4. Numerical Methods.

(P) $\quad \min\limits_{\underset{\sim}{x \in K}} f(x).$ $\qquad K \subseteq \mathbb{R}^n. \longrightarrow K := \left\{ x \in \mathbb{R}^n : \begin{matrix} g_i(x) \le 0 \\ h_j(x) = 0 \end{matrix} \right\}$

K is a closed set.

4.1. Gradient projection.

Definition. Let $K \subseteq \mathbb{R}^n$ be a closed set, we say x^* is a projection of point $y \in \mathbb{R}^n$ to the set K if $x^* \in K$ is a solution to

$$\underline{\min_{x \in K} \|x - y\|^2}$$

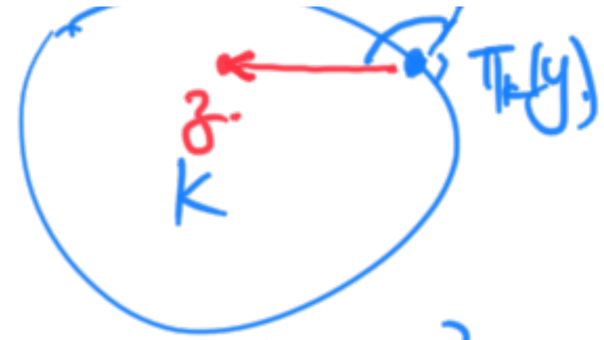
Proposition. If K is a convex closed subset in $\mathbb{R}^n$. Then ~~the~~ there exists a unique projection of y to K, denoted by $\Pi_K(y)$.

Moreover, $\Pi_K(y)$ is the unique point in K satisfying.

$$\underline{\langle y - \Pi_K(y), \; z - \Pi_K(y) \rangle \le 0, \quad \forall z \in K.}$$

And, $\|\Pi_K(y_1) - \Pi_K(y_2)\| \le \|y_1 - y_2\| \quad \forall y_1, y_2 \in \mathbb{R}^n.$

$\cdots \| \le \|y_1 - y_2\|. \quad \forall y_1, y_2 \cdots$



Proof: ① $x \mapsto \|x-y\|^2$ is strictly convex.

If x_1, x_2 be two different solutions to $\min_{x \in K} \|x-y\|^2$.

then $\underbrace{x_3 := \frac{x_1+x_2}{2}}_{\in K}$ will satisfy $\|x_3 - y\|^2 < \frac{\|x_1-y\|^2 + \|x_2-y\|^2}{2}$.

which is a contradiction!

So $\min_{x \in K} \|x-y\|^2$ has a unique solution.

② $\forall z \in K$, then $t \underset{\in K}{z} + (1-t) \underset{\in K}{\Pi_K(y)} \in K$. $t \in [0,1]$.

$$\Rightarrow \cancel{\|y - \Pi_K(y)\|^2} \le \|y - \underbrace{(tz + (1-t)\Pi_K(y))}_{\in K}\|^2.$$

$$= \cancel{\|y - \Pi_K(y)\|^2} - 2t \langle y - \Pi_K(y), z - \Pi_K(y) \rangle + t^2 \|z - \Pi_K(y)\|^2.$$

$$\Rightarrow 2t \langle y - \Pi_K(y), z - \Pi_K(y) \rangle \le t^2 \|z - \Pi_K(y)\|^2. \quad \forall t \in (0,1]$$

$\Rightarrow \langle y-\Pi_K(y), z-\Pi_K(y)\rangle \le \frac{t}{2}\|z-\Pi_K(y)\|^2, \quad \forall t\in(0,1]$

$\Rightarrow \langle y-\Pi_K(y), z-\Pi_K(y)\rangle \le 0.$

(2.2). If $x^*\in K$ satisfies $\langle y-x^*, z-x^*\rangle \le 0 \quad \forall z\in K.$

Then: $\forall z\in K$, $\|y-z\|^2 - \|y-x^*\|^2 = \|y-x^*+x^*-z\|^2 - \|y-x^*\|^2$

