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_2 \leq r\}$ is a ball with center the zero vector $\mathbf{0}$ and radius equal to r.

For $n = 2$, set $\mathbf{B}(\mathbf{0}, r)$ is a circular disk, as depicted bellow :



For $n = 3$, set $\mathbf{B}(\mathbf{0}, r)$ is a sphere, as depicted bellow :



(b) Optimization problem to be solved

For a convex set $\mathbf{B} \subset \mathbb{R}^n$ and a point $\mathbf{x}_0 \in \mathbb{R}^n$, the projection of $\mathbf{x}_0$ onto the set $\mathbf{B}$ is the solution of the following problem :

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) := \frac{1}{2} \|\mathbf{x}_0 - \mathbf{x}\|_2^2 \\ \text{subject to} & f_1(\mathbf{x}) = \frac{1}{2} \left(\|\mathbf{x}\|_2^2 - r^2 \right) \leq 0 \end{array}$$

Set $\mathbf{B}$ is convex since for two points $\mathbf{x_1}, \mathbf{x_2} \in \mathbf{B}(\mathbf{0}, r)$ and for $\theta \in (0, 1)$ we get :

$$\mathbf{x_1} \in \mathbf{B}(\mathbf{0}, r) \Rightarrow ||\mathbf{x_1}||_2 \leq r,$$
$$\mathbf{x_2} \in \mathbf{B}(\mathbf{0}, r) \Rightarrow ||\mathbf{x_2}||_2 \leq r$$

$$||(1-\theta)\mathbf{x_1} + \theta \mathbf{x_2}||_2 \leq ||(1-\theta)\mathbf{x_1}||_2 + ||\theta \mathbf{x_2}||_2 = |1-\theta| \, ||\mathbf{x_1}||_2 + |\theta| \, ||\mathbf{x_2}||_2 =$$

$$= (1-\theta)||\mathbf{x_1}||_2 + \theta ||\mathbf{x_2}||_2$$

$$\leq (1-\theta)r + \theta r = r,$$

which means that $(1-\theta)\mathbf{x_1} + \theta \mathbf{x_2} \in \mathbf{B}(\mathbf{0}, r)$.

(c) KKT conditions

The KKT conditions we will use are the following :

- $\nabla f_0 \left(\mathbf{x}_*\right) + \lambda_* \nabla f_1 \left(\mathbf{x}_*\right) = 0 \quad (\mathbf{1})$
- $\lambda_* \geq 0 \quad (\mathbf{2})$
- $f_1(\mathbf{x}_*) \leq 0 \quad (\mathbf{3})$
- $\lambda_* f_1(\mathbf{x}_*) = 0 \quad (\mathbf{4})$

Computing $\nabla f_0 \left(\mathbf{x}\right)$

$$\nabla f_0 \left(\mathbf{x}\right) = \frac{d}{d\mathbf{x}} \left(f_0 \left(\mathbf{x}\right)\right) = \frac{d}{d\mathbf{x}} \left(\tfrac{1}{2} \,||\mathbf{x_0} - \mathbf{x}||_2^2\right) = \frac{d}{d\mathbf{x}} \left(\tfrac{1}{2}(\mathbf{x_0} - \mathbf{x})^T (\mathbf{x_0} - \mathbf{x})\right) =$$

$$= \frac{d}{d\mathbf{x}} \left(\frac{1}{2}(\mathbf{x_0}^T - \mathbf{x}^T)(\mathbf{x_0} - \mathbf{x})\right) =$$

$$= \frac{1}{2} \left(\frac{d}{d\mathbf{x}}(\mathbf{x_0}^T \mathbf{x_0}) - \frac{d}{d\mathbf{x}}(\mathbf{x_0}^T \mathbf{x}) - \frac{d}{d\mathbf{x}}(\mathbf{x}^T \mathbf{x_0}) + \frac{d}{d\mathbf{x}}(\mathbf{x}^T \mathbf{x})\right) =$$

$$= \frac{1}{2} \left(0 - \mathbf{x_0} - \mathbf{x_0} + (\mathbf{I} + \mathbf{I^T})\mathbf{x}\right) = \frac{1}{2} \left(2\,\mathbf{x} - 2\,\mathbf{x_0}\right) = \mathbf{x} - \mathbf{x_0}$$

Computing $\nabla f_1(\mathbf{x})$

$$\nabla f_1(\mathbf{x}) = \frac{d}{d\mathbf{x}}(f_1(\mathbf{x})) = \frac{d}{d\mathbf{x}} \left(\tfrac{1}{2} \left(\|\mathbf{x}\|_2^2 - r^2\right)\right) = \frac{1}{2} \frac{d}{d\mathbf{x}} \left(\|\mathbf{x}\|_2^2\right) = \frac{1}{2} \frac{d}{d\mathbf{x}} \left(\mathbf{x}^T \mathbf{x}\right) =$$

$$= \frac{1}{2} \,(\mathbf{I} + \mathbf{I^T})\mathbf{x} = \frac{1}{2} \,(2\,\mathbf{I})\mathbf{x} = \mathbf{x}$$

$(\star)$: Identities $\dfrac{d}{d\mathbf{y}}(\mathbf{y}^T \mathbf{a}) = \mathbf{a}$ and $\dfrac{d}{d\mathbf{y}}(\mathbf{y}^T \mathbf{A} \mathbf{y}) = (\mathbf{A} + \mathbf{A^T})\mathbf{y}$ were used.

