Fixed point iterations of non-expansive operators in algorithms for convex optimization

L. M. Briceño-Arias
Universidad Técnica Federico Santa María

ENS Lyon 2023

Convex optimization problems

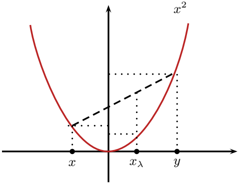
Problem (P)

$$\min_{x\in\mathcal{H}} F(x).$$

- $(\mathcal{H}, \langle\cdot \mid \cdot\rangle)$ is a real Hilbert space with norm $\|\cdot\| = \sqrt{\langle\cdot \mid \cdot\rangle}$.
- $F\colon \mathcal{H} \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ is
  - **proper:** $\operatorname{dom} F = \big\{x \in \mathcal{H} \mid F(x) < +\infty\big\} \neq \varnothing$.
  - **convex :** $(\forall x, y \in \mathcal{H})(\forall \lambda \in [0,1])$ $F(x + \lambda(y-x)) \leq F(x) + \lambda(F(y) - F(x))$.
  - **lower semicontinuous (l.s.c):** $(\forall x_n \to x \in \operatorname{dom} F) \quad F(x) \leq \liminf_{n\to+\infty} F(x_n)$.
- $\Gamma_0(\mathcal{H})$: Class of functions satisfying above conditions.

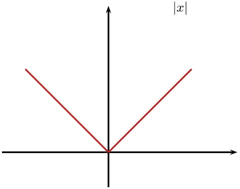
Examples of functions in $\Gamma_0(\mathcal{H})$

- **Differentiable convex functions:** $x \mapsto e^x$, $x \mapsto \|x\|_2^2$,...



$$F(\underbrace{x+\lambda(y-x)}_{x_\lambda}) \leq F(x) + \lambda(F(y)-F(x))$$

- **Non-smooth convex functions:** $x \mapsto |x|$, $x \mapsto \max\{0, x\}$, $x \mapsto \|x\|_1$...

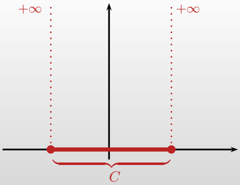


Examples of functions in $\Gamma_0(\mathcal{H})$

- **Discontinuous convex functions:** $\varnothing \neq C \subset \mathcal{H}$ is closed and convex.

Indicator function

$$\iota_C(x) = \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{otherwise.} \end{cases}$$



- **Constrained convex functions:** Let $f \in \Gamma_0(\mathcal{H})$ and C be closed and convex:

$$F(x) = \begin{cases} f(x), & \text{if } x \in C \\ +\infty, & \text{otherwise.} \end{cases} = f(x) + \iota_C(x)$$

Examples of functions in $\Gamma_0(\mathcal{H})$

- $f, g \in \Gamma_0(\mathcal{H})$, $\operatorname{dom} f \cap \operatorname{dom} g \neq \varnothing \Rightarrow f + g \in \Gamma_0(\mathcal{H})$.
- If $g \in \Gamma_0(\mathcal{G})$, and $L\colon \mathcal{H} \to \mathcal{G}$ is linear bounded s.t. $\underbrace{\operatorname{ran} L}_{L(\mathcal{H})} \cap \operatorname{dom} g \neq \varnothing$, then $g \circ L \in \Gamma_0(\mathcal{H})$.

Exercise

$(\forall \lambda \in [0,1]) \quad \|(1-\lambda)x + \lambda y\|^2 = (1-\lambda)\|x\|^2 + \lambda\|y\|^2 - \lambda(1-\lambda)\|x-y\|^2.$

- Hence, $\|\cdot\|^2$ is in $\Gamma_0(\mathcal{H})$.
- In particular, if $\mathcal{H} = \mathbb{R}^N$,

  $$\|x\|_1 = |x_1| + |x_2| + \cdots + |x_N| \quad \text{and} \quad \|x\|_2^2 = x_1^2 + x_2^2 + \cdots + x_N^2$$

  are in $\Gamma_0(\mathcal{H})$.
- $F\colon x \mapsto f(x) + g(Lx)$ is also in $\Gamma_0(\mathcal{H})$.

Concrete example 1: Image processing

Recover image x of $N = n \times p$ pixels (or wavelet coefficients) from an observation

$$z = Ax + \epsilon,$$

- A: $m \times N$ real matrix (e.g., blur)
- ϵ: Gaussian noise.

Figure: Original $\overline{x}$

Figure: $z = A\overline{x} + \epsilon$

Concrete example 1: Image processing

A priori we assume that the image is piecewise constant.

- D: discrete derivative (usually $D = [H, V]$).
- $\lambda > 0$: parameter.

TV minimization

$$\min_{x \in \mathbb{R}^N} F(x) := \underbrace{\lambda \|Dx\|_1}_{\substack{non-smooth \\ composite(D)}} + \underbrace{\|Ax - z\|_2^2}_{smooth}$$

Figure: Degraded z

Figure: Restored x

Concrete example 2: MFG

Mean Field Games (MFG)[1]: Static example

- N players choose their positions on a set Q (compact).
- $\mathcal{P}(Q)$ is the set of Borel probability measures.
- They minimize their distance to a place $P \in Q$.
- Players are congestion-averse.
- The cost of player i can be modeled by

$$\begin{aligned} f_i(x_1, \dots, x_i, \dots, x_N) &= \alpha|x_i - P| - \frac{\beta}{N-1} \sum_{j \neq i} |x_j - x_i| \\ &= \alpha|x_i - P| - \beta \int_Q |x - x_i| \mathrm{d}\left(\frac{1}{N-1} \sum_{j \neq i} \delta_{x_j} \right) \\ &= f\left(x_i, \frac{1}{N-1} \sum_{j \neq i} \delta_{x_j} \right). \end{aligned}$$

[1] J.-M. Lasry, P.-L. Lions. Mean field games. *Jpn. J. Math.* 2007
M. Huang, R. P. Malhamé, P. E. Caines. Large population stochastic dynamic games. *Commun. Inf. Syst.* 2006.

- Suppose that for each N, $(\bar{x}_1^N, \dots, \bar{x}_N^N)$ is a Nash equilibrium of the previous game, i.e., for every $i \in \{1, \dots, N\}$,

$$(\forall x_i \in Q) \quad f\left(\bar{x}_i^N, \frac{1}{N-1}\sum_{j\neq i}\delta_{\bar{x}_j^N}\right) \leq f\left(x_i, \frac{1}{N-1}\sum_{j\neq i}\delta_{\bar{x}_j^N}\right)$$

- Then, $\exists\, \bar{m} \in \mathcal{P}(Q)$ such that, up to some sub-sequence,

$$\frac{1}{N}\sum_{i=1}^{N}\delta_{\bar{x}_i^N} \stackrel{*}{\rightharpoonup} \bar{m}.$$

- The equilibrium $\bar{m}$ satisfies the fixed point equation

$$\operatorname{supp}(\bar{m}) \subseteq \operatorname{argmin}\left\{f(x, \bar{m}) \mid x \in Q\right\}.$$

Dynamic & deterministic case

- Differential game with N players, where Player i minimizes

$$\int_0^T \left[\frac{|\alpha(t)|^2}{2} + f\Big(x_i(t), \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j(t)}\Big)\right] dt + g\left(x_i(T), \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j(T)}\right)$$

$$\text{s.t.} \quad \dot{x}_i(t) = \alpha(t) \qquad \forall t\in[0,T],$$
$$x_i(0) = \bar{x}_{0,i}^N.$$

- Suppose that $(\bar{x}_1^N,\ldots,\bar{x}_N^N)$ is a Nash equilibrium and that

$$\frac{1}{N}\sum_{i=1}^N \delta_{\bar{x}_{0,i}^N} \overset{*}{\rightharpoonup} m_0.$$

Then for each $t\in[0,T]$, $\exists\, m(t)\in\mathcal{P}(Q)$ such that

$$\frac{1}{N}\sum_{i=1}^N \delta_{\bar{x}_i^N(t)} \overset{*}{\rightharpoonup} m(t)$$

Dynamic & deterministic case

- Any equilibrium m solves the MFG system

$$
\begin{aligned}
(HJB) \quad -\partial_t u + \frac{|\nabla u|^2}{2} &= f(x, m(t)) \\
u(x,T) &= g(x, m(T)), \\
(FP) \quad \partial_t m - \operatorname{div}(m \nabla u) &= 0, \\
m(0) &= m_0.
\end{aligned}
$$

- At (x,t) the solution u of the HJB equation is given by

$$
\begin{aligned}
u(x,t) = \inf_{\alpha} \int_t^T \left[\frac{|\alpha(s)|^2}{2} + f(x(s), m(s)) \right] ds + g(x(T), m(T)) \\
\text{s.t.} \quad \dot{x}(s) = \alpha(s) \qquad \forall s \in (t,T), \\
x(t) = x.
\end{aligned}
$$

Dynamic & stochastic case

$$\int_0^T \left[\frac{|\alpha(t)|^2}{2} + f\left(x_i(t), \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j(t)}\right)\right] dt + g\left(x_i(T), \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j(T)}\right)$$

$$\text{s.t.} \quad dx_i(t) = \alpha(t)dt + \sigma dW_i(t) \qquad \forall t \in [0,T],$$

$$x_i(0) = \bar{x}_{0,i}^N.$$

As before, $\frac{1}{N}\sum_{i=1}^N \delta_{\bar{x}_i^N(t)} \to m(t)$ which now solves ($t \in [0,T]$, $x \in Q$):

MFG

$$-\partial_t u - \frac{\sigma^2}{2}\Delta u + \frac{|\nabla u|^2}{2} = f(x, m(t))$$

$$u(x,T) = g(x, m(T))$$

$$\partial_t m - \frac{\sigma^2}{2}\Delta m - \operatorname{div}(m\nabla u) = 0$$

$$m(0) = m_0$$

Variational formulation[3]

Considering the change of variables[2] $w = -m\nabla u$ and

$$b(x,m,w) = \begin{cases} \frac{|w|^2}{2m}, & \text{if } m > 0; \\ 0, & \text{if } (m,w) = (0,0); \\ +\infty, & \text{otherwise,} \end{cases} \qquad \begin{cases} F(x,m) = \int_0^m f(x,m')\mathrm{d}m' \\ G(x,m) = \int_0^m g(x,m')\mathrm{d}m' \end{cases}$$

MFG = optimality condition of the convex problem

$$\min_{(m,w)} \int_0^T \int_Q [b(x,m(x,t),w(x,t)) + F(x,m(x,t))]\mathrm{d}x + \int_Q G(x,m(x,T))\mathrm{d}x$$

subject to

$$\partial_t m - \frac{\sigma^2}{2}\Delta m + \mathrm{div}(w) = 0 \text{ in } Q \times (0,T),$$
$$m(\cdot,0) = m_0(\cdot) \quad \text{in } Q,$$

[2] Benamou-Brenier, 2000.

[3] Lasry-Lions, 2007

Variational formulation

Convex problem

$$\min_{(m,w)} F(m,w) := \int_0^T \int_Q \underbrace{b(x,m(x,t),w(x,t))\mathrm{d}x\mathrm{d}t + \iota_C(m,w)}_{convex,\, non-smooth}$$
$$+ \underbrace{\int_0^T \int_Q F(x,m(x,t))\mathrm{d}x\mathrm{d}t + \int_Q G(x,m(x,T))\mathrm{d}x}_{convex,\, smooth}$$

where $(m,w) \in C$ if and only if

$$\partial_t m - \frac{\sigma^2}{2}\Delta m + \mathrm{div}(w) = 0 \text{ in } Q \times (0,T),$$
$$m(\cdot,0) = m_0(\cdot) \quad \text{ in } Q,$$

Goals...