$= 2\langle y-x^*, x^*-z\rangle + \|x^*-z\|.$

$= -2\langle y-x^*, z-x^*\rangle + \|x^*-z\|$

$\ge 0. \quad \forall z\in K.$

$\Longrightarrow \boxed{x^* = \Pi_K(y)}$

(3). $\|\Pi_K(y_1)-\Pi_K(y_2)\|^2$

$= \langle \Pi_K(y_1)-\Pi_K(y_2), \Pi_K(y_1)-\Pi_K(y_2)\rangle$

$\le \langle y_1-y_2, \Pi_K(y_1)-\Pi_K(y_2)\rangle \quad \Leftarrow$

$\begin{cases} \langle y_1-\Pi_K(y_1), \Pi_K(y_2)-\Pi_K(y_1)\rangle \le 0 \\ \langle y_2-\Pi_K(y_2), \Pi_K(y_1)-\Pi_K(y_2)\rangle \le 0. \end{cases}$

$z\in K.$

$z\in K$

$\Leftrightarrow$

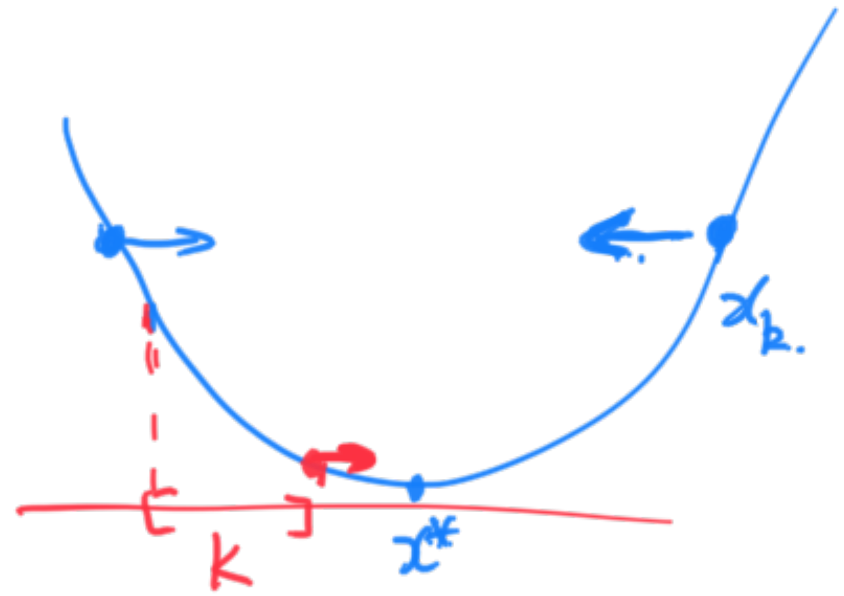
Cauchy-Schwarz.

$\le \|y_1 - y_2\| \, \|\Pi_K(y_1) - \Pi_K \ldots$

$= \langle y_1 - \Pi_K(y_1) \ldots$

$$- \|y_1 - y_2\| \cdot \underline{\|\Pi_K(y_1) - \Pi_K(y_2)\|}. \qquad \langle y_2 - \Pi_K(y_2),\ \Pi_K(y_2) - \Pi_K(y_1)\rangle \le 0$$

$$\Longrightarrow \|\Pi_K(y_1) - \Pi_K(y_2)\| \le \|y_1 - y_2\|. \qquad \#$$



Algorithm: $x_{k+1} = \Pi_K\left(x_k - \rho \nabla f(x_k)\right).$

Theorem, Assume that K is closed convex subset of $\mathbb{R}^n$.

$f: \mathbb{R}^n \to \mathbb{R} \in C^2$. and $\underline{\mathrm{Hess}(f) \geq \alpha I_n}$. for some $\alpha > 0$.

$$\begin{pmatrix} \partial^2_{x_1 x_1} f & & \\ & \ddots & \partial^2_{x_i x_j} f \\ & & \ddots \end{pmatrix}$$

$\Rightarrow$ f is strictly convex. Coercive.

and ∇f is Lipschitz. with Lipschitz constant $M > 0$.

Moreover, $\rho \in \left(0, \frac{2\alpha}{M^2}\right)$.