By substituting the above expressions in (**1**), we get :

$$\nabla f_0\left(\mathbf{x}_*\right) + \lambda_* \nabla f_1\left(\mathbf{x}_*\right) = 0 \Rightarrow \mathbf{x}_* - \mathbf{x_0} + \lambda_* \mathbf{x}_* = 0 \Rightarrow \mathbf{x}_* = \frac{\mathbf{x_0}}{1+\lambda_*}, \quad (\mathbf{5})$$

(d) Case when $\lambda_* > 0$

From (**4**), for $\lambda_* > 0$, we get that :

$$\lambda_* f_1(\mathbf{x}_*) = 0 \Rightarrow f_1(\mathbf{x}_*) = 0 \Rightarrow \tfrac{1}{2}\left(\|\mathbf{x}_*\|_2^2 - r^2\right) = 0 \Rightarrow ||\mathbf{x}_*||_2 = r, \quad (\mathbf{6})$$

From (**5**), we calculate the norm of $\mathbf{x}_*$: $||\mathbf{x}_*||_2 = ||\dfrac{\mathbf{x_0}}{(1+\lambda_*)}||_2 = \dfrac{||\mathbf{x_0}||_2}{1+\lambda_*} \overset{(\mathbf{6})}{=\!=} r$

Solving for λ_* : $\dfrac{1}{1+\lambda_*}||\mathbf{x_0}||_2 = r \Rightarrow ||\mathbf{x_0}||_2 = r(1+\lambda_*) \Rightarrow ||\mathbf{x_0}||_2 = r + r\,\lambda_*$

$$\Rightarrow r\,\lambda_* = ||\mathbf{x_0}||_2 - r \Rightarrow \boxed{\lambda_* = \frac{||\mathbf{x_0}||_2 - r}{r}} \quad (\mathbf{7})$$

By substituting (**7**) in (**5**), we get that :

$$\mathbf{x}_* = \frac{\mathbf{x_0}}{\left(1 + \dfrac{||\mathbf{x_0}||_2 - r}{r}\right)} = \frac{r\,\mathbf{x_0}}{r + ||\mathbf{x_0}||_2 - r} \Rightarrow \boxed{\mathbf{x}_* = \frac{r}{||\mathbf{x_0}||_2}\mathbf{x_0}}$$

The projected point is the point $\mathbf{x}_*$ (solution of the initial problem).

(e) Case when $\lambda_* = 0$

Then (**5**) leads to : $\boxed{\mathbf{x}_* = \mathbf{x_0}}$

The projected point is the initial point (the projection of $\mathbf{x_0}$ is itself, which is absolutely correct for the internal and boundary points of the ball - the ones where $||\mathbf{x_0}||_2 \leq r$).

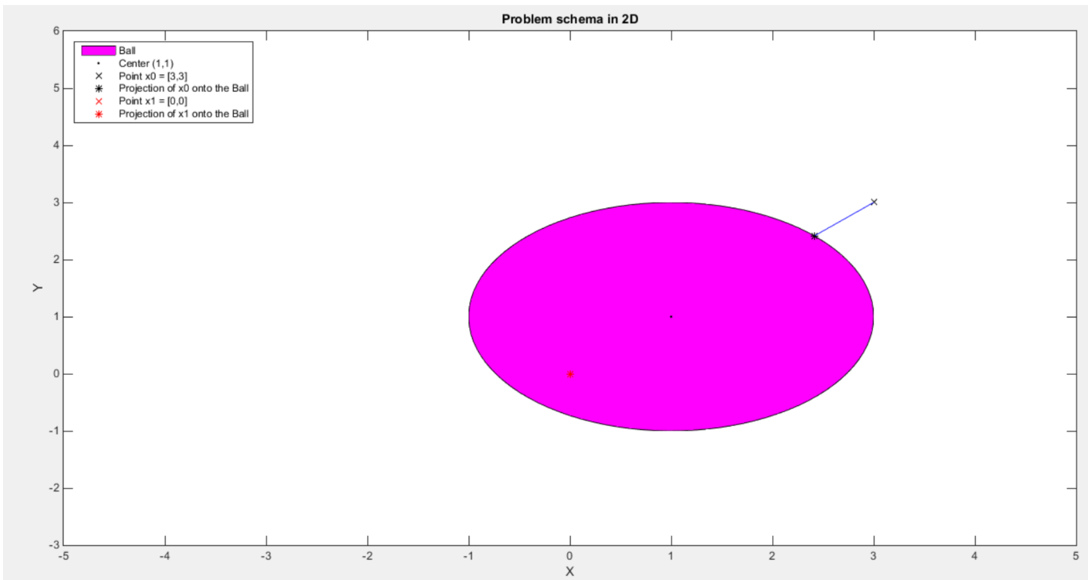
So, generally we have that : $\mathbf{x}_* = \dfrac{r}{\max\left\{r, \|\mathbf{x_0}\|_2\right\}}\mathbf{x_0}$

(2) Computing the projection of $\mathbf{x}_0 \in \mathbb{R}^n$ onto the set $\mathbf{B}(\mathbf{y}, r) := \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{y}\|_2 \leq r\}$.

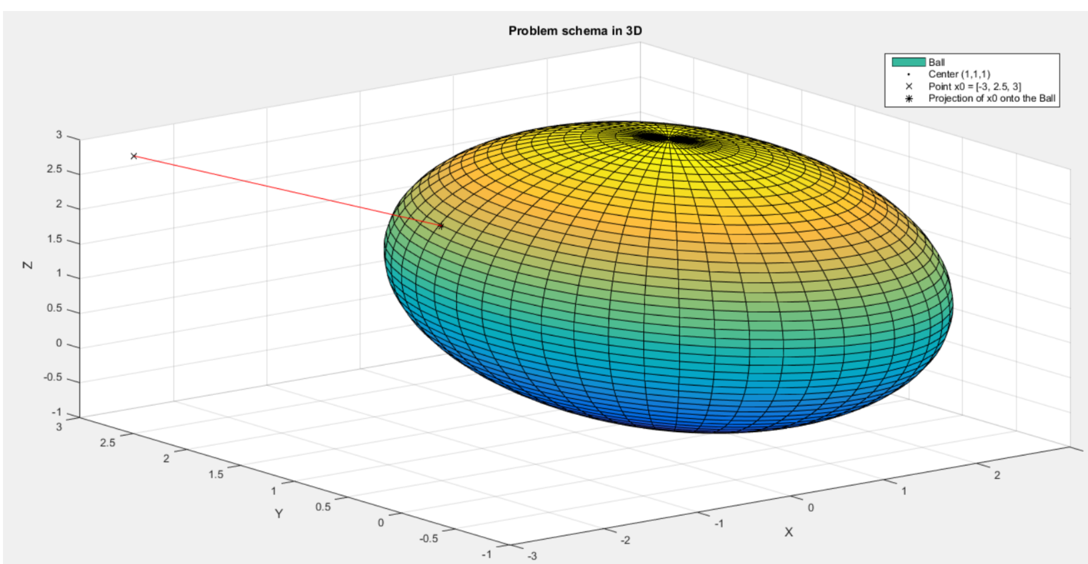
(a) Drawing a scheme of the problem

The problem remains the same but the ball now has a different center $\mathbf{y} \in \mathbb{R}^n$.