Given $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$, we aim at solving

$$\min_{x\in\mathcal{H}} F(x) = f(x) + g(x)$$

$$x_0 \in \mathcal{H}, \quad (\forall n \in \mathbb{N}) \quad x_{n+1} = \Phi(x_n).$$

1. Part I: Case when f and g are diff. with Lipschitzian gradient: Gradient method: $\Phi = \mathrm{Id} - \gamma(\nabla f + \nabla g)$
2. Part II: Case when g is differentiable with Lipschitzian gradient: Prox-gradient algorithm: $\Phi = \mathrm{prox}_{\gamma f}(\mathrm{Id} - \gamma\nabla g)$
3. Part III: General case: Douglas/Peaceman-Rachford algorithm: $\Phi = R_{\gamma f} R_{\gamma g}$
4. Part IV: $g = h \circ L$ with $L\colon \mathcal{H} \to \mathcal{G}$ linear bounded: Primal-dual splitting methods: $\Phi\colon \mathcal{H} \times \mathcal{G} \to \mathcal{H} \times \mathcal{G}$.
5. Part V: Numerical experiments and extensions.

- f and/or g strongly convex: Φ strict-contractions.
- General case: Φ averaged non-expansive operators.

Part 0: Convex functions

1. Existence and uniqueness
2. F smooth: gradient
   - F convex
   - F strongly convex
3. F nonsmooth: subdifferential
   - Proximity operator
   - Examples

Existence of solutions

Banach-Alaoglu

Every bounded sequence in $\mathcal{H}$ has a weakly convergent subsequence.

Let $F \in \Gamma_0(\mathcal{H})$ be coercive, i.e.,

$$\lim_{\|x\| \to +\infty} F(x) = +\infty.$$

Then $\arg\min F = \{x^* \in \mathcal{H} \mid (\forall x \in \mathcal{H})\ F(x^*) \leq F(x)\} \neq \varnothing$.

Dem. Take $(x_n)_{n\in\mathbb{N}}$ be a minimizing sequence, i.e., $F(x_n) \to \inf_{x\in\mathcal{H}} F(x) =: \nu$. Since coercivity implies $\{x \in \mathcal{H} \mid F(x) \leq \gamma\}$ is bounded, for every $\gamma > \nu$, there exists a subsequence $x_{n_k} \rightharpoonup x \in \mathcal{H}$ by Banach-Alaoglu. By l.s.c. + convexity,

$$\nu \leq F(x) \leq \liminf_{k\in\mathbb{N}} F(x_{n_k}) = \nu.$$

Strong convexity and uniqueness of solutions

- F is β-strongly convex ($\beta > 0$) if $F - \frac{\beta}{2}\|\cdot\|^2$ is convex.
- Strongly convex functions are coercive.

Exercise

F is β-strongly convex $\Leftrightarrow$ $(\forall x, y \in \mathcal{H})(\forall \lambda \in [0,1])$
$F(x + \lambda(y - x)) \leq F(x) + \lambda(F(y) - F(x)) - \frac{\beta}{2}\lambda(1-\lambda)\|x - y\|^2$.

Uniqueness of solutions

Suppose that $F \in \Gamma_0(\mathcal{H})$ is β-strongly convex ($\beta > 0$). Then $\arg\min F$ is a singleton.

Dem. Existence is ok. Suppose that $\{x^*, y^*\} \subset \arg\min F$, $x^* \neq y^*$.

$$F(x^* + \lambda(y^* - x^*)) \leq F(x^*) + \lambda(F(y^*) - F(x^*)) - \frac{\beta}{2}\lambda(1-\lambda)\|x^* - y^*\|^2$$
$$< F(x^*) = F(y^*) \Rightarrow\Leftarrow \quad \square.$$

F G-differentiable and convex

G-differentiable

$F\colon \mathcal{H} \to \mathbb{R}$ with $\operatorname{dom} F$ open. For every $x \in \operatorname{dom} F$ and $h \in \mathcal{H}$,

$$\lim_{\lambda \to 0} \frac{F(x+\lambda h) - F(x)}{\lambda} = \langle \nabla F(x) \mid h \rangle.$$

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \begin{cases} F(x) + \langle \nabla F(x) \mid y - x \rangle \leq F(y) \\ \langle \nabla F(x) - \nabla F(y) \mid x - y \rangle \geq 0. \end{cases}$$

Dem. Let x and y in $\mathcal{H}$. From convexity, we have, for every $\lambda \in [0,1]$,

$$\underbrace{\frac{F(x+\lambda(y-x)) - F(x)}{\lambda}}_{\to \langle \nabla F(x) | y - x \rangle} \leq F(y) - F(x). \quad \square$$

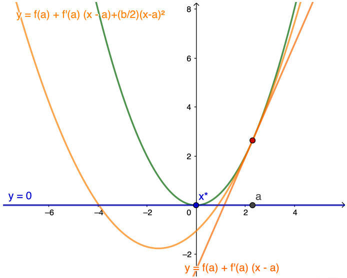
F G-differentiable and strongly convex

Since $F - \frac{\beta}{2}\|\cdot\|^2$ is convex and G-differentiable:

Exercise

Suppose that F is G-differentiable. Prove that F is β-strongly convex $\Leftrightarrow$

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \begin{cases} F(x) + \nabla\langle F(x) \mid y - x\rangle + \frac{\beta}{2}\|x - y\|^2 \leq F(y) \\ \langle \nabla F(x) - \nabla F(y) \mid x - y\rangle \geq \beta\|x - y\|^2 \end{cases}$$



Baillon-Haddad thm[4]

Theorem (Baillon-Haddad, 1977)

Let $F\colon \mathcal{H} \to \mathbb{R}$ be convex and differentiable. Then, the following are equivalent:

1. ∇F is L-Lipschitz continuous.
2. $(\forall x, y \in \mathcal{H}) \quad F(x) \leq F(y) + \langle \nabla F(y) \mid x - y \rangle + \frac{L}{2}\|x - y\|^2$.
3. $(\forall x, y \in \mathcal{H}) \quad \langle \nabla F(x) - \nabla F(y) \mid x - y \rangle \leq L\|x - y\|^2$.
4. $(\forall x, y \in \mathcal{H}) \quad \langle \nabla F(x) - \nabla F(y) \mid x - y \rangle \geq \frac{1}{L}\|\nabla F(x) - \nabla F(y)\|^2$.

[4] see, e.g. Bauschke–Combettes (2017)

Baillon–Haddad thm. $(1 \Rightarrow 2 \Rightarrow 3)$

Dem. $1 \Rightarrow 2$: Given x and y in $\mathcal{H}$, define $\phi\colon t \mapsto F(y + t(x - y))$, which is differentiable, $\phi'(t) = \langle \nabla F(y + t(x - y)) \mid x - y\rangle$ (exercise), $\phi(0) = F(y)$ and $\phi(1) = F(x)$. Hence, from C-S, Lipschitz cont., and FTC we have

$$
\begin{aligned}
F(x) - F(y) &= \int_0^1 \langle \nabla F(y + t(x - y)) \mid x - y\rangle dt \\
&= \int_0^1 \langle \nabla F(y + t(x - y)) - \nabla F(y) \mid x - y\rangle dt + \langle \nabla F(y) \mid x - y\rangle \\
&\leq \int_0^1 \|\nabla F(y + t(x - y) - \nabla F(y)\| \|x - y\| dt + \langle \nabla F(y) \mid x - y\rangle \\
&\leq L\|x - y\|^2 \int_0^1 t dt + \langle \nabla F(y) \mid x - y\rangle \\
&= \frac{L}{2}\|x - y\|^2 + \langle \nabla F(y) \mid x - y\rangle.
\end{aligned}
$$

Dem. $2 \Rightarrow 3$: Change the roles of x and y and sum. Dem. $3 \Rightarrow 2$: Exercise.

Baillon–Haddad ($3 \Rightarrow 4 \Rightarrow 1$)

$3 \Rightarrow 4$: Using 2 and convexity we have, for every x, y, z in $\mathcal{H}$

$$\begin{aligned} F(x) + \langle \nabla F(x) \mid z - x \rangle &\leq F(z) \\ F(z) &\leq F(y) + \langle \nabla F(y) \mid z - y \rangle + \frac{L}{2}\|z - y\|^2 \end{aligned}$$

which leads to

$$F(x) + \langle \nabla F(x) \mid y - x \rangle \leq F(y) + \underbrace{\langle \nabla F(y) - \nabla F(x) \mid z - y \rangle + \frac{L}{2}\|z - y\|^2}_{\varphi(z)}.$$

Since $\varphi \colon \mathcal{H} \to \mathbb{R}$ is strongly convex and differentiable, admits a unique minimizer satisfying (Fermat)

$$0 = \nabla\varphi(z^*) = \nabla F(y) - \nabla F(x) + L(z^* - y) \;\Leftrightarrow\; z^* - y = \frac{1}{L}(\nabla F(x) - \nabla F(y))$$

obtaining $F(x) + \langle \nabla F(x) \mid y - x \rangle \leq F(y) - \frac{1}{2L}\|\nabla F(x) - \nabla F(y)\|^2$. Changing the roles of x and y, the result follows.

$4 \Rightarrow 1$: C-S. □.

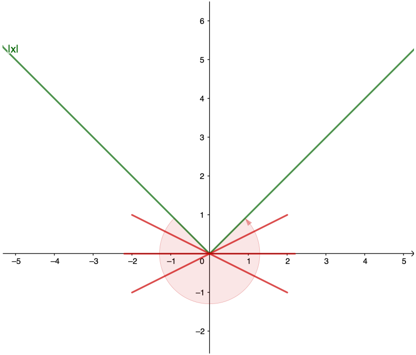
F convex nonsmooth: Subdifferential

Subdifferential of F

$$\partial F\colon \mathcal{H} \to 2^{\mathcal{H}} = \mathcal{P}(\mathcal{H})$$
$$x \mapsto \{u \in \mathcal{H} \mid (\forall y \in \mathcal{H})\ F(x) + \langle u \mid y - x\rangle \le F(y)\}$$

Example: $|x| = \begin{cases} x, & \text{if } x \ge 0; \\ -x & \text{if } x < 0. \end{cases}$

$$\partial|\cdot|\colon x \mapsto \begin{cases} \{1\}, & \text{if } x > 0; \\ [-1,1], & \text{if } x = 0; \\ \{-1\}, & \text{if } x < 0. \end{cases}$$



F convex non-smooth: Subdifferential

If F is G-differentiable, then, for every $x \in \mathcal{H}$, $\partial F(x) = \{\nabla F(x)\}$.

Dem. Let $u \in \partial F(x)$, let $h \in \mathcal{H}$, and let $t > 0$. Set $y = x + th$

$$\langle u \mid h \rangle \leq \frac{F(x+th) - F(x)}{t} \to \langle \nabla F(x) \mid h \rangle \Rightarrow \langle u - \nabla F(x) \mid h \rangle \leq 0. \quad \square$$

Chain's rule: Moreau-Rockafellar

Suppose that $F = f + g$ and g is continuous in a point of $\operatorname{dom} f$. Then, for every $x \in \mathcal{H}$, $\partial F(x) = \partial f(x) + \partial g(x)$.

Monotonicity of ∂F

For every x and y in $\mathcal{H}$, $u \in \partial F(x)$, and $v \in \partial F(y)$,

$$\langle u - v \mid x - y \rangle \geq 0.$$

When F is G-differentiable, it reduces to monotonicity of the gradient.

