6 – Fixed point algorithms

- $2^{\mathcal{H}}$ is the power of $\mathcal{H}$ i.e the family of all subsets of $\mathcal{H}$
- Let $\Phi : \mathcal{H} \to 2^{\mathcal{H}}$
  - Set of fixed points of Φ : $\mathrm{Fix}\,\Phi = \{x \in \mathcal{H} \mid x \in \Phi x\}$
  - Set of zeros of Φ : $\mathrm{Zer}\,\Phi = \{x \in \mathcal{H} \mid 0 \in \Phi x\}$

- Convergence of a sequence $(x_k)_{k\in\mathbb{N}}$ in $\mathcal{H}$ and $\hat{x} \in \mathcal{H}$
  - Strong convergence of $(x_k)_{k\in\mathbb{N}}$ to $\hat{x}$ if

$x_k \to \hat{x}$ $$\lim_{k\to+\infty} \|x_k - \hat{x}\| = 0$$

  - Weak convergence

$x_k \rightharpoonup \hat{x}$ $$(\forall y \in \mathcal{H}) \quad \lim_{n\to+\infty} \langle y \mid x_k - \hat{x} \rangle = 0$$

Remark : In finite dim. Hilbert space, strong and weak convergences are equivalent.

- Lipschitz continuous operator
  - $\Phi : \mathcal{H} \to \mathcal{H}$ is ω-Lipschitz continuous for some $\omega \in [0, +\infty[$ if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|\Phi x - \Phi y\| \leq \omega \|x - y\|$$

  - Φ nonexpansive if 1-Lipschitz continuous.

- Banach – Picard Theorem

Hyp Let $\omega \in [0,1[$. $\Phi : \mathcal{H} \to \mathcal{H}$ be a ω-Lipschitz continuous operator
Let $x_0 \in \mathcal{H}$.
Set $(\forall k)$ $x_{k+1} = \Phi x_k$.

Then $\mathrm{Fix}\,\Phi = \{\hat{x}\}$ for some $\hat{x} \in \mathcal{H}$ and we have $\|x_k - \hat{x}\| \leq \omega^k \|x_0 - \hat{x}\|$
Moreover, $(x_k)_{k\in\mathbb{N}}$ converges strongly to $\hat{x}$ with linear convergence rate ω.

* **Class $C^{1,1}_{\zeta}(\mathcal{H})$**

- **Def**: For every $\zeta > 0$, the Class $C^{1,1}_{\zeta}(\mathcal{H})$ of functions $f : \mathcal{H} \to \mathbb{R}$ satisfying
  - f is Gâteaux differentiable in $\mathcal{H}$
  - $\nabla f : \mathcal{H} \to \mathcal{H}$ is ζ-Lipschitz continuous

- **Prop** $f \in C^{1,1}_{\zeta}(\mathcal{H}) \iff \begin{cases} f \text{ Fréchet diff} \\ \langle x-y \mid \nabla f(x) - \nabla f(y) \rangle \leq \zeta \|x-y\|^2 \end{cases} \quad \forall x, y$

* **Cocoercive operators**

For every $\eta > 0$, we define the class $\mathcal{C}_\eta$ of η-cocoercive op $\mathcal{M}$ satisfying,

$$\boxed{\forall x, y \in \mathcal{H} \qquad \eta \|\mathcal{M}x - \mathcal{M}y\|^2 \leq \langle x-y, \mathcal{M}x - \mathcal{M}y \rangle}$$

Prop: $f \in C^{1,1}_{\zeta}(\mathcal{H}) \iff \begin{cases} f \text{ Fréchet diff.} \\ \nabla f \in \mathcal{C}_{1/\zeta} \end{cases}$

Prop: If $\mathcal{M} \in \mathcal{C}_\eta$ then $\mathcal{M}$ is η^{-1}-Lipschitz continuous

- Averaged nonexpansive operator

× $\Phi : \mathcal{H} \to \mathcal{H}$ is μ-averaged non-expansive for some $\mu \in]0,1]$ if
$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \|\Phi x - \Phi y\|^2 \leqslant \|x - y\|^2 - \left(\frac{1-\mu}{\mu}\right)\|(Id - \Phi)x - (Id - \Phi)y\|_2^2$

× Φ firmly nonexpansive if it is $\frac{1}{2}$ averaged

× Φ nonexpansive $\Longleftrightarrow$ Φ is 1-averaged

× Theorem

Hyp : $\mu \in]0,1[$
$\Phi : \mathcal{H} \to \mathcal{H}$ be a μ-averaged nonexpansive operator such that $\mathrm{Fix}\,\Phi \neq \emptyset$
Let $x_0 \in \mathcal{H}$.
Set $x_{k+1} = \Phi x_k$.
Then $(x_k)_{k \in \mathbb{N}}$ converges weakly to a point in $\mathrm{Fix}\,\Phi$.

× M is η-cocoercive $\Longleftrightarrow$ $Id - \gamma M$ is $\frac{\gamma}{2\eta}$-averaged.

Examples. $F \in C_{\xi}^{1,1}$

- $\Phi = I - \nabla F$ α-averaged.

, Recall descent lemma

$\forall (x,y) \in \mathcal{H}$ and $t \in \mathbb{R}$

Let $\varphi(t) = f(x + t(y-x))$

$\rightarrow$ φ diff and $\varphi'(t) = \langle y - x \mid \nabla f(x + t(y-x)) \rangle$

$\rightarrow$ $\varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)\, dt$

$\Leftrightarrow f(y) - f(x) = \int_0^1 \langle y - x \mid \nabla f(x + t(y-x)) \rangle dt$

$\Leftrightarrow f(y) - f(x) - \langle y - x, \nabla f(x) \rangle = \int_0^1 \langle y - x \mid \nabla f(x + t(y-x)) - \nabla f(x) \rangle dt$

$\rightarrow$ According to Cauchy-Schwarz inequality

$\langle y - x \mid \nabla f(x + t(y-x)) - \nabla f(x) \rangle$

$\leq \|y - x\| \, \|\nabla f(x + t(y-x)) - \nabla f(x)\|$

∇f ξ Lipschitz continuous

$\leq \xi \|y - x\| \, \|t(y-x)\|$

$= \xi t \|y - x\|^2$

$$\Rightarrow \boxed{f(y) \leq f(x) + \langle y - x, \nabla f(x) \rangle + \frac{\xi}{2} \|y - x\|^2}$$

Descent lemma.

Conjugate $\quad f^*(u) = \sup_x \langle x, u\rangle - f^*(x)$

$$\forall (x,y,z) \in \mathcal{H}^3, \quad f^*(\nabla f(y)) \geq \langle z \mid \nabla f(y)\rangle - f(z)$$

$$\geq \langle z \mid \nabla f(y)\rangle - f(x) - \langle z - x \mid \nabla f(x)\rangle - \frac{\xi}{2}\|z-x\|^2$$

$$= \langle z \mid \nabla f(y) - \nabla f(x)\rangle + \langle x \mid \nabla f(x)\rangle - f(x) - \frac{\xi}{2}\|z-x\|^2$$

$\rightarrow$ Fenchel-Young inequality

$$\langle x \mid \nabla f(x)\rangle - f(x) = f^*(\nabla f(x))$$

$$= \langle z \mid \nabla f(y) - \nabla f(x)\rangle + f^*(\nabla f(x)) - \frac{\xi}{2}\|z-x\|^2$$

$$= f^*(\nabla f(x)) + \langle x \mid \nabla f(y) - \nabla f(x)\rangle + \langle z - x \mid \nabla f(y) - \nabla f(x)\rangle - \frac{\xi}{2}\|z-x\|^2$$

$(\alpha f)^* = \alpha f^*(\cdot/\alpha)$

$+\ f = \frac{1}{2}\|\cdot\|^2 \Rightarrow f^* = f$

for z s.t $\nabla f(z) = \nabla f(y)$ + Fenchel Young inequality.

$$= f^*(\nabla f(x)) + \langle x \mid \nabla f(y) - \nabla f(x)\rangle + \left(\frac{\xi}{2}\|\cdot\|^2\right)^*(\nabla f(y) - \nabla f(x))$$