Then: $\underline{x_k \longrightarrow x^*}$, where x^* is the unique solution to $\min_{x \in K} f(x)$

Lemma: In the context of the theorem, one has

$$\langle \nabla f(y) - \nabla f(x^*), \; y - x^* \rangle \geq \alpha \|y - x^*\|.$$

and: $\Pi_K (x^* - \rho \nabla f(x^*)) = x^*.$

Proof: ① Let. $\phi(t) \triangleq \langle \nabla f (t y + (1-t) x^*), \; y - x^* \rangle = \begin{cases} \langle \nabla f(y), y - x^* \rangle, & t = 1. \\ \langle x^*, y - x^* \rangle, & t = 0. \end{cases}$

$t \in [0, 1]$

$$\Rightarrow \langle \nabla f(y) - \nabla f(x^*), y - x^* \rangle = \phi(1) - \phi(0) = \int_0^1 \phi'(t) \, dt. \geq \alpha \cdot \|y - x^*\|^2.$$

Since

$$\phi'(t) = \langle \, Hess(f) \cdot (t y + (1-t) x^*) (y - x^*)^T, \quad y - x^* \rangle$$

$$\geq \alpha \langle I_n (y - x^*)^T, y - x^* \rangle = \alpha \|y - x^*\|^2.$$

---

② Notice that x^* is the solution of $\min_{x \in K} f(x)$.

Then. for any ~~y~~ $z \in K$, $(1-t) x^* + t \cancel{y}\, z \in K. \quad \forall t \in [0, 1]$

$$\Rightarrow f((1-t) x^* + t \cancel{y}\, z) - f(x^*) \geq 0$$

$$\underbrace{\qquad\qquad}_{t} \xrightarrow{t \downarrow 0}$$

taking the limit $t \downarrow 0$. $\Rightarrow \rho \langle \nabla f(x^*), \cancel{y}^{\,z} - x^* \rangle \geq 0$

$\Rightarrow \forall \cancel{y}^{\,z} \in K$, $\langle \underbrace{(x^* - \rho \nabla f(x^*))}_{y} - x^*, \cancel{y}^{\,z} - x^* \rangle \leq 0$, $\forall z \in K$.

$\Rightarrow x^* = \Pi_K(y) = \Pi_K(x^* - \rho \nabla f(x^*))$. #

Proof of Theorem: $\| x_{k+1} - x^* \|^2 = \| \Pi_K (x_k - \rho \nabla f(x_k)) - x^* \|^2$.

$$= \| \Pi_K (x_k - \rho \nabla f(x_k)) - \Pi_K (x^* - \rho \nabla f(x^*)) \|^2$$

$\| \Pi_K(y_1) - \Pi_K(y_2) \| \leq \| y_1 - y_2 \|$

$$\leq \| (x_k - \rho \nabla f(x_k)) - (x^* - \rho \nabla f(x^*)) \|^2$$

$$= \| (x_k - x^*) - \rho (\nabla f(x_k) - \nabla f(x^*)) \|^2$$

$\| y_1 - y_2 \|^2 = \| y_1 \|^2 - 2 \langle y_1, y_2 \rangle + \| y_2 \|^2$

$$= \| x_k - x^* \|^2 - 2\rho \langle x_k - x^*, \nabla f(x_k) - \nabla f(x^*) \rangle + \rho^2 \| \nabla f(x_k) - \nabla f(x^*) \|^2$$

$$\underbrace{\langle x_k - x^*, \nabla f(x_k) - \nabla f(x^*)\rangle}_{\geq \alpha \|x_k - x^*\|^2} \qquad \underbrace{\|\nabla f(x_k) - \nabla f(x^*)\|^2}_{\leq M^2 \cdot \|x_k - x^*\|^2}$$

$$\leq (1 - 2\rho\alpha + \rho^2 M^2)\, \|x_k - x^*\|^2.$$

$$\boxed{\beta := 1 - 2\rho\alpha + \rho^2 M^2 < 1}$$

since $\rho < \frac{2\alpha}{M^2} \Leftrightarrow \rho^2 M^2 < 2\alpha\rho$.

$$\Rightarrow \|x_{k+1} - x^*\|^2 \leq \beta^{k+1} \|x_0 - x^*\|^2$$

$$\longrightarrow 0 \quad \text{as } k \to +\infty \quad \#$$

Remark: The convergence rate is $C \cdot \beta^k$, where k is the number of iteration.

4.2. Uzawa Algorithm. (Affine equality constraints.)

$$(P) \quad \min_{\{x:\, Ax = b\}} f(x).$$

$K = \{x \in \mathbb{R}^n : Ax = b\}$.

A is matrix $m \times n$. $m \leq n$.

$b \in \mathbb{R}^m$. $\boxed{\mathrm{Rank}(A) = m.}$

$f : \mathbb{R}^n \to \mathbb{R}$ convex. $\in C^1$.

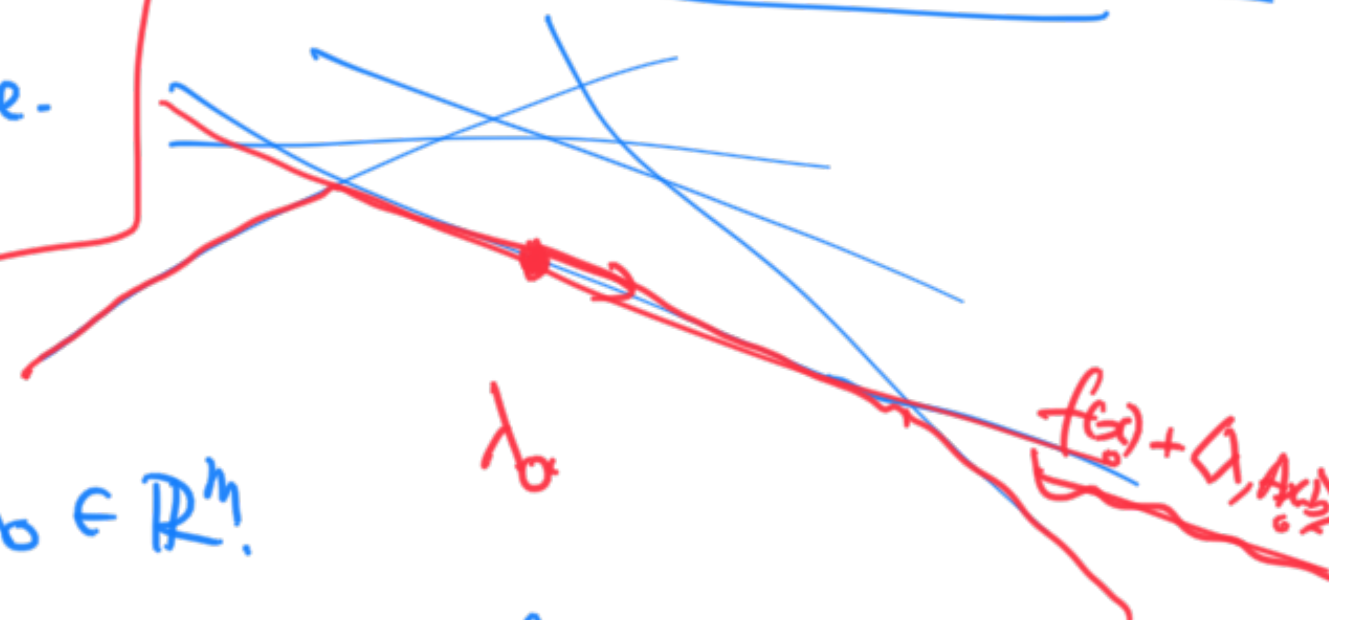
$$\boxed{(D) \quad \max_{\lambda \in \mathbb{R}^m} d(\lambda)}, \quad \text{where } d(\lambda) = \min_{x \in \mathbb{R}^n} \left(f(x) + \langle \lambda, Ax-b \rangle \right).$$

Algorithm of Uzawa:

$$\begin{cases} \lambda_{k+1} = \lambda_k + \rho \nabla d(\lambda_k) \\ \qquad = \lambda_k + \rho (Ax_k - b) \\ x_k \text{ is solution to } d(\lambda_k) = \min_{x \in \mathbb{R}^n} \left(f(x) + \langle \lambda_k, Ax-b \rangle \right) \end{cases}$$

Remark: $\lambda \longmapsto f(x) + \langle \lambda, Ax-b \rangle$ is affine.

so. $\lambda \mapsto d(\lambda)$ is concave.



Lemma: Assume that d is differentiable at $\lambda_0 \in \mathbb{R}^m$.

and $x_0 \in \mathbb{R}^n$ be a solution to $d(\lambda_0) = \min_{x \in \mathbb{R}^n} \left(f(x) + \langle \lambda_0, Ax-b \rangle \right)$

Then: $\nabla d(\lambda_0) = Ax_0 - b.$

Proof: Take $v \in \mathbb{R}^m$, $h > 0$.

