



UNIVERSITÀ DEGLI STUDI
DI **CAGLIARI**



Introduction to Open Multi-Agent Systems – Part IV

Splitting algorithms via operator theory

Diego Deplano*

*Department of Electrical and Electronic Engineering, University of Cagliari, Italy

SIDRA Ph.D. Summer School • July 13–15, 2026 • Bertinoro, Italy

Outline

- ① Recap
- ② The Peaceman-Rachford splitting (PRS) algorithm
- ③ The distributed PRS algorithm in OMAS (Open ADMM)
- ④ The federated PRS algorithm in OMAS (Open FedPRS)
- ⑤ A numerical example on the EMNIST dataset
- ⑥ Conclusions, future directions, and debate

Outline

- 1 Recap
- 2 The Peaceman-Rachford splitting (PRS) algorithm
- 3 The distributed PRS algorithm in OMAS (Open ADMM)
- 4 The federated PRS algorithm in OMAS (Open FedPRS)
- 5 A numerical example on the EMNIST dataset
- 6 Conclusions, future directions, and debate

In **open multi-agent systems** the set of agents $\mathcal{V}_k$ may change at each step $k \in \mathbb{N}$, and each agent $i \in \mathcal{V}_k$ holds:

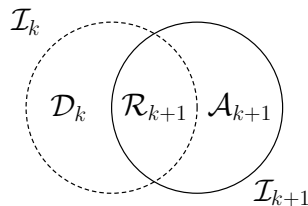
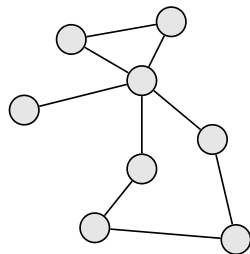
- $x_i \in \mathbb{R}^p$ a local decision variable;
- $f_i : \mathbb{R}^p \rightarrow \mathbb{R}$ a local objective function;
- $\mathcal{X}_i \subseteq \mathbb{R}^p$ a local admissible decision space (constraints);

so that the distributed optimization problem is:

$$\begin{aligned} & \min_{\{x_i\}_{i \in \mathcal{V}_k}} \sum_{i \in \mathcal{V}_k} f_i(x_i) \\ & x_i = x_j, \quad \forall (i, j) \in \mathcal{E}_k \text{ (consensus constraints)}. \end{aligned}$$

The set of agents is partitioned into:

- **Remaining labels** $\mathcal{R}_{k+1} = \mathcal{I}_{k+1} \cap \mathcal{I}_k$: labels present both at time k and time $k+1$;
- **Arriving labels** $\mathcal{A}_{k+1} = \mathcal{I}_{k+1} \setminus \mathcal{I}_k$: labels present at time $k+1$ but not at time k ;
- **Departing labels** $\mathcal{D}_k = \mathcal{I}_k \setminus \mathcal{I}_{k+1}$: labels present at time k but not at time $k+1$.



Theorem 1 in [R2]

Consider a sequence of finite label sets $\{\mathcal{I}_k\}_{k \in \mathbb{N}}$ and spaces $\{\mathcal{X}_k\}_{k \in \mathbb{N}}$ labeled by those sets. Consider the iteration $x(k)$ ruled by an open operator $T_k : \mathcal{X}_k \rightarrow \mathcal{X}_{k+1}$, and split the state $x(k) = [x_k^R; x_k^A]$ into two vectors:

- $x_k^R \in \mathbb{R}^{\mathcal{R}_k}$ is the vector of the remaining components;
- $x_k^A \in \mathbb{R}^{\mathcal{A}_k}$ is the vector of the arriving components.

Assume that

- (a) F_k is paracontractive with $\gamma \in (0, 1)$;
- (b) the TSI has bounded variation $B \geq 0$, i.e., $d_{\text{SH}}(\hat{\mathcal{X}}_{k+1}, \hat{\mathcal{X}}_k) / \sqrt{|\mathcal{I}_{k+1}|} \leq B$;
- (c) the arrival process is bounded with $H \geq 0$, i.e., $\exists \hat{x}_k \in \text{proj}(x_k^R, \hat{\mathcal{X}}_k)$ s.t. $d(x_k^A, \hat{x}_k) / \sqrt{|\mathcal{I}_k|} \leq H$;
- (d) the departure process is bounded with $\beta \in (\gamma, 1)$, i.e., $\sqrt{|\mathcal{I}_{k+1}|} / \sqrt{|\mathcal{I}_k|} \geq \beta$.