For $n = 2$, set $\mathbf{B}(\mathbf{y}, r)$ is a circular disk, as depicted bellow (let $\mathbf{y} = [1, 1]$):



For $n = 3$, set $\mathbf{B}(\mathbf{y}, r)$ is a sphere, as depicted bellow (let $\mathbf{y} = [1, 1, 1]$):



(b) Optimization problem to be solved

For a convex set $\mathbf{B} \subset \mathbb{R}^n$ and a point $\mathbf{x}_0 \in \mathbb{R}^n$, the projection of $\mathbf{x}_0$ onto the set $\mathbf{B}$ is the solution of the following problem :

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) := \frac{1}{2} \|\mathbf{x}_0 - \mathbf{x}\|_2^2 \\ \text{subject to} & f_1(\mathbf{x}) = \frac{1}{2} \left(\|\mathbf{x} - \mathbf{y}\|_2^2 - r^2 \right) \leq 0 \end{array}$$

Set $\mathbf{B}$ is convex since for two points $\mathbf{x_1}, \mathbf{x_2} \in \mathbf{B}(\mathbf{y}, r)$ and for $\theta \in (0, 1)$ we get :

$$\mathbf{x_1} \in \mathbf{B}(\mathbf{y}, r) \Rightarrow ||\mathbf{x_1} - \mathbf{y}||_2 \leq r,$$
$$\mathbf{x_2} \in \mathbf{B}(\mathbf{y}, r) \Rightarrow ||\mathbf{x_2} - \mathbf{y}||_2 \leq r$$

$$\begin{aligned} ||(1-\theta)\mathbf{x_1} + \theta\mathbf{x_2} - \mathbf{y}||_2 &= ||(1-\theta)(\mathbf{x_1} - \mathbf{y}) + \theta(\mathbf{x_2} - \mathbf{y})||_2 \\ &\leq ||(1-\theta)(\mathbf{x_1} - \mathbf{y})||_2 + ||\theta(\mathbf{x_2} - \mathbf{y})||_2 = \\ &= (1-\theta)\,||\mathbf{x_1} - \mathbf{y}||_2 + \theta\,||\mathbf{x_2} - \mathbf{y}||_2 = \\ &\leq (1-\theta)r + \theta r = r, \end{aligned}$$

which means that $(1-\theta)\mathbf{x_1} + \theta\mathbf{x_2} \in \mathbf{B}(\mathbf{y}, r)$.

(c) KKT conditions

The KKT conditions we will use are the following :

- $\nabla f_0\left(\mathbf{x}_*\right) + \lambda_* \nabla f_1\left(\mathbf{x}_*\right) = 0 \quad (\mathbf{1})$
- $\lambda_* \geq 0 \quad (\mathbf{2})$
- $f_1(\mathbf{x}_*) \leq 0 \quad (\mathbf{3})$
- $\lambda_* f_1(\mathbf{x}_*) = 0 \quad (\mathbf{4})$

Computing $\nabla f_0\left(\mathbf{x}\right)$

$\nabla f_0\left(\mathbf{x}\right) = \dfrac{d}{d\mathbf{x}}\left(f_0\left(\mathbf{x}\right)\right) = \dfrac{d}{d\mathbf{x}}\left(\frac{1}{2}\,||\mathbf{x_0} - \mathbf{x}||_2^2\right) = \mathbf{x} - \mathbf{x_0}$, calculated like before.

Computing $\nabla f_1(\mathbf{x})$

$$\nabla f_1(\mathbf{x}) = \frac{d}{d\mathbf{x}}(f_1(\mathbf{x})) = \frac{d}{d\mathbf{x}}\left(\tfrac{1}{2}\left(\|\mathbf{x}-\mathbf{y}\|_2^2 - r^2\right)\right) = \frac{1}{2}\frac{d}{d\mathbf{x}}\left(\|\mathbf{x}-\mathbf{y}\|_2^2\right) =$$

$$= \frac{1}{2}\frac{d}{d\mathbf{x}}\left((\mathbf{x}-\mathbf{y})^T(\mathbf{x}-\mathbf{y})\right) = \frac{1}{2}\left(\frac{d}{d\mathbf{x}}(\mathbf{x}^T\mathbf{x}) - \frac{d}{d\mathbf{x}}(\mathbf{x}^T\mathbf{y}) - \frac{d}{d\mathbf{x}}(\mathbf{y}^T\mathbf{x}) + \frac{d}{d\mathbf{x}}(\mathbf{y}^T\mathbf{y})\right) =$$

$$= \frac{1}{2}\left((\mathbf{I}+\mathbf{I^T})\mathbf{x} - \mathbf{y} - \mathbf{y} + 0\right) = \frac{1}{2}(2\,\mathbf{x} - 2\,\mathbf{y}) = \mathbf{x} - \mathbf{y}$$

$(\star)$: Identities $\frac{d}{d\mathbf{y}}(\mathbf{y}^T\mathbf{a}) = \mathbf{a}$ and $\frac{d}{d\mathbf{y}}(\mathbf{y}^T\mathbf{A}\mathbf{y}) = (\mathbf{A}+\mathbf{A^T})\mathbf{y}$ were used.

