Denote by $\Gamma_0(\mathcal{X})$ the set of proper convex l.s.c. functions on $\mathcal{X} \to (-\infty, +\infty]$.

Douglas-Rachford algorithm

Let A and B be maximal monotone operators on $\mathcal{X} \to 2^{\mathcal{X}}$. Denote by J_A, J_B their resolvents. Define

$$C_A = 2J_A - I$$

and $C_B = 2J_B - I$. The Douglas-Rachford (DR) operator is given by

$$T = \frac{I + C_A C_B}{2}.$$

Exercise 1 (Fixed points of the DR-operator)**.** The aim is to prove the following points.

i) If $\xi \in \operatorname{fix} T$, then $J_B(\xi) \in \operatorname{zer}(A+B)$.
ii) If $\operatorname{zer}(A+B) \neq \emptyset$, then $\operatorname{fix} T \neq \emptyset$.

1. Show that $\operatorname{fix} T = \operatorname{fix} C_A C_B$.
2. Choose $\xi \in \mathcal{X}$. Define $u = J_B(\xi)$, $b = \xi - u$. Prove that
$$C_B(\xi) = u - b\,.$$
3. Assume that $\xi \in \operatorname{fix} C_A C_B$. Prove that $u = J_A(u - b)\,.$
4. Deduce that $0 \in b + A(u)$.
5. Using that $J_B(\xi) = u$, show that $b \in B(u)$.
6. Deduce that $u \in \operatorname{zer}(A+B)$ and conclude about point *i)*.
7. Consider an arbitrary $\bar{u} \in \operatorname{zer}(A+B)$. Prove that $\exists \bar{b} \in B(\bar{u})$ s.t. $-\bar{b} \in A(\bar{u})$.
8. Using that $\bar{b} \in B(\bar{u})$, prove that $\bar{u} = J_B(\bar{u} + \bar{b})$. Deduce that $C_B(\bar{u} + \bar{b}) = \bar{u} - \bar{b}$.
9. Using that $-\bar{b} \in A(\bar{u})$, prove that $\bar{u} = J_A(\bar{u} - \bar{b})$. Deduce that $C_A(\bar{u} - \bar{b}) = \bar{u} + \bar{b}$.
10. From the previous two questions, deduce that $\bar{u} + \bar{b} \in \operatorname{fix} C_A C_B$ and conclude.

Exercise 2 (DR as an averaged operator)**.** The aim is to prove the following result.

$T = \frac{I + C_A C_B}{2}$ is an averaged operator.

We recall the inequality: $\|J_B(x) - J_B(y)\|^2 \leq \|x - y\|^2 - \|(I - J_B)(x) - (I - J_B)(y)\|^2$.

1. Using the above inequality, prove that C_B is non-expansive *i.e.*, $\|C_B x - C_B y\| \leq \|x - y\|$ for every x, y.

2. Show that $C_A C_B$ is non-expansive.

3. Prove that $\|Tx - Ty\|^2 \leq \|x - y\|^2 - \|(I - T)(x) - (I - T)(y)\|^2$ and conclude.
Hint: Use the parallelogram identity $2\|a\|^2 + 2\|b\|^2 = \|a + b\|^2 + \|a - b\|^2$ *with* $a = Tx - Ty$ *and* $b = (I - T)(x) - (I - T)(y)$*, along with question 2.*

Exercise 3. We assume that $\text{zer}(A + B) \neq \emptyset$. We consider the iterates $\xi^{k+1} = T(\xi^k)$.

1. Show that the iterates ξ^k are generated by the following algorithm:

$$\boxed{\begin{aligned} u^k &= J_B(\xi^k) \\ v^k &= J_A(2u^k - \xi^k) \\ \xi^{k+1} &= \xi^k + v^k - u^k\,. \end{aligned}}$$

2. Prove that the above sequence (u^k) converges to a point in $\text{zer}(A + B)$.

3. Consider the problem
$$\text{minimize } f + g$$
where $f, g \in \Gamma_0(\mathcal{X})$. Let $\gamma > 0$. Using Question 2, propose an algorithm which requires one call to $\text{prox}_{\gamma f}$ and $\text{prox}_{\gamma g}$ at every iteration.

4. Under what sufficient condition on $f + g$ does this algorithm indeed converge to a minimizer?

Exercise 4 (Parallel programming)**.** Consider the problem

$$\text{minimize } \sum_{i=1}^{n} f_i(x) \text{ w.r.t.} x \in \mathcal{X} \tag{1}$$

where $f_1, \ldots, f_n \in \Gamma_0(\mathcal{X})$. Let $C \subset \mathcal{X}^n$ be the linear space $C := \{(x, \ldots, x) : x \in \mathcal{X}\}$.

1. In what sense is the following problem equivalent to (1) ?

$$\text{minimize } \sum_{i=1}^{n} f_i(x_i) + \iota_C(x_1, \ldots, x_n) \text{ w.r.t.} (x_1, \ldots, x_n) \in \mathcal{X}^n$$

2. Write the Douglas-Rachford algorithm associated to this problem.

3. In what sense does this algorithm deserves the name of *parallel algorithm*?

Smoothness, strong convexity and Baillon Haddad's lemma

- We say that $f : \mathcal{X} \to \mathbb{R}$ is L-smooth if it is differentiable and if ∇f is L-lipschitz continuous.
- We say that $f \in \Gamma_0(\mathcal{X})$ is μ-strongly convex ($\mu > 0$) if $f - \frac{\mu}{2}\| . \|^2$ is convex.
- We say that a operator $A : \mathcal{X} \to 2^{\mathcal{X}}$ is μ-strongly monotone if

$$\forall (x, y) \in \mathrm{dom}(A) \times \mathrm{dom}(A),\ \forall (u, v) \in A(x) \times A(y),\ \langle u - v, x - y \rangle \geq \mu \|x - y\|^2 .$$

Exercise 5. The aim is to prove the statement:

If $f \in \Gamma_0(\mathcal{X})$ is μ-strongly convex, then f^* is μ^{-1}-smooth.

1. Let $f \in \Gamma_0(\mathcal{X})$ be μ-strongly convex. Prove that ∂f is μ-strongly monotone.
   Hint: set $g := f - \frac{\mu}{2}\| . \|^2$ and use that ∂g is monotone.
2. Justify that $\mathrm{dom}(f^*) = \mathcal{X}$.
3. Show that for all $\varphi \in \mathcal{X}$, $\partial f^*(\varphi) = \arg\min_{x \in \mathcal{X}} f(x) - \langle \varphi, x \rangle$.
4. Deduce that f^* is differentiable.
5. Using 1), prove that for every $\varphi, \lambda \in \mathcal{X}$,

$$\langle \varphi - \lambda, \nabla f^*(\varphi) - \nabla f^*(\lambda) \rangle \geq \mu \|\nabla f^*(\varphi) - \nabla f^*(\lambda)\|^2 . \tag{2}$$

6. Conclude.

As a matter of fact, the above result has a converse, which we admit (the proof can be found in Hiriart-Urruty and LeMaréchal, Fundamentals of Convex Analysis.).

If f is convex and L-smooth, then f^* is L^{-1}-strongly convex

7. Using Question 5, recover the Baillon-Haddad Lemma.

Correction 1

1. *Evident.*
2. $C_B(\xi) = 2u - (u+b) = u - b.$
3. *If* $\xi = C_A C_B(\xi)$, *then* $\xi = C_A(u-b)$. *This reads* $u + b = 2J_A(u-b) - u + b$. *Thus,* $u = J_A(u-b)$.
4. *Consequently,* $u - b \in (I + A)(u)$, *which reads* $-b \in A(u)$.
5. *As* $J_B(\xi) = u$, *we have* $\xi \in u + B(u)$, *thus* $b = \xi - u \in B(u)$.
6. *Given* $\xi \in \operatorname{fix} T$, *we have found some* $b \in B(u)$ *s.t.* $0 \in b + A(u)$. *That is,* $0 \in B(u) + A(u)$. *Thus,* $u \in \operatorname{zer}(A+B)$.
7. *Evident from the definition of* $\operatorname{zer}(A+B)$.
8. *One has* $\bar b \in B(\bar u)$, *thus* $\bar u + \bar b \in (I+B)(\bar u)$, *hence* $\bar u = J_A(\bar u + \bar b)$. *Thus* $C_A(\bar u + \bar b) = 2\bar u - (\bar u + \bar b) = \bar u - \bar b$.
9. *One has* $-\bar b \in A(\bar u)$, *thus* $\bar u - \bar b \in (I+A)(\bar u)$, *hence* $\bar u = J_A(\bar u - \bar b)$. *Thus* $C_A(\bar u - \bar b) = 2\bar u - (\bar u - \bar b) = \bar u + \bar b$.
10. *One has* $\bar u + \bar b = C_A(\bar u - \bar b) = C_A C_B(\bar u + \bar b)$.

Correction 2

1. *Set* $C = C_B$. *As* $J_B = \frac{I+C}{2}$, *the inequality given in the text reads:*

$$\left\| \frac{x - y + Cx - Cy}{2} \right\|^2 \le \|x-y\|^2 - \left\| \frac{x - y - (Cx - Cy)}{2} \right\|^2 .$$

   The result follows by simply expanding the squared-norms in both handsides of the above inequality.

2. *Straightforward by composition of two non-expansive mappings.*
3. *Setting* a *and* b *as suggested, we obtain* $a + b = x - y$ *and*

$$a - b = (2T - I)(x) - (2T - I)(y) = C_A C_B x - C_A C_B y\,.$$

   In particular $\|a - b\| \le \|x - y\|$. *The parallelogram identity yields*

$$2\|Tx - Ty\|^2 + 2\|(I-T)(x) - (I-T)(y)\|^2 \le 2\|x-y\|^2\ ,$$

   which is the expected result.

Correction 3

1. *Just expand T as a function of J_A, J_B.*
2. *As* $\operatorname{zer}(A+B) \neq \emptyset$*, we have* $\operatorname{fix} T \neq \emptyset$ *(see exercise 1). As T is averaged, ξ^k converges to a point $\xi^\star$ in* $\operatorname{fix} T$*. By continuity of J_B, u^k converges to $J_B(\xi^\star)$. By exercise 1,* $J_B(\xi^\star) \in \operatorname{zer}(A+B)$*.*
3. *Just replace J_A, J_B by* $\operatorname{prox}_{\gamma f}$ *and* $\operatorname{prox}_{\gamma g}$ *respectively.*
4. *A sufficient condition is*

$$\begin{aligned} &\arg\min f + g \neq \emptyset \\ &0 \in ri(\operatorname{dom} f - \operatorname{dom} g)\ , \end{aligned}$$

where the last condition ensures that $\arg\min f + g = \operatorname{zer}(\partial f + \partial g)$*.*

Correction 4

1. *If x is a feasible point for* (1)*, then $(x, \ldots, x) \in \mathcal{X}^n$ is a feasible point for the constrained problem and its objective value is $\sum_{i=1}^n f_i(x) + \iota_C(x, \ldots, x) = \sum_{i=1}^n f_i(x)$.*

   Reciprocally, if $(x_1, \ldots, x_n)$ is a feasible point for the constrained problem, then $x_i = x_1$ for all i and x_1 is a feasible point for (1)*. Moreover, its objective value is $\sum_{i=1}^n f_i(x_1) = \sum_{i=1}^n f_i(x_i) + \iota_C(x_1, \ldots, x_n)$.*

   Hence, there is a bijection between the sets of feasible points of both problems that keeps the objective value unchanged. In that sense, both problems are equivalent.

2. *We choose $B = \partial\iota_C$ and $A = \partial f_1 \times \ldots \times \partial f_n$.*

   As f is separable, $(J_A(x))_i = (\operatorname{prox}_f(x))_i = \operatorname{prox}_{f_i}(x_i)$*.*

   $J_B(y) = \operatorname{prox}_{\iota_C}(y) = \arg\min_{x\in C} \frac{1}{2}\|y - x\|^2$ *so* $\operatorname{prox}_{\iota_C}$ *is the projection on C. This projection amounts to computing for all i, $p_i = \frac{1}{n}\sum_{i=1}^n y_i$ and* $\operatorname{prox}_{\iota_C}(y) = p$*.*

   Hence, the Douglas-Rachford algorithm writes

$$\begin{aligned} &\forall i,\ u_i^k = \frac{1}{n}\sum_{j=1}^n \xi_j^k \\ &\forall i,\ v_i^k = \operatorname{prox}_{f_i}(2u_i^k - \xi_i^k) \\ &\forall i,\ \xi_i^{k+1} = \xi_i^k + v_i^k - u_i^k \end{aligned}$$

3. *The updates for u^k amounts to a sum over n vectors. The sum is commutative and associative so this operation can be performed in parallel.*

   The updates for v^k and ξ^{k+1} are independent component-wise. Hence, these updates can be performed in parallel, each i being allocated to a single computing unit.

   All the operations of the algorithm can be performed in parallel, so one can say that this is a parallel algorithm.

Correction 5

1. *Let* $g := f - \frac{\mu}{2}\|\,.\|^2$. *One has* $g \in \Gamma_0(\mathcal{X})$. *Thus,*

$$\partial f = \partial(g + \frac{\mu}{2}\|\,.\|^2) = \partial g + \mu I\,.$$

Let $(x, y) \in \mathrm{dom}(\partial f)\times\mathrm{dom}(\partial f)$ *and choose* $(u, v) \in \partial f(x)\times\partial f(y)$. *One has* $u-\mu x \in \partial g(x)$ *and* $v - \mu y \in \partial g(y)$. *Thus,* $\langle u - \mu x - (v - \mu y), x - y\rangle \geq 0$, *which is the required inequality.*

2. $f^*(\varphi) = -\inf_x(f(x) - \langle\varphi, x\rangle)$. *The infemum is attained as* f *is strongly convex. Thus,* $f^*(\varphi) = \langle\varphi, x\rangle - f(x)$ *for some fixed* x. *Hence,* $f^*(\varphi) < +\infty$. *As this holds for every* $\varphi \in \mathcal{X}$, *we deduce that* $\mathrm{dom}(f^*) = \mathcal{X}$.

3. $x \in \partial f^*(\varphi) \Leftrightarrow \varphi \in \partial f(x) \Leftrightarrow 0 \in \partial f(x) - \varphi$. *By Fermat's rule, this is equivalent to* $x \in \arg\min_{y\in\mathcal{X}} f(y) - \langle\varphi, y\rangle$.

4. *As* f *is strongly convex, the argument of the minimum is attained (for any* $\varphi \in \mathcal{X}$*) and is a singleton. Thus* ∂f^* *is single-valued, meaning that* f^* *is differentiable.*

5. *Choose* $(\varphi, \lambda) \in \mathcal{X}\times\mathcal{X}$. *One has* $\varphi \in \partial f(\nabla f^*(\varphi))$ *and similarly for* λ. *By strong-monotonicity of* ∂f, *Eq. (2) is proven.*

6. *The conclusion follows from Cauchy-Schwartz inequality.*

7. *Choose* $f : \mathcal{X} \to \mathbb{R}$ *as convex and* L*-smooth. As* f^* *is* L^{-1}*-strongly convex, Question 5 states that for every* $(x, y) \in \mathcal{X}\times\mathcal{X}$,

$$\langle x - y, \nabla f^{**}(x) - \nabla f^{**}(y)\rangle \geq L^{-1}\|\nabla f^{**}(x) - \nabla f^{**}(y)\|^2\,.$$

As $f = f^{**}$, *we obtain Baillon-Haddad's lemma.*