Then, the open sequence $\{x(k) \in \mathcal{X}_k\}_{k \in \mathbb{N}}$ converges linearly with rate $\theta = \gamma/\beta \in (0, 1)$ to the TSI within a radius

$$R = \frac{B + H}{1 - \theta}, \quad (1)$$

according to the following pointwise upper bound

$$\frac{d(x(k), \hat{\mathcal{X}}_k)}{\sqrt{|\mathcal{I}_k|}} \leq \theta^k \frac{d(x(0), \hat{\mathcal{X}}_0)}{\sqrt{|\mathcal{I}_0|}} + \frac{1 - \theta^k}{1 - \theta} (B + H).$$

[R2] D. Deplano, N. Bastianello, M. Franceschelli, K.H. Johansson (2026). Optimization and learning in Open Multi-Agent Systems, IEEE Transactions on Automatic Control.

The local state updated of the DGD in open multi-agent systems is given by:

$$x_i(k+1) = \begin{cases} \sum_{j \in \mathcal{N}_{i,k} \cup \{i\}} w_{ij(k)} x_j(k) - \rho \nabla f_i(x_i(k)) & \text{if } i \in \mathcal{V}_{k+1}^R, \\ x_{i,k+1}^* & \text{if } i \in \mathcal{V}_{k+1}^A, \end{cases}$$

where $w_{ii}(k) = 1 - \varepsilon |\mathcal{N}_{i,k}|$ and $w_{ij}(k) = \varepsilon$ if $j \in \mathcal{N}_{i,k}$.

Assumptions:

- 1 local costs f_i are proper, lsc, m -strongly convex, L -smooth, differentiable for all agents $i \in \mathcal{V}_k$;
- 2 There exists a uniform upper bound $\bar{\lambda}$ on the largest eigenvalue $\lambda_{\max}$ of the Laplacian matrix;
 - Assumptions 1-2 guarantee the contractivity of the standard iteration, with factor
 $\gamma = \max\{|1 - \rho m|, |1 - \rho L|, |1 - \rho L - \varepsilon \bar{\lambda}|\} \in (0, 1)$
- 3 the distance between two consecutive (unique) global solutions x_k^* and x_{k+1}^* is upper bounded by a constant $\sigma \geq 0$;
- 4 the distance between each (unique) local solution $x_{i,k}^*$ and the (unique) global solution x_k^* is upper bounded by $\omega \geq 0$;
 - Assumptions 3-4 guarantee: (i) the trajectory of points of interest (TPI) has bounded variation with
 $B = \frac{1}{\sqrt{p}}(\frac{L\omega}{m} + \sigma + \frac{L\omega}{m\beta})$; (ii) the arrival process is bounded with $H = \frac{1}{\sqrt{p}}(1 + \frac{L}{m})\omega$.
- 5 The departure process is bounded by $\beta \in (\gamma, 1)$;
 - Assumption 5 guarantees the linear convergence of the state trajectory to the trajectory of point of interest, with rate
 $\theta = \gamma/\beta \in (0, 1)$

Remark 1. The standard iteration of the distributed DGD in open networks admits a unique fixed point at each step k , thus yielding a TPI and allowing to prove its contractivity.

Question 1. How to deal with algorithms that do not admit a TPI but a trajectory of sets of interest (TSI)?

Answer 1. Since contractivity cannot hold by definition, one must resort to the less strict property of paracontractivity.

Remark 2. The TPI of the distributed DGD in open networks does not coincide with the set of optimal solutions (i.e., it converges to the wrong point).

Question 2. Are there alternatives to distributed DGD?

Answer 2. Many. In this set of slides a distributed version of ADMM in open network will be discussed and characterized.

Outline

- 1 Recap
- 2 The Peaceman-Rachford splitting (PRS) algorithm
- 3 The distributed PRS algorithm in OMAS (Open ADMM)
- 4 The federated PRS algorithm in OMAS (Open FedPRS)
- 5 A numerical example on the EMNIST dataset
- 6 Conclusions, future directions, and debate

We consider unconstrained, composite optimization problems:

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

A solution to the problem can be computed by means of the Peaceman-Rachford algorithm

$$\begin{aligned} z(k+1) &= \text{refl}_g^\rho(\text{refl}_f^\rho(z(k))), \\ x(k+1) &= \text{prox}_f^\rho(z(k)), \end{aligned}$$

if the objective function satisfies the following conditions:

- convexity: both f and g ;
- properness: both f and g ;
- lower semicontinuous (lsc): both f and g .