F convex non-smooth: Fermat

- If F is convex and non necessarily G-differentiable, we have

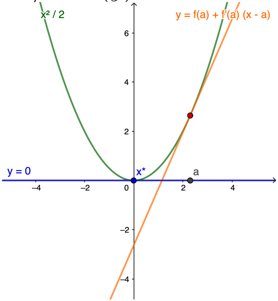
Fermat's Theorem

$$x \in \arg\min F \quad \Leftrightarrow \quad 0 \in \partial F(x)$$

Dem. $0 \in \partial F(x) \quad \Leftrightarrow \quad (\forall y \in \mathcal{H}) \quad F(x) + \langle 0 \mid y - x \rangle \leq F(y)$ $\square$.

Fermat's Theorem (F G-diff)

$$x^* \in \arg\min F \quad \Leftrightarrow \quad \nabla F(x^*) = 0$$



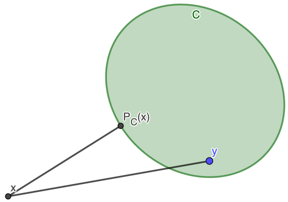
Proximity operator

Suppose that $F \in \Gamma_0(\mathcal{H})$.

Proximity operator of F

$$\operatorname{prox}_F \colon x \mapsto \underset{y \in \mathcal{H}}{\operatorname{argmin}} \; F(y) + \frac{1}{2}\|y - x\|^2$$

Example: $F = \iota_C$



Projection

$$\operatorname{prox}_{\iota_C} = P_C x = \underset{y \in C}{\operatorname{argmin}} \; \frac{1}{2}\|y - x\|^2.$$

Proximity operator

- Since $F+\|\cdot-x\|^2/2 \in \Gamma_0(\mathcal{H})$ is strongly convex, $\arg\min(F+\|\cdot-x\|^2/2) = \{p^*\}$ and prox_F is well defined.
- By Fermat's and Moreau-Rockafellar's Theorems:

$$0 \in \partial(F+\|\cdot-x\|^2/2)(p^*) = \partial F(p^*) + \{p^* - x\}$$

- Then $p^* = \operatorname{prox}_F(x)$ is the unique solution to the inclusion

$$x \in p^* + \partial F(p^*) = (\mathrm{Id} + \partial F)(p^*)$$

or, equivalently,

$$\operatorname{prox}_F(x) = (\mathrm{Id} + \partial F)^{-1}(x).$$

Example: smooth thresholder

Set $\lambda > 0$ and $F = \lambda|\cdot|$.

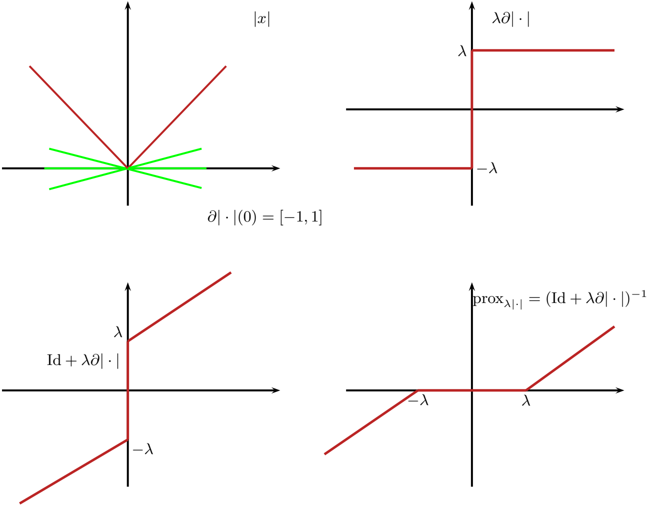
Proximity operator of $\lambda|\cdot|$

$$p = \text{prox}_{\lambda|\cdot|}x \quad \Leftrightarrow \quad x - p \in \lambda\partial|\cdot|(p) = \begin{cases} \{\lambda\}, & \text{if } p > 0; \\ [-\lambda, \lambda], & \text{if } p = 0; \\ \{-\lambda\}, & \text{if } p < 0. \end{cases}$$

- If $p > 0$, then $x - p = \lambda$ and $p = x - \lambda > 0$.
- If $p < 0$, then $x - p = -\lambda$ and $p = x + \lambda < 0$.
- If $p = 0$, then $x - p = x \in [-\lambda, \lambda]$.

$$\text{prox}_{\lambda|\cdot|}x = \begin{cases} x - \lambda, & \text{if } x > \lambda; \\ 0, & \text{if } x \in [-\lambda, \lambda]; \\ x + \lambda, & \text{if } x < -\lambda \end{cases} = \text{sign}(x)\max\{0, |x| - \lambda\}.$$

Example: smooth thresholder



Exercise

For every $i \in \{1, \ldots, N\}$, let $f_i \in \Gamma_0(\mathbb{R})$. Define $F \colon \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ as

$$(\forall x = (x_i)_{1 \leq i \leq N} \in \mathcal{H}) \quad F(x) = \sum_{i=1}^{N} f_i(x_i).$$

Prove that

$$(\forall x = (x_i)_{1 \leq i \leq N} \in \mathcal{H}) \quad \mathrm{prox}_F x = (\mathrm{prox}_{f_i} x_i)_{1 \leq i \leq N}.$$

Part I: f and g are G-differentiable

1. Gradient operator $G_{\gamma F}$.
2. $\arg\min F = \operatorname{Fix} G_{\gamma F}$.
3. F strongly convex and ∇F Lipschitz
   - $G_{\gamma F}$ strict contraction.
   - Banach-Picard's theorem.
4. F with ∇F Lipschitz
   - $G_{\gamma F}$ averaged nonexpansive.
   - Opial's lemma.

From Fermat to fixed points

Problem (P)

$$\min_{x\in\mathcal{H}} F(x) = f(x) + g(x).$$

- f and g differentiable and convex and $\gamma > 0$.

$$\begin{aligned} x^* \in \arg\min F \quad &\Leftrightarrow \quad 0 = \nabla F(x^*) \\ &\Leftrightarrow \quad x^* = x^* - \gamma \nabla F(x^*) \\ &\Leftrightarrow \quad x^* = G_{\gamma F} x^* \end{aligned}$$

Gradient operator

$$G_{\gamma F} x = x - \gamma \nabla F(x).$$

$F = f + g$ strongly convex

Definition

Let $L > 0$ and let $T\colon \mathcal{H} \to \mathcal{H}$.

- T is L-Lipschitz if $(\forall x, y \in \mathcal{H})\ \|Tx - Ty\| \leq L\|x - y\|$.
- T is a strict contraction if it is L-Lipschitz with $L \in \,]0, 1[$.
- $\operatorname{Fix} T = \big\{x \in \mathcal{H} \mid x = Tx\big\}$: fixed points of T.

Theorem

- F is G-differentiable, ρ-strongly convex, ∇F is L-Lipschitz.
- $0 < \gamma < \frac{2}{L}$.

Then, $G_{\gamma F}$ is $r_G(\gamma)$-strict contraction, where

$$r_G(\gamma) = \max\{|1 - \gamma\rho|, |1 - \gamma L|\} \in \,]0, 1[\,.$$

Moreover, $r_G(\frac{2}{L+\rho}) = \min_{\gamma>0} r_G(\gamma) = \frac{L-\rho}{L+\rho}$.

Proof $G_{\gamma F}$ strict contraction

- F is ρ-strongly convex $\Leftrightarrow$ $h = F - \frac{\rho}{2}\|\cdot\|^2$ is convex.
- F differentiable $\Leftrightarrow$ h differentiable and $\nabla h = \nabla F - \rho \mathrm{Id}$.
- ∇F is L-Lipschitz continuous (B-H)

$$\begin{aligned}
&\Leftrightarrow \langle \nabla F(x) - \nabla F(y) \mid x - y\rangle \leq L\|x-y\|^2 \\
&\Leftrightarrow \langle \nabla h(x) - \nabla h(y) \mid x - y\rangle \leq (L-\rho)\|x-y\|^2 \\
&\Leftrightarrow \nabla h \text{ is } (L-\rho)\text{-Lipschitz continuous.}
\end{aligned}$$

For every x and y in $\mathcal{H}$,

$$\begin{aligned}
\|G_{\gamma F}x - G_{\gamma F}y\|^2 &= \|x - y - \gamma(\nabla F(x) - \nabla F(y))\|^2 \\
&= \|(1-\gamma\rho)(x-y) - \gamma(\nabla h(x) - \nabla h(y))\|^2 \\
&= (1-\gamma\rho)^2\|x-y\|^2 + \gamma^2\|\nabla h(x) - \nabla h(y)\|^2 \\
&\qquad - 2\gamma(1-\gamma\rho)\langle \nabla h(x) - \nabla h(y) \mid x-y\rangle \\
&\leq (1-\gamma\rho)^2\|x-y\|^2 \\
&\qquad + \gamma(\gamma(L+\rho) - 2)\langle \nabla h(x) - \nabla h(y) \mid x-y\rangle
\end{aligned}$$

Proof $G_{\gamma F}$ strict contraction

Two cases:

- If $\gamma < \frac{2}{L+\rho}$: Since h is convex and differentiable, ∇h is monotone, i.e., $\langle \nabla h(x) - \nabla h(y) \mid x - y \rangle \geq 0$, obtaining

$$\|G_{\gamma F}x - G_{\gamma F}y\|^2 \leq \underbrace{(1-\gamma\rho)^2}_{\leq r_G(\gamma)^2} \|x-y\|^2.$$

- If $\gamma \geq \frac{2}{L+\rho}$: h is convex and ∇h is $(L-\rho)$-Lipschitz, B-H implies $\langle \nabla h(x) - \nabla h(y) \mid x - y \rangle \leq (L-\rho)\|x-y\|^2$, which yields

$$\begin{aligned}\|G_{\gamma F}x - G_{\gamma F}y\|^2 &\leq \big((1-\gamma\rho)^2 + \gamma(L-\rho)(\gamma(L+\rho)-2)\big)\|x-y\|^2 \\ &= \underbrace{(1-\gamma L)^2}_{\leq r_G(\gamma)^2} \|x-y\|^2. \quad \square\end{aligned}$$

Banach-Picard's theorem

- Suppose that $T\colon \mathcal{H} \to \mathcal{H}$ is a strict contraction with constant $L \in [0,1[$.
- Let $x_0 \in \mathcal{H}$ and $(\forall n \in \mathbb{N}) \quad x_{n+1} = Tx_n$.

Then $\operatorname{Fix} T = \{x^*\}$ and $(\forall n \in \mathbb{N})\ \|x_n - x^*\| \leq L^n \|x_0 - x^*\|$.
Hence, $x_n \to x^*$ with linear convergence rate L.

- In particular, the gradient method converges with a linear rate of $r_G(\gamma)$.

Proof of Banach-Picard's Theorem

Dem. For every $m > n$,

$$\begin{aligned}\|x_m - x_n\| &\leq \|x_m - x_{m-1}\| + \cdots + \|x_{n+1} - x_n\| \\ &\leq \|Tx_{m-1} - Tx_{m-2}\| + \cdots + \|Tx_n - Tx_{n-1}\| \\ &\leq (L^{m-2} + \cdots + L^{n-1})\|x_1 - x_0\| \\ &= (L^{n-1} - L^{m-1})/(1-L)\|x_1 - x_0\| \to 0, m, n \to +\infty\end{aligned}$$

- $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, hence it converges to x^* and

$$\begin{aligned}\|x^* - Tx^*\| &\leq \|x^* - x_n\| + \|Tx_{n-1} - Tx^*\| \\ &\leq \|x^* - x_n\| + \|x_{n-1} - x^*\| \to 0.\end{aligned}$$

- Then $x^* \in \operatorname{Fix} T$. Uniqueness (exercise).
- $\|x_n - x^*\| = \|Tx_{n-1} - Tx^*\| \leq L\|x_{n-1} - x^*\| \leq \cdots \leq L^n\|x_0 - x^*\|$. □

Gradient operator without strong conv.