$$\underline{d(\lambda_0 + h v)} \le \underline{f(x_0) + \langle \lambda_0 + h v, \ A x_0 - b \rangle}$$

$$= \underline{f(x_0) + \langle \lambda_0, A x_0 - b \rangle} + h \cdot \langle v, A x_0 - b \rangle$$

$$= d(\lambda_0) + h \langle v, A x_0 - b \rangle$$

$$\Longrightarrow \frac{d(\lambda_0 + h v) - d(\lambda_0)}{h} \ge \langle v, A x_0 - b \rangle$$

$$h \downarrow 0 \Longrightarrow \left.\begin{array}{l} \langle \nabla d(\lambda_0), v \rangle \ge \langle v, A x_0 - b \rangle \\ \langle \nabla d(\lambda_0), -v \rangle \ge \langle -v, A x_0 - b \rangle \end{array}\right\} \Longrightarrow \nabla d(\lambda_0) = A x_0 - b. \ \#$$

Theorem: Assume that $\underline{f \text{ is convex}}$ and $\underline{\mathrm{Hess}(f) \ge \alpha I_d}$ for some $\alpha > 0$.

and let $\rho \in \left(0, \frac{2\alpha}{\|A\|^2}\right)$. with $\|A\|^2 := \max_{\|x\|=1} x^T A^T A x$.

Then: ① $x_k \longrightarrow x^*$.

② $\lambda_k \longrightarrow \lambda^*$ (λ^* is solution to $\max_{\lambda \in \mathbb{R}^m} d(\lambda)$)

Proof: ① Let x^* be the solution of (P). and λ^* be such that.

$$\nabla f(x^*) + A^T\lambda^* = 0. \qquad \boxed{Ax^* = b}$$

$$\Leftrightarrow A\nabla f(x^*) + \underset{m\times m}{AA^T}\lambda^* = 0$$

$$\Leftrightarrow \lambda^* = -(AA^T)^{-1}A\nabla f(x^*)$$

Give λ_k, x_k is solution to $\min_{x\in\mathbb{R}^n} f(x) + \langle\lambda_k, Ax-b\rangle$

$$\Rightarrow \boxed{\nabla f(x_k) + A^T\lambda_k = 0} \quad (\text{first order necessary condition})$$

$$\Rightarrow \nabla f(x_k) - \nabla f(x^*) + A^T(\lambda_k - \lambda^*) = 0$$

② $\lambda_{k+1} - \lambda_* = \lambda_k + \rho(Ax_k - b) - \lambda_*$.

$$= \lambda_k - \lambda_* + \rho\cdot A(x_k - x^*) \qquad (b = Ax^*)$$

$$\Rightarrow \|\lambda_{k+1} - \lambda^*\|^2 = \|\lambda_k - \lambda^*\|^2 \ldots$$

$$\|\lambda_{k+1}-\lambda^*\|^2 = \|\lambda_k-\lambda^*\|^2 + 2\rho\langle \lambda_k-\lambda^*, A(x_k-x^*)\rangle + \rho^2\|A(x_k-x^*)\|^2.$$

$$\leq \|\lambda_k-\lambda^*\|^2 + \underline{2\rho\langle A^T(\lambda_k-\lambda^*), x_k-x^*\rangle} + \underline{\rho^2\|A\|^2\|x_k-x^*\|^2}$$

$$= \|\lambda_k-\lambda^*\|^2 - 2\rho \underbrace{\langle \nabla f(x_k)-\nabla f(x^*), x_k-x^*\rangle}_{\geq \alpha \|x_k-x^*\|^2} + \rho^2\|A\|^2\cdot\|x_k-x^*\|^2$$

$$\leq \underline{\|\lambda_k-\lambda^*\|} - \underbrace{(2\rho\alpha-\rho^2\|A\|^2)}_{>0}\cdot\|x^k-x^*\|^2$$

$$\Rightarrow \left(\|\lambda_{k+1}-\lambda^*\|^2\right)_{k\geq 1} \text{ is decreasing.} \Rightarrow \|\lambda_{k+1}-\lambda^*\|^2 \xrightarrow{k\to+\infty} C$$

$$\Rightarrow \underline{\|\lambda_k-\lambda^*\|^2 - \|\lambda_{k+1}-\lambda^*\|^2} \longrightarrow 0 \quad \text{as } k\to+\infty$$

$$\Rightarrow \underline{(2\rho\alpha-\rho^2\|A\|^2)}\|x^k-x^*\|^2 \leq \|\lambda_k-\lambda^*\|^2 - \|\lambda_{k+1}-\lambda^*\|^2 \longrightarrow 0$$

$$\Rightarrow \|x^k-x^*\|^2 \longrightarrow 0. \quad \text{as } k\to+\infty$$

③ Remember: $\lambda_k = -(AA^T)^{-1} A \nabla f(x_k) \longrightarrow -(AA^T)^{-1} A \nabla f(x^*) = \lambda^*$.

