Variational Analysis and Optimisation Webinar Series

Adaptive Splitting Algorithms

Hung Phan
Mathematical Sciences
University of Massachusetts Lowell

May 11 or 12, 2021

Definitions

Let X be a Hilbert spaces and let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a proper lsc convex function.
The *subdifferential* of f at x: $\partial f(x) = \Big\{\text{all subgradients of } f \text{ at } x\Big\}$, where a vector u is called a *subgradient* of f at x if

$$\forall y \in X, \quad f(y) \geq f(x) + \langle u, y - x \rangle \,.$$

The *indicator function* of a set $\Omega \subset X$ is $\iota_\Omega(x) := \begin{cases} 0, & \text{if } x \in \Omega, \\ +\infty, & \text{otherwise.} \end{cases}$

The subdifferential of ι_Ω is the normal cone operator of Ω

$$\partial(\iota_\Omega)(x) = N_\Omega(x) = \Big\{ u \in X, \langle u, z - x \rangle \leq 0, \forall z \in \Omega \Big\}$$

Fermat's Stationary

Let $f, g : X \to \mathbb{R} \cup \{+\infty\}$ be proper lsc convex functions. The Fermat's stationary condition:

$$\bar{x} \text{ solves } \min_{x \in X} f(x) \iff 0 = \nabla f(\bar{x}) \quad (f \text{ is differentiable})$$

$$\bar{x} \text{ solves } \min_{x \in X} f(x) \iff 0 \in \partial f(\bar{x}) \quad (\text{f is not differentiable})$$

$$\bar{x} \text{ solves } \min f(x) + g(x) \Longleftarrow 0 \in \partial f(\bar{x}) + \partial g(\bar{x})$$

$$\bar{x} \text{ solves } \min_{x \in \Omega} f(x) \iff \bar{x} \text{ solves } \min f(x) + \iota_\Omega(x) \Longleftarrow 0 \in \partial f(\bar{x}) + N_\Omega(\bar{x})$$

$$\bar{x} \in \Omega_1 \cap \Omega_2 \iff \bar{x} \text{ solves } \min_{x \in X} \iota_{\Omega_1}(x) + \iota_{\Omega_2}(x) \Longleftarrow 0 \in N_{\Omega_1}(\bar{x}) + N_{\Omega_2}(\bar{x})$$

So we may consider the inclusion problem: find an x such that

$$0 \in Ax + Bx \quad \text{where } A, B : X \rightrightarrows X \text{ are set-valued operators.}$$

Resolvent and Relaxed Resolvent

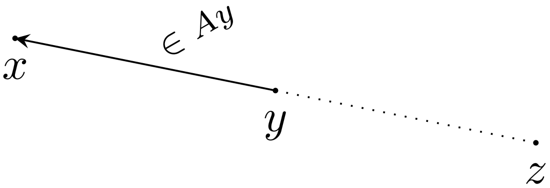
Let $A : X \rightrightarrows X$ be an operator.

The resolvent of A is defined by $J_A := (\mathrm{Id} + A)^{-1}$

The reflected resolvent of A is defined by $R_A := J_A^2 = 2J_A - \mathrm{Id}$

Let $\lambda > 0$, the λ-relaxed resolvent of A is defined by $J_A^\lambda := (1-\lambda)\,\mathrm{Id} + \lambda J_A$

$$y \in J_A x \iff y = (\mathrm{Id} + A)^{-1} x \iff x \in y + Ay$$



The resolvent of the normal cone operator is the projection:

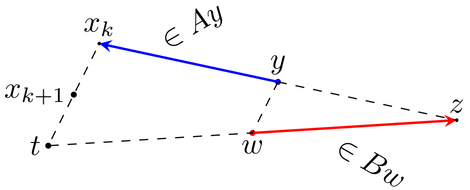
$$J_{N_\Omega}(x) = P_\Omega(x) = \left\{ y \in \Omega\,,\ \|x - y\| = \min_{z \in \Omega} \|x - z\| \right\}$$

The Douglas–Rachford (DR) Algorithm

$$x_{k+1} = Tx_k \quad \text{where} \quad T = \frac{1}{2}(\mathrm{Id} + R_B R_A).$$

Illustration:

$$y = J_A x_k, \quad z = R_A x, \quad w = J_B z, \quad t = R_B z, \quad x_{k+1} = \frac{1}{2}(x_k + t).$$



- If $x_{k+1} = x_k$, then $y = w$ and $0 \in Ay + Bw$. i.e., y is a solution.

