Optimization 1 - 098311

Winter 2021 - HW 3

Ido Czerninski 312544596, Asaf Gendler 301727715

November 12, 2020

Problem 1:

(a)

Denote:

$$A = \begin{pmatrix} B & b \\ b^T & c \end{pmatrix}, B \succ 0$$

$$\begin{aligned} p(x) &= \begin{pmatrix} x^T & 1 \end{pmatrix} A \begin{pmatrix} x \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} x^T & 1 \end{pmatrix} \begin{pmatrix} B & b \\ b^T & c \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} x^T & 1 \end{pmatrix} \begin{pmatrix} Bx + b \\ b^T x + c \end{pmatrix} \\ &= x^T B x + x^T b + b^T x + c \\ &= x^T B x + 2b^T x + c \end{aligned}$$

$p(x)$ is a quadratic function of x ,B is symmetric and positive definite, thus from the lemma shown in the lecture:

$$x^* = -B^{-1}b$$

is a strict global minimum point of p over $\mathbb{R}^n$

Since x^* is a strict global minimum:

$$\begin{aligned} \forall x \in \mathbb{R}^n : p(x) &\geq p(x^*) \\ &= x^{*T} B x^* + 2b^T x^* + c \\ &= \left(-B^{-1}b\right)^T B \left(-B^{-1}b\right) + 2b^T \left(-B^{-1}b\right) + c \\ (*) &= b^T B^{-1} B B^{-1} b - 2b^T B^{-1} b + c \\ &= b^T B^{-1} b - 2b^T B^{-1} b + c \\ &= -b^T B^{-1} b + c \end{aligned}$$

Note:

$(*)\, B = B^T \Rightarrow \left(B^{-1}\right)^T = B^{-1}$

if:

$$c > b^T B^{-1} b$$
$$\Rightarrow - b^T B^{-1} b + c > 0$$

hence:

$$p\left(x\right) \geq -b^T B^{-1} b + c > 0$$
$$\Rightarrow \forall x \in \mathbb{R}^n : p\left(x\right) > 0$$

On the other hand:

if:

$$\forall x \in \mathbb{R}^n : p\left(x\right) > 0$$

In particular for $x = x^*$:

$$\begin{aligned} 0 &< p\left(x^*\right) \\ &= -b^T B^{-1} b + c \end{aligned}$$
$$\Rightarrow c > b^T B^{-1} b$$

To conclude:

$$\boxed{\forall x \in \mathbb{R}^n : p\left(x\right) > 0 \iff c > b^T B^{-1}}$$

(b)

$$\begin{aligned}
\begin{pmatrix} I & d \\ 0 & 1 \end{pmatrix}^T A \begin{pmatrix} I & d \\ 0 & 1 \end{pmatrix} &= \begin{pmatrix} I & d \\ 0 & 1 \end{pmatrix}^T \begin{pmatrix} B & b \\ b^T & c \end{pmatrix} \begin{pmatrix} I & d \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} I & 0 \\ d^T & 1 \end{pmatrix} \begin{pmatrix} B & Bd + b \\ b^T & b^T d + c \end{pmatrix} \\
&= \begin{pmatrix} B & Bd + b \\ d^T B + b^T & d^T Bd + d^T b + b^T d + c \end{pmatrix} \\
&= \begin{pmatrix} B & Bd + b \\ d^T B + b^T & d^T Bd + 2b^T d + c \end{pmatrix}
\end{aligned}$$

Let:

$$d = -B^{-1}b$$

We get:

$$\begin{pmatrix} I & d \\ 0 & 1 \end{pmatrix}^T A \begin{pmatrix} I & d \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} B & B\left(-B^{-1}b\right) + b \\ \left(-B^{-1}b\right)^T B + b^T & \left(-B^{-1}b\right)^T B\left(-B^{-1}b\right) + 2b^T\left(-B^{-1}b\right) + c \end{pmatrix}$$

$$= \begin{pmatrix} B & -b + b \\ -b^T + b^T & b^T B^{-1} b - 2b^T B^{-1} b + c \end{pmatrix}$$

$$= \begin{pmatrix} B & 0 \\ 0 & c - b^T B^{-1} b \end{pmatrix}$$

Note:

$$det\left[\begin{pmatrix} I & d \\ 0 & 1 \end{pmatrix}\right] = det\left[\begin{pmatrix} I & d \\ 0 & 1 \end{pmatrix}^T\right]$$

$$(*) = 1 \cdot (-1)^{2(n+1)} \cdot det\,(I) = 1$$

$(*)$ by computing the determinant with the last row

hence:

$$det\left[\begin{pmatrix} I & d \\ 0 & 1 \end{pmatrix}^T A \begin{pmatrix} I & d \\ 0 & 1 \end{pmatrix}\right] = det\left[\begin{pmatrix} I & d \\ 0 & 1 \end{pmatrix}^T\right] det\,[A]\, det\left[\begin{pmatrix} I & d \\ 0 & 1 \end{pmatrix}\right]$$

$$= 1 \cdot det\,[A] \cdot 1$$

$$= det\,[A]$$

thus:

$$det\,[A] = det\left[\begin{pmatrix} I & d \\ 0 & 1 \end{pmatrix}^T A \begin{pmatrix} I & d \\ 0 & 1 \end{pmatrix}\right]$$

$$= det\left[\begin{pmatrix} B & 0 \\ 0 & c - b^T B^{-1} b \end{pmatrix}\right]$$

$$(*) = (-1)^{2(n+1)} \cdot det\,[B] \cdot \left(c - b^T B^{-1} b\right) =$$

$$= det\,[B] \cdot \left(c - b^T B^{-1} b\right)$$

$(*)$ by computing the determinant with the last row

$$\boxed{det\left[A\right] = det\left[B\right] \cdot \left(c - b^T B^{-1} b\right)}$$

$\blacksquare$

(c)

$\Rightarrow$:

We know:

$$A \in \mathbb{R}^{n+1 \times n+1} : A \succ 0$$

we need to show:

$$M_1\left(A\right) > 0, ..., M_{n+1}\left(A\right) > 0$$

Proof by induction on the dimension of A:

Base $n = 1$

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \succ 0$$

A is positive definite thus, its diagonal elements are positive:

$$a > 0, c > 0$$

Moreover, since A is a two by two matrix, it is positive definite if and only if:

$$det\left(A\right) > 0$$

The minors of A are given by:

$$M_1 = a > 0$$

$$M_2 = det\left(A\right) > 0$$

Hence, the statement is true for $n = 1$

Assumption:

Let $B \in \mathbb{R}^{n \times n}$ symmetric matrix. If B is positive definite then all its principal minors are positive

Step

Let $A \in \mathbb{R}^{n+1 \times n+1}$ symmetric positive definite matrix.

Lets write A in the following (general) way:

$$A = \begin{pmatrix} B & b \\ b^T & c \end{pmatrix}, B \in \mathbb{R}^{n \times n}$$

if A is symmetric than B is symmetric, let's prove it's also positive definite

let $v \in R^n$, and define $\tilde{v} \in R^{n+1}$:

$$\tilde{v} = \begin{pmatrix} v \\ 0 \end{pmatrix}$$

notice that because A is positive definite:

$$0 < \tilde{v}^T A \tilde{v} = \tilde{v}^T \begin{pmatrix} B & b \\ b^T & c \end{pmatrix} \tilde{v} = \begin{pmatrix} v^T & 0 \end{pmatrix} \begin{pmatrix} B & b \\ b^T & c \end{pmatrix} \begin{pmatrix} v \\ 0 \end{pmatrix} = v^T B v$$

we proved that:

$$\forall v \in R^n : v^T B v > 0$$

hence B is positive definite by definition.

using the induction assumption, because $B \in \mathbb{R}^{n \times n}$ is symmetric positive definite matrix, then all its principal minors are positive.

One might notice that:

$$M_1(A) = M_1(b) > 0$$

$$.$$

$$.$$

$$.$$

$$M_n(A) = M_n(B) > 0$$

in addition, because A is positive definite:

$$M_{n+1}(A) = det(A) > 0$$

thus all of A principal minors are positive.