$\Rightarrow \nabla d(\lambda^*) = Ax^* - b = 0.$

$\Rightarrow \lambda^*$ is a solution to $\max_{\lambda \in \mathbb{R}^m} d(\lambda)$ #.

4.3. Gradient Algorithm for optimization without constraint.

$$\min_{x \in \mathbb{R}^n} f(x).$$

f is convex.

$\Rightarrow \nabla f(x)$ is a subgradient.

Algo: $- \ x_{k+1} = \underline{\left(x_k - \rho_k \nabla f(x_k)\right)}$

~~① $\mathrm{Hess}(f) \geq \alpha I_n$~~

Assumption: ① $\sum_{k=1}^{+\infty} \rho_k = +\infty$, $\sum_{k=1}^{+\infty} \rho_k^2 < +\infty$, $\rho_k > 0 \cdot \forall k$

② $\underline{\|\nabla f\|_\infty < +\infty}$, $\quad \underline{\langle x - x^*, \nabla f(x) \rangle > 0} \quad \forall x \neq x^*$.

∇f is continuous.

Theorem: $x_k \longrightarrow x^*$ as $k \to +\infty$

Proof: ① $\|x_{k+1} - x^*\|^2 = \| x_k - x^* - \rho_k \nabla f(x_k)\|^2$

$$= \|x_k - x^*\|^2 - \underbrace{2\rho_k \langle x_k - x^*, \nabla f(x_k)\rangle}_{>0} + \rho_k^2 \|\nabla f(x_k)\|^2$$

$$\leq \|x_k - x^*\|^2 + \underline{\rho_k^2 \|\nabla f\|_\infty}$$

$\Rightarrow$ $y_k := \|x_k - x^*\| - \sum_{i=1}^{k} \rho_i^2 \|\nabla f\|_\infty \downarrow$ in k.

and $y_k \geq \underline{0 - \left(\sum_{i=1}^{\infty} \rho_i^2\right) \|\nabla f\|_\infty} > -\infty \quad \forall k$.

$\Rightarrow$ $y_k \longrightarrow y_\infty \in \mathbb{R}$.

Besides: $\underline{y_k} = \|x_k - x^*\| - \underline{\sum_{i=1}^{k} \rho_i^2 \|\nabla f\|_\infty}$

and $\sum_{i=1}^{k} \rho_i^2 \|\nabla f\|_\infty \longrightarrow \sum_{i=1}^{\infty} \rho_i^2 \|\nabla f\|_\infty$

$\Rightarrow \underline{\|x_k - x^*\|^2 \to l \in \mathbb{R}}$

(2) It is enough to prove that $l = 0$.

If $l > 0$ $\emptyset$

Then $\eta := \min\limits_{\frac{l}{2} \le \|x - x^*\|^2 \le 2l} \langle x - x^*, \nabla f(x) \rangle > 0$

Since $\|x_k - x^*\|^2 \longrightarrow l$. then for k large enough, one has

$$\frac{l}{2} \le \|x_k - x^*\|^2 \le 2l.$$

$$\Longrightarrow \sum_{k \ge 1}^{\infty} \rho_k \underbrace{\langle x_k - x^*, \nabla f(x_k) \rangle}_{\ge \eta} \ge \eta \cdot \underbrace{\left(\sum_{k \ge 1}^{\infty} \rho_k \right)}_{} = +\infty$$

However, $\sum_{k=1}^{+\infty} \rho_k \langle x_k - x^*, \nabla f(x_k) \rangle = - \sum^{\infty} \langle x \ldots$

$$\sum_{k\geq 1} \left(\quad \cdots \quad \right) \quad \cdots \quad \underline{\sum_{k\geq 1} \langle x_k - x^*, \; x_{k+1} - x_k \rangle}$$

(since $x_{k+1} - x_k = -\rho \nabla f(x_k)$)

$$= -\frac{1}{2} \sum_{k\geq 1}^{\infty} \left(\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 - \|x_{k+1} - x_k\|^2 \right)$$

$$= \frac{1}{2} \sum_{k=1}^{+\infty} \left(\rho_k^2 \|\nabla f(x_k)\|^2 - l + \|x_0 - x^*\|^2 \right) < +\infty$$

Contradiction!

$$\Longrightarrow l = 0. \quad \Longrightarrow \quad \|x_k - x^*\|^2 \to 0.$$

#