Sum of Finitely Many Operators

Consider the problem of finding an x such that

$$0 \in A_1 x + A_2 x + \cdots + A_m x.$$

Let $\mathrm{x} := (x_1, \ldots, x_m)$. Define

$$\boldsymbol{A} := A_1 \times \cdots \times A_m : \mathrm{x} \mapsto A_1 x_1 \times \cdots \times A_m x_m$$
$$\text{and} \quad \boldsymbol{B}(\mathrm{x}) := N_\Delta(\mathrm{x}) \quad \text{where} \quad \Delta := \Big\{(x, \ldots, x) \in X^m\Big\}.$$

Then

$$0 \in A_1 x + A_2 x + \cdots + A_m x \quad \Longleftrightarrow \quad 0 \in \boldsymbol{A}(\mathrm{x}) + \boldsymbol{B}(\mathrm{x}).$$

The resolvents

$$J_{\boldsymbol{A}}(\mathrm{x}) = J_{A_1} x_1 \times \cdots \times J_{A_m} x_m,$$
$$J_{\boldsymbol{B}}(\mathrm{x}) = (x, \ldots, x) \quad \text{where} \quad x = \frac{1}{m}(x_1 + \cdots + x_m).$$

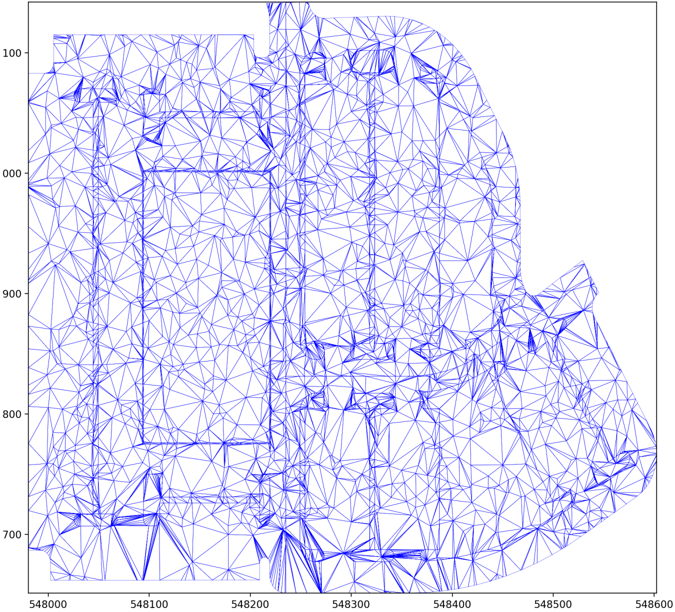
An Application [Koch, Ph '19]

A common problem in civil engineering design is the grading of a parking lot or a building pad. Within a given area, the engineer has to define grading slopes such that

- the grading site fits with existing structures.
- the drainage requirements on the surface are met.
- safety and comfort are taken into account.
- the engineer would like to change the existing surface as little as possible, in order to save on earthwork costs.

The grading site is usually represented as a Triangulated Irregular Network (TIN). The engineer is interested in *adjusting the heights* of the vertices in the triangulated grid, so that the newly obtained mesh-grid satisfies the above requirements.

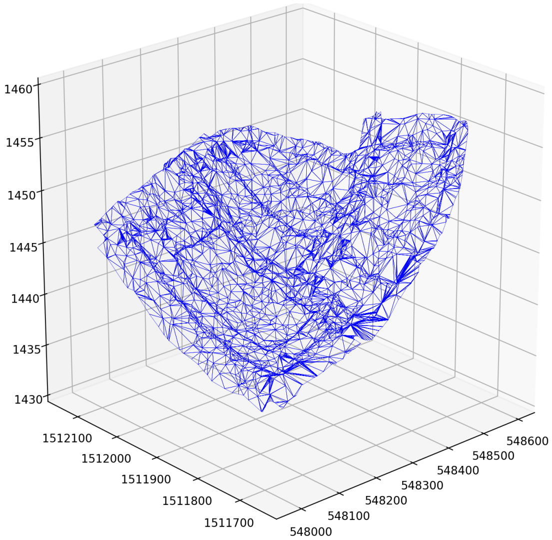
2D View of a Construction Site



$\#$ of vertices $\approx 5,000$

$\#$ of triangles: $\approx 7,000$

3D View of a Construction Site



The Triangular Mesh:

$$V = \{p_j = (p_{j1}, p_{j2}, z_j) \in \mathbb{R}^3\}, \ |V| = n,$$

$$E \subset \left\{ p_i p_j \;\middle|\; p_i, p_j \in V \right\},$$

$$T \subset \left\{ p_i p_j p_k \;\middle|\; p_i p_j, p_j p_k, p_k p_i \in E \right\}.$$

The variables are the **elevations** of the vertices, written as a vector

$$z = (z_1, z_2, \ldots, z_n) \in \mathbb{R}^n$$

Constraints and Costs

- Interpolation constraints, e.g., several values z_i's are predetermined.

$$C_{\text{interpolation}} := \left\{ z \in \mathbb{R}^n \ , \ z_j = y_j \text{ for some vertex } j \right\}.$$

- Edge-slope constraints, e.g., slopes of several edges must be within a range.

$$C_{\text{edge-slope}} := \left\{ z \in \mathbb{R}^n \ , \ \alpha \leq \text{slope}(e) \leq \beta \text{ for some edge } e \right\}.$$

- Edge-alignment constraints, e.g., slopes of several edges must equal.

$$C_{\text{edge-alignment}} := \left\{ z \in \mathbb{R}^n \ , \ \text{slope}(e_1) = \text{slope}(e_2) \text{ for some edges } e_1, e_2 \right\}.$$

- Low-point constraints, e.g., minimum slope at drainage points.