$\blacksquare$

$\Leftarrow$:

We know:

$$M_1(A) > 0, ..., M_{n+1}(A) > 0$$

we need to show:

$$A \in \mathbb{R}^{n+1 \times n+1} : A \succ 0$$

Proof by induction on the dimension of A:

Base $n = 1$:

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

Moreover:

$$M_1(A) = a > 0$$
$$M_2(A) = ac - b^2 > 0 \Rightarrow ac > b^2 \geq 0$$

since $a > 0$ and $ac > 0$:

$$c \geq 0$$

hence:

$$Tr(A) = a + c > 0 + c \geq 0$$

A is a two by two matrix that has the following properties:

$$det(A) = M_2(A) > 0, \; Tr(A) > 0$$

thus:

$$A \succ 0$$

Assumption:

Let $B \in \mathbb{R}^{n \times n}$ symmetric matrix. If all its principal minors are positive then B is positive definite

Step

Let $A \in \mathbb{R}^{n+1 \times n+1}$ symmetric matrix, such that all its principal minors are positive

Lets write A in the following (general) way:

$$A = \begin{pmatrix} B & b \\ b^T & c \end{pmatrix}, B \in \mathbb{R}^{n \times n}$$

One might notice that:

$$0 < M_1 (A) = M_1 (b)$$

$$\cdot$$

$$\cdot$$

$$\cdot$$

$$0 < M_n (A) = M_n (B)$$

Thus, all the principal minors of B are positive as well, in addition if A is symmetric than B is symmetric.

Hence, By the induction assumption B is positive definite.

by using the statement that we have proved in section b:

$$det [A] = det [B] \cdot \left(c - b^T B^{-1} b\right)$$

in addition:

$$M_{n+1} (A) = det [A] > 0$$

meaning:

$$det [B] \cdot \left(c - b^T B^{-1} b\right) > 0$$

but also notice:

$$M_n (A) = det [B] > 0$$

thus:

$$c - b^T B^{-1} b > 0 \longrightarrow c > b^T B^{-1} b$$

by using the statment that we have proved in section a:

$$\forall x \in R^n : p (x) > 0$$

$$\forall x \in R^n : \begin{pmatrix} x^T & 1 \end{pmatrix} A \begin{pmatrix} x \\ 1 \end{pmatrix} > 0$$

let $v \in R^{n+1}$ and define $\tilde{v} \in R^n$:

$$\tilde{v} = \frac{v}{v_n}$$

where v_n is the n-th coordinate of v (assuming $v_n \neq 0$).

notice that $\tilde{v}$ is a vector of the form:

$$\tilde{v} = \begin{pmatrix} x \\ 1 \end{pmatrix}$$

thus:

$$v^T Av = (v_n \tilde{v})^T A (v_n \tilde{v}) = \underbrace{v_n^2}_{>0} \underbrace{\tilde{v}^T A \tilde{v}}_{>0} > 0$$

if $v_n = 0$ then:

$$v^T Av = v^T \begin{pmatrix} B & b \\ b^T & c \end{pmatrix} v = v^T Bv > 0$$

where the last step is true because B is positive definite from the induction assumption.

we have proved that:

$$\forall v \in R^{n+1} : v^T Av > 0$$

thus A is positive definite.

■

Problem 2:

(a)

assume $d_0 = 1$, from the given relation we get:

$$y_i = \frac{c_0 + c_1 x_i + c_2 x_i^2}{1 + d_1 x_i + d_2 x_i^2}$$

$$y_i + y_i x_i d_1 + y_i x_i^2 d_2 = c_0 + c_1 x_i + c_2 x_i^2$$

$$c_0 + c_1 x_i + c_2 x_i^2 - x_i y_i d_1 - x_i^2 y_i d_2 = y_i$$

$$\begin{pmatrix} 1 & x_i & x_i^2 & -x_i y_i & -x_i^2 y_i \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ d_1 \\ d_2 \end{pmatrix} = y_i$$

and that's true for all $1 \leq i \leq m$ thus:

$$\underbrace{\begin{pmatrix} 1 & x_1 & x_1^2 & -x_1 y_1 & -x_1^2 y_1 \\ 1 & x_2 & x_2^2 & -x_2 y_2 & -x_2^2 y_2 \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & x_m & x_m^2 & -x_m y_m & -x_m^2 y_m \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ d_1 \\ d_2 \end{pmatrix}}_{\tilde{u}} = \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_m \end{pmatrix}}_{b}$$

and the least square problem is:

$$\min_{\tilde{u} \in R^5} \|A\tilde{u} - b\|^2$$

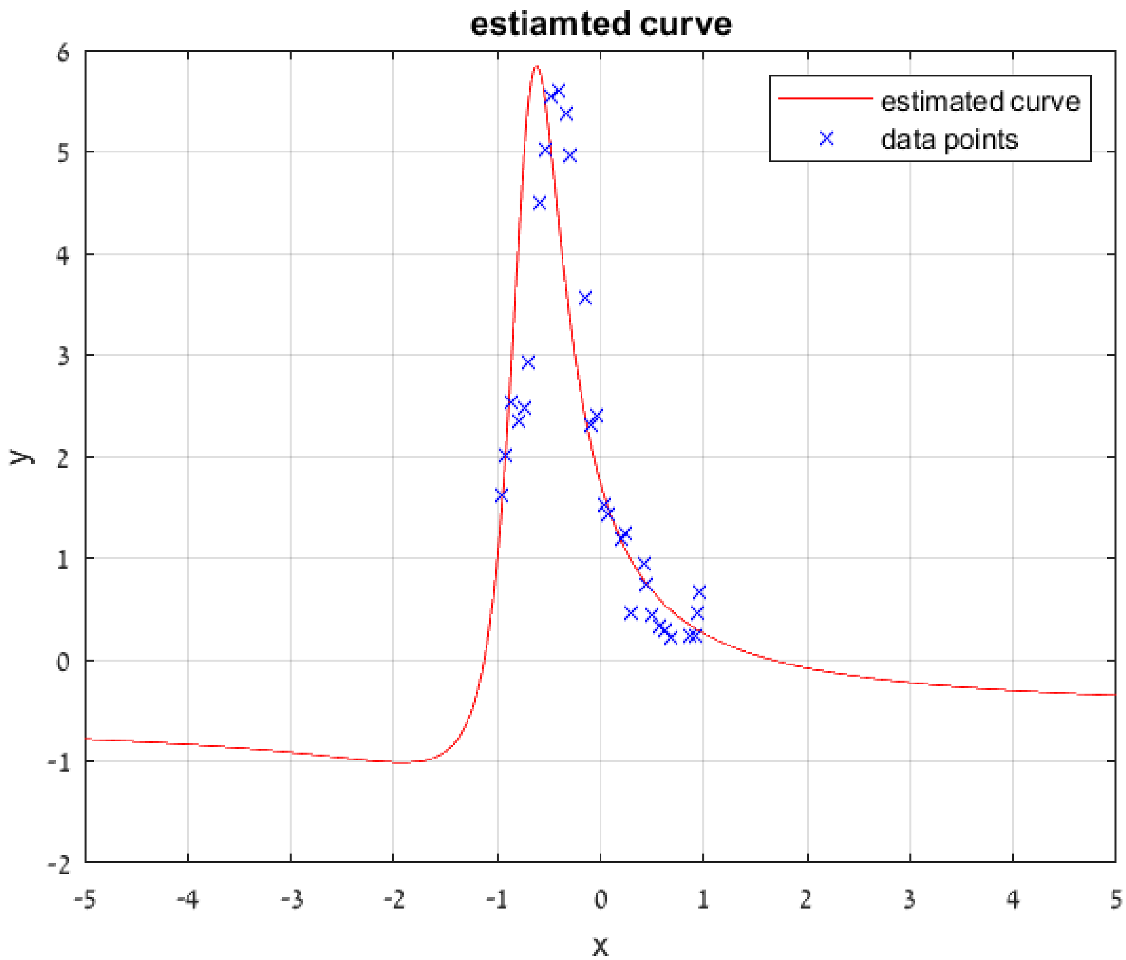
(b)

The coefficients that we get from solving the least squares problem on the given set of points is:

$$\begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} 1.7402 \\ 0.4932 \\ -0.9342 \\ 2.3461 \\ 1.6613 \end{pmatrix}$$

the estimated curve:

Figure 1:



(c)

now we assume $\|u\| = 1$, from the given relation we get:

$$y_i = \frac{c_0 + c_1 x_i + c_2 x_i^2}{d_0 + d_1 x_i + d_2 x_i^2}$$

$$y_i d_0 + y_i x_i d_1 + y_i x_i^2 d_2 = c_0 + c_1 x_i + c_2 x_i^2$$

$$c_0 + c_1 x_i + c_2 x_i^2 - y_i d_0 - x_i y_i d_1 - x_i^2 y_i d_2 = 0$$

$$\begin{pmatrix} 1 & x_i & x_i^2 & -y_i & -x_i y_i & -x_i^2 y_i \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ d_0 \\ d_1 \\ d_2 \end{pmatrix} = 0$$

and that's true for all $1 \leq i \leq m$ thus:

$$\underbrace{\begin{pmatrix} 1 & x_1 & x_1^2 & -y_1 & -x_1 y_1 & -x_1^2 y_1 \\ 1 & x_2 & x_2^2 & -y_2 & -x_2 y_2 & -x_2^2 y_2 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & x_m & x_m^2 & -y_m & -x_m y_m & -x_m^2 y_m \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ d_0 \\ d_1 \\ d_2 \end{pmatrix}}_{u} = 0_m$$

and the optimization problem is:

$$\min_{u \in R^6} \left\{ \|Au\|^2 : \|u\| = 1 \right\} = \min_{u \in R^6} \left\{ u^T A^T A u : \|u\| = 1 \right\}$$

since $A^T A$ is a symmetric matrix, this is a quadratic function of u.

we actually solved this problem in tutorial 1, the answer is the normalized eigen vector that corresponds to the minimal eigen value of $A^T A$.

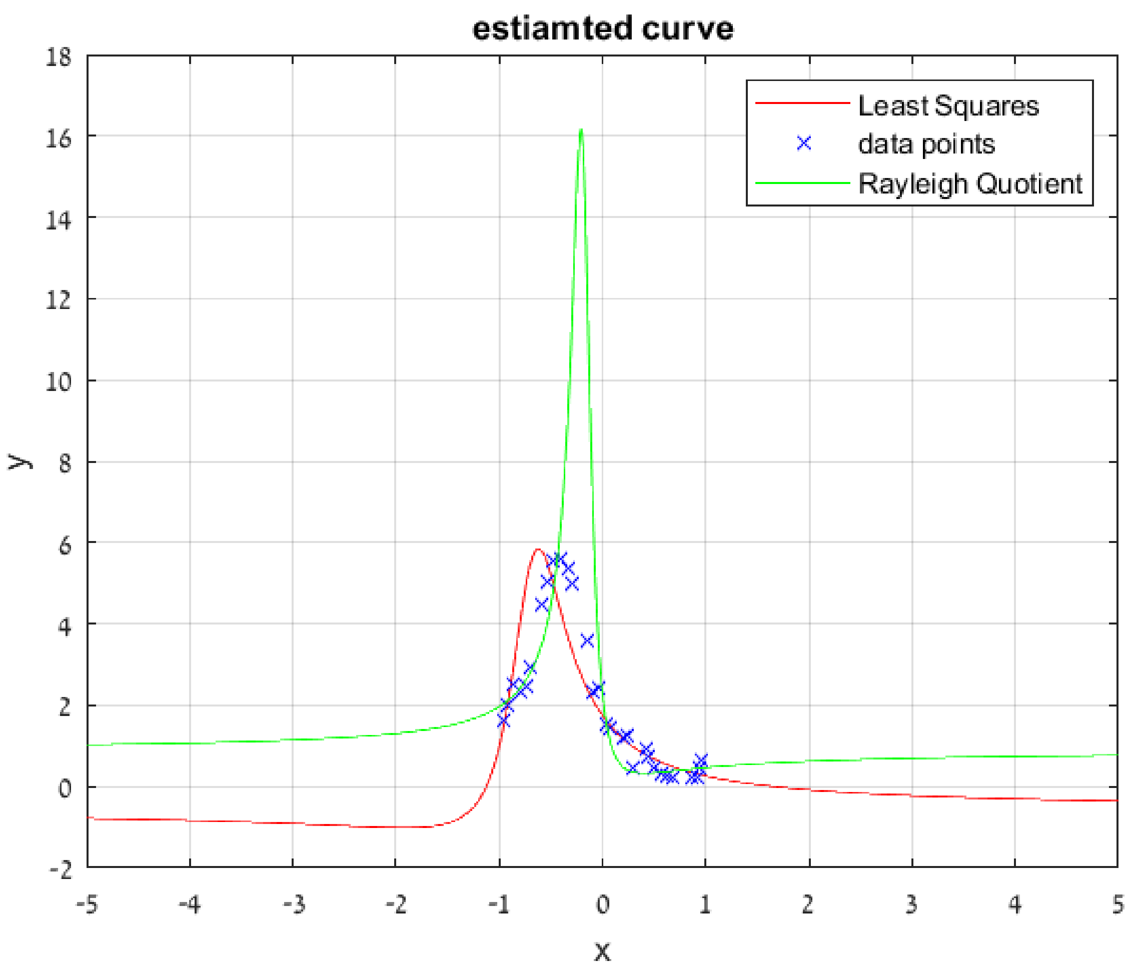
(d)

The coefficients that we get from solving the least squares problem on the given set of points is:

$$\begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ d_0 \\ d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} 0.0761 \\ -0.2328 \\ 0.6177 \\ 0.0334 \\ 0.2611 \\ 0.6994 \end{pmatrix}$$

the estimated curve:

Figure 2:



It seems that the green curve is closer to the dataset than the red curve. But, the peak around -0.5 does not appear to represent the relation correctly. Although it seems the red curve has higher error, it does not have any singular peak. Hence, it is easier to trust the red curve.

Problem 3:

Section a

$$f(x,y) = (x-2y)^4 + 64xy$$

Let's examine the function:

$$\begin{aligned}
&\left((x-2y)^2 - \alpha\right)^2 \geq 0 \\
&(x-2y)^4 - 2\alpha(x-2y)^2 + \alpha^2 \geq 0 \\
&(x-2y)^4 \geq 2\alpha(x-2y)^2 - \alpha^2
\end{aligned}$$

$$\begin{aligned}
f(x,y) &= (x-2y)^4 + 64xy \\
&\geq 2\alpha(x-2y)^2 - \alpha^2 + 64xy \\
&= 2\alpha\left(x^2 - 4xy + 4y^2\right) - \alpha^2 + 64xy
\end{aligned}$$

for $\alpha = 8$

$$\begin{aligned}
f(x,y) &\geq 2\cdot 8\left(x^2 - 4xy + 4y^2\right) - 64 + 64xy \\
&= 16x^2 - 64xy + 64y^2 - 64 + 64xy \\
&= 16x^2 + 64y^2 - 64 \\
&= \begin{pmatrix} x & y \end{pmatrix} \underbrace{\begin{pmatrix} 16 & 0 \\ 0 & 64 \end{pmatrix}}_{A} \begin{pmatrix} x \\ y \end{pmatrix} - 64 = g(x,y)
\end{aligned}$$

$g(x,y)$ is a quadratic form with a positive definite matrix A $(Tr, det > 0)$ hence, it is coercive. Moreover, $g(x,y)$ bounds $f(x,y)$ from below. Thus, $f(x,y)$ is coercive as well.

Section b

$$\begin{aligned}
f(x,y) &= \begin{cases} x^2 - 2xy + y^2 & x \neq y \\ x^2 + y^2 & x = y \end{cases} \\
&= \begin{cases} (x-y)^2 & x \neq y \\ x^2 + y^2 & x = y \end{cases}
\end{aligned}$$

$$f\left(t,t+e^{-t}\right)\underbrace{=}_{(*)}+\left(t-\left(t+e^{-t}\right)\right)^{2}$$
$$=e^{-2t}\xrightarrow[t\to\infty]{}0$$

$\boxed{\text{We found a direction } \left(t,t+e^{-t}\right) \text{, which its norm goes to infinity as } t\to\infty \text{, such that } f\to 0. \text{ Thus, } f \text{ is not coercive.}}$

$(*)$ $t+e^{-t}>t$ for all $t\in\mathbb{R}$ hence, the upper definition holds for all t in this direction.

Section c

$$f\left(x\right)=\frac{x^{T}Ax}{||x||+1},\ A\in\mathbb{R}\text{ is PD}$$

$$\begin{aligned}
f\left(x\right) &= \frac{x^{T}Ax}{||x||+1}\\
&= \frac{||x||^{2}}{||x||^{2}}\cdot\frac{x^{T}Ax}{||x||+1}\\
&= \frac{\left(\frac{x}{||x||}\right)^{T}A\left(\frac{x}{||x||}\right)}{\frac{||x||+1}{||x||^{2}}}\underbrace{\geq}_{(*)}\frac{\lambda_{min}}{\frac{||x||+1}{||x||^{2}}}\\
&= \frac{\lambda_{min}\,||x||^{2}}{||x||+1}\xrightarrow[||x||\to\infty]{(**)\text{since }\lambda_{min}>0}\infty
\end{aligned}$$

$\boxed{\text{We found a coercive function that bounds } f \text{ from below, hence, } f \text{ is coercive as well.}}$

$(*)$ since $\left|\left|\frac{x}{||x||}\right|\right|=1$, we need to show that the solution for:

$$\min_{v\in\mathbb{R}^{n}}v^{T}Av$$
$$s.t\ ||v||=1$$

is λ_{min}

We prove that the solution it attained for the normalized eigen vector that corresponds to the minimal eigen value (denoted as v_{min})

Hence, the minimal value of the function is given by:

$$\begin{aligned}
v_{min}^{T}Av_{min} &= v_{min}^{T}v_{min}\lambda_{min}\\
&= ||v_{min}||^{2}\lambda_{min}\\
&= \lambda_{min}
\end{aligned}$$

$(**)$ A is PD, hence, all its eigen value are positive, in particular, λ_{min}.

Problem 4:

Let:

$$f\left(x\right)=x^{T}Ax+2b^{T}x+c,\ A\succeq 0$$

$\Rightarrow$:

We know

$$b\in\text{Image}\left(A\right)$$

We need to prove that f is bounded from below over $\mathbb{R}^{n}$.

proof

The gradient of f is given by:

$$\nabla f\left(x\right)=2Ax+2b$$

The hessian of f is given by:

$$\nabla^{2}f\left(x\right)=2A\succeq 0$$

Since the hessian is contious in x and semi positive definite $\forall x\in\mathbb{R}^{n}$:

$$\nabla f\left(x^{*}\right)=0\Rightarrow x^{*}\text{ is a global minimun of }f$$

Now, we need to prove that x^{*} exists.

$$\nabla f\left(x^{*}\right)=0$$

$$\iff 2Ax^{*}+2b=0$$

$$\iff Ax^{*}=-b$$

Since $b\in\text{Image}\left(A\right)$:

$$\exists y\in\mathbb{R}^{n}:\sum_{i=1}^{n}y_{i}\cdot a_{i}=b$$

where a_{i} are i'th column for A

In other words:

$$\exists y\in\mathbb{R}^{n}:Ay=b$$

by choosing $x^* = -y$ we get:

$$\begin{aligned}\nabla f(x^*) &= 2Ax^* + 2b \\ &= -2Ay + 2b \\ &= -2b + 2b = 0\end{aligned}$$

$\Rightarrow x^*$ is a global minimum of f

hence:

$$\forall x \in \mathbb{R}^n : f(x) \geq f(x^*)$$

meaning $f(x)$ is bounded from below over $\mathbb{R}^n$

$\Leftarrow$

We know f is bounded from below over $\mathbb{R}^n$.

We need to prove that

$$b \in \operatorname{Image}(A)$$

Proof

Assuming by contradiction that

$$b \notin \operatorname{Image}(A)$$

$\Rightarrow \operatorname{Image}(A) \neq \mathbb{R}^n \Rightarrow A$ does not have a full rank$\Rightarrow \lambda_{min}(A) = 0$

We might have more than one singular eigen vector corresponding to $\lambda_{min}(A)$, but, since $b \notin \operatorname{Image}(A)$, at least one of them , (denoted as v_{min})holds:

$$b^T v_{min} = \langle b, v_{min} \rangle \neq 0$$

Otherwise, b could have been correctly presented as the linear combination of the non-singular eigen vector of A, which means :

$$b \in \operatorname{Image}(A)$$

.

