Part III : Nonlinear optimization

$$\begin{cases} \text{minimize} & f(x) \\ \text{s.t.} & g_1(x) \le 0 \\ & \vdots \\ & g_m(x) \le 0 \end{cases}$$

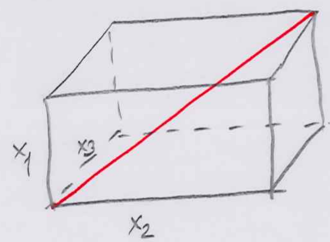
Example: Packing our furniture in our furniture company.

Cost is proportional to the cardboard material wed.

So we want to minimize the surface area.

But there are constraints : Volume constraint

Diagonal constraint.



volume $\geq$ 100 dm^3

spatial diagonal $\geq$ 9dm.

So our problem is :

$$\left.\begin{cases} \text{minimize} & 2(x_1x_2 + x_2x_3 + x_3x_1) \\ \text{s.t} & x_1x_2x_3 \geq 100 \\ & x_1^2 + x_2^2 + x_3^2 \geq 9^2 \\ & x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0. \end{cases}\right\}$$

Nonlinear optimization problem!

Three types of problems:

$$\begin{cases} \text{minimize} & f(x) \\ \text{s.t.} & x \in \mathbb{R}^n \end{cases}$$

no constraints.

Calculus

Numerical method for finding approximate solutions

$$\begin{cases} \text{minimize} & f(x) \\ \text{s.t.} & h_1(x) = 0 \\ & \vdots \\ & h_m(x) = 0 \end{cases}$$

Equality constraints.

Calculus
+
linear algebra

Lagrange multiplier method

$$\begin{cases} \text{minimize} & f(x) \\ \text{s.t.} & g_1(x) \leq 0 \\ & \vdots \\ & g_m(x) \leq 0 \end{cases}$$

Calculus
+
linear algebra

Karush - Kuhn - Tucker conditions

(KKT - conditions)

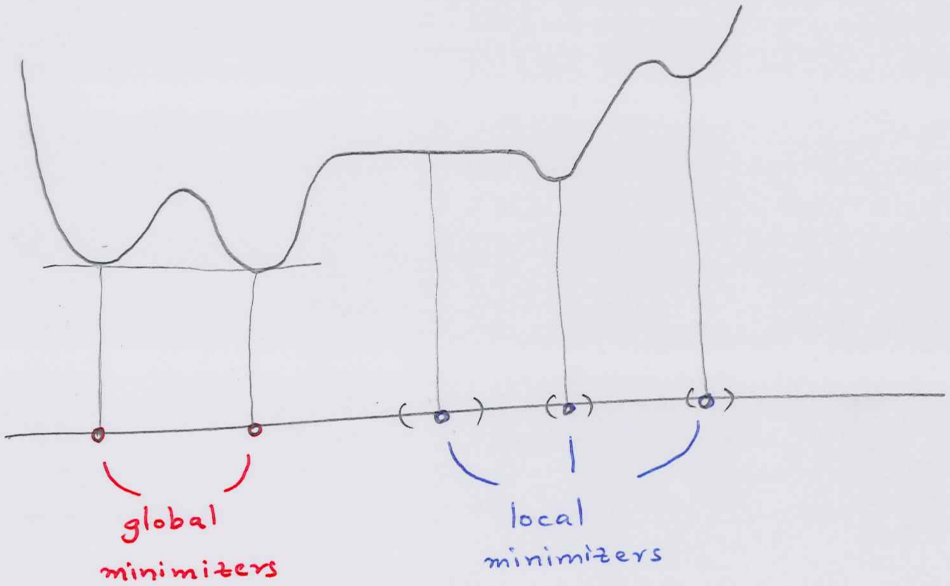
$$(P)\begin{cases}\text{minimize} & f(x)\\ \text{s.t.} & x\in\mathcal{X}\end{cases}$$

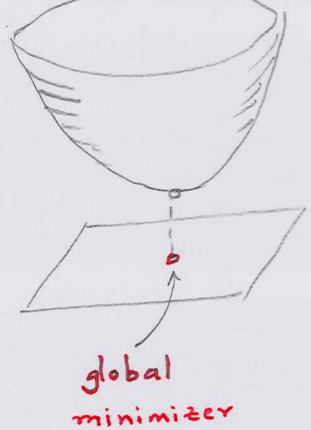
A point $\hat{x}\in\mathcal{X}$ is called a <u>local minimizer for (P)</u> if

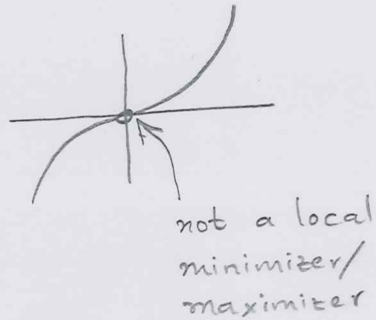
$\exists\ \delta>0$ s.t. $\forall\ x\in\mathcal{X}$ with $\|x-\hat{x}\|<\delta$, $f(\hat{x})\le f(x)$.

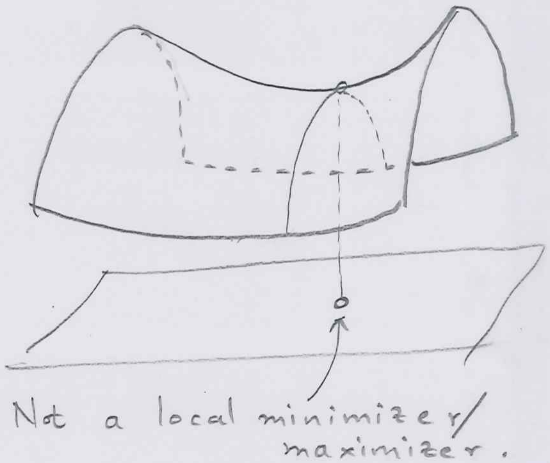
A point $\hat{x}\in\mathcal{X}$ is called a <u>global minimizer for (P)</u> if

$\forall\ x\in\mathcal{X}$, $f(\hat{x})\le f(x)$









How do we solve $\left\{ \begin{array}{ll} \text{minimize} & f(x) \\ \text{s.t.} & x \in \mathbb{R}^n \end{array} \right\}$?

Assume f is twice continuously differentiable

($\frac{\partial f}{\partial x_j}(x)$, $\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$ exist at each $x \in \mathbb{R}^n$ and the maps $x \mapsto \frac{\partial f(x)}{\partial x_j}$, $x \mapsto \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$ are continuous on $\mathbb{R}^n$)

Gradient of f at $x \in \mathbb{R}^n$: $\quad \nabla f(x) := \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) & \cdots & \frac{\partial f}{\partial x_n}(x) \end{bmatrix}$

Hessian of f at $x \in \mathbb{R}^n$:

$$F(x) := \begin{array}{cc} & \quad j \\ i & \begin{bmatrix} & & \\ & \boxed{\frac{\partial^2 f(x)}{\partial x_i \partial x_j}} & \\ & & \end{bmatrix} \end{array} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

<u>Example</u> $\quad f(x_1, x_2) = e^{x_1} x_2$

$$\nabla f(x_1, x_2) = \begin{bmatrix} e^{x_1} x_2 & e^{x_1} \end{bmatrix} \qquad \nabla f(0,0) = \begin{bmatrix} 0 & 1 \end{bmatrix}$$

$$F(x_1, x_2) = \begin{bmatrix} e^{x_1} x_2 & e^{x_1} \\ e^{x_1} & 0 \end{bmatrix} \qquad F(0,0) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$(P): \begin{cases} \text{minimize} & f(x) \\ \text{s.t.} & x \in \mathbb{R}^n \end{cases}$$

<u>Theorem</u> (Necessary condition)

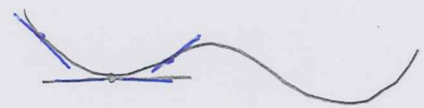
$$x_0 \text{ local minimizer for } (P) \implies \begin{cases} \nabla f(x_0) = 0 \\ F(x_0) \text{ p.s.d.} \end{cases}$$

<u>Theorem</u> (Sufficient condition)

$$\left.\begin{array}{l} \nabla f(x_0) = 0 \\ F(x_0) \text{ p.d.} \end{array}\right\} \implies x_0 \text{ local minimizer for } (P).$$

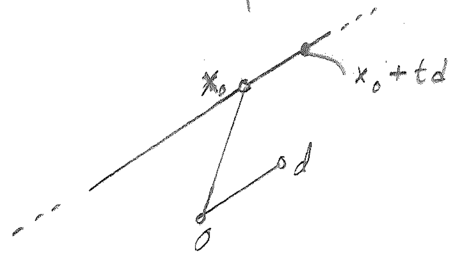
We will use the <u>1 variable result:</u> $f : \mathbb{R} \to \mathbb{R}$

$$x_0 \text{ local minimizer for } f \implies \begin{cases} f'(x_0) = 0 \\ f''(x_0) \geq 0 \end{cases}$$



$f : \mathbb{R}^n \longrightarrow \mathbb{R}$

How does f change in a particular direction?

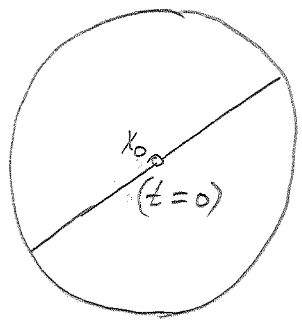


Define $\varphi(t) = f(x_0 + td)$.

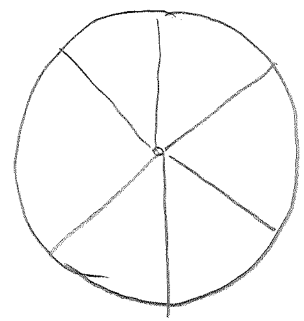
By the chain rule, $\varphi'(t) = \nabla f(x_0 + td) \cdot d$ $\quad (\varphi'(0) = \nabla f(x_0) \cdot d)$,

$\varphi''(t) = d^T F(x_0 + td)\, d$ $\quad (\varphi''(0) = d^T F(x_0)\, d)$.

<u>Observation:</u> x_0 local minimizer of f $\Rightarrow$ $t = 0$ is a local minimizer of φ.



Note that d can be arbitrary:



Theorem (Necessary condition)

x_0 local minimizer of f $\Longrightarrow$ $\begin{cases} \nabla f(x_0) = 0 \\ F(x_0) \text{ p.s.d.} \end{cases}$

Proof Take a $d \in \mathbb{R}^n$.

Know: $t = 0$ is a local minimizer of φ $\quad (\varphi(t) := f(x_0 + td))$.

So by the 1 variable result, $\varphi'(0) = 0$
$\varphi''(0) \geq 0$.

Thus $\nabla f(x_0) \cdot d = 0$

If we had taken $d = (\nabla f(x_0))^T$, then we get $\nabla f(x_0) \cdot (\nabla f(x_0))^T = 0$

Hence $\nabla f(x_0) = 0$.

Also from $\varphi''(0) \geq 0$, we get $d^T F(x_0) d \geq 0$.

But d was arbitrary. Hence $F(x_0)$ is p.s.d. $\square$

Use: Narrows down the set of possible optimal solutions. (If we know $\exists$ a minimizer, this is very useful!)

Helps rule out non-optimal points.

Example

$$\begin{cases} \text{minimize} & f(x_1, x_2) = +x_1 x_2 \\ \text{s.t.} & (x_1, x_2) \in \mathbb{R}^2 \end{cases}$$

$$\nabla f(x) = [x_2 \quad x_1].$$

If there is an optimal solution x, we must have $\nabla f(x) = 0$ and so $x_1 = x_2 = 0$.

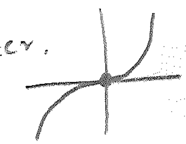
But $F(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and so $F(0) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is not p.s.d.

So $(0,0)$ is not a local minimizer, and hence not an optimal solution.

Remark Is it true that the condition is sufficient too, i.e., does $\left\{ \begin{matrix} \nabla f(x_0) = 0 \\ F(x_0) \text{ p.s.d} \end{matrix} \right\}$ imply x_0 is a local minimizer?

Answer: No. Take $f(x) = x^3$.

$$\left. \begin{aligned} f(0) &= 3x^2 \big|_{x=0} = 0 \\ f''(0) &= 6x \big|_{x=0} = 0 \geq 0 \end{aligned} \right\}$$ But 0 is not a local minimizer.

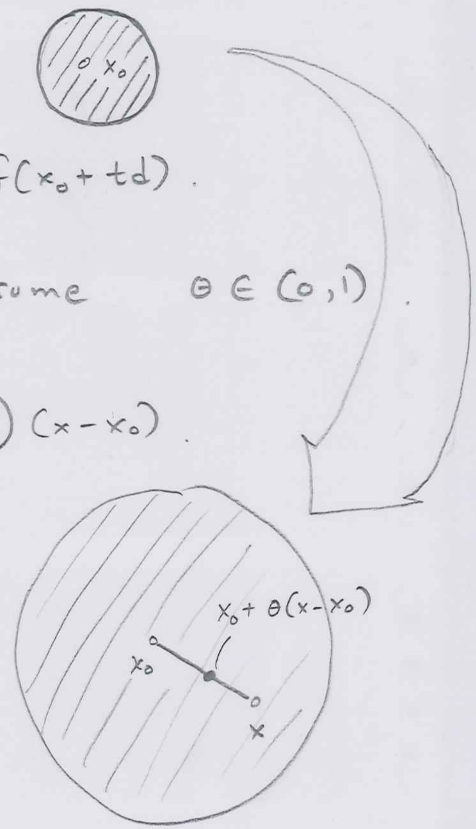


But we do have:

<u>Theorem (sufficient condition)</u>

$$\left.\begin{array}{l}\nabla f(x_0) = 0 \\ F(x_0) \text{ p.d.}\end{array}\right\} \Longrightarrow x_0 \text{ is a local minimizer of } f.$$

<u>Proof</u> $F(x_0)$ p.d. $\Rightarrow$ $F(x)$ p.d. for x near x_0.



Fix an x. Take $d := x - x_0$. Consider $\varphi(t) = f(x_0 + td)$.

$$\varphi(1) = \varphi(0) + \varphi'(0)(1-0) + \frac{1}{2}\varphi''(\theta)(1-0)^2 \quad \text{for some } \theta \in (0,1).$$

$$f(x) = f(x_0) + \underbrace{\nabla f(x_0)}_{=0} \cdot (x - x_0) + \frac{1}{2}(x - x_0)^T F(x_0 + \theta(x - x_0))(x - x_0).$$

$$f(x) - f(x_0) = \frac{1}{2}(x - x_0)^T \underbrace{F(x_0 + \theta(x - x_0))}_{\text{p.d.}}(x - x_0) \geqslant 0.$$

So $f(x) \geqslant f(x_0)$.

Hence x_0 is a local minimizer of f. $\square$

Example

$$\begin{cases} \text{minimize} & \sin x_1 + \cos x_2 \\ \text{s.t.} & (x_1, x_2) \in \mathbb{R}^2 . \end{cases}$$

$$\nabla f(x) = \begin{bmatrix} \cos x_1 & -\sin x_2 \end{bmatrix}.$$

$$\nabla f(x)\Big|_{(x_1,x_2) = (3\pi/2,\, \pi)} = \begin{bmatrix} 0 & -0 \end{bmatrix} = 0 .$$

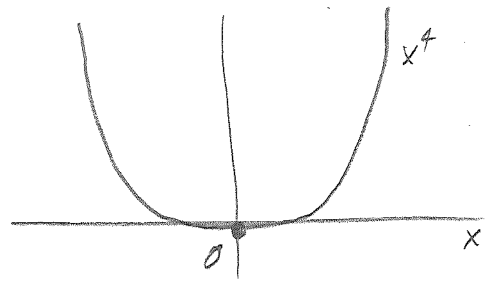
$$F(x) = \begin{bmatrix} -\sin x_1 & 0 \\ 0 & -\cos x_2 \end{bmatrix}$$

$$F(x)\Big|_{(x_1,x_2) = (3\pi/2,\, \pi)} = \begin{bmatrix} -(-1) & 0 \\ 0 & -(-1) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ is positive definite.}$$

So $(3\pi/2, \pi)$ is a local minimizer.

Remark. Is the condition $\begin{Bmatrix} \nabla f(x_0) = 0 \\ F(x_0) \text{ p.d.} \end{Bmatrix}$ necessary too? I.e, does x_0 loc. min $\Downarrow$ $\begin{Bmatrix} \nabla f(x_0) = 0 \\ F(x_0) \text{ p.d.} \end{Bmatrix}$?

Answer: No!, For example, take



$f(x) = x^4, \quad x \in \mathbb{R}$.

$f'(x) = 4x^3$ and $f''(x) = 12x^2$

$f'(0) = 0$ and $f''(0) = 0$.

So $f''(0)$ not p.d., but 0 is a local minimizer.

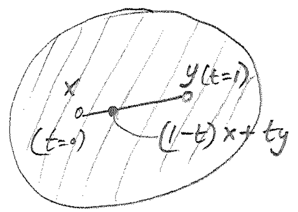
Magic of convex functions.

Theorem Let $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ be "convex".

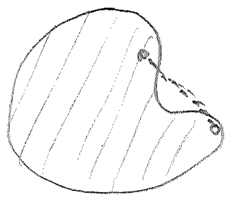
$$\boxed{\hat{x} \in \mathbb{R}^n \text{ is a global minimizer}} \Longleftrightarrow \boxed{\nabla f(\hat{x}) = 0}.$$

What is a convex function?

A set $C \subset \mathbb{R}^n$ is called convex if $\forall x, y \in C$ and $\forall t \in (0,1)$, $(1-t)x + ty \in C$.



convex


not convex.

Examples

① $\mathbb{R}^n$ is convex

② $\{ x \in \mathbb{R}^n \mid \|x\| \leq 1 \}$ is convex

③ $\{ x \in \mathbb{R}^n \mid \|x\| = 1 \}$ is <u>not</u> convex

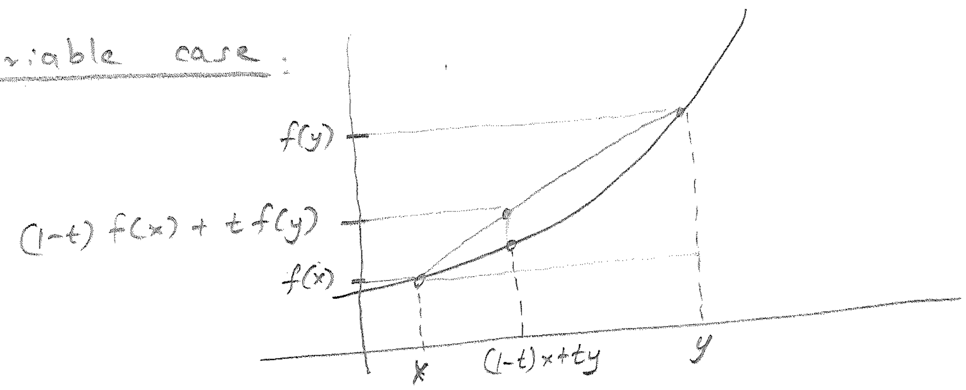
④ $\left\{ x \in \mathbb{R}^n \;\middle|\; \begin{matrix} Ax = b \\ x \geq 0 \end{matrix} \right\}$ is convex.

Let $C \subset \mathbb{R}^n$ be convex.

$f : C \longrightarrow \mathbb{R}$ is called <u>convex</u> if $\forall x, y \in C$ and $\forall t \in (0,1)$,

$$f((1-t)x + ty) \leq (1-t) f(x) + t f(y).$$

<u>Picture in the 1-variable case:</u>



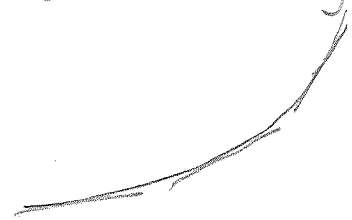
<u>How do we check convexity of a function?</u>

<u>Theorem</u>. Suppose C is a convex subset of $\mathbb{R}^n$ s.t. C has an "interior" point. Then:

$$\boxed{f : C \to \mathbb{R} \text{ is convex}} \iff \boxed{F(x) \text{ is p.s.d. } \forall x \in C}$$

It makes sense:

In the 1-variable case, f is convex $\iff$ f' is increasing

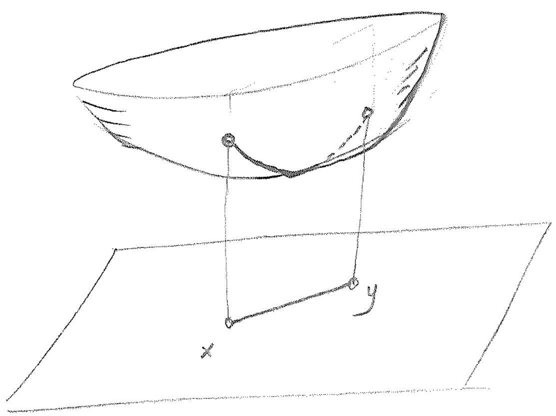


$\iff$ $f''(x) \geq 0 \quad \forall\, x$.

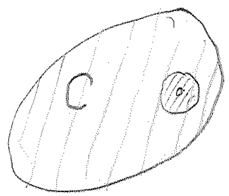
Multivariable case, same sort of a thing, in all directions

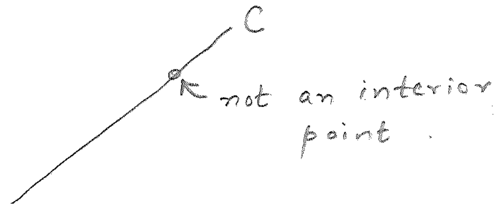
$\Downarrow$

$F(x)$ p.s.d. $\forall\, x$



Interior point of C means a point x_0 for which there is a small ball centered at x_0 comprising points only from C.





<u>Example</u> $C = \{ x \in \mathbb{R}^2 : x_1 = x_2 \}$ is convex.

But it has no interior point

Consider $f : C \longrightarrow \mathbb{R}$ given by $f(x) = x_1 x_2$.

$F(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is not p.s.d. at any $x \in C$!

But f is convex on C!

<u>Reason</u>:

$$f\Big((1-t)\underbrace{\begin{bmatrix}\alpha\\ \alpha\end{bmatrix}}_{x} + t\underbrace{\begin{bmatrix}\beta\\ \beta\end{bmatrix}}_{y}\Big) = f\left(\begin{bmatrix}(1-t)\alpha + t\beta \\ (1-t)\alpha + t\beta\end{bmatrix}\right) = \big((1-t)\alpha + t\beta\big)^2$$

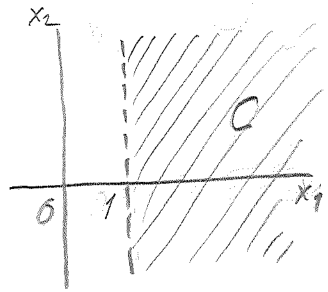
$$\leq (1-t)\alpha^2 + t\beta^2 \quad \text{convexity of } \theta \longmapsto \theta^2$$

$$= (1-t) f\left(\begin{bmatrix}\alpha\\ \alpha\end{bmatrix}\right) + t f\left(\begin{bmatrix}\beta\\ \beta\end{bmatrix}\right).$$

So the previous theorem is applicable only when C has interior points!

<u>Example</u>

$C = \{ x \in \mathbb{R}^2 : x_1 > 1 \}$



C has interior points.

(E.g. $x = (2,0)$ is an interior point).

Let $f : C \longrightarrow \mathbb{R}$ be given by

$$f(x_1, x_2) = x_1^3 + x_1 x_2 + x_2^2 .$$

$$\nabla f(x_1, x_2) = \begin{bmatrix} 3x_1^2 + x_2 & x_1 + 2x_2 \end{bmatrix}$$

$$F(x) = \begin{bmatrix} 6x_1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$E_1 F(x) E_1^T \underset{(x_1>1)}{=} \begin{bmatrix} 1 & 0 \\ 0 & 2 - \frac{1}{6x_1} \end{bmatrix}$$

$2 - \frac{1}{6x_1} \geq 0$?

$\Updownarrow$

$2 \geq \frac{1}{6x_1}$?

$\Updownarrow$

$x_1 \geq \frac{1}{12}$? Yes: $x_1 > 1$.

So f is convex on C.