$$C_{\text{low-point}} := \left\{ z \in \mathbb{R}^n \ , \ \text{slope}(e) \geq \alpha \text{ for all } e \text{ connected to a low-point} \right\}.$$

Constraints and Costs

- Surface-alignment constraints, e.g., slopes of several triangles must equal.

$$C_{\text{surface-alignment}} := \left\{ z \in \mathbb{R}^n \ , \ \text{slope}(\Delta_1) = \text{slope}(\Delta_2) \text{ for some triangles } \Delta_1, \Delta_2 \right\}.$$

- Surface orientation constraints

$$C_{\text{surface-orientation}} := \left\{ z \in \mathbb{R}^n \ , \ \text{slope}(\Delta) = \angle(\vec{n}_\Delta, \vec{q}) \leq \alpha \text{ for some triangle } \Delta \right\}.$$

Special case: surface maximum slope: $\angle(\vec{n}_\Delta, \vec{e}_3) \leq \alpha$, $\vec{e}_3 = (0, 0, 1) \in \mathbb{R}^3$.

Special case: surface minimum slope: $\angle(\vec{n}_\Delta, \vec{d}) \leq \alpha$, $\vec{d} = (d_1, d_2, 0) \in \mathbb{R}^3$.

The cost function F can be a linear combination of

- Earth work total volume (i.e., cut and fill).
- Earth work net volume (dirt from cutting can be used for filling).
- Curvatures between adjacent triangles.

The Optimization Problem

$$\min \quad \sum_i \alpha_i F_i(z) \quad \text{subject to} \quad z \in C := \bigcap_i C_i.$$

By replacing C_i's with the indicator functions, this is equivalent to

$$\min_z \quad \sum_{j=1}^{m} f_j(z) \quad \text{where} \quad f_j \in \{\alpha_i F_i, \iota_{C_i}\}.$$

Given $\mathrm{z}_k = (z_{k,i}) \in X^m$, the DR iteration (in product space) is defined by

$$\overline{x}_k := \frac{1}{m}\sum_i x_{k,i},$$

$$\forall i = 1, \dots, m: \quad y_{k,i} := J_{\gamma \partial f_i}(2\overline{x}_{k,i} - x_{k,i}) = \mathrm{prox}_{\gamma f_i}(2\overline{x}_{k,i} - x_{k,i}),$$

$$\forall i = 1, \dots, m: \quad x_{k+1,i} := x_{k,i} - \overline{x}_k + y_{k,i},$$

$$\text{(new iteration)} \quad \mathrm{z}_{k+1} := (x_{k+1,i})_{i \in m}.$$

Then the $(\overline{x}_k)_{k \in \mathbb{N}}$ converges to a solution.

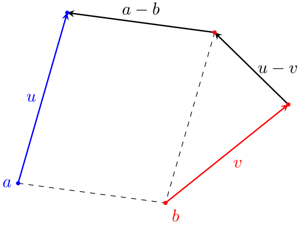
Monotonicity and Firm Nonexpansiveness

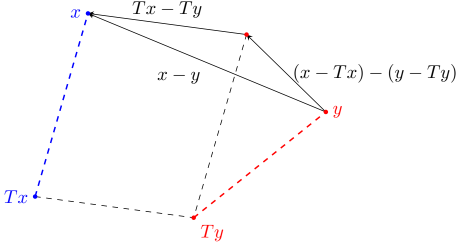
An operator A is monotone if $\forall (a,u),(b,v) \in \operatorname{gr} A, \quad \langle a-b, u-v\rangle \geq 0$.

A is maximally monotone if there is no monotone operator $\hat{A}$ such that $\operatorname{gr} A \subsetneq \operatorname{gr} \hat{A}$.

An operator T is firmly expansive (on its domain) if for all $x, y \in \operatorname{dom} T$,

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





$$A \text{ is monotone} \iff T = (\mathrm{Id}+A)^{-1} \text{ is firmly nonexpansive}$$

$$A \text{ is maximally monotone} \iff \operatorname{dom}(\mathrm{Id}+A)^{-1} = X$$

Convergence Analysis

Theorem ([Lions-Mercier 1979])

Let $A, B : X \rightrightarrows X$ *be two maximally monotone operators such that* $\operatorname{zer}(A+B) \neq \varnothing$*. Let* (x_k) *be a sequence generated by the Douglas–Rachford algorithm*

$$x_{k+1} = Tx_k \quad , \quad T = \frac{1}{2}(\mathrm{Id} + R_B R_A).$$

Then x_k *converges weakly to a fixed point* $\overline{x} \in \operatorname{Fix} T = \operatorname{Fix} R_B R_A$ *and* $J_A \overline{x} \in \operatorname{zer}(A+B)$*.*

Theorem ([Svaiter '11])

The sequence $J_A x_k$ *converges weakly to* $J_A \overline{x}$*.*

Theorem ([Bauschke '13])

The sequence $J_A x_k$ *converges weakly to* $J_A \overline{x}$*. (The proof is based on Demiclosedness Principle).*

The Adaptive DR Algorithm (aDR)

$$x_{k+1} = Tx_k \quad , \quad T = (1-\kappa)\,\mathrm{Id} + \kappa R_2 R_1,$$

where $J_1 := J_{\gamma A}$, $J_2 := J_{\delta B}$

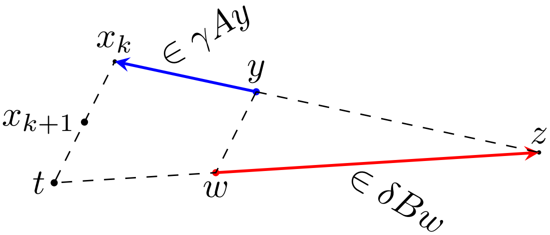
$R_1 := (1-\lambda)\,\mathrm{Id} + \lambda J_1$, $R_2 := (1-\mu)\,\mathrm{Id} + \mu J_2$,

$\gamma > 0$, $\delta > 0$, $(\lambda - 1)(\mu - 1) = 1$, $\delta = \gamma(\lambda - 1)$, $\kappa \in]0,1[$.

Illustration:

$y = J_1 x_k$, $z = R_1 x_k$, $w = J_2 z$, $t = R_2 z$,

$x_{k+1} = (1-\kappa)x_k + \kappa t$.



If $x_{k+1} = x_k \in \mathrm{Fix}\, T$, then

$$y = w \text{ and } 0 \in Ay + Bw,$$

i.e., y is a solution.

- If $\lambda = \mu = 2$, $\gamma = \delta > 0$, then the adaptive DR becomes the classical DR.

Generalized Monotonicity and Comonotonicity

Let $A : X \rightrightarrows X$ and $\alpha \in \mathbb{R}$. We say that A is

$$\alpha\text{-monotone if} \quad \forall (x,u),(y,v) \in \operatorname{gr} A, \quad \langle x-y, u-v\rangle \geq \alpha \|x-y\|^2,$$
$$\alpha\text{-comonotone if} \quad \forall (x,u),(y,v) \in \operatorname{gr} A, \quad \langle x-y, u-v\rangle \geq \alpha \|u-v\|^2,$$

and maximally α-monotone/comonotone if there is no α-monotone/comonotone operator whose graph strictly contains $\operatorname{gr} A$.

- $\alpha = 0$: monotone.
- $\alpha > 0$: strongly monotone / strongly comonotone (= cocoercive).
- $\alpha < 0$: weakly monotone/ weakly comonotone.

Apply the aDR to the problem: find x such that $0 \in Ax + Bx$ where

- A and B are maximally α- and β- monotone with $\alpha + \beta \geq 0$.
- A and B are maximally α- and β- comonotone with $\alpha + \beta \geq 0$.

Weak and Strong Monotonicity

Note that:

$$A \text{ is } \alpha_1\text{-monotone} \iff A + \alpha_2 \operatorname{Id} \text{ is } (\alpha_1 + \alpha_2)\text{-monotone.}$$

So, if A is α-monotone and B is β-monotone with $\alpha + \beta \geq 0$, then

$$A + B = \left(A - \frac{\alpha - \beta}{2} \operatorname{Id}\right) + \left(B + \frac{\alpha - \beta}{2} \operatorname{Id}\right) =: \widetilde{A} + \widetilde{B}.$$

Here, $\widetilde{A}$ and $\widetilde{B}$ are both $(\frac{\alpha+\beta}{2})$-monotone, in particular, monotone.

So, one can simply solve the problem

$$0 \in \widetilde{A}x + \widetilde{B}x$$

using available tools for monotone operators, e.g., the classical DR algorithm.

Convergence of the Adaptive DR Algorithm

Theorem [Dao-Ph '19]: Let X be a Euclidean space. Assume $A, B : X \rightrightarrows X$ are respectively maximally α-monotone and maximally β-monotone with $\operatorname{zer}(A+B) \neq \varnothing$. Let $\gamma > 0, \delta > 0, \lambda > 1, \mu > 1,\ \kappa \in \,]0,1[$, and suppose further that

$$\begin{aligned} &\alpha + \beta \geq 0, \quad 1 + 2\gamma\alpha > 0, \\ &2 - 2\gamma\beta \leq \mu \leq 2 + 2\gamma\alpha, \\ &(\lambda - 1)(\mu - 1) = 1, \quad \delta = (\lambda - 1)\gamma. \end{aligned}$$

Let $(x_k)_{k\in\mathbb{N}}$ be a sequence generated by the adaptive DR algorithm. Then (x_k) converges weakly to a point $\overline{x} \in \operatorname{Fix} T$ with $J_1\overline{x} \in \operatorname{zer}(A+B)$.

Theorem [Dao-Ph '19]: Let $\lambda = \mu = 2$ and $\gamma = \delta > 0$. Suppose that

$$\alpha + \beta > 0 \quad , \quad 1 + \gamma\frac{\alpha\beta}{\alpha+\beta} > \kappa > 0.$$

Let $(x_k)_{k\in\mathbb{N}}$ be generated by the classical DR algorithm. Then (x_k) converges weakly to a point $\overline{x} \in \operatorname{Fix} T$ with $J_1\overline{x} \in \operatorname{zer}(A+B)$.

Sketch of the Proof

Under the assumptions, we derive

$$\begin{aligned}\|Tx - Ty\|^2 \leq \|x - y\|^2 &- \frac{1-\kappa}{\kappa}\|(\mathrm{Id} - T)x - (\mathrm{Id} - T)y\|^2 \\ &- \kappa\mu\big(2 + 2\gamma\alpha - \mu\big)\|J_1 x - J_1 y\|^2 \\ &- \kappa\mu\big(\mu - (2 - 2\gamma\beta)\big)\|J_2 R_1 x - J_2 R_1 y\|^2.\end{aligned}$$

Since $2 - 2\gamma\beta \leq \mu \leq 2 + 2\gamma\alpha$, we obtain

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

which allows for the convergence of the adaptive DR algorithm via the Krasnosel'skiĭ–Mann Theorem.

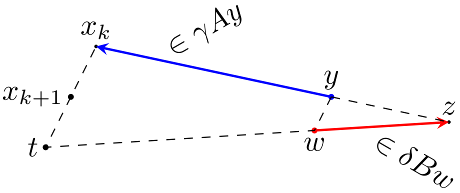
Remark: Under- and Over-Reflecting the Resolvents

Let $\alpha > 0$ and suppose that A is maximally α-monotone ("strong"),
B is maximally $(-\alpha)$-monotone ("weak").

Then

$$\mu = 2 + 2\gamma\alpha > 2 \quad \text{and} \quad \lambda = \frac{\mu}{\mu - 1} < 2.$$

- Under-reflect the resolvent of the strongly monotone operator A (use $\lambda < 2$).
- Over-reflect the resolvent of the weakly monotone operator B (use $\mu > 2$).



Convergence Analysis via Conical Averagedness

Let $\theta > 0$, we say that an operator $T : X \to X$ is conically θ-averaged if

$$T = (1-\theta)\,\mathrm{Id} + \theta N \quad \text{for some nonexpansive operator } N.$$

$$\begin{array}{lcl} \theta = 1 & : & \text{nonexpansive} \\ \theta = \frac{1}{2} & : & \text{firmly nonexpansive} \\ \theta \in \,]0,1[& : & \text{averaged} \end{array}$$

Proposition (Compositions of two conically averaged operators) [Bartz-Dao-Ph '19]

Let $T_1, T_2 : X \to X$ be conically θ_1-averaged and conically θ_2-averaged. Suppose that either $\theta_1 = \theta_2 = 1$ or $\theta_1\theta_2 < 1$. Let also $\omega \in \mathbb{R} \smallsetminus \{0\}$. Then

$$T := \left(\frac{1}{\omega} T_2\right)(\omega T_1) \quad \text{is conically } \theta\text{-averaged with} \quad \theta := \begin{cases} 1, & \theta_1 = \theta_2 = 1, \\ \frac{\theta_1+\theta_2-2\theta_1\theta_2}{1-\theta_1\theta_2}, & \theta_1\theta_2 < 1. \end{cases}$$

Theorem [Bartz-Dao-Ph '19]:

Assume A, B are maximally α-monotone and maximally β-monotone, $1 + 2\gamma\alpha > 0$, $\mu > 1$, and

$$\alpha + \beta \geq 0 \quad \text{and} \quad 2 + 2\gamma\alpha - \varepsilon \leq \mu \leq 2 + 2\gamma\alpha + \varepsilon \quad \text{with} \quad \varepsilon = 2\sqrt{\gamma(1 + \gamma\alpha)(\alpha + \beta)},$$

and either three strict inequalities happen simultaneously or none of them happens. Define

$$\lambda = \frac{\mu}{\mu - 1} \quad , \quad \delta = \frac{\gamma}{\mu - 1} \quad , \quad 0 < \kappa < \kappa^*,$$

where

$$\kappa^* := \begin{cases} 1, & \alpha + \beta = 0, \\ \frac{4\gamma\delta(1+\gamma\alpha)(1+\delta\beta)-(\gamma+\delta)^2}{2\gamma\delta(\gamma+\delta)(\alpha+\beta)}, & \alpha + \beta > 0. \end{cases}$$

Let $(x_k)_{k\in\mathbb{N}}$ be a sequence generated by the aDR algorithm.

Then (x_k) converges weakly to a fixed point $\overline{x}$ and $J_1\overline{x} \in \operatorname{zer}(A + B)$.

Theorem [Bartz-Dao-Ph '19]:

Assume A, B are maximally α-comonotone and maximally β-comonotone, $\gamma + 2\alpha > 0$, and

$$\alpha + \beta \geq 0 \quad \text{and} \quad \gamma + 2\alpha - \varepsilon \leq \delta \leq \gamma + 2\alpha + \varepsilon \quad \text{with} \quad \varepsilon = 2\sqrt{(\gamma+\alpha)(\alpha+\beta)},$$

and either three strict inequalities happen simultaneously or none of them happens. Define

$$\lambda = 1 + \frac{\delta}{\gamma} \quad , \quad \mu = 1 + \frac{\gamma}{\delta} \quad , \quad 0 < \kappa < \kappa^*,$$

where

$$\kappa^* := \begin{cases} 1, & \alpha + \beta = 0, \\ \frac{4(\gamma+\alpha)(\delta+\beta)-(\gamma+\delta)^2}{2(\gamma+\delta)(\alpha+\beta)}, & \alpha + \beta > 0. \end{cases}$$

Let $(x_k)_{k\in\mathbb{N}}$ be a sequence generated by the aDR algorithm.

Then (x_k) converges weakly to a fixed point $\overline{x}$ and $J_1\overline{x} \in \mathrm{zer}(A+B)$.

Theorem [Bauschke '13] (Demiclosedness principle for firmly nonexpansive operators)

Let $T_1, T_2 : X \to X$ be firmly nonexpansive operators, let $(x_n)_{n\in\mathbb{N}}$ and $(z_n)_{n\in\mathbb{N}}$ be sequences in X. Suppose that as $n \to +\infty$,

$$\begin{aligned} x_n \rightharpoonup x^*, &\quad z_n \rightharpoonup z^*, \\ T_1x_n \rightharpoonup y^*, &\quad T_2z_n \rightharpoonup y^*, \\ (x_n - T_1x_n) + (z_n - T_2z_n) &\to (x^* - y^*) + (z^* - y^*), \\ T_1x_n - T_2z_n &\to 0. \end{aligned}$$

Then $y^* = T_1x^* = T_2z^*$.

Theorem [Bartz-Campoy-Ph ’20] (Demiclosedness principle for cocoercive operators)
Let $T_1 : X \to X$ and $T_2 : X \to X$ be respectively σ_1- and σ_2-cocoercive[1], let $(x_n)_{n\in\mathbb{N}}$ and $(z_n)_{n\in\mathbb{N}}$ be sequences in X, and let $\rho_1, \rho_2 \in \mathbb{R}_{++}$ be such that

$$\frac{\rho_1\sigma_1 + \rho_2\sigma_2}{\rho_1 + \rho_2} \geq 1.$$

Suppose that as $n \to +\infty$,

$$\begin{aligned} x_n \rightharpoonup x^*, \quad z_n \rightharpoonup z^*, \\ T_1 x_n \rightharpoonup y^*, \quad T_2 z_n \rightharpoonup y^*, \\ \rho_1(x_n - T_1 x_n) + \rho_2(z_n - T_2 z_n) \to \rho_1(x^* - y^*) + \rho_2(z^* - y^*), \\ T_1 x_n - T_2 z_n \to 0. \end{aligned}$$

Then $y^* = T_1 x^* = T_2 z^*$.

[1] Firm nonexpansiveness is equivalent to 1-cocoercivity

Theorem [Bartz-Campoy-Ph '20] (Demiclosedness principle for averaged operators)

Let $T_1, T_2 : X \to X$ be respectively θ_1- and θ_2-averaged where $\theta_1, \theta_2 \in]0, 1[$. Let $(x_n)_{k\in\mathbb{N}}$ and $(z_n)_{k\in\mathbb{N}}$ be sequences in X and let $\rho_1, \rho_2 > 0$ be such that

$$\theta_1 \leq \frac{\rho_2}{\rho_1 + \rho_2} \quad \text{and} \quad \theta_2 \leq \frac{\rho_1}{\rho_1 + \rho_2}.$$

Suppose that as $n \to +\infty$,

$$\begin{aligned} x_n \rightharpoonup x^* \quad &\text{and} \quad z_n \rightharpoonup z^*, \\ T_1(x_n) \rightharpoonup y^* \quad &\text{and} \quad T_2(z_n) \rightharpoonup y^*, \\ \rho_1(x_n - T_1(x_n)) + \rho_2(z_n - T_2(z_n)) &\to 0, \\ T_1(x_n) - T_2(z_n) &\to 0. \end{aligned}$$

Then $T_1(x^*) = T_2(z^*) = y^*$.

Theorem [Bartz-Campoy-Ph '20] (Monotone operators)

Suppose that A and B are maximally α-monotone and maximally β-monotone, respectively, where $\alpha+\beta \geq 0$ and $\operatorname{zer}(A+B) \neq \varnothing$. Suppose the parameters $\gamma, \delta, \lambda, \mu, \kappa > 0$ are appropriately chosen. Let $(x_k)_{k\in\mathbb{N}}$ be generated the aDR. Then

$$J_{\gamma A}(x_k) \rightharpoonup J_{\gamma A}(x^*) \in \operatorname{zer}(A+B), \quad \text{where } x^* \text{ is the weak limit of } x_k.$$

Theorem [Bartz-Campoy-Ph '20] (Comonotone operators)

Suppose that A and B are maximally α-comonotone and maximally β-comonotone, respectively, where $\alpha+\beta \geq 0$ and $\operatorname{zer}(A+B) \neq \varnothing$. Suppose the parameters $\gamma, \delta, \lambda, \mu, \kappa > 0$ are appropriately chosen. Let $(x_k)_{k\in\mathbb{N}}$ be generated the aDR. Then

$$J_{\gamma A}(x_k) \rightharpoonup J_{\gamma A}(x^*) \in \operatorname{zer}(A+B), \quad \text{where } x^* \text{ is the weak limit of } x_k.$$

An Adaptive Alternating Directions Method of Multipliers (aADMM)

The Alternating Directions Method of Multipliers (ADMM) is a well studied splitting algorithm for the optimization problem

$$\begin{aligned} \min \quad & f(x) + g(z) \\ \text{s.t.} \quad & Mx = z, \quad x \in \mathbb{R}^n \quad, z \in \mathbb{R}^m, \end{aligned}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^m \to \mathbb{R}$ are proper, lsc, convex functions, and $M \in \mathbb{R}^{m \times n}$.