- Since F is differentiable and convex, we already know that

$$x^* \in \arg\min F \quad \Leftrightarrow \quad x^* \in \operatorname{Fix} G_{\gamma F} = (\mathrm{Id} - \gamma \nabla F)$$

- Since F is not strongly convex, $G_{\gamma F}$ is no longer a strict contraction and Banach-Picard's theorem does not guarantee the convergence of the gradient method.
- Still, we have

$$\begin{aligned}
\|G_{\gamma F}x - G_{\gamma F}y\|^2 &= \|x-y\|^2 + \gamma^2\|\nabla F(x) - \nabla F(y)\|^2 \\
&\qquad - 2\gamma\langle \nabla F(x) - \nabla F(y) \mid x - y\rangle \\
&\leq \|x-y\|^2 - \gamma\left(\frac{2}{L} - \gamma\right)\|\nabla F(x) - \nabla F(y)\|^2 \\
&= \|x-y\|^2 - \left(\frac{1-\frac{L\gamma}{2}}{\frac{L\gamma}{2}}\right)\|\underbrace{\gamma\nabla F}_{\mathrm{Id}-G_{\gamma F}}(x) - \underbrace{\gamma\nabla F}_{\mathrm{Id}-G_{\gamma F}}(y)\|^2
\end{aligned}$$

Gradient operator without strong conv.

Definition

T is α-averaged nonexpansive ($\alpha \in]0,1[$) if for every $x, y \in \mathcal{H}$,

$$\|Tx - Ty\|^2 \leq \|x-y\|^2 - \frac{1-\alpha}{\alpha}\|(\mathrm{Id}-T)x - (\mathrm{Id}-T)y\|^2.$$

In particular, T is nonexpansive (i.e., 1-Lipschitz cont.)

- Then $G_{\gamma F}$ is $L\gamma/2$-averaged nonexpansive if $0 < \gamma < 2/L$.

Theorem

- T is α-averaged nonexpansive with $\operatorname{Fix} T \neq \emptyset$.
- Let $x_0 \in \mathcal{H}$ and $(\forall n \in \mathbb{N}) \quad x_{n+1} = Tx_n$.

Then, there exists $x^* \in \operatorname{Fix} T$ such that $x_n \rightharpoonup x^*$.

- In particular, the gradient method converges weakly to a solution.

Fixed point convergence: averaged nonexpansive

Dem.

- For every $x^* \in \operatorname{Fix} T$ and $n \in \mathbb{N}$, the av. nonexpansive property implies

$$\|x_{n+1}-x^*\|^2 = \|Tx_n-Tx^*\|^2 \leq \|x_n-x^*\|^2-\frac{1-\alpha}{\alpha}\|Tx_n-x_n\|^2.$$

- We obtain that $(\|x_n - x^*\|)_{n\in\mathbb{N}}$ is decreasing and bounded from below (by 0). Thus,

(1)

$$(\forall x^* \in \operatorname{Fix} T) \quad (\|x_n - x^*\|)_{n\in\mathbb{N}} \text{ converges.}$$

Fixed point convergence: averaged nonexpansive

- As a by-product we obtain $(1-\alpha)/\alpha\|Tx_n - x_n\|^2 \leq \|x_n - x^*\|^2 - \|x_{n+1} - x^*\|^2$ and

$$\frac{1-\alpha}{\alpha}\sum_{n=0}^{N}\|Tx_n - x_n\|^2 \leq \|x_0 - x^*\|^2 - \|x_{N+1} - x^*\|^2$$

- Hence, the series $\sum_{n\geq 0}\|Tx_n - x_n\|^2$ converges implying $x_n - Tx_n \to 0$.
- Since $(x_{n_k})_{n\in\mathbb{N}}$ is bounded, let y^* be any weak accumulation point, i.e., $x_{n_k} \rightharpoonup y^*$. Since T is 1-Lipschitz,

$$\begin{aligned}
\|x_{n_k} - y^*\|^2 + \|y^* - Ty^*\|^2 + 2\langle y^* - Ty^* \mid x_{n_k} - y^*\rangle
&= \|x_{n_k} - Ty^*\|^2 \\
&= \|x_{n_k} - Tx_{n_k}\|^2 + \|Tx_{n_k} - Ty^*\|^2 + 2\langle x_{n_k} - Tx_{n_k} \mid Tx_{n_k} - Ty^*\rangle \\
&\leq \underbrace{\|x_{n_k} - Tx_{n_k}\|^2}_{\to 0} + \|x_{n_k} - y^*\|^2 + 2\left\langle \underbrace{x_{n_k} - Tx_{n_k}}_{\to 0} \,\middle|\, \underbrace{Tx_{n_k} - Ty^*}_{\text{bounded}} \right\rangle
\end{aligned}$$

Fixed point convergence: averaged nonexpansive

(2)

Any weak accumulation point of $(x_n)_{n\in\mathbb{N}}$ is in $\operatorname{Fix} T$.

- With (1) and (2), we can conclude the uniqueness of the accumulation point.[5]
- Suppose that $x_{n_k} \rightharpoonup x^*$ and $x_{n_m} \rightharpoonup y^*$.
- (2) $\Rightarrow$ x^* and y^* are in $\operatorname{Fix} T$.
- (1) $\Rightarrow$ $\|x_n - x^*\| \to L_1$ and $\|x_n - y^*\| \to L_2$.
- $\underbrace{\|x_n - x^*\|^2}_{\to L_1^2} = \underbrace{\|x_n - y^*\|^2}_{\to L_2^2} + \|y^* - x^*\|^2 + 2\langle x_n - y^* \mid y^* - x^*\rangle$
- $\langle y^* - x^* \mid x_n\rangle \to L = \frac{1}{2}(L_1^2 - L_2^2 - \|y^* - x^*\|^2) + \langle y^* - x^* \mid y^*\rangle$.
- $\langle y^* - x^* \mid x^*\rangle \underset{n=n_k}{\leftarrow} \langle y^* - x^* \mid x_n\rangle \underset{n=n_m}{\rightarrow} \langle y^* - x^* \mid y^*\rangle$
- $\|y^* - x^*\|^2 = 0 \quad \Rightarrow \quad x^* = y^*$. $\square$

[5]The argument is known as Opial's Lemma

1. Motivation
2. Part 0: Convex functions
3. Part I: f and g are G-differentiable
4. Part II: g is G-differentiable
5. Part III: General case
6. Part IV: Case when $g = h \circ L$ with $L\colon \mathcal{H} \to \mathcal{G}$ linear bounded
7. Numerical experiments and extensions

Part II: g are G-differentiable

1. Proximal-gradient operator $\operatorname{prox}_{\gamma f} G_{\gamma g}$.
2. $\arg\min F = \operatorname{Fix} \operatorname{prox}_{\gamma f} G_{\gamma g}$.
3. g strongly convex and ∇g Lipschitz: $G_{\gamma g}$ strict contraction.
4. f strongly convex and ∇g Lipschitz: $\operatorname{prox}_{\gamma f}$ strict contraction.
5. g with ∇g Lipschitz: $\operatorname{prox}_{\gamma f} G_{\gamma g}$ av. nonexpansive.

From Fermat to fixed points

Problem (P)

$$\min_{x\in\mathcal{H}} F(x) = f(x) + g(x).$$

- g differentiable and convex and $\gamma > 0$.

$$\begin{aligned} x^* \in \arg\min F \quad &\Leftrightarrow \quad 0 \in \partial f(x^*) + \nabla g(x^*) \\ &\Leftrightarrow \quad x^* - \gamma\nabla g(x^*) \in \gamma\partial f(x^*) + x^* \\ &\Leftrightarrow \quad x^* = \mathrm{prox}_{\gamma f}(G_{\gamma g} x^*) \end{aligned}$$

Forward-backward operator

$$T_{\gamma f,\gamma g} = \mathrm{prox}_{\gamma f} \circ G_{\gamma g}$$

Proximity operator revisited

Let $f \in \Gamma_0(\mathcal{H})$ and $\gamma > 0$

Proximity operator

$$p = \operatorname{prox}_{\gamma f} x \quad \Leftrightarrow \quad \frac{x-p}{\gamma} \in \partial f(p)$$

Property

$\operatorname{prox}_{\gamma f}$ is 1/2-averaged nonexpansive

Dem. Let $x, y \in \mathcal{H}$, set $p = \operatorname{prox}_{\gamma f} x$, and $q = \operatorname{prox}_{\gamma f} y$. The monotonicity of ∂f implies

$$\begin{aligned} 0 &\leq \left\langle \frac{x-p}{\gamma} - \frac{y-q}{\gamma} \,\middle|\, p-q \right\rangle \\ &= \frac{1}{\gamma}\big(\langle x-y \mid p-q\rangle - \|p-q\|^2\big). \end{aligned}$$

g differentiable and strongly convex

Suppose that g is ρ-strongly convex and ∇g is L_g-Lipschitz.

- Then $G_{\gamma g}$ is a $\max\{|1-\gamma\rho|, |1-\gamma L_g|\}$-strict contraction for $\gamma \in]0, 2/L_g[$.
- $\text{prox}_{\gamma f}$ is nonexpansive (1-Lipschitz).

$T_{\gamma f,\gamma g} = \text{prox}_{\gamma f} \circ G_{\gamma g}$ is $\max\{|1-\gamma\rho|, |1-\gamma L_g|\}$-strict cont.

Forward-backward splitting (FBS)

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \text{prox}_{\gamma f}(x_n - \gamma \nabla g(x_n))$$

Then $x_n \to x^* \in \arg\min[F = (f+g)]$ linearly.

g differentiable and f strongly convex

Proximity operator

$$p = \mathrm{prox}_{\gamma f} x \quad \Leftrightarrow \quad \frac{x-p}{\gamma} \in \partial f(p)$$

Suppose that f is ρ-strongly convex.

Property

$\mathrm{prox}_{\gamma f}$ is $\frac{1}{1+\gamma\rho}$-strict contraction

Dem. Let $x, y \in \mathcal{H}$, set $p = \mathrm{prox}_{\gamma f} x$, and $q = \mathrm{prox}_{\gamma f} y$. The strong monotonicity of ∂f (monotonicity of $\partial f - \rho \mathrm{Id} = \partial(f - \frac{\rho}{2}\|\cdot\|^2)$) implies

$$\rho\|p-q\|^2 \leq \frac{1}{\gamma}\big(\langle x-y \mid p-q\rangle - \|p-q\|^2\big),$$

from which we obtain by C.S. $(1+\gamma\rho)\|p-q\| \leq \|x-y\|$ $\square$.

g differentiable and f strongly convex

Suppose that f is ρ-strongly convex and ∇g is L_g-Lipschitz.

- Then $G_{\gamma g}$ is $\gamma L_g/2$-averaged nonexpansive for $\gamma \in]0, 2/L_g[$ (it is nonexpansive for $\gamma = 2/L_g$).
- $\text{prox}_{\gamma f}$ is $\frac{1}{1+\gamma\rho}$-strict contraction.

$$T_{\gamma f,\gamma g} = \text{prox}_{\gamma f} \circ G_{\gamma g} \text{ is } \frac{1}{1+\gamma\rho}\text{-strict contraction.}$$

Forward-backward splitting (FBS)

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \text{prox}_{\gamma f}(x_n - \gamma \nabla g(x_n))$$

Then $x_n \to x^* \in \arg\min[F = (f+g)]$ linearly.

General case: ∇g is L_g-Lipschitz.

- Let $\gamma \in]0, 2/L_g[$. Then $G_{\gamma g}$ is $\gamma L_g/2$-averaged nonexpansive.
- Let $\gamma > 0$. Then $\mathrm{prox}_{\gamma f}$ is $1/2$-averaged nonexpansive.
- Composition of averaged nonexpansive operators is averaged nonexpansive ?

Combettes-Yamada (2015)

Let S_1 and S_2 be averaged nonexpansive with constants $\alpha_1 \in \,]0,1[$ and $\alpha_2 \in \,]0,1[$, respectively. Then $T = S_1 \circ S_2$ is α-averaged nonexpansive, where

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

General case: ∇g is L_g-Lipschitz.

$T_{\gamma f,\gamma g} = \mathrm{prox}_{\gamma f} \circ G_{\gamma g}$ is averaged nonexpansive.

Forward-backward splitting (FBS)

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \mathrm{prox}_{\gamma f}(x_n - \gamma \nabla g(x_n))$$

Then $x_n \rightharpoonup x^* \in \arg\min(f+g)$.

1. Motivation
2. Part 0: Convex functions
3. Part I: f and g are G-differentiable
4. Part II: g is G-differentiable
5. Part III: General case
6. Part IV: Case when $g = h \circ L$ with $L\colon \mathcal{H} \to \mathcal{G}$ linear bounded
7. Numerical experiments and extensions

Part III: General case

1. Reflection operator $R_{\gamma f}$.
2. $\arg\min F = \operatorname{prox}_{\gamma g}(\operatorname{Fix} R_{\gamma f} R_{\gamma g})$.
3. g strongly convex and ∇g Lipschitz: $R_{\gamma g}$ strict contraction.
4. $R_{\gamma f} R_{\gamma g}$ nonexpansive: KM iterations.

Reflections

Let f and g be functions in $\Gamma_0(\mathcal{H})$.

Problem

$$\min_{x\in\mathcal{H}} f(x)+g(x)$$

Definition: Reflection operator

$$R_f = 2\text{prox}_f - \text{Id}$$

Proposition

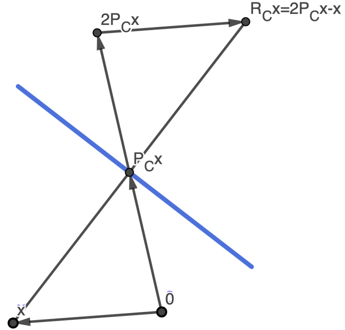
$(\forall\gamma>0)$ $R_{\gamma f}$ is nonexpansive (1-Lipschitz)

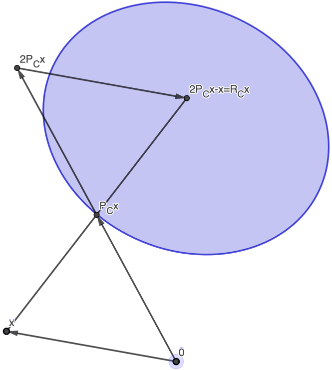
Dem.

$$\begin{aligned}\|R_f x - R_f y\|_2^2 &= 4(\|\text{prox}_f x - \text{prox}_f y\|_2^2 - \langle \text{prox}_f x - \text{prox}_f y \mid x-y\rangle) \\ &\quad + \|x-y\|_2^2 \\ &\leq \|x-y\|_2^2 \quad \square\end{aligned}$$

Reflections

Example: $f = \iota_C$. Then $\mathrm{prox}_f = P_C$ y $R_f = 2P_C - \mathrm{Id}$.





Reflections

Proposition

If $z^* \in \operatorname{Fix} R_f R_g$, then $\operatorname{prox}_g z^* \in \arg\min(f+g)$

Dem.

$$
\begin{aligned}
z^* = R_f R_g z^* \quad &\Leftrightarrow \quad z^* = 2\operatorname{prox}_f(2\operatorname{prox}_g z^* - z^*) - 2\operatorname{prox}_g z^* + z^* \\
&\Leftrightarrow \quad \operatorname{prox}_f(2\operatorname{prox}_g z^* - z^*) = \operatorname{prox}_g z^* \\
&\Rightarrow \quad \begin{cases} (2\operatorname{prox}_g z^* - z^*) - \operatorname{prox}_g z^* \in \partial f(\operatorname{prox}_g z^*) \\ z^* - \operatorname{prox}_g z^* \in \partial g(\operatorname{prox}_g z^*) \end{cases} \\
&\Leftrightarrow \quad \begin{cases} \operatorname{prox}_g z^* - z^* \in \partial f(\operatorname{prox}_g z^*) \\ z^* - \operatorname{prox}_g z^* \in \partial g(\operatorname{prox}_g z^*) \end{cases} \\
&\Rightarrow \quad 0 \in \partial f(\operatorname{prox}_g z^*) + \partial g(\operatorname{prox}_g z^*) \\
&\Rightarrow \quad \operatorname{prox}_g z^* \in \arg\min(f+g).
\end{aligned}
$$

More generally, we have $\operatorname{prox}_g(\operatorname{Fix} R_f R_g) = \arg\min(f+g)$.

Reflections: ∇g Lipschitz and $f+g$ strongly convex[6]

Suppose that

- g is G-differentiable with L-Lipschitz gradient.
- g is ρ-strongly convex.

Then $R_{\gamma g}$ is $r_g(\gamma)$-strict contraction, where

$$r_g(\gamma) = \max\left\{\frac{\gamma L - 1}{\gamma L + 1}, \frac{1-\gamma\rho}{1+\gamma\rho}\right\}.$$

Since $R_{\gamma f}$ is nonexpansive, we have:

Peaceman-Rachford algorithm

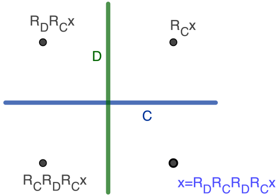
$(\forall n \in \mathbb{N}) \quad z_{n+1} = R_{\gamma f} R_{\gamma g} z_n$. Then $z_n \to z^* \in \operatorname{Fix} R_{\gamma f} R_{\gamma g}$ (unique) linearly, and $\operatorname{prox}_{\gamma g} z^* \in \arg\min(f+g)$.

Rate is optimized at $\gamma^* = 1/\sqrt{\rho L}$.

[6] Giselsson, Boyd (2017).

Reflections: general case

- Problem: $z_{n+1} = R_f R_g z_n$ does not converge.
- Example: $f = \iota_D$ and $g = \iota_C$.



- $T = R_f R_g$ is (merely) nonexpansive (1-Lipschitz)

T nonexpansive with $\operatorname{Fix} T \neq \varnothing$

- Let $T\colon \mathcal{H} \to \mathcal{H}$ be nonexpansive.
- $\operatorname{Fix} T \neq \varnothing$.

Then, defining

$$(\forall \alpha \in \,]0,1[) \quad T_\alpha = (1-\alpha)\mathrm{Id} + \alpha T,$$

we have (exercise)

- T_α is α-averaged nonexpansive.
- $\operatorname{Fix} T_\alpha = \operatorname{Fix} T \neq \varnothing$.

Krasnoselskii-Mann (KM)

$$(\forall n \in \mathbb{N}) \quad z_{n+1} = (1-\alpha) z_n + \alpha T z_n,$$

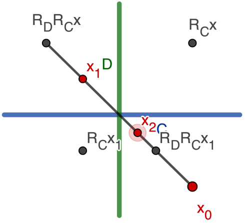
There exists $z^* \in \operatorname{Fix} T$, such that $z_n \rightharpoonup z^*$.

Douglas-Rachford splitting (DRS)

DRS: KM with $T = R_{\gamma f} R_{\gamma g}$ and $\alpha = 1/2$

$$(\forall n \in \mathbb{N}) \quad z_{n+1} = \frac{z_n + R_{\gamma f} R_{\gamma g} z_n}{2}$$
$$= \mathrm{prox}_{\gamma f}(2\mathrm{prox}_{\gamma g} z_n - z_n) + z_n - \mathrm{prox}_{\gamma g} z_n$$

There exists $z^* \in \operatorname{Fix} T$, such that $z_n \rightharpoonup z^*$ and $\mathrm{prox}_{\gamma g} z^* \in \arg\min(f+g)$.



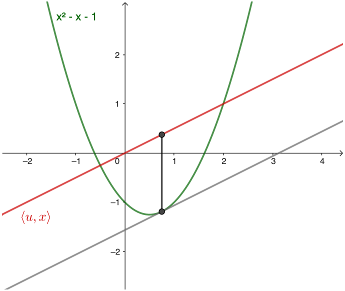
Part IV: Case when $g = h \circ L$

1. Fenchel conjugate and Fenchel Rockafellar duality.
2. ADMM.
3. Maximally monotone operators and non standard-metrics.
4. Primal-dual algorithm.

Duality: Fenchel conjugate and its properties

Fenchel conjugate

$$(\forall y \in \mathcal{H}) \quad f^*(y) = \sup_{x \in \mathcal{H}} \big(\langle x \mid y \rangle - f(x) \big)$$



- $f \in \Gamma_0(\mathcal{H}) \quad \Leftrightarrow \quad f^* \in \Gamma_0(\mathcal{H})$.
- $\partial f^* = (\partial f)^{-1}$.
- $f^{**} = f$.
- $\operatorname{prox}_{\gamma f^*} x = x - \gamma \operatorname{prox}_{f/\gamma}(x/\gamma)$.

Case $g = h \circ A$: Fenchel-Rockafellar duality

- $h \in \Gamma_0(\mathcal{G})$, $A\colon \mathcal{H} \to \mathcal{G}$ linear bounded, $\operatorname{ran} A \cap \operatorname{dom} h \neq \varnothing$.

Then $g = h \circ A \in \Gamma_0(\mathcal{H})$. $A^*\colon \mathcal{G} \to \mathcal{H}$ is the adjoint [7] of A

Primal problem

$$(P) \quad \min_{x \in \mathcal{H}} f(x) + h(Ax)$$

Dual problem

$$(D) \quad \min_{u \in \mathcal{G}} f^*(-A^*u) + h^*(u)$$

- Both values coincide.
- u^* solves (D), then any $x^* \in \partial f^*(-A^*u^*)$ solves (P).[8]

---

[7] $\langle Ax \mid u \rangle = \langle x \mid A^*u \rangle$

[8] Proposition 19.4, Bauschke-Combettes (2017)

Duality and ADMM[9]

If we use DRS with functions $f^* \circ (-A^*)$ and h^* we obtain

Alternating direction method of multipliers (ADMM)

$$x_{n+1} = \arg\min_{x \in \mathcal{H}} \left\{ f(x) + \langle u_n \mid Ax \rangle + \frac{\gamma}{2} \|Ax - y_n\|^2 \right\}$$

$$y_{n+1} = \mathrm{prox}_{h/\gamma}(u_n/\gamma + Ax_{n+1})$$

$$u_{n+1} = u_n + \gamma(Ax_{n+1} - y_{n+1}).$$

The name follows from the fact that ADMM is also the alternating minimization-maximization of

Augmented Lagrangian

$$\mathcal{L}_\gamma(x, y, u) = f(x) + h(y) + \langle u \mid Ax - y \rangle + \frac{\gamma}{2} \|Ax - y\|^2$$

[9] Glowinski-Morocco (1975), Gabay-Mercier (1976), Gabay (1983)

Splitting the linear operator A

Primal problem

$$(P) \quad \min_{x\in\mathcal{H}} f(x) + h(Ax)$$

Under qualification conditions, a more general chain rule holds:

$$\begin{aligned} x \text{ solves } (P) \quad &\Leftrightarrow \quad 0 \in \partial f(x) + A^*\partial h(Ax) \\ &\Leftrightarrow \quad (\exists\, u \in \partial h(Ax)) \quad 0 \in \partial f(x) + A^* u \\ &\Leftrightarrow \quad (\exists\, u \in \mathbb{R}^M) \quad \begin{cases} 0 \in \partial f(x) + A^* u \\ Ax \in (\partial h)^{-1} u = \partial h^*(u) \end{cases} \\ &\Leftrightarrow \quad \begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \underbrace{\begin{bmatrix} \partial f & A^* \\ -A & \partial h^* \end{bmatrix}}_{\boldsymbol{M}} \underbrace{\begin{bmatrix} x \\ u \end{bmatrix}}_{\boldsymbol{x}} \end{aligned}$$

Monotone inclusions[10]

(MI) Find $\boldsymbol{x} \in \boldsymbol{\mathcal{H}}$ such that $\mathbf{0} \in \boldsymbol{M}\boldsymbol{x}$.

In our case

- $\boldsymbol{\mathcal{H}} = \mathcal{H} \oplus \mathcal{G}$: $\langle (x,u) \mid (y,v) \rangle_{\boldsymbol{\mathcal{H}}} = \langle x \mid y \rangle + \langle u \mid v \rangle$.
- $\boldsymbol{M} \colon \boldsymbol{x} = (x,u) \mapsto (\partial f(x) + A^*u) \times (\partial h^*(u) - Ax)$.

Proposition

$\boldsymbol{M}$ is monotone.

Dem: Let $(z,w) \in \boldsymbol{M}(x,u)$ and $(r,s) \in \boldsymbol{M}(y,v)$. Then $z - A^*u \in \partial f(x)$, $r - A^*v \in \partial f(y)$, $w + Ax \in \partial h^*(u)$, $s + Ay \in \partial h^*(v)$ and we have

$$\begin{aligned}\langle (x,u) - (y,v) \mid (z,w) - (r,s) \rangle_{\boldsymbol{\mathcal{H}}} &= \langle x - y \mid z - r \rangle + \langle u - v \mid w - s \rangle \\ &= \langle x - y \mid z - r - A^*(u - v) \rangle \\ &\quad + \langle u - v \mid w - s + A(x - y) \rangle \geq 0. \quad \square\end{aligned}$$

[10] Note the similarity with Fermat.

Maximally monotone operators

Denote $\operatorname{ran} \boldsymbol{A} = \cup_{\boldsymbol{x}\in\boldsymbol{\mathcal{H}}} \boldsymbol{A}\boldsymbol{x}$ and $\boldsymbol{A}^{-1}\colon \boldsymbol{u} \mapsto \{\boldsymbol{x} \in \boldsymbol{\mathcal{H}} \mid \boldsymbol{u} \in \boldsymbol{A}\boldsymbol{x}\}$.

Definition

$\boldsymbol{M}\colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}$ is maximally monotone if it is monotone and $\operatorname{ran}(\boldsymbol{I} + \boldsymbol{M}) = \boldsymbol{\mathcal{H}}$.

- Example: $\boldsymbol{M} = \partial \boldsymbol{f}$ for $\boldsymbol{f} \in \Gamma_0(\boldsymbol{\mathcal{H}})$. Convexity $\Rightarrow$ monotonicity. L.s.c. $\Rightarrow$ maximality (Exercise).
- Example: $\boldsymbol{M}\colon (x,u) \mapsto (\partial f(x) + A^* u) \times (\partial h^*(u) - Ax)$.

Resolvent of maximally monotone operators

Definition: resolvent

Given $\boldsymbol{M}\colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}$, the resolvent of $\boldsymbol{M}$ is $J_{\boldsymbol{M}} = (\boldsymbol{I}+\boldsymbol{M})^{-1}$.

- Maximal monotonicity implies 1/2-averaged nonexpansiveness of $J_{\boldsymbol{M}}$ and $\operatorname{dom} J_{\boldsymbol{M}} = \boldsymbol{\mathcal{H}}$. (Exercise)
- $\operatorname{Fix} J_{\boldsymbol{M}} = \boldsymbol{M}^{-1}\boldsymbol{0}$ (solution to (MI)).

Proximal point algorithm (PPA)

$$(\forall n \in \mathbb{N}) \quad \boldsymbol{x}_{n+1} = J_{\boldsymbol{M}}\boldsymbol{x}_n \rightharpoonup \boldsymbol{x}^* \in \boldsymbol{M}^{-1}\boldsymbol{0}$$

- Example: $\boldsymbol{M} = \partial \boldsymbol{f}$ for $\boldsymbol{f} \in \Gamma_0(\boldsymbol{\mathcal{H}})$,
  $J_{\boldsymbol{M}} = (\boldsymbol{I}+\partial\boldsymbol{f})^{-1} = \operatorname{prox}_{\boldsymbol{f}}$, $\operatorname{Fix} J_{\boldsymbol{M}} = (\partial \boldsymbol{f})^{-1}\boldsymbol{0} = \arg\min \boldsymbol{f}$.
- Example: $\boldsymbol{M}\colon (x,u) \mapsto (\partial f(x) + A^*u) \times (\partial h^*(u) - Ax)$.
  $J_{\boldsymbol{M}}$ is averaged nonexpansive and $\operatorname{Fix} J_{\boldsymbol{M}} = \boldsymbol{M}^{-1}\boldsymbol{0}$, but...
  $J_{\boldsymbol{M}}$ is not explicit ! (try !)

To some primal dual algorithms

- Let $\delta > 0$ and let $\boldsymbol{U}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}$ be such that $\boldsymbol{U}^* = \boldsymbol{U}$ and $\langle \boldsymbol{U}\boldsymbol{x} \mid \boldsymbol{x}\rangle \geq \delta\|\boldsymbol{x}\|^2$ (we denote $\boldsymbol{U} \in \mathcal{S}_\delta$).
- $\langle \boldsymbol{x} \mid \boldsymbol{y}\rangle_{\boldsymbol{U}} = \langle \boldsymbol{x} \mid \boldsymbol{U}\boldsymbol{y}\rangle$ is an inner product for $\boldsymbol{\mathcal{H}}$ $((\boldsymbol{\mathcal{H}}, \langle \cdot \mid \cdot\rangle_{\boldsymbol{U}})$ is a Hilbert space).
- $\boldsymbol{U}^{-1}\boldsymbol{M}$ is maximally monotone in $(\boldsymbol{\mathcal{H}}, \langle \cdot \mid \cdot\rangle_{\boldsymbol{U}})$[11].
- $\boldsymbol{M}^{-1}\mathbf{0} = (\boldsymbol{U}^{-1}\boldsymbol{M})^{-1}\mathbf{0}$ and we have the averaged nonexpansiveness of $J_{\boldsymbol{U}^{-1}\boldsymbol{M}}$ in this new twisted space (PPA).
- **Question:** $J_{\boldsymbol{U}^{-1}\boldsymbol{M}}$ is explicit for some $\boldsymbol{U}$?

[11] $U^{-1}A\colon x \mapsto \{v \in \mathcal{H} \mid Uv \in Ax\}$

To some primal dual algorithms

- Let $\boldsymbol{M}\colon (x,u) \mapsto (\partial f(x) + A^* u) \times (\partial h^*(u) - Ax)$ which is maximally monotone in $(\boldsymbol{\mathcal{H}}, \langle \cdot \mid \cdot \rangle_{\boldsymbol{\mathcal{H}}})$.
- Let $\sigma > 0$ and $\tau > 0$ and set

$$\boldsymbol{U} = \begin{bmatrix} I/\tau & -A^* \\ -A & I/\sigma \end{bmatrix}$$

- Note that for $\sigma\tau\|A\|^2 < 1$, $\boldsymbol{U} \in \mathcal{S}_\delta$ for some $\delta > 0$.
- Moreover, by setting $\boldsymbol{p} = (p,q)$ and $\boldsymbol{x} = (x,u)$,

$$\begin{aligned} \boldsymbol{p} = J_{\boldsymbol{U}^{-1}\boldsymbol{M}}\boldsymbol{x} \quad & \Leftrightarrow \quad \boldsymbol{U}(\boldsymbol{x} - \boldsymbol{p}) \in \boldsymbol{M}\boldsymbol{p} \\ & \Leftrightarrow \quad \begin{cases} \frac{x-p}{\tau} - A^*(u-q) \in \partial f(p) + A^* q \\ \frac{u-q}{\sigma} - A(x-p) \in \partial h^*(q) - Ap \end{cases} \\ & \Leftrightarrow \quad \begin{cases} p = \operatorname{prox}_{\tau f}(x - \tau A^* u) \\ q = \operatorname{prox}_{\sigma h^*}(u + \sigma A(2p - x)) \end{cases} \end{aligned}$$

Primal-dual splitting[12]

Primal-dual problems

$$(P) \quad \min_{x\in\mathcal{H}} f(x) + h(Ax)$$

$$(D) \quad \min_{u\in\mathcal{G}} f^*(-A^*u) + h^*(u)$$

PPA applied to $\boldsymbol{U}^{-1}\boldsymbol{M}$ in $(\boldsymbol{\mathcal{H}}, \langle\cdot \mid \cdot\rangle_{\boldsymbol{U}})$

Let $x_0 \in \mathcal{H}$, $u_0 \in \mathcal{G}$, and $\tau\sigma\|A\|^2 < 1$.

$$\begin{aligned} x_{n+1} &= \operatorname{prox}_{\tau f}(x_n - \tau A^* u_n) \\ u_{n+1} &= \operatorname{prox}_{\sigma h^*}(u_n + \sigma A(2x_{n+1} - x_n)). \end{aligned}$$

Then, $x_n \rightharpoonup x^*$ and $u_n \rightharpoonup u^*$, x^* solves (P) and u^* solves (D).

- If f or h^* strongly convex: rate $O(1/n^2)$ in values (τ_n, σ_n).
- If f and h^* strongly convex: linear rate.

[12] Chambolle-Pock (2011), Condat (2013), Vũ(2011)

1. Motivation
2. Part 0: Convex functions
3. Part I: f and g are G-differentiable
4. Part II: g is G-differentiable
5. Part III: General case
6. Part IV: Case when $g = h \circ L$ with $L \colon \mathcal{H} \to \mathcal{G}$ linear bounded
7. Numerical experiments and extensions

Numerical experiments and extensions

1. Variable steps/metrics.
2. Numerical experiments in image processing.
3. Numerical experiments in stationary mean field games.
4. Discussion.

Extensions: Variable steps/metrics/preconditioners

- In all previous methods, we can use non-standard metrics instead of step-sizes: for $\Upsilon \in \mathcal{S}_\tau$ and $\Sigma \in \mathcal{S}_\sigma$,

$$I/\tau \to \Upsilon^{-1} \quad \text{and} \quad I/\sigma \to \Sigma^{-1}.$$

- In primal-dual algorithm, the condition on this metrics is generalized:

$$\tau\sigma\|A\|^2 < 1 \quad \to \quad \|\sqrt{\Sigma}A\sqrt{\Upsilon}\| < 1.$$

- It is also possible to prove convergence when Σ_n and Υ_n, leading to $\boldsymbol{U}_n$. But needs strong compatibility conditions as

$$\boldsymbol{U}_{n+1} \preccurlyeq (1+\eta_n)\boldsymbol{U}_n \quad \text{and} \quad \sup_{n\in\mathbb{N}} \|\boldsymbol{U}_n\| < +\infty,$$

for $(\eta_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{R})$. **Idea:** There exists $\boldsymbol{U}$ such that $\boldsymbol{U}_n \to \boldsymbol{U}$ pointwise, and use Opial in $(\boldsymbol{\mathcal{H}}, \langle \cdot \mid \cdot \rangle_{\boldsymbol{U}})$.

Numerics in image processing

- $z = Ax + \varepsilon$: A blur, ε noise.
- Image piecewise constant.

TV minimization

$$\min_{x \in \mathbb{R}^N} \lambda \|Dx\|_1 + \frac{1}{2}\|Ax - z\|_2^2 + \frac{\epsilon}{2}\|x\|_2^2$$

Original

Degraded

Figure: Blur of size 3×3 and a Gaussian noise with standard deviation $\sigma = 0.012$.

Alternative formulations

Primal fully-split formulation

$$\min_{x\in\mathbb{R}^N} f(x) + g(Lx)$$

- $f\colon x \mapsto \frac{1}{2}\|Ax - z\|_2^2 + \frac{\epsilon}{2}\|x\|_2^2 \to \operatorname{prox}_f$
- $g = \lambda\|\cdot\|_1 \to \operatorname{prox}_{g^*}$
- $L = D$

Primal-dual splitting

$\tau\gamma\|D\|^2 < 1$

For $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} x_{n+1} = (\tau A^\top A + (\tau\epsilon + 1)\mathrm{Id})^{-1}(x_n - \tau D^\top u_n + \tau A^\top y) \\ u_{n+1} = \operatorname{prox}_{\gamma g^*}\big(y_n + \gamma D(2x_{n+1} - x_n)\big) \end{array}\right.$$

Alternative formulations

Primal-dual formulation

$$\min_{\substack{x\in\mathbb{R}^N\\ u\in\mathbb{R}^M}} f(x)+g(u)+\iota_A(x,u)$$

- $f\oplus g\colon (x,u)\mapsto \frac{1}{2}\|Ax-z\|_2^2+\frac{\epsilon}{2}\|x\|_2^2+\lambda\|u\|_1$
  $\to (\operatorname{prox}_f, \operatorname{prox}_g)$
- $A=\ker[L,-\mathrm{Id}]\to P_A$.

Primal-dual Douglas-Rachford

For $n=0,1,\ldots$

$$\left\lfloor\begin{array}{l} x_n=(\tau A^\top A+(\tau\epsilon+1)\mathrm{Id})^{-1}(\tau A^\top y+z_n)\\ v_n=\operatorname{prox}_{\tau g}(\tilde{z}_n)\\ u_n=(DD^\top+\mathrm{Id})^{-1}(D(2x_n-z_n)-2v_n+\tilde{z}_n)\\ z_{n+1}=x_n-D^\top u_n\\ \tilde{z}_{n+1}=v_n+u_n \end{array}\right.$$

Alternative formulations

Dual unsplit formulation

$$\min_{u \in \mathbb{R}^M} g^*(u) + f^*(-L^*u)$$

- $g^* \to \operatorname{prox}_g$
- $f^* \circ (-L^*) \to \operatorname{prox}_{f^* \circ (-L^*)}$

ADMM

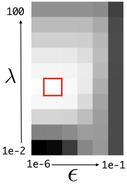
For $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} x_{n+1} = (A^\top A + \epsilon \mathrm{Id} + \tau D^\top D)^{-1}(A^\top y + D^\top(\tau u_n - y_n)) \\ u_{n+1} = \operatorname{prox}_{g/\tau}\big(Dx_{n+1} + y_n/\tau\big) \\ y_{n+1} = y_n + \tau\big(Dx_{n+1} - u_{n+1}\big) \end{array} \right.$$

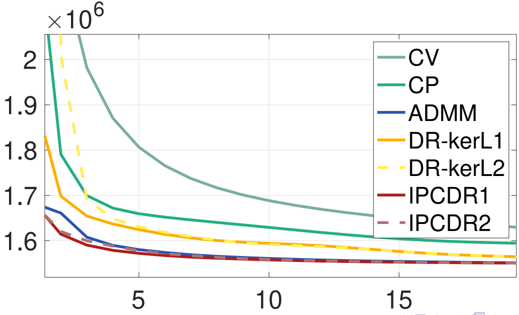
Results

Degraded

SNR MAP



Restored



Numerics in Stationary mean field games

MFG

$$-\partial_t u - \frac{\sigma^2}{2}\Delta u + \frac{|\nabla u|^2}{2} = f(x, m(t))$$
$$u(x,T) = g(x, m(T))$$
$$\partial_t m - \frac{\sigma^2}{2}\Delta m - \operatorname{div}(m\nabla u) = 0$$
$$m(0) = m_0$$

Numerics in Stationary mean field games

Stationary MFG (SMFG)

$$-\frac{\sigma^2}{2}\Delta u + \frac{|\nabla u|^2}{2} + \lambda = f(x,m),$$
$$-\frac{\sigma^2}{2}\Delta m - \operatorname{div}(m\nabla u) = 0,$$
$$\int_Q u(x)dx = 0, \quad m \geq 0, \; \int_Q m(x)dx = 1.$$

The solution (u, m, λ) of SMFG describes the long time average of solutions (u^T, m^T) of MFG as $T \to \infty$ [13].

[13] P. Cardaliaguet, J.-M. Lasry, P.-L. Lions, A. Porretta, *Long time average of mean field games with a nonlocal coupling*, *SIAM J. Control Optim.*, 2013

Variational Approach

$$b(m,w)=\begin{cases}\frac{|w|^2}{2m}, & \text{if } m>0;\\ 0, & \text{if } (m,w)=(0,0);\\ +\infty, & \text{otherwise,}\end{cases} \qquad F(x,m)=\begin{cases}\int_0^m f(x,m')dm', & \text{if } m\geq 0;\\ +\infty, & \text{otherwise.}\end{cases}$$

The SMFG is (formally) the FOC of the optimization problem

Optimization Problem (P)

$$\inf_{m,w}\int_{\mathbb{T}^2}\left[b(m(x),w(x))+F(x,m(x))\right]dx$$

$$\text{s.t}\quad \begin{cases}-\nu\Delta m(x)+\operatorname{div}\big(w(x)\big)=0, & \text{in } \mathbb{T}^2\\ \int_{\mathbb{T}^2} m(x)dx=1,\end{cases}$$

where u and λ are Lagrange multipliers and $w=-m\nabla u$ (see Lasry & Lions, 2007).

Discrete SMFG

DSMFG (Achdou & Capuzzo Dolcetta, 2010)

$$-\nu(\Delta_h u^h)_{i,j} + \frac{1}{2}|\widehat{[D_h u^h]}_{i,j}|^2 + \lambda^h = f(x_{i,j}, m^h_{i,j}) \quad \forall\, 0\le i,j\le N_h-1$$

$$-\nu(\Delta_h m^h)_{i,j} - \left(\mathrm{div}_h(m^h \widehat{[D_h u^h]})\right)_{i,j} = 0 \quad \forall\, 0\le i,j\le N_h-1$$

$$m^h_{i,j} \geq 0,\ h^2\sum_{i,j} m^h_{i,j} = 1,\ \sum_{i,j} u^h_{i,j} = 0.$$

- $h>0$, $N_h = 1/h$, $\mathcal{M}_h = \mathbb{R}^{N_h\times N_h}$, $\mathcal{W}_h = \mathbb{R}^{4(N_h\times N_h)}$.
- $\widehat{[D_h u]}_{i,j} = \left((D_1u)^-_{i,j}, -(D_1u)^+_{i-1,j}, (D_2u)^-_{i,j}, -(D_2u)^+_{i,j-1}\right) \in \mathbb{R}^4$, where $(D_1u)_{i,j} := \frac{u_{i+1,j}-u_{i,j}}{h}$, $(D_2u)_{i,j} := \frac{u_{i,j+1}-u_{i,j}}{h}$.
- Δ_h and div_h are linear operators defined by

$$(\Delta_h m)_{i,j} := -\frac{1}{h^2}(4m_{i,j}-m_{i+1,j}-m_{i-1,j}-m_{i,j+1}-m_{i,j-1})$$

$$\left(\mathrm{div}_h(w)\right)_{i,j} := (D_1w^1)_{i-1,j} + (D_1w^2)_{i,j} + (D_2w^3)_{i,j-1} + (D_2w^4)_{i,j}.$$

SMFG discretization

- If $\nu > 0$, $f(x, \cdot)$ is increasing and we suppose that the stationary system admits a unique classical solution, in Achdou, Camilli & Capuzzo Dolcetta (2013) the convergence of DSMFG (unif-L^2) to the unique solution to the stationary system as $h \to 0$ is proved.

Optimization Problem (P_h)

Discrete optimization problem (P_h)

$$\inf_{(m,w)\in\mathcal{M}_h\times\mathcal{W}_h} h^2 \sum_{i,j=0}^{N_h-1} \left[\hat{b}(m_{i,j}, w_{i,j}) + F(x_{i,j}, m_{i,j})\right]$$

$$\text{s.t.} \quad \begin{cases} -\nu(\Delta_h m)_{i,j} + (\mathrm{div}_h w)_{i,j} = 0, & \forall\, 0 \le i, j, \le N_h - 1 \\ h^2 \sum_{i,j} m_{i,j} = 1. \end{cases}$$

- $-\nu\Delta_h \colon \mathcal{M}_h \to \mathcal{M}_h$ and $\mathrm{div}_h \colon \mathcal{W}_h \to \mathcal{M}_h$ are linear.
- $\hat{b} \colon \mathbb{R} \times \mathbb{R}^4$ is given by

$$\hat{b} \colon (m, w) \mapsto \begin{cases} \dfrac{|w|^2}{2\,m}, & \text{if } m > 0,\ w \in K, \\ 0, & \text{if } (m, w) = (0, 0), \\ +\infty, & \text{otherwise.} \end{cases}$$

- $K := \mathbb{R}_+ \times \mathbb{R}_- \times \mathbb{R}_+ \times \mathbb{R}_-$.

(P_h)'s structure

- Assume $f(x,\cdot)$ increasing ($F(x,\cdot)$ convex).
- $\varphi\colon (m,w) \mapsto \sum_{i,j} \phi_{i,j}(m_{i,j},w_{i,j})$ where, $\forall\, 0 \le i,j \le N_h - 1$, $\phi_{i,j}(m_{i,j},w_{i,j}) = \hat{b}(m_{i,j},w_{i,j}) + F(x_{i,j},m_{i,j})$ is proper, convex, l.s.c., non-smooth.
- Denote $-\nu\Delta_h = A$ and $\mathrm{div}_h = B$.

(P_h)

$$\min_{(m,w)\in\mathcal{M}_h\times\mathcal{W}_h} \varphi(m,w)$$

$$\text{s.t.} \quad \begin{cases} Am + Bw = 0, \\ h^2 \mathbf{1}^\top m = 1. \end{cases}$$

(P_h)'s unsplit reformulation

- $\Xi\colon (m,w) \mapsto (Am + Bw, h^2\mathbf{1}^\top m)$.

(P_h^0)

$$\min_{(m,w)\in\mathcal{M}_h\times\mathcal{W}_h} \Phi^0(m,w) := \varphi(m,w) + \iota_{\Xi^{-1}(0,1)}(m,w)$$

- We call it unsplit formulation.
- We can use Douglas-Rachford or Primal-dual splitting ($L = \mathrm{Id}$):
  1. $P_{\Xi^{-1}(0,1)}$, which need the inversion of $(AA^* + BB^*)$.
  2. prox_φ.

(P_h)'s fully-split reformulation

- $\Xi\colon (m,w) \mapsto (Am + Bw, h^2 \mathbf{1}^\top m)$.

(P_h^1)

$$\min_{(m,w)\in\mathcal{M}_h\times\mathcal{W}_h} \Phi^1(m,w) = \varphi(m,w) + \iota_{(0,1)}\big(\Xi(m,w)\big)$$

- We call it fully-split formulation.
- Primal-dual splitting:
  1. prox_φ.
  2. Explicit activations of Ξ and Ξ^*.
  3. $\psi = \iota_{\{(0,1)\}}$, $\mathrm{prox}_{\gamma\psi^*} = \mathrm{Id} - \gamma(0,1)$, and $L = \Xi$.
  4. Then, it includes a Lagrange multiplier step of the form

$$u^{n+1} = u^n + \gamma(\Xi x^n - (0,1))$$

  5. The primal iterates $(x_k)_{k\in\mathbb{N}}$ are not feasible !
  6. Very slow...

Projected Chambolle-Pock splitting[14]

We avoid matrix inversions along with ensuring primal iterates to satisfy some of the constraints.

Projected Chambolle-Pock (PCP)

Let $x_0 \in \mathcal{H}$, $u_0 \in \mathcal{G}$ and $\sigma\tau\|L\|^2 < 1$.

$$(\forall n \in \mathbb{N}) \quad \begin{cases} p_{n+1} = \mathrm{prox}_{\tau\varphi}(x_n - \tau L^* u_n) \\ x_{n+1} = P_C p_{n+1} \\ u_{n+1} = \mathrm{prox}_{\sigma\psi^*}(u_n + \sigma L(x_{n+1} + p_{n+1} - x_n)). \end{cases}$$

$(x^k)_{k\in\mathbb{N}} \subset C$ converges to a solution in C.

- In particular, we use C as the mass constraint.
- C can change deterministically/randomly among iterations.

[14] with J. Deride, S. López Rivera, and C. Vega

$\text{prox}_{\gamma\varphi}$ computation

- In all previous methods we need to compute $\text{prox}_{\gamma\varphi}(m, w)$.
- Recall that $\varphi\colon (m, w) \mapsto \sum_{i,j} \phi_{i,j}(m_{i,j}, w_{i,j})$, where $\phi_{i,j}\colon (\mu, \omega) \mapsto \hat{b}(\mu, \omega) + F_{i,j}(\mu)$ and $F_{i,j} = F(x_{i,j}, \cdot)$.
- We have $\text{prox}_{\gamma\varphi}(m, w) = (\text{prox}_{\gamma\phi_{i,j}}(m_{i,j}, w_{i,j}))_{i,j}$.

Prox computation

$$\text{prox}_{\gamma\phi_{i,j}} : (\mu, \omega) \mapsto \begin{cases} (0, 0), & \text{if } \mu + \frac{1}{2\gamma}|P_C\omega|^2 \leq \gamma F'(0); \\ (p^*, p^* \, P_C\omega/(p^* + \gamma)), & \text{otherwise} , \end{cases}$$

where $p^* \geq 0$ is the unique solution to

$$(p + \gamma F'(p) - m)(p + \gamma)^2 - \frac{\gamma}{2}|P_K w|^2 = 0.$$

- We extend prox in Papadakis, Peyre & Oudet (2014) used in the context of optimal transport (we include F and K).

Test 1[15]

We consider the first-order stationary MFG system (Almulla-Ferreira-Gomes, 2015)

$$\frac{1}{2}|\nabla u|^2 - \lambda = \log m - \sin(2\pi x) - \sin(2\pi y),$$

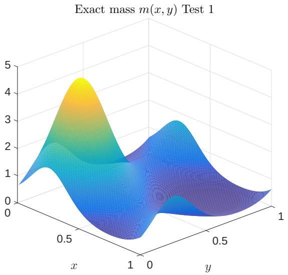
$$\operatorname{div}(m\nabla u) = 0, \quad \int_{\mathbb{T}^2} m dx = 1, \quad \int_{\mathbb{T}^2} u dx = 0,$$

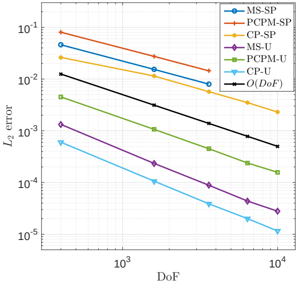
with explicit solution

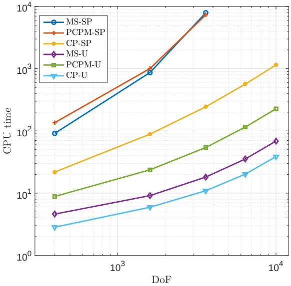
$$u(x,y) = 0, \quad m(x,y) = e^{\sin(2\pi x)+\sin(2\pi y)-\lambda},$$

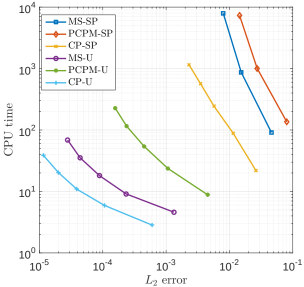
$$\lambda = \log\left(\int_{\mathbb{T}^2} e^{\sin(2\pi x)+\sin(2\pi y)} dxdy\right).$$

[15] with F.J Silva and D. Kalise


Exact mass $m(x, y)$ Test 1
x
y


MS-SP
PCPM-SP
CP-SP
MS-U
PCPM-U
CP-U
$O(DoF)$
L_2 error
DoF


MS-SP
PCPM-SP
CP-SP
MS-U
PCPM-U
CP-U
CPU time
DoF


MS-SP
PCPM-SP
CP-SP
MS-U
PCPM-U
CP-U
CPU time
L_2 error

Discussion

- All the algorithms include inertial and/or relaxation steps and errors.
- All the revised algorithms have their monotone counterparts.
- Which method I should use for my problem ? See the seminar tomorrow...

References:

- Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, second edn. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, Cham (2017).
- Nesterov, Y. Introductory lectures on convex optimization: A basic course (Vol. 87). Springer Science & Business Media. (2003).

References: Part I&II

- Aujol, J.-F., Gilboa, G., Chan, T., Osher, S.J.: Structure-texture image decomposition—Modeling, algorithms, and parameter selection. Int. J. Comput. Vis. 67, 111–136 (2006)
- Briceño-Arias, L. M. and Pustelnik, N., Convergence rate comparison of proximal algorithms for non-smooth convex optimization. Application to texture segmentation. IEEE Signal Processing Letters, vol. 22, pp. 1337–1341, 2022.
- Briceño-Arias, L. M., Combettes, P. L., Pesquet, J.-C., and Pustelnik, N., Proximal algorithms for multicomponent image processing, J. Math. Imaging Vision, vol. 41, pp. 3–22 (2011).
- Briceño-Arias, L. M., Pustelnik, N., Theoretical and numerical comparison of first-order algorithms for cocoercive equations and smooth convex optimization, Signal Processing, vol. 206, art. 108900 (2023).
- Combettes, P. L., Solving monotone inclusions via compositions of nonexpansive averaged operators, Optimization, vol. 53, no. 5–6, pp. 475–504 (2004).
- Combettes, P. L., Dũng, D., Vũ, B. C., Dualization of signal recovery problems, Set-Valued and Variational Analysis, vol. 18, pp. 373–404 (2010).

References: Part I&II

- Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forwardbackward splitting. Multiscale Model. Simul. 4, 1168–1200 (2005)
- Daubechies, I., Defrise, M., De Mol, C.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Commun. Pure Appl. Math. 57, 1413–1457 (2004)

References: Part III&IV

- Chambolle, A., Pock, T., A first-order primal-dual algorithm for convex problems with applications to imaging, J. Math. Imaging Vision, 40, pp. 120–145 (2011).
- Combettes, P.L., Yamada, I., Compositions and convex combinations of averaged nonexpansive operators, Journal of Mathematical Analysis and Applications, vol. 425, no. 1, pp. 55–70, (2015).
- Condat, L. A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms, J. Optim. Theory Appl., vol. 158, pp. 460–479 (2013).
- Gabay, D., Chapter IX applications of the method of multipliers to variational inequalities, in Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, M. Fortin and R. Glowinski, eds., vol. 15 of Studies in Mathematics and Its Applications, Elsevier, (1983) pp. 299–331.
- Gabay, D., Mercier, B., A dual algorithm for the solution of nonlinear variational problems via finite element approximation, Computers & Mathematics with Applications, 2, pp. 17–40 (1976).

References: Part III&IV

- Giselsson, P., Boyd, S., Linear convergence and metric selection for Douglas-Rachford splitting and ADMM, IEEE Trans. Automat. Control vol. 62, pp. 532–544, (2017).
- Glowinski, R., Marrocco, A., Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires, Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge Anal. Numér., 9, pp. 41–76 (1975).
- Lions, P.-L., Mercier, B., Splitting algorithms for the sum of two nonlinear operators, SIAM J. Numer. Anal., 16, pp. 964–979, (1979).
- Vũ., B. C., A splitting algorithm for dual monotone inclusions involving cocoercive operators, Adv. Comput. Math. 38, pp. 667–681, (2013).

References: Part V

- Achdou, Y., Capuzzo Dolcetta, I., Mean field games: Numerical methods, SIAM J. Numer. Anal., vol. 48(3), pp. 1136–1162 (2010).
- Almulla, N., Ferreira, R., Gomes, D., Two Numerical Approaches to Stationary Mean-Field Games, Dyn. Games Appl., vol. 7, pp. 657–682 (2017).
- Briceño-Arias, L. M., Pustelnik, N., Infimal post-composition approach for composite convex optimization. Application to image restoration, preprint, 2023.
- Briceño-Arias, L. M., Kalise, D., Silva, F. J., Proximal methods for stationary Mean Field Games with local couplings, SIAM J. Control Optim., vol. 56, no. 2, pp. 801-836 (2018).
- Briceño-Arias, L. M., Kalise, D., Kobeissi, Z. ,Laurière, M. , González, A.M., Silva, F. J., On the implementation of a primal-dual algorithm for second order time-dependent mean field games with local couplings, ESAIM Proc. Surveys, vol. 65, pp. 330–348 (2019).
- Briceño-Arias, L. M., Deride, J., López Rivera, S., Silva, F. J., A Primal-dual partial inverse splitting for constrained monotone inclusions: applications to stochastic programming and mean field games, Appl. Math. Optim., vol. 87, art. 21, 36 pp. (2023).

References: Part V

- Combettes, P. L., Vũ, B. C., Variable metric quasi-Fejér monotonicity, Nonlinear Analysis: Theory, Methods, and Applications, vol. 78, pp. 17–31, (2013).
- Lasry, J.-M., Lions, P.-L., Mean field games. Jpn. J. Math., vol. 2, pp. 229–260 (2007).
- Lasry, J.-M., Lions, P.-L., Jeux à champ moyen I. Le cas stationnaire. C. R. Math. Acad. Sci. Paris, vol. 343, pp. 619–625 (2006).
- Huang, M., Caines, P. E., Malhamé, R. P., Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized Nash equilibria. IEEE Trans. Automat. Control, vol. 52(9), pp. 1560–1571 (2007).