By substituting the above expressions in (**1**), we get :

$$\nabla f_0(\mathbf{x}_*) + \lambda_*\nabla f_1(\mathbf{x}_*) = 0 \Rightarrow \mathbf{x}_* - \mathbf{x_0} + \lambda_*(\mathbf{x}_* - \mathbf{y}) = 0 \Rightarrow \mathbf{x}_*(1+\lambda_*) = \mathbf{x_0} + \lambda_*\mathbf{y}$$

$$\Rightarrow \mathbf{x}_* = \frac{\mathbf{x_0} + \lambda_*\mathbf{y}}{1+\lambda_*}, \quad \mathbf{(5)}$$

(d) Case when $\lambda_* > 0$

From (**4**), for $\lambda_* > 0$, we get that :

$$\lambda_* f_1(\mathbf{x}_*) = 0 \Rightarrow f_1(\mathbf{x}_*) = 0 \Rightarrow \tfrac{1}{2}\left(\|\mathbf{x}_* - \mathbf{y}\|_2^2 - r^2\right) = 0 \Rightarrow ||\mathbf{x}_* - \mathbf{y}||_2 = r, \quad \mathbf{(6)}$$

In (**6**) we substitute (**5**) and we have:

$$||\mathbf{x}_* - \mathbf{y}||_2 = ||\frac{\mathbf{x_0} + \lambda_*\mathbf{y}}{1+\lambda_*} - \mathbf{y}||_2 = ||\frac{\mathbf{x_0} + \lambda_*\mathbf{y} - \mathbf{y} - \lambda_*\,\mathbf{y}}{1+\lambda_*}||_2 = ||\frac{\mathbf{x_0} - \mathbf{y}}{1+\lambda_*}||_2 =$$

$$= \frac{||\mathbf{x_0} - \mathbf{y}||_2}{1+\lambda_*} \overset{\mathbf{(6)}}{=\!=} r$$

Solving for λ_* :

$$\frac{||\mathbf{x_0} - \mathbf{y}||_2}{1+\lambda_*} = r \Rightarrow \frac{||\mathbf{x_0} - \mathbf{y}||_2}{r} = 1 + \lambda_* \Rightarrow \boxed{\lambda_* = \frac{||\mathbf{x_0} - \mathbf{y}||_2 - r}{r}} \quad \mathbf{(7)}$$

By substituting (**7**) in (**5**), we get that :

$$\mathbf{x}_* = \frac{\mathbf{x_0} + \lambda_* \mathbf{y}}{1+\lambda_*} = \frac{\mathbf{x_0} + \dfrac{||\mathbf{x_0}-\mathbf{y}||_2 - r}{r}\mathbf{y}}{1 + \dfrac{||\mathbf{x_0}-\mathbf{y}||_2 - r}{r}} = \frac{r\,\mathbf{x_0} + (||\mathbf{x_0}-\mathbf{y}||_2 - r)\mathbf{y}}{r + ||\mathbf{x_0}-\mathbf{y}||_2 - r} =$$

$$= \frac{r\,\mathbf{x_0} + (||\mathbf{x_0}-\mathbf{y}||_2 - r)\mathbf{y}}{||\mathbf{x_0}-\mathbf{y}||_2} = \mathbf{y} + \frac{r\,\mathbf{x_0} - r\mathbf{y}}{||\mathbf{x_0}-\mathbf{y}||_2}$$

$$\Rightarrow \boxed{\mathbf{x}_* = \mathbf{y} + r\frac{\mathbf{x_0}-\mathbf{y}}{||\mathbf{x_0}-\mathbf{y}||_2}}$$

The projected point is the point $\mathbf{x}_*$ (solution of the initial problem).

(e) Case when $\lambda_* = 0$

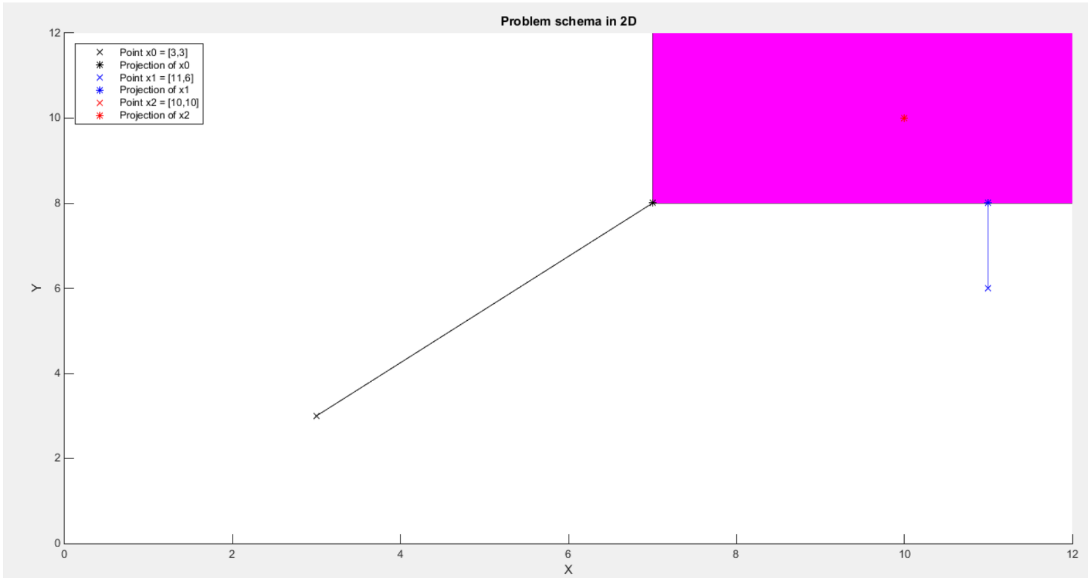
Then (**5**) leads to : $\boxed{\mathbf{x}_* = \mathbf{x_0}}$

The projected point is the initial point (the projection of $\mathbf{x_0}$ is $\mathbf{x_0}$).

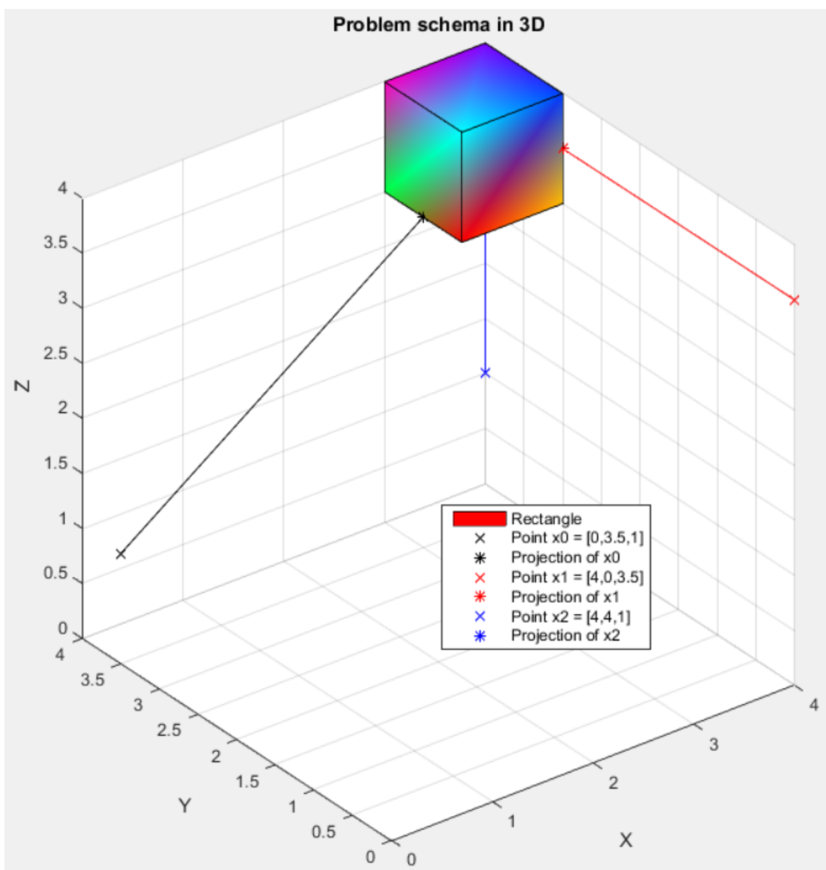
(3) Computing the projection of $\mathbf{x}_0 \in \mathbb{R}^n$ onto the set $\mathbb{S} := \{\mathbf{x} \in \mathbb{R}^n | \mathbf{a} \leq \mathbf{x}\}$.

(a) Drawing a scheme of the problem

For $n = 2$, set $\mathbb{S} := \{\mathbf{x} \in \mathbb{R}^2 | \mathbf{a} \leq \mathbf{x}\}$ is a rectangle. The bottom and the left edge of the rectangle are both restricted by the coordinates of $\mathbf{a}$, while the other two edges stretch to infinity, as depicted bellow (let $\mathbf{a} = [7, 8]$):



For $n = 3$, set $\mathbb{S} := \{\mathbf{x} \in \mathbb{R}^3 | \mathbf{a} \leq \mathbf{x}\}$ is a rectangular box, as depicted bellow :



(b) Optimization problem to be solved

For a convex set $\mathbb{S} \subset \mathbb{R}^n$ and a point $\mathbf{x}_0 \in \mathbb{R}^n$, the projection of $\mathbf{x}_0$ onto the set $\mathbb{S}$ is the solution of the following problem :

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) := \frac{1}{2}\left\|\mathbf{x}_0 - \mathbf{x}\right\|_2^2 \\ \text{subject to} & f_i(\mathbf{x}) = a_i - x_i \leq 0,\ for\, i = 1, 2, \ldots n \end{array}$$

Set $\mathbb{S}$ is convex since for two points $\mathbf{x_1}, \mathbf{x_2} \in \mathbb{S}$ and for $\theta \in (0,1)$ we get :

$$\mathbf{x_1} \in \mathbb{S} \Rightarrow \mathbf{x_1} \geq \mathbf{a} \Rightarrow (1-\theta)\mathbf{x_1} \geq (1-\theta)\mathbf{a},$$
$$\mathbf{x_2} \in \mathbb{S} \Rightarrow \mathbf{x_2} \geq \mathbf{a} \Rightarrow \theta\mathbf{x_2} \geq \theta\mathbf{a}$$

Adding the equations above we get that :

$$(1-\theta)\mathbf{x_1} + \theta\mathbf{x_2} \geq (1-\theta)\mathbf{a} + \theta\mathbf{a} = \mathbf{a},$$

which means that $(1-\theta)\mathbf{x_1} + \theta\mathbf{x_2} \in \mathbb{S}$.

(c) KKT conditions

The KKT conditions we will use are the following :

- $\nabla f_0(\mathbf{x}_*) + \sum_{i=1}^n \lambda_i \nabla f_i(\mathbf{x}_*) = 0 \quad (\mathbf{1})$
- $\lambda_i \geq 0, \quad i = 1, \ldots, n \quad (\mathbf{2})$
- $f_i(\mathbf{x}_*) \leq 0 \Rightarrow a_i - x_{*,i} \leq 0, \quad i = 1, \ldots, n \quad (\mathbf{3})$
- $\lambda_i f_i(\mathbf{x}_*) = 0, \quad i = 1, \ldots, n \quad (\mathbf{4})$

The gradient $\nabla f_0(\mathbf{x}) = \mathbf{x} - \mathbf{x_0}$ was calculated before.