Given an initial point (x^0, y^0, y^0) and a parameter $\gamma > 0$, the ADMM generates

$$\begin{aligned} x^{k+1} &= \operatorname*{argmin}_{x \in \mathbb{R}^n} L_\gamma(x, z^k, y^k), \\ z^{k+1} &= \operatorname*{argmin}_{z \in \mathbb{R}^m} L_\gamma(x^{k+1}, z, y^k), \\ y^{k+1} &= y^k + \gamma(Mx^{k+1} - z^{k+1}), \end{aligned}$$

where $L_\gamma(x, z, y) = f(x) + g(z) + \langle y, Mx - z \rangle + \frac{\gamma}{2}\|Mx - z\|^2$ is the *augmented Lagrangian* associated with (P).

The aADMM [Bartz-Campoy-Ph '21]

Let $(x^0, z^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ be an initial point and let $\gamma, \delta > 0$. The aADMM iterates as follows

$$
\begin{aligned}
x^{k+1} &= \operatorname*{argmin}_{x \in \mathbb{R}^n} L_\gamma(x, z^k, y^k) \quad = \operatorname*{argmin}_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{\gamma}{2} \left\| Mx - z^k + \frac{y^k}{\gamma} \right\|^2 \right\}, \\
z^{k+1} &= \operatorname*{argmin}_{z \in \mathbb{R}^m} L_\delta(x^{k+1}, z, y^k) = \operatorname*{argmin}_{z \in \mathbb{R}^m} \left\{ g(z) + \frac{\delta}{2} \left\| Mx^{k+1} - z + \frac{y^k}{\delta} \right\|^2 \right\}, \\
y^{k+1} &= y^k + \delta(Mx^{k+1} - z^{k+1}),
\end{aligned}
$$

where the *augmented Lagrangian* is

$$
L_\gamma(x, z, y) = f(x) + g(z) + \langle y, Mx - z \rangle + \frac{\gamma}{2} \| Mx - z \|^2.
$$

Weak and Strong Convexity

We say that f is *α-convex* if $f - \frac{\alpha}{2}\|\cdot\|^2$ is convex, equivalently, if $\forall x, y \in \mathbb{R}^n$, $\lambda \in [0,1]$,

$$f((1-\lambda)x + \lambda y) \leq \lambda f(x) + (1-\lambda) f(y) - \frac{\alpha}{2}\lambda(1-\lambda)\|x-y\|^2.$$

$\alpha > 0$: We also say that f is *strongly convex*.
$\alpha < 0$: We also say that f is *weakly convex* (or *hypoconvex*).

The function f is *coercive* if

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

and *supercoercive* if

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

It is known that

$$\text{strong convexity} \implies \text{supercoercivity} \implies \text{coercivity.}$$

More Definitions

The *Fréchet subdifferential* of f at x is the set

$$\widehat{\partial} f(x) := \left\{ u \in \mathbb{R}^n : \liminf_{\substack{y \to x \\ y \neq x}} \frac{f(y) - f(x) - \langle u, y - x \rangle}{\|y - x\|} \geq 0 \right\}.$$

The *recession function* of f is defined by

$$\operatorname{rec} f : \mathbb{R}^n \to]-\infty, +\infty] : y \mapsto \sup_{x \in \operatorname{dom} f} \{ f(x+y) - f(x) \},$$

The *Fenchel conjugate* of f is defined by

$$f^* : \mathbb{R}^n \to]-\infty, +\infty] : u \mapsto \sup_{x \in \mathbb{R}^n} \{ \langle u, x \rangle - f(x) \}.$$

Convergence of the aADMM [Bartz-Campoy-Ph '21]

Let $M \in \mathbb{R}^{m\times n}$ be a nonzero matrix, let $f : \mathbb{R}^n \to]-\infty, +\infty]$ be proper, lsc and α-convex, and let $g : \mathbb{R}^m \to]-\infty, +\infty]$ be proper, lsc and β-convex with

$$\alpha \geq 0 \quad \text{and} \quad \alpha + \beta\|M\|^2 \geq 0.$$

Suppose that one of the following conditions holds:

(A.1) the Lagrangian L_0 has a critical point,

(A.2) the Lagrangian L_0 has a saddle point,

(A.3) problem (P) has an optimal solution and $0 \in \operatorname{ri}(\operatorname{dom} g - M(\operatorname{dom} f))$;

and that one of the following conditions holds:

(B.1) $0 \in \operatorname{ri}(\operatorname{dom} f^* - \operatorname{ran} M^T)$,

(B.2) $\operatorname{ri}(\operatorname{ran} \partial f) \cap \operatorname{ran} M^T \neq \varnothing$,

(B.3) $(\operatorname{rec} f)(x) > 0$ for all $x \in \ker M \setminus \{x \in \mathbb{R}^n : -(\operatorname{rec} f)(-x) = (\operatorname{rec} f)(x) = 0\}$,

(B.4) f is coercive (in particular, supercoercive),

(B.5) $\alpha > 0$ (i.e., f is strongly convex),

(B.6) $M^T M$ is invertible.