Let's examine $f(\alpha v_{min})$ for $\alpha \in \mathbb{R}$

$$\begin{aligned}f(\alpha v_{min}) &= \alpha^2 v_{min}^T A v_{min} + 2\alpha b^T v_{min} + c \\ &= \alpha^2 v_{min}^T 0 \cdot v_{min} + 2\alpha b^T v_{min} + c \\ &= 2\alpha b^T v_{min} + c\end{aligned}$$

for $b^T v_{min} > 0$:

$$f\left(\alpha v_{min}\right) = 2\alpha b^T v_{min} + c \xrightarrow[\alpha \to -\infty]{} -\infty$$

for $b^T v_{min} < 0$

$$f\left(\alpha v_{min}\right) = 2\alpha b^T v_{min} + c \xrightarrow[\alpha \to \infty]{} -\infty$$

$\boxed{\text{Hence, the function } f \text{ is not bounded. This is a contradiction, thus, } b \in \text{Image}\left(A\right)}$

Problem 5:

$$
\begin{aligned}
f(x) &= ||Ax-b||^2 + \lambda ||Lx||^2 \\
&= \langle Ax-b, Ax-b\rangle + \lambda \langle Lx, Lx\rangle \\
&= (Ax-b)^T (Ax-b) + \lambda (Lx)^T (Lx) \\
&= x^T A^T Ax - x^T A^T b - b^T Ax + ||b||^2 + \lambda x^T L^T Lx \\
&= x^T A^T Ax - 2b^T Ax + ||b||^2 + \lambda x^T L^T Lx \\
&= x^T \underbrace{\left(A^T A + \lambda \cdot L^T L\right)}_{B} x - 2b^T Ax + ||b||^2
\end{aligned}
$$

$$Bx = A^T b$$

$$
\begin{aligned}
f(x) &= ||b||^2 - b^T Ax \\
&= ||b||^2
\end{aligned}
$$

$\Rightarrow$

We know:

$$\operatorname{Ker}(A) \cap \operatorname{Ker}(L) = \{0\}$$

We need to prove

f has a strict global minimum

Proof

Let $x \neq 0_n \in \mathbb{R}^n$:

$$
\begin{aligned}
x^T Bx &= x^T \left(A^T A + \lambda \cdot L^T L\right) x \\
&= x^T A^T Ax + \lambda \cdot x^T L^T Lx \\
&= \underbrace{||Ax||^2}_{\geq 0} + \lambda \underbrace{||Lx||^2}_{\geq 0} \geq 0
\end{aligned}
$$

Let's prove that the equality is never attained

Assuming by contradiction that for some $x \neq 0_n \in \mathbb{R}^n$ the equality holds:

$$\Rightarrow \begin{cases} ||Ax||^2 = 0 \\ ||Lx||^2 = 0 \end{cases} \iff \begin{cases} Ax = 0_n \\ Lx = 0_n \end{cases} \iff 0_n \neq x \in \operatorname{Ker}(A) \cap \operatorname{Ker}(L)$$

This is a contradiction to:

$$\text{Ker}\left(A\right) \cap \text{Ker}\left(L\right) = \{0_n\}$$

hence:

$$x^T B x = \left|\left|Ax\right|\right|^2 + \lambda \left|\left|Lx\right|\right|^2 > 0$$

$\Rightarrow B \succ 0$ by definition

$\boxed{\text{Since } f \text{ quadratic with } B \succ 0 \text{ it has a strict global minimum}}$

$\Leftarrow$

We know that f has a **strict** global minimum attained at x^*

We need to prove

$$\text{Ker}\left(A\right) \cap \text{Ker}\left(L\right) = \{0_n\}$$

Proof

Let's assume by contradiction that

$$\text{Ker}\left(A\right) \cap \text{Ker}\left(L\right) \neq \{0_n\}$$

hence:

$$\exists x_k \neq 0 \in \mathbb{R}^n : x_k \in \text{Ker}\left(A\right) \cap \text{Ker}\left(L\right)$$

denote $y = x_k + x^* \underbrace{\neq}_{x_k \neq 0} x^*$

but:

$$\begin{aligned} f\left(x_k + x^*\right) &= \left|\left|A\left(x_k + x^*\right) - b\right|\right|^2 + \lambda \left|\left|L\left(x_k + x^*\right)\right|\right|^2 \\ &= \left|\left|\underbrace{Ax_k}_{=0_n} + Ax^* - b\right|\right|^2 + \lambda \left|\left|\underbrace{Lx_k}_{=0_n} + Lx^*\right|\right|^2 \\ &= \left|\left|Ax^* - b\right|\right|^2 + \lambda \left|\left|Lx^*\right|\right|^2 = f\left(x^*\right) \end{aligned}$$

and this is a contradiction to the uniqueness of the global minimum, thus:

$$\boxed{\text{Ker}\left(A\right) \cap \text{Ker}\left(L\right) = \{0_n\}}$$