Furthermore, we define the relaxed version (Krasnoselskij iteration) by

$$z(k+1) = (1 - \alpha)z(k) + \alpha \text{refl}_g^\rho(\text{refl}_f^\rho(z(k))), \quad \alpha \in (0, 1).$$

Corollary 27.6 in [R1] (Consequence of the Fermat's rule)

Let $f : \mathcal{X} \rightarrow \mathbb{R}$ and $g : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper, lsc, and convex functions. Then, if f and g satisfy a qualification condition^o, minimizers x^* of $f + g$ satisfy

$$x^* \in \operatorname{argmin}(f + g) \quad \Longleftrightarrow \quad 0 \in (\partial f + \partial g)(x^*).$$

[R1] Bauschke HH and Combettes PL (2017). Convex analysis and monotone operator theory in Hilbert spaces, second ed., CMS books in mathematics, Springer, Cham.
^o The most basic qualification condition (see Cor. 27.6(b) in [R1]) is that the intersection between the relative interiors of the domains of f and g is nonempty.

Remark. We make the standing assumption that f and g satisfy a qualification condition (i.e., we assume it holds without mentioning it every time).

Operator splitting

Our problem is, again, that of finding a zero of the subgradients, namely:

$$0 \in (\partial f + \partial g)(x),$$

which we know are maximal monotone under convexity and lower semicontinuity of f and g .

Idea: Transform this problem into a fixed-point iteration with operators constructed from ∂f , ∂g :

$$0 \in (\partial f + \partial g)(x)$$

$$0 \in (\rho \partial f + \rho \partial g)(x)$$

$$\exists u \in \partial f(x) \text{ such that } -u \in \partial g(x) \xrightarrow{\rho > 0} -\rho u \in \rho \partial g(x)$$

$$\text{for } z := x + \rho u \Rightarrow x - z \in \rho \partial g(x),$$

$$2x - z \in (I + \rho \partial g)(x),$$

$$J_{\rho \partial g}(2x - z) = x,$$

$$\text{since } z := x + \rho u \in x + \rho \partial f(x) \Rightarrow x = J_{\rho \partial f}(z) \Rightarrow J_{\rho \partial g}(2J_{\rho \partial f}(z) - z) = J_{\rho \partial f}(z),$$

$$2J_{\rho \partial g}((2J_{\rho \partial f} - I)(z)) = (2J_{\rho \partial f} - I)(z) + z,$$

$$\underbrace{(2J_{\rho \partial g} - I)}_{R_{\rho \partial g}} \underbrace{(2J_{\rho \partial f} - I)}_{R_{\rho \partial f}}(z) = z.$$

Where $R_{\rho \partial h} := 2J_{\rho \partial h} - I$ is known as the **reflected resolvent operator**.

Proposition 16.44 in [R1]

Let $g : \mathbb{R}^p \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper, lsc, and convex. Then, the reflected resolvent operator becomes the so-called reflective operator defined by

$$R_{\rho \partial g}(x) = \text{refl}_g^\rho := 2 \text{prox}_g^\rho(x) - x.$$

[R1] Bauschke HH and Combettes PL (2017). Convex analysis and monotone operator theory in Hilbert spaces, second ed., CMS books in mathematics, Springer, Cham.

Proposition 4.4 and 12.28 in [R1]

Let $g : \mathbb{R}^p \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper, lsc, and convex. Since the proximal operator is firmly nonexpansive, then the reflective operator refl_g^ρ is nonexpansive, i.e.,

$$\|\text{refl}_g^\rho(x) - \text{refl}_g^\rho(y)\| \leq \|x - y\|, \quad \forall x, y \in \mathbb{R}^p.$$

[R1] Bauschke HH and Combettes PL (2017). Convex analysis and monotone operator theory in Hilbert spaces, second ed., CMS books in mathematics, Springer, Cham.

The Peaceman-Rachford algorithm via the reflected resolvent operator

Idea: Design an operator T whose fixed points z^* can be projected into the minimizer x^* of f .

A possible solution can be found by means of the (averaged) Peaceman-Rachford operator

$$T(z) = (1 - \alpha)z + \alpha \operatorname{refl}_g^\rho(\operatorname{refl}_f^\rho(z)), \quad \rho > 0,$$

which is such that

$$0 \in (\partial f + \partial g)(x^*) \iff x^* = \operatorname{prox}_f^\rho(z^*) \text{ with } T(z^*) = z^*.$$

i.e., minimizers of $f + g$ are proximals of fixed points of T , as we wanted.

Does it work? (Is the operator contractive or averaged?)

The Peaceman-Rachford algorithm: the main result for convex functions

Theorem (Relaxed Peaceman-Rachford Algorithm)

Consider the unconstrained composite optimization

$$\min_{x \in \mathbb{R}^p} f(x) + g(x),$$

and assume that f and g are proper, lsc, **convex**, that $f + g$ admits at least one minimizer. Then, the iteration

$$\begