$$f(x) = \frac{1}{2}\|Ax-b\|_2^2$$

I.

1. • $Ax=0 \Rightarrow A^tAx=0 \Rightarrow \operatorname{Ker} A \subset \operatorname{Ker} A^tA$

• si $A^tAx=0$, alors $0=\langle A^tAx, x\rangle = \|Ax\|_2^2$

Donc $x \in \operatorname{Ker} A$. Donc $\operatorname{Ker} A^tA \subset \operatorname{Ker} A$

D'où $\operatorname{Ker} A = \operatorname{Ker} A^tA$

2. on a clairement $\operatorname{Im}(A^tA) \subset \operatorname{Im} A^t$

$\operatorname{rg} A^t = \operatorname{rg} A = n - \dim \operatorname{Ker}(A)$ ~~$= n - \dim$~~

$= n - \dim \operatorname{Ker}(A^tA)$ d'après 1)

$= \operatorname{rg}(A^tA)$

donc $\operatorname{Im}(A^tA) = \operatorname{Im} A^t$

II. * si A est injective, alors $\operatorname{Ker} A^tA = \{0\}$

Alors A^tA est symétrique (car $(A^tA)^t = A^tA$), positive

(car $\langle A^tAh, h\rangle = \langle Ah, Ah\rangle = \|Ah\|^2 \geq 0$) et inversible

Elle est donc diagonalisable en bon, et sa plus petite

vp λ_1 est >0. ~~Donc~~ Donc

$$f(x) = \frac{1}{2}\langle A^tAx, x\rangle - \langle x, A^tb\rangle + \|b\|^2$$

$$\geq \frac{1}{2}\lambda_1\|x\|^2 - \|x\|\|A^tb\| + \|b\|^2$$

car $\langle A^tAx, x\rangle = \langle A^tA \sum \lambda_i x_i e_i, \sum \lambda_i x_i e_i\rangle = \sum \lambda_i x_i^2 \geq \lambda_1 \|x\|^2$

et en utilisant Cauchy-Schwarz

Le polynome en $\|x\|$ à droite tend vers $+\infty$ lorsque $\|x\| \to +\infty$

donc f est coercive

* si A n'est pas injective. Soit $e \neq 0$, $e \in \operatorname{Ker} A$, et $x_n = ne$

alors $\|x_n\| \to +\infty$ lorsque $n \to \infty$ mais $f(x_n) = \frac{1}{2}\|b\|_2$

ne tend pas vers $+\infty$. Donc f n'est pas coercive

En utilisant le th. du cours, on en déduit que si A est injective alors f a au moins un minimiseur global.

3. f est polynomiale en les coordonnées de x donc $\mathcal{C}^\infty$

$f(x+h) = \frac{1}{2} \langle Ax + Ah - b, Ax + Ah - b \rangle$

$= \frac{1}{2} \|Ax-b\|_2^2 + \frac{1}{2} \langle Ah, Ax-b \rangle + \frac{1}{2} \langle Ax-b, Ah \rangle + \frac{1}{2} \|Ah\|_2^2$

$= f(x) + \langle A^t(Ax-b), h \rangle + \frac{1}{2} \|Ah\|_2^2$

. $\|Ah\|_2^2 \leq \|A\|_2^2 \|h\|^2$ donc $\|Ah\|_2^2 = o(\|h\|)$

. $h \to \langle A^t(Ax-b), h \rangle$ est linéaire

Donc $d_x f(h) = \langle A^tAx - A^tb, h \rangle$ et $\nabla f(x) = A^tAx - A^tb$

. $\nabla f(x+h) = \nabla f(x) + A^tAh$, donc $\mathrm{Hess} f(x) = A^tA$

4. A^tA est symétrique semi-définie positive, ~~donc~~ car
$\langle A^tAh, h \rangle = \|Ah\|^2 \geq 0$
donc f est convexe sur $\mathbb{R}^n$

5. x^* minimiseur local de f ~~sur $\mathbb{R}^n$~~ $\Rightarrow \nabla f(x^*) = 0$
$\Rightarrow A^tAx^* = A^tb$

Comme f est convexe, cette condition est suffisante et le minimiseur est global.

6. si A est injective, alors comme $\operatorname{Ker} A^tA = \operatorname{Ker} A$
A^tA est inversible. Donc $A^tAx^* = A^tb$ a une unique solution, donc f a un unique minimiseur global x^*
donné par $x^* = (A^tA)^{-1}A^tb$

7. Comme ~~Im~~ $A^tb \in \operatorname{Im} A^t = \operatorname{Im} A^tA$, ~~$\exists x$~~ il existe (au moins) un $x^* \in \mathbb{R}^n$ vérifiant $A^tAx^* = A^tb$
Donc il existe un ou des minimiseurs globaux de f, qui sont solution du système $A^tAx^* = A^tb$.