From ($\mathbf{1}$), we get :

$$\nabla f_0(\mathbf{x}_*) + \sum_{i=1}^n \lambda_i \nabla f_i(\mathbf{x}_*) = 0 \Rightarrow \mathbf{x}_* - \mathbf{x_0} - \lambda = 0 \Rightarrow x_{*,i} = x_{0,i} + \lambda_i, \quad (\mathbf{5})$$

(d) Case when $\lambda_i > 0$

From (**4**), for $\lambda_i > 0$, we get that : $\lambda_i f_i(\mathbf{x}_*) = 0 \Rightarrow f_i(\mathbf{x}_*) = 0 \Rightarrow \boxed{x_{*,i} = a_i}$

The projected point is the point $x_{*,i} = a_i$ (solution of the initial problem).

(e) Case when $\lambda_i = 0$

For $\lambda_i = 0$, then (**5**) leads to : $\boxed{x_{*,i} = x_{0,i}}$

The projected point is the initial point (the projection of $\mathbf{x_0}$ is $\mathbf{x_0}$).

(4) Let $\mathbf{0} \neq \mathbf{a} \in \mathbb{R}^n, b \in \mathbb{R}$ and consider the problem

$$\text{(P)} \qquad \min_{\mathbf{x}} f_0(\mathbf{x}) := \tfrac{1}{2}\|\mathbf{x}\|_2^2, \text{ subject to } \mathbf{x} \in \mathbb{H} := \left\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T\mathbf{x} = b\right\}.$$

(a) Writing and solving the KKT for problem (P)

The problem above can be equally written as the following convex optimization problem :

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) := \frac{1}{2}\|\mathbf{x}\|_2^2 \\ \text{subject to} & f_1(\mathbf{x}) = \mathbf{a}^T\mathbf{x} - b = 0 \end{array}$$

The KKT conditions for this problem are the following :

- $\nabla f_0(\mathbf{x}_*) + v_1 \nabla f_1(\mathbf{x}_*) = 0$ (**1**)
- $\mathbf{a^T x}_* = b$ (**2**)

Computing $\nabla f_0(\mathbf{x})$: $\nabla f_0(\mathbf{x}) = \dfrac{d}{d\mathbf{x}}(f_0(\mathbf{x})) = \dfrac{d}{d\mathbf{x}}\left(\tfrac{1}{2}||\mathbf{x}||_2^2\right) = \dfrac{1}{2}\dfrac{d}{d\mathbf{x}}\left(\mathbf{x}^T\mathbf{x}\right) = \mathbf{x}$

Computing $\nabla f_1(\mathbf{x})$: $\nabla f_1(\mathbf{x}) = \dfrac{d}{d\mathbf{x}}(f_1(\mathbf{x})) = \dfrac{d}{d\mathbf{x}}\left(\mathbf{a}^T\mathbf{x} - b\right) = \mathbf{a}$

Replacing in $(\mathbf{1})$, we get : $\nabla f_0(\mathbf{x}_*) + v_1 \nabla f_1(\mathbf{x}_*) = 0 \Rightarrow \mathbf{x}_* = -v_1 \mathbf{a}, \quad (\mathbf{5})$

<u>Now let's calculate v_1 :</u>

$$(\mathbf{5}) \Rightarrow \mathbf{x}_* = -v_1 \mathbf{a} \Rightarrow \mathbf{a}^T \mathbf{x}_* = -v_1 \mathbf{a}^T \mathbf{a} \Rightarrow \mathbf{a}^T \mathbf{x}_* = -v_1 ||\mathbf{a}||_2^2$$

From $(\mathbf{2})$ we have that $\mathbf{a}^T \mathbf{x}_* = b$. So, $-v_1 ||\mathbf{a}||_2^2 = b \Rightarrow \boxed{v_1 = \frac{-b}{||\mathbf{a}||_2^2}}$

And the projected point will be $\mathbf{x}_* = \dfrac{b}{||\mathbf{a}||_2^2} \mathbf{a}$ (solution of the initial problem).

(b) <u>Computing the solution of (P) using the projected gradient descent method</u>

$$\mathbf{x}_{k+1} = \mathbf{P}_{\mathrm{H}}\left(\mathbf{x}_k - \tfrac{1}{L} \nabla f_0(\mathbf{x}_k)\right), \text{ where } L := max(eig(\nabla^2 f_0(\mathbf{x})))$$

Let's first calculate the projection of $\mathbf{x_0}$ onto the set $\mathbb{H}$ (i.e. $\mathbf{P}_{\mathrm{H}}(\mathbf{x_0})$).

The projection of $\mathbf{x_0}$ onto the set $\mathbb{H}$ is the solution of the following problem :

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) := \frac{1}{2} \|\mathbf{x_0} - \mathbf{x}\|_2^2 \\ \text{subject to} & f_1(\mathbf{x}) = \mathbf{a}^T \mathbf{x} - b = 0 \end{array}$$

The KKT conditions for this problem are the following :

- $\nabla f_0(\mathbf{x}_*) + v_1 \nabla f_1(\mathbf{x}_*) = 0 \quad (\mathbf{1})$
- $\mathbf{a^T} \mathbf{x}_* = b \quad (\mathbf{2})$

<u>Computing $\nabla f_0(\mathbf{x})$:</u> $\nabla f_0(\mathbf{x}) = \dfrac{d}{d\mathbf{x}} \left(\frac{1}{2} ||\mathbf{x_0} - \mathbf{x}||_2^2\right) = \mathbf{x} - \mathbf{x_0}$

<u>Computing $\nabla f_1(\mathbf{x})$:</u> $\nabla f_1(\mathbf{x}) = \dfrac{d}{d\mathbf{x}}(f_1(\mathbf{x})) = \dfrac{d}{d\mathbf{x}} \left(\mathbf{a}^T \mathbf{x} - b\right) = \mathbf{a}$

Replacing in $(\mathbf{1})$, we get : $\nabla f_0(\mathbf{x}_*) + v_1 \nabla f_1(\mathbf{x}_*) = 0 \Rightarrow \mathbf{x}_* = \mathbf{x_0} - v_1 \mathbf{a}, \quad (\mathbf{5})$

Now let's calculate v_1 :

$$(\mathbf{5}) \Rightarrow \mathbf{x}_* = \mathbf{x_0} - v_1 \mathbf{a} \Rightarrow \mathbf{a}^T \mathbf{x}_* = \mathbf{a^T x_0} - v_1 \mathbf{a}^T \mathbf{a} \Rightarrow \mathbf{a}^T \mathbf{x}_* = \mathbf{a^T x_0} - v_1 ||\mathbf{a}||_2^2$$

From $(\mathbf{2})$ we have that $\mathbf{a}^T \mathbf{x}_* = b$.