Then $\quad f^*(\nabla f(y)) \geq f^*(\nabla f(x)) + \langle x \mid \nabla f(y) - \nabla f(x)\rangle + \frac{1}{2\xi}\|\nabla f(y) - \nabla f(x)\|^2$

$$f^*(\nabla f(x)) \geq f^*(\nabla f(y)) + \langle y \mid \nabla f(x) - \nabla f(y)\rangle + \frac{1}{2\xi}\|\nabla f(y) - \nabla f(x)\|^2$$

$+$

$$\langle x - y \mid \nabla f(y) - \nabla f(x)\rangle + \frac{1}{\xi}\|\nabla f(y) - \nabla f(x)\|^2 \leq 0$$

$$\frac{1}{2}\|(I - \nabla f)x - (I - \nabla f)y\|^2 - \frac{1}{2}\|x-y\|^2 - \frac{1}{2}\|\nabla f(y) - \nabla f(x)\|^2 + \frac{1}{\xi}\|\nabla f(y) - \nabla f(x)\|^2 \leq 0$$

$$\|(I - \nabla f)x - (I - \nabla f)y\|^2 + \frac{(1 - \xi/2)}{\xi/2}\|\nabla f(x) - \nabla f(y)\|^2 \leq \|x - y\|^2$$

$\Rightarrow \frac{\xi}{2}$ - averaged.

× $\boxed{\text{prox}_{\partial f} \text{ is } \frac{1}{2} \text{ averaged.}}$

Recall : def Subdifferential

$$\partial F(x) = \{ u \mid \langle y - x \mid u \rangle + f(x) \leq f(y) \}$$

, Let $u_1 \in \partial F(x_1)$ and $u_2 \in \partial f(x_2)$

(1) $\langle x_2 - x_1 \mid u_1 \rangle + f(x_1) \leq f(x_2)$

(2) $\langle x_1 - x_2 \mid u_2 \rangle + f(x_2) \leq f(x_1)$

(1+2) $\Rightarrow$ $\quad 0 \geq \langle x_2 - x_1 , u_1 - u_2 \rangle$

∂f is monotone $\leftarrow$ $\quad 0 \leq \langle x_1 - x_2 , u_1 - u_2)$

$$\|x_1 - x_2\|^2 \leq \langle x_1 - x_2 , u_1 - v_2 + x_1 - x_2 \rangle$$

,

, Let $u_1' \in (Id + \partial f) x_1$

$u_2' \in (Id + \partial f) x_2$

$$\langle x_1 - x_2 , u_1' - u_2' \rangle \geq \|x_1 - x_2\|^2$$

By def of the prox. $\quad x_1 = \text{prox}_f(u_1')$

$x_2 = \text{prox}_f(u_2')$

$$\underbrace{\langle \text{prox}_f(u_1') - \text{prox}_f(u_2')}_{} , u_1' - u_2' \rangle \geq \|\text{prox}_f(v_1') - \text{prox}_f(u_2')\|^2$$

$$= -\frac{1}{2}\|(Id - \text{prox}_f)(u_1') - (Id - \text{prox}_f)(u_2')\|^2 + \frac{1}{2}\|\text{prox}_f(u_1') - \text{prox}_f(u_2')\|^2 + \frac{1}{2}\|u_1' - u_2'\|^2$$

$$\cancel{\frac{1}{2}}\|u_1' - u_2'\|^2 \geq \cancel{\frac{1}{2}}\|\text{prox}_f(u_1') - \text{prox}_f(u_2')\|^2 + \cancel{\frac{1}{2}}\|(Id - \text{prox}_f)u_1' - (Id - \text{prox}_f)u_2'\|^2$$

2/ Gradient method

<u>Hyp</u> $f \in \Gamma_0(\mathcal{H})$

$f \in C^{1,1}_{\xi}$

$$\boxed{\Phi = I - \tau \nabla f}$$

<u>Iterations</u> $(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n - \tau \nabla f(x_n)$

<u>Theorem</u> The sequence $(x_n)_{n\in\mathbb{N}}$ converges weakly to $\hat{x} \in \operatorname{zer} \nabla F$ ($\Leftrightarrow \hat{x} \in \operatorname{Argmin} F$)

- $\operatorname{Fix} \Phi = \{x \mid x \in \Phi x\}$

$$x \in (I - \tau \nabla F)x \Leftrightarrow 0 \in \nabla F(x) \Leftrightarrow x \in \operatorname{zer} \nabla F$$

$$\boxed{\operatorname{zer} \nabla F = \operatorname{Fix} \Phi}$$

- <u>Φ averaged nonexpansive</u>

Φ is $\frac{\tau \xi}{2}$-averaged.

$\Rightarrow (x_n)_{n\in\mathbb{N}}$ weakly converge to a point in $\operatorname{Fix} T$

3/ Forward - backward [Bauschke - Combettes 2017]

Th. 26.14

Hyp $g \in C^{1,1}_{\xi}(\mathcal{H})$ et $\Gamma_0(\mathcal{H})$

$f \in \Gamma_0(\mathcal{H})$

$\phi = \mathrm{prox}_{\tau f} \circ (I - \tau \nabla g)$

Iterations $x_{n+1} = \mathrm{prox}_{\tau f}(x_n - \tau \nabla g(x_n))$

Theorem $\boxed{\text{The sequence } (x_n)_{n \in \mathbb{N}} \text{ generated by FB converges weakly to a point in } \mathrm{zer}(\nabla g + \partial f).}$

$\times \quad x = (I + \tau \partial f)^{-1}(I - \tau \nabla g)$

$\iff (\cancel{I} + \tau \partial f) x \ni \cancel{I} - \tau \nabla g$.

$\iff 0 \in \nabla g + \partial f$

$\nabla f(x) + \partial f(x) = 0$

α $\mathrm{prox}_{\tau f}$ is $\frac{1}{2}$-averaged

$(I - \tau \nabla g)$ is $\frac{\tau \xi}{2}$-averaged.

Prop: Let D be a nonempty subset of $\mathcal{H}$, let $\alpha_1 \in]0,1[$, let $\alpha_2 \in]0,1[$, let $T_1 : D \to D$ α_1 averaged and $T_2 : D \to \mathcal{H}$ α_2 averaged.
Then $T = T_1 T_2$ is α-averaged with

$$\alpha = \frac{\alpha_1 + \alpha_2 - 2\alpha_1\alpha_2}{1 - \alpha_1\alpha_2} \in]0,1[$$

$$\alpha = \frac{\frac{1}{2} + \cancel{\frac{\tau\xi}{2}} - \cancel{\frac{\tau\xi}{2}}}{1 - \frac{\tau\xi}{4}} = \frac{1}{2}\left(\frac{4}{4 - \tau\xi}\right)$$

$$= \frac{2}{4 - \tau\xi}$$

$$= \frac{1}{2 - \frac{\tau\xi}{2}} \in]0,1[$$

$$1 < 2 - \frac{\tau\xi}{2} \iff \frac{\tau\xi}{2} < 1$$

$$\iff \tau < \frac{2}{\xi}$$

4/ Douglas_Rachford

$$\left\| \begin{array}{l} y_n = \text{prox}_{\gamma f}(x_n) \\ x_{n+1} = \text{prox}_{\gamma g}(2y_n - x_n) + x_n - y_n \end{array} \right.$$

$\forall \gamma > 0 \qquad \text{zer}(\nabla f + \partial g) = \text{prox}_{\gamma f}(\text{Fix}\, \phi)$

Weak convergence [Lions_Mercier 1979]
Linear convergence [Giselsson 2017]

$a + b \leq$

$a - b \leq a$

5/ Convergence rate

- A function $h \in C_\xi^{1,1}(\mathcal{H})$ is ρ-strongly convex for some $\rho > 0$ if $h - \frac{\rho}{2}\|\cdot\|_2^2$ is convex

- A function $h \in C_\xi^{1,1}(\mathcal{H})$ is ρ-strongly convex for some $\rho > 0$ if

$$\boxed{\langle \nabla h(x) - \nabla h(y) \,|\, x - y \rangle \geq \rho \|x - y\|^2}$$

- Linear convergence of gradient descent.

/ Since $f \in C^{1,1}_{\xi}(\mathcal{H})$ ρ strongly convex then

$$\phi = f - \frac{\rho}{2}\|\cdot\|_2^2 \text{ is convex and Fréchet diff.}$$

A function $h \in C^{1,1}_{\xi}(\mathcal{H})$ is ρ-strongly convex for some $\rho > 0$ if

$$\boxed{\langle \nabla h(x) - \nabla h(y) \mid x - y \rangle \geq \rho \|x-y\|^2}$$

- Prop $f \in C^{1,1}_{\xi}(\mathcal{H}) \iff \begin{cases} f \text{ Fréchet diff} \\ \langle x-y \mid \nabla h(x) - \nabla h(y) \rangle \leq \xi \|x-y\|^2 \end{cases} \quad \forall x, y$

For every $\eta > 0$, we define the class $\mathcal{C}_\eta$ of η-cocoercive op M satisfying,

$$\boxed{\forall x, y \in \mathcal{H} \quad \eta \|Mx - My\|^2 \leq \langle x-y, Mx - My \rangle}$$

f

/ Since ∇f is ξ-Lipschitz continuous

Since $\xi > \rho$

$$\langle x-y \mid \nabla\phi(x) - \nabla\phi(y) \rangle \leq \xi \|x-y\|^2$$

$$= \langle x-y \mid \nabla f(x) - \nabla f(y) - \rho x + \rho y \rangle$$

$$\leq \xi \|x-y\|^2 - \rho \|x-y\|^2$$

$$= (\xi - \rho)\|x-y\|^2$$

Then $\phi \in C^{11}_{(\xi-\rho)}(\mathcal{H})$

/ $\rho \leq \xi$

/ $\langle x-y \mid \nabla\phi(x) - \nabla\phi(y) \rangle$

$$= \langle x-y \mid \nabla f(x) - \rho x - \nabla f(y) + \rho y \rangle$$

$$= \langle x-y \mid \nabla f(x) - \nabla f(y) \rangle - \rho \|x-y\|^2$$

$$= (\xi - \rho)\|x-y\|^2 \quad \left(+\text{Frechet diff}\right) \implies \phi \in C^{1,1}_{(\xi-\rho)}(\mathcal{H})$$

$$\nabla\phi \in \mathcal{C}_{\frac{1}{\xi-\rho}} \iff \phi \in C^{1,1}_{(\xi-\rho)}(\mathcal{H})$$

/ $G_{\tau\nabla f} = I - \tau\nabla f = I - \tau\nabla\phi - \tau\rho I = (1 - \tau\rho)\,\mathrm{Id} - \tau\nabla\phi$

/ (Let $\tau \in \left]0, \frac{2}{\xi}\right[$) and $\begin{cases} p = G_{\tau\nabla f}\, x \\ q = G_{\tau\nabla f}\, y \end{cases}$

$$\|p-q\|^2 = \|(1-\tau\rho)(x-y) - \tau\nabla\phi(x-y)\|^2$$

$$= (1-\tau\rho)^2\|x-y\|^2 + \tau^2\|\nabla\phi(x) - \nabla\phi(y)\|^2$$

$\phi \in \mathcal{C}_{\frac{1}{\xi-\rho}}$

$$- 2(1-\tau\rho)\tau \langle x-y \mid \nabla\phi(x) - \nabla\phi(y)\rangle$$

$$\leq (1-\tau\rho)^2\|x-y\|^2 + \tau^2(\xi-\rho)\langle x-y \mid \nabla\phi(x) - \nabla\phi(y)\rangle$$

$\tau^2\xi - \tau^2\rho$
$-2\tau + 2\tau^2\rho$
$= \tau^2\xi + \tau^2\rho$
-2τ
$= \tau(\tau(\xi+\rho) - 2)$

$$- 2(1-\tau\rho)\tau\langle x-y \mid \nabla\phi(x) - \nabla\phi(y)\rangle$$

$$\leq (1-\tau\rho)^2\|x-y\|^2 + \tau(\tau(\xi+\rho)-2)\langle x-y \mid \nabla\phi(x) - \nabla\phi(y)\rangle$$

$$\leq (1-\tau\rho)^2\|x-y\|^2 + \max\{0, \tau(\tau(\xi+\rho)-2)\}(\xi-\rho)\|x-y\|^2$$

$$\leq \underbrace{\max\{(1-\tau\rho)^2, (1-\tau\xi)^2\}}_{\omega^2}\|x-y\|^2$$

$\omega = \max\{|1-\tau\rho|, |1-\tau\xi|\}$
$\in \left]0,1\right[$

Theorem Let $\tau > 0$, $\xi > 0$, $\rho > 0$, $f \in C_{\xi}^{1,1}(\mathcal{H})$ and f ρ-strongly convex.