III

8. la derivée de $x \mapsto \cancel{f(x-t\nabla}$ φ_x est

$$\varphi_x'(t) = \langle \nabla f(x - t\nabla f(x)), -\nabla f(x) \rangle$$

donc $\varphi_x'(0) = -\|\nabla f(x)\|$

si $x \neq x^*$, $\nabla f(x^*) \neq 0$ (car f convexe nous permet de dire que x^* minimiseur $\Leftrightarrow \nabla f(x) = 0$)

donc $\varphi_x'(0) < 0$

donc $\exists\, \varepsilon > 0$ tq $\forall\, u \in [0, \varepsilon[$, $\varphi_x'(u) < 0$

$$\varphi_x(t) = \varphi_x(0) + \int_0^t \varphi_x'(u)\,du < \varphi_x(0) \quad \text{pour } t \in]0, \varepsilon[$$

cad $f(x - t\nabla f(x)) < f(x)$ pour $t \in]0, \varepsilon[$

cela signifie que $d = -\nabla f(x)$ est une direction de descente.

9. $$x_{n+1} = x_n - t_n \nabla f(x_n)$$
$$= x_n - t_n (A^t A x_n - A^t b)$$

10. (a) $$x_{n+1} = (Id - \tau A^t A) x_n + \tau A^t b$$

(b) x^* verifie $A^t A x^* = A^t b$ donc
$$x^* = (Id - \tau A^t A) x^* + \tau A^t b$$

$$\Rightarrow x_{n+1} - x^* = (Id - \tau A^t A)(x_n - x^*)$$

$$\Rightarrow x_n - x^* = (Id - \tau A^t A)^n (x_0 - x^*)$$

$A^t A$ def > 0 donc diag en bon. Soit $0 < \lambda_1 < \dots < \lambda_n$ ses vp

$$Id - \tau A^t A = P\left(Id - \tau \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}\right) P^{-1} = P \begin{pmatrix} 1 - \tau\lambda_1 & & \\ & \ddots & \\ & & 1 - \tau\lambda_n \end{pmatrix} P^{-1}$$

$$\Rightarrow (Id - \tau A^t A)^n = P \begin{pmatrix} (1-\tau\lambda_1)^n & & \\ & \ddots & \\ & & (1-\tau\lambda_n)^n \end{pmatrix} P^{-1}$$

si $0 < \tau < \frac{1}{\lambda_n}$ et $0 < \tau < \frac{2}{\lambda_1}$, alors $0 < 1 - \tau\lambda_i < 1$ pour tout i

et $(1 - \tau\lambda_i)^n \to 0$

Dans ce cas ~~oo~~ on a $x_n \xrightarrow[n\to\infty]{} x^*$

11. (a) φ_x est ~~infor~~ coercive, par $x \neq x^*$:

si $|t| \to \infty$, alors $\| x - t \underbrace{\nabla f(x)}_{\neq 0} \| \to +\infty$

et $\varphi_x(t) = f(x - t\nabla f(x)) \longrightarrow +\infty$

donc φ_x admet au moins un minimiseur

$$\varphi_x'(t) = \langle \nabla f(x - t\nabla f(x)), -\nabla f(x) \rangle$$
$$= \langle A^tA(x - t\nabla f(x)) - A^tb, -\nabla f(x) \rangle$$
$$= +t \langle A^tA\nabla f(x), \nabla f(x) \rangle + \langle A^tAx - A^tb, -\nabla f(x) \rangle$$

donc $\varphi_x''(t) = \langle A^tA\nabla f(x), \nabla f(x) \rangle = \|A\nabla f(x)\|_2^2$

> 0 car $\nabla f(x) \neq 0$

(unicité de x^* ~~$\Rightarrow A^tA$ inversible~~ $\Rightarrow \mathrm{Ker}\, A = \mathrm{Ker}\, A^tA = \{0\}$)

donc φ_x est strictement convexe

$\Rightarrow$ unicité du minimiseur

(b). ~~on a aussi~~ ~~$\varphi_x''(t) \geq \lambda_1 \underbrace{\|\nabla f(x)\|}_{>0}$ où λ_1 plus petite~~ vp de ~~A^tA~~

~~donc t_0~~

Methode du gradient à pas optimal

(c) $\mathrm{Hess} f(x) = A^tA$, donc $\langle \mathrm{Hess} f(x) h, h \rangle \geq \underbrace{\lambda_1}_{>0} \|h\|^2$

λ_1 plus petite vp de A^tA

donc f est λ_1-convexe

Le theoreme du cours garantit donc la convergence de la methode du gradient à pas optimal.