Let $\delta > \max\{0, -2\beta\}$ and set

$$\begin{aligned} &\gamma = \delta + 2\beta, && \text{if } \alpha + \beta\|M\|^2 = 0,\\ &\gamma \in \left]\max\{0, \delta + 2\beta - \Delta_\delta\}, \delta + 2\beta + \Delta_\delta\right[, && \text{if } \alpha + \beta\|M\|^2 > 0; \end{aligned}$$

where

$$\Delta_\delta := \frac{1}{\|M\|}\sqrt{2\left(\alpha + \beta\|M\|^2\right)\left(\delta + 2\beta\right)}.$$

Set $(x^0, z^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ and let $(x^k, z^k, y^k)_{k\in\mathbb{N}}$ be generated by the aADMM. Then

$$Mx^k \to Mx^\star, \quad z^k \to z^\star \quad \text{and} \quad y^k \to y^\star$$

where $(x^\star, z^\star, y^\star)$ is a critical point of $L_0(x, z, y)$. Consequently, $(x^\star, z^\star)$ is a solution of (P). If, in particular, (B.5) or (B.6) holds, then $x^k \to x^\star$.

Convergence of the aADMM: Sketch of the Proof

Define

$$Q : \mathbb{R}^m \rightrightarrows \mathbb{R}^m : y \mapsto \left\{-Mx : -M^T y \in \partial f(x)\right\} = (-M) \circ (\partial f)^{-1} \circ (-M^T)(y),$$
$$S : \mathbb{R}^m \rightrightarrows \mathbb{R}^m : y \mapsto \left\{z : y \in \widehat{\partial} g(z)\right\} = (\widehat{\partial} g)^{-1}(y),$$

Then the sequence $w^k := y^k + \delta z^k$ is generated by the aDR algorithm with parameters γ, δ applied to S and Q.

Under the assumptions made:

- $\operatorname{zer}(Q + S) \neq \varnothing$.
- Q is maximally $\frac{\alpha}{\|M\|^2}$-comonotone, S is maximally β-comonotone, and $\frac{\alpha}{\|M\|^2} + \beta \geq 0$.

Finally, apply the convergence result of the aDR for two comonotone operators.

Thank you!

Bartz, S., Dao, M.N., Phan, H.M.: Conical averagedness and convergence analysis of fixed point algorithms, preprint 2019, arXiv:1910.14185.

Bartz, S., Campoy, R., Phan, H.M.: Demiclosedness principles for generalized nonexpansive mappings, *J. Optim. Theory Appl.* **186**(3), 759–778 (2020)

Bauschke, H.H.: New demiclosedness principles for (firmly) nonexpansive operators. In: Bailey D. et al. (eds) *Computational and Analytical Mathematics*, Springer Proceedings in Mathematics & Statistics. Springer, New York (2013)

Dao, M.N., Phan, H.M.: Adaptive Douglas–Rachford splitting algorithm for the sum of two operators, *SIAM J. Optim.* **29**(4), 2697–2724 (2019)

Douglas, J., Rachford, H.H.: On the numerical solution of the heat condition problem in 2 and 3 space variables, *Trans. AMS.* **82**, 421–439 (1956)

Guo, K., Han, D., Yuan, X.: Convergence analysis of Douglas–Rachford splitting method for "strongly + weakly" convex programming, *SIAM J. Numer. Anal.* **55**(4), 1549–1577 (2017).

Koch, V.R., Phan, H.M.: Optimization of triangular networks with spatial constraints, *Optim. Methods Softw.* (2019), DOI:10.1080/10556788.2019.1604703.

Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal.* **16**, 964–979 (1979)

Svaiter, B.F.: On weak convergence of the Douglas–Rachford method. *SIAM J. Control Optim.* **49**(1), 280–287 (2011)

Thank you!

- Bartz, S., Campoy, R., Phan, H.M: An adaptive alternating directions method of multipliers, preprint 2021, arXiv:2103.07159
- Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**(1), 1–122 (2011)
- Chen, L., Sun, D., Toh, K.C.: A note on the convergence of ADMM for linearly constrained convex optimization problems. *Comput. Optim. Appl.* **66**(2), 327–343 (2017)
- Eckstein, J., Yao, W.: Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives. *Pac. J. Optim.* **11**(4), 619–644 (2015)
- Fortin, M., Glowinski, R.: On decomposition-coordination methods using an augmented Lagrangian. In M. Fortin and R. Glowinski (eds.), *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems. Studies in Mathematics and Its Applications* **15**, 97–146 (1983)
- Gabay, D.: Applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski (eds.), *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems. Studies in Mathematics and Its Applications* **15** 299–331 (1983)
- Gabay, G., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximations. *Comput. Math. Appl.* **2**(1), 17–40 (1976)
- Glowinski, R., Marroco, 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. *ESAIM-Math. Model. Num.* **9**(R2), 41–76 (1975)