Suppose that $\tau \in]0, 2/\xi[$

Then $\phi = (Id - \tau \nabla f)$ is ω-Lipschitz continuous

where $\omega(\tau) = \max\{|1 - \tau\rho|, |1 - \tau\xi|\}$
$\in]0,1[$.

→ Same rate for Forward - backward.

$$\begin{cases} \tilde{p} = \mathrm{prox}_{\tau f} \circ (I - \tau \nabla g)\, x \\ \tilde{q} = \mathrm{prox}_{\tau f} \circ (I - \tau \nabla g)\, y \end{cases}$$

$$\|\tilde{p} - \tilde{q}\| \leq \|p - q\|$$

nonexpansive prox.

Reminder

- f Gâteaux diff if there exists $\nabla f(x) \in \mathcal{H}$ s.t. $x \in \operatorname{dom} f$

$$(\forall y \in \mathcal{H}) \quad \langle \nabla f(x) \mid y \rangle = \lim_{\substack{\alpha \to 0 \\ \alpha \neq 0}} \frac{f(x+\alpha y) - f(x)}{\alpha}$$

- $\boxed{f \text{ convex iff } \quad f(y) \geqslant f(x) + \langle \nabla f(x), y - x \rangle}$

$(\Rightarrow)$ f convex thus by def.

$$\alpha f(x) + (1-\alpha) f(y) \geqslant f(\alpha x + (1-\alpha) y)$$

$$f(x) \geqslant \frac{f(\alpha x + (1-\alpha) y) - (1-\alpha) f(y)}{\alpha}$$

$$\cancel{\alpha} \left(f(x) - f(y)\right) \geqslant \frac{f\left(y + \alpha(x-y)\right) - f(y)}{\alpha} \quad \lim_{\substack{\alpha \to 0 \\ \alpha \neq 0}}$$

$$f(x) - f(y) \geqslant \langle \nabla f(y), x - y \rangle$$

$$f(x) \geqslant f(y) + \langle \nabla f(y), x - y \rangle$$

$(\Leftarrow)$ $(x, y) \in \operatorname{dom} f \times \operatorname{dom} f$

$\alpha \in [0,1]$ $\quad \alpha x + (1-\alpha) y \in \operatorname{dom} f$

(1) $$f(x) \geqslant f(\alpha x + (1-\alpha) y) + \langle \nabla f(\alpha x + (1-\alpha) y), \underbrace{x - \alpha x - (1-\alpha) y}_{(1-\alpha)(x-y)} \rangle$$

(2) $$f(y) \geqslant f(\alpha x + (1-\alpha) y) + \langle \nabla f(\alpha x + (1-\alpha) y), \underbrace{y - \alpha x - (1-\alpha) y}_{\alpha(y-x)} \rangle$$

$\alpha(1) + (1-\alpha)(2) \Rightarrow$ $$\alpha f(x) + (1-\alpha) f(y) \geqslant f(\alpha x + (1-\alpha) y)$$

$\Rightarrow$ convexity

① ISTA

Goal: $\min \Psi(x) := f(x) + g(y)$

where $f, g \in \Gamma_0(\mathcal{H})$

g β-Lipschitz diff $\beta > 0$.

- Fundamental property for a smooth function in $C^{1,1}$

g continuously diff. with Lipschitz gradient and Lipschitz constant $\overline{\beta}$

$$\forall \beta > \overline{\beta} \quad g(x) \leq g(y) + \langle x - y, \nabla F(y) \rangle + \frac{\beta}{2} \|x - y\|^2$$

- Quadratic approximation of Ψ

$$Q_\beta(x, y) = \underbrace{g(y) + \langle x - y, \nabla g(y) \rangle + \frac{\beta}{2} \|x - y\|^2}_{g(x) \leq} + f(x) \qquad \forall \beta \geq \overline{\beta}$$

$$\Psi(x) \leq Q(x, y)$$

- Forward - backward iterations

$\forall k$ $\quad x_{n+1} = \text{prox}_{\gamma F}(x_n - \gamma \nabla g(x_n)) \qquad \gamma < \frac{2}{\beta}$

$$x_{n+1} = \arg\min_x \frac{1}{2} \|x - x_n + \gamma \nabla g(x_n)\|^2 + \gamma F(x)$$

IF $\gamma = \frac{1}{\beta}$ $\quad x_{n+1} = \arg\min_x \frac{\beta}{2} \|x - x_n + \frac{1}{\beta} \nabla g(x_n)\|^2 + F(x)$

$= p_\beta(x_n)$

- Preliminary result

Hyp $\quad \forall y \in \mathbb{R}^n \quad \Psi(p_\beta(y)) \leq Q(p_\beta(y), y)$

Condition toujours satisfaite pour $\beta > \overline{\beta}$

Prove that $(\forall x \in \mathbb{R}^n)$

$$\Psi(x) - \Psi(p_\beta(y)) \geq \frac{\beta}{2} \|p_\beta(y) - y\|^2 + \beta \langle y - x, p_\beta(y) - y \rangle$$

× Par hyp. $\Psi(p_\beta(y)) \leq Q(p_\beta(y), y)$

donc

$$\Psi(x) - \Psi(p_\beta(y)) \geq \Psi(x) - Q(p_\beta(y), y) \quad (3)$$

× Par prop. de convexité

$$\forall y \quad g(x) \geq g(y) + \langle x-y, \nabla g(y) \rangle \quad (1)$$

$$\forall y \quad f(x) \geq f(y) + \langle x-y, u \rangle \quad \text{avec } u \in \partial f(y)$$

et donc

$$(2) \quad f(x) \geq f(p_\beta(y)) + \langle x - p_\beta(y), \tilde{u} \rangle \text{ avec } \tilde{u} \in \partial f(p_\beta(y))$$

$$(1)+(2) \quad \Psi(x) \geq g(y) + f(p_\beta(y)) + \langle x-y, \nabla g(y) \rangle + \langle x - p_\beta(y), \tilde{u} \rangle \quad (4)$$

$(3)+(4)$ + def $Q(p_\beta(y), y)$

$$\boxed{\Psi(x) - \Psi(p_\beta(y))} \geq \Psi(x) - Q(p_\beta(y), y)$$

$$\geq \cancel{g(y)} + \cancel{f(p_\beta(y))} + \langle x-y, \nabla g(y) \rangle + \langle x - p_\beta(y), \tilde{u} \rangle$$

$$- \cancel{g(y)} - \langle p_\beta(y) - y, \nabla g(y) \rangle$$

$$- \frac{\beta}{2} \| p_\beta(y) - y \|^2 - \cancel{f(p_\beta(y))}$$

$$\geq \langle x - p_\beta(y), \nabla g(y) \rangle + \langle x - p_\beta(y), \tilde{u} \rangle - \frac{\beta}{2} \| p_\beta(y) - y \|^2$$

$$\boxed{= \langle x - p_\beta(y), \nabla g(y) + \tilde{u} \rangle} - \frac{\beta}{2} \| p_\beta(y) - y \|^2 \quad \tilde{u} \in \partial f(p_\beta(y))$$

By definition of $p_\beta(y)$ and characterization of prox:

$$z = p_\beta(y) = \underset{x}{\operatorname{argmin}} \frac{\beta}{2} \| x - y + \frac{1}{\beta} \nabla g(y) \|^2 + F(x)$$

$$\beta \left(z - y + \frac{1}{\beta} \nabla g(y) \right) + \tilde{u} = 0 \quad \tilde{u} \in \partial F(z)$$

$$\beta(p_\beta(y) - y) + \nabla g(y) + \tilde{u} = 0 \quad \tilde{u} \in \partial F(p_\beta(y))$$

$$\Psi(x) - \Psi(p_\beta(y)) \geq \beta \langle x \overset{-y+y}{} - p_\beta(y), y - p_\beta(y) \rangle - \frac{\beta}{2} \| p_\beta(y) - y \|^2$$

$$\geq \beta \langle x - y, y - p_\beta(y) \rangle + \frac{\beta}{2} \| p_\beta(y) - y \|^2$$

Theorem: Let $\{x_n\}$ be a sequence generated by ISTA.
Then, for any $K \geq 1$

$$F(x_K) - F(x^*) \leq \frac{\beta \|x_0 - x^*\|^2}{2K}$$

ISTA has a worst case complexity result of $O(\frac{1}{n})$

Proof

* From previous lemma:

$$\frac{2}{\beta}\left(\Psi(x^*) - \Psi(x_{n+1})\right) \geq \|x_{n+1} - x_n\|^2 + 2\langle x_n - x^*, x_{n+1} - x_n\rangle$$

$$= \cancel{\|x_{n+1} - x_n\|^2} + \|x_{n+1} - x^*\|^2 - \|x_n - x^*\|^2 - \cancel{\|x_{n+1} - x_n\|^2}$$

$$= \|x_{n+1} - x^*\|^2 - \|x_n - x^*\|^2$$

→ Summing this inequality over $n = 0, \dots, K-1$

(1) $$\boxed{\frac{2}{\beta}\left(K\Psi(x^*) - \sum_{n=0}^{K-1}\Psi(x_{n+1})\right) \geq \|x_K - x^*\|^2 - \|x_0 - x^*\|^2}$$

× From previous lemma

$$\frac{2}{\beta}\left(\Psi(x_n) - \Psi(x_{n+1})\right) \geq \|x_{n+1} - x_n\|^2$$

→ multiply by k and summing over $k = 0, \dots, K-1$

$$\frac{2}{\beta}\sum_{k=0}^{K-1}\left\{k\Psi(x_k) - (k+1)\Psi(x_{k+1}) + \Psi(x_{n+1})\right\} \geq \sum_{k=0}^{K-1} k\|x_k - x_{k+1}\|^2$$

$$\cancel{-\Psi(x_1)} + \cancel{\Psi(x_1)} + \Psi(x_1) - \cancel{2\Psi(x_2)} + \Psi(x_2) + \cancel{2\Psi(x_2)} - 3\Psi(x_3) + \Psi(x_3)$$

(2) $$\boxed{\frac{2}{\beta}\sum_{k=0}^{K-1}\Psi(x_{n+1}) - K\Psi(x_K) \geq \sum_{n=0}^{K-1} k\|x_n - x_{n+1}\|^2}$$

(1+2) $$\frac{2}{\beta}K\left(\Psi(x^*) - \Psi(x_K)\right) \geq \|x_K - x^*\|^2 - \|x_0 - x^*\|^2 + \sum_{k=0}^{K-1} k\|x_n - x_{n+1}\|^2$$

$$\Longrightarrow \Psi(x_K) - \Psi(x^*) \leq \frac{\beta}{2K}\|x_0 - x^*\|^2$$

② FISTA

- Iterations

Set $y_1 = x_0 \in \mathbb{R}^N$

Set $t_1 = 1$

$$\left| \begin{aligned} x_k &= p_\beta(y_k) \qquad\qquad \text{avec } \beta \geq \bar{\beta} \\ t_{k+1} &= \frac{1 + \sqrt{1 + 4t_k^2}}{2} \\ y_{k+1} &= x_k + \frac{t_k - 1}{t_{k+1}} (x_k - x_{k-1}) \end{aligned} \right.$$

- Theorem [Beck _ Teboulle 2009]

For any $k \geq 1$

$$\Psi(x_k) - \Psi(x^*) \leq \frac{2\alpha\beta \|x_0 - x^*\|^2}{(k+1)^2}$$

→ FISTA has a worst case complexity of $O\left(\frac{1}{k^2}\right)$

→ Convergence of the sequence $(x_k)_{k \geq 1}$ not secured theoretically.

③ • [Chambolle & Dossal, 2015]

- Iterations

Set $y_1 = x_0 \in \mathbb{R}^N$

Set ~~$t_1 = 1$~~, $a > 2$

$$\left| \begin{aligned} x_k &= p_\beta(y_k) . \qquad\qquad \text{avec } \beta \geq \bar{\beta} \\ d_k &= \frac{n - 1}{n + a} \\ y_{k+1} &= x_k + d_k (x_k - x_{k-1}) \end{aligned} \right.$$

→ Same rate as FISTA

→ convergence of the iterates.