So, $\mathbf{a^T x_0} - v_1 ||\mathbf{a}||_2^2 = b \Rightarrow v_1 ||\mathbf{a}||_2^2 = \mathbf{a^T x_0} - b \Rightarrow \boxed{v_1 = \frac{\mathbf{a^T x_0} - b}{||\mathbf{a}||_2^2}}$

And the projected point will be $\mathbf{x}_* = \mathbf{x_0} - \dfrac{\mathbf{a^T x_0} - b}{||\mathbf{a}||_2^2} \mathbf{a}$.

In MATLAB we create a function calculating the projection of $\mathbf{x_0}$ onto $\mathbb{H}$ as $\mathbf{P}_{\mathbb{H}}(\mathbf{x_0}) = \mathbf{x_0} - \dfrac{\mathbf{a^T x_0} - b}{||\mathbf{a}||_2^2} \mathbf{a}$.

In a while loop, we compute the next value of xk based on this form (increasing the number of iterations k every time). The loop is terminated when the value of xk between two iterations remains approximately constant.

Last, we also solve the problem using the cvx just to make sure our answer is correct.

```
EXERCISE 4

  Iter =    1, f(xkk) = 0.550292, norm(grad) = 0.585522
  Iter =    2, f(xkk) = 0.550138, norm(grad) = 0.009654
  Iter =    3, f(xkk) = 0.550100, norm(grad) = 0.004828
  Iter =    4, f(xkk) = 0.550090, norm(grad) = 0.002414
  Iter =    5, f(xkk) = 0.550088, norm(grad) = 0.001207
  Iter =    6, f(xkk) = 0.550087, norm(grad) = 0.000603
  Optimal value = 0.550087
  Number of iterations =  6

Confirming the accuracy of our solution by solving with cvx
Optimal value = 0.550087
```

(5) Let $\mathbf{A} \in \mathbb{R}^{p\times n}$, with $\text{rank}(\mathbf{A}) = p$ and $\mathbf{b} \in \mathbb{R}^p$.

(a) Projection of $\mathbf{x_0}$ onto $\mathbb{S} := \{\mathbf{x} \in \mathbb{R}^n | \mathbf{A}\mathbf{x} = \mathbf{b}\}$

For a convex set $\mathbb{S} \subset \mathbb{R}^n$ and a point $\mathbf{x}_0 \in \mathbb{R}^n$, the projection of $\mathbf{x}_0$ onto the set $\mathbb{S}$ is the solution of the following problem :

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) := \frac{1}{2}\left\|\mathbf{x_0} - \mathbf{x}\right\|_2^2 \\ \text{subject to} & f_i(\mathbf{x}) = \mathbf{a_i}^T\mathbf{x} - b_i = 0 \end{array}$$

The KKT conditions for this problem are the following :

- $\nabla f_0\left(\mathbf{x}_*\right) + \sum_{i=1}^n v_i \nabla f_i\left(\mathbf{x}_*\right) = 0 \quad (\mathbf{1})$
- $\mathbf{a_i^T}\mathbf{x}_* = b_i$ (where $\mathbf{a_i^T}$ is one of the p rows of $\mathbf{A}$ and has size $(1 \times n)$) $\quad (\mathbf{2})$

Computing $\nabla f_0\left(\mathbf{x}\right)$: $\nabla f_0\left(\mathbf{x}\right) = \dfrac{d}{d\mathbf{x}}\left(f_0\left(\mathbf{x}\right)\right) = \dfrac{d}{d\mathbf{x}}\left(\frac{1}{2}\left|\left|\mathbf{x_0} - \mathbf{x}\right|\right|_2^2\right) = \mathbf{x} - \mathbf{x_0}$

Computing $\nabla f_i(\mathbf{x})$: $\nabla f_i(\mathbf{x}) = \dfrac{d}{d\mathbf{x}}(f_i(\mathbf{x})) = \dfrac{d}{d\mathbf{x}}\left(\mathbf{a_i}^T\mathbf{x} - b_i\right) = \mathbf{a_i}$, for $i = 1, .., p$

Substituting the above in $(\mathbf{1})$, we get :

$$\nabla f_0\left(\mathbf{x}_*\right) + \sum_{i=1}^n v_i \nabla f_i\left(\mathbf{x}_*\right) = 0 \Rightarrow \mathbf{x}_* - \mathbf{x_0} + \sum_{i=1}^n v_i \mathbf{a_i} = 0$$

$$\Rightarrow \mathbf{x}_* - \mathbf{x_0} + \mathbf{A^T}\mathbf{v} = 0, \quad (\mathbf{3})$$

Now let's calculate $\mathbf{v}$:

$$(\mathbf{3}) \Rightarrow \mathbf{x}_* = \mathbf{x_0} - \mathbf{A^T}\mathbf{v} \Rightarrow \mathbf{A}\mathbf{x}_* = \mathbf{A}\mathbf{x_0} - \mathbf{A}\mathbf{A^T}\mathbf{v} \Rightarrow \mathbf{A}\mathbf{A^T}\mathbf{v} = \mathbf{A}\mathbf{x_0} - \mathbf{A}\mathbf{x}_*$$

$$\Rightarrow \mathbf{v} = (\mathbf{A}\mathbf{A^T})^{-\mathbf{1}}(\mathbf{A}\mathbf{x_0} - \mathbf{A}\mathbf{x}_*)$$

Therefore, the projected point (solution of the initial problem) will be :

$$\mathbf{x}_* = \mathbf{x_0} - \mathbf{A^T}\mathbf{v} = \mathbf{x_0} - \mathbf{A^T}(\mathbf{A}\mathbf{A^T})^{-1}(\mathbf{A}\mathbf{x_0} - \mathbf{A}\mathbf{x}_*)$$

$$\Rightarrow \boxed{\mathbf{x}_* = \mathbf{x_0} - \mathbf{A^T}(\mathbf{A}\mathbf{A^T})^{-\mathbf{1}}(\mathbf{A}\mathbf{x_0} - \mathbf{b})}$$

($*$) Matrix $(\mathbf{AA^T})^{\mathbf{-1}}$ exists since the following holds for every real matrix :

$$rank(\mathbf{AA^T}) = rank(\mathbf{A^TA}) = rank(\mathbf{A}) = rank(\mathbf{A}^T) = p$$

Also, $\mathbf{AA^T} \in \mathbb{R}^{p\times p}$, which means that $\mathbf{AA^T}$ is full rank (rank = dimension).

Therefore, $\mathbf{AA^T}$ is invertible.

(b) Consider the problem (Q) $\min_{\mathbf{x}\in\mathbb{S}} f_0(\mathbf{x}) := \frac{1}{2}\mathbf{x^TPx} + \mathbf{q^Tx}$.

(i) First, we solve the problem using cvx. In MATLAB we create a positive definite matrix $\mathbf{P} = \mathbf{P^T}$ and a random vector $\mathbf{q} \in \mathbb{R}^n$, just like we did on the previous project. Then we call the cvx to solve the convex optimization problem. The solution is stored in the variable *x_star_cvx*.

(ii) Then, we solve the problem by writing the KKT conditions and computing the optimal solution by solving them in MATLAB.

The problem is equally written as :

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) := \dfrac{1}{2}\mathbf{x^TPx} + \mathbf{q^Tx} \\ \text{subject to} & f_i(\mathbf{x}) = \mathbf{a_i}^T\mathbf{x} - b_i = 0 \end{array}$$

The KKT conditions for this problem are the following :

- $\nabla f_0\left(\mathbf{x}_*\right) + \sum_{i=1}^n v_i \nabla f_i\left(\mathbf{x}_*\right) = 0 \quad (\mathbf{1})$
- $\mathbf{a_i^T}\mathbf{x}_* = b_i$ (where $\mathbf{a_i^T}$ is one of the p rows of $\mathbf{A}$ with size $(1\times n)$) $\quad (\mathbf{2})$

<u>Computing $\nabla f_0\left(\mathbf{x}\right)$:</u> $\nabla f_0\left(\mathbf{x}\right) = \dfrac{d}{d\mathbf{x}}\left(f_0\left(\mathbf{x}\right)\right) = \dfrac{d}{d\mathbf{x}}\left(\frac{1}{2}\mathbf{x^TPx} + \mathbf{q^Tx}\right) = \mathbf{Px} + \mathbf{q}$

<u>Computing $\nabla f_1(\mathbf{x})$:</u> $\nabla f_1(\mathbf{x}) = \dfrac{d}{d\mathbf{x}}\left(\mathbf{a_i}^T\mathbf{x} - b\right) = \mathbf{a_i}$, for $i = 1, .., p$

Substituting the above in (**1**), we get :

$$\nabla f_0 (\mathbf{x}_*) + \sum_{i=1}^{n} v_i \nabla f_i (\mathbf{x}_*) = 0 \Rightarrow \mathbf{P}\mathbf{x}_* + \mathbf{q} + \sum_{i=1}^{n} v_i \mathbf{a_i} = 0$$

$$\Rightarrow \mathbf{P}\mathbf{x}_* + \mathbf{A^T}\mathbf{v} = -\mathbf{q}, \ (\mathbf{3})$$

Combining (**2**) and (**3**), the KKT conditions can be written in the form :

$$\begin{bmatrix} \mathbf{P} & \mathbf{A^T} \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_* \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} -\mathbf{q} \\ \mathbf{b} \end{bmatrix} \Rightarrow \begin{bmatrix} \mathbf{x}_* \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} \mathbf{P} & \mathbf{A^T} \\ \mathbf{A} & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} -\mathbf{q} \\ \mathbf{b} \end{bmatrix}$$

Solving these in MATLAB we can see that the result is the same with the one produced by the cvx.

(iii) Last, we compute the solution of (Q) using the projected gradient descent method.

$$\mathbf{x}_{k+1} = \mathbf{P}_{\mathrm{S}} \left(\mathbf{x}_k - \tfrac{1}{L} \nabla f_0 (\mathbf{x}_k) \right), \text{ where } L := max(eig(\nabla^2 f_0(\mathbf{x})))$$

The projection of $\mathbf{x_0}$ onto the set $\mathbb{S}$ has been calculated on the first question of the exercise and is equal to $\mathbf{P}_{\mathbb{S}} = \mathbf{x_0} - \mathbf{A^T}(\mathbf{A}\mathbf{A^T})^{\mathbf{-1}}(\mathbf{A}\mathbf{x_0} - \mathbf{b})$.

We create a function representing the form above. In a while loop, we compute the next value of $\mathbf{x_k}$ based on this form (increasing the number of iterations k every time). The loop is terminated when the value of $\mathbf{x_k}$ between two iterations remains approximately constant.

It can be noticed that the projected gradient descent method requires only a few iterations to be completed. This is because of the constraint $\mathbf{Ax} = \mathbf{b}$, since it's easy to minimize smooth functions when the constraint is a linear equality.

Here, we can see that all the methods applied above lead to the same result :

```
You are using p = 3, n = 2 and K = 10.

  (i)    Using the cvx
         Optimal value = 1.185101

  (ii)   Using the KKT conditions
         Optimal value = 1.185101

  (iii)  Using the projected gradient descent method
         Iter =   1, f(xkk) = 1.185101, norm(grad) = 0.947402
         Iter =   2, f(xkk) = 1.185101, norm(grad) = 0.000000
         Optimal value = 1.185101
         Number of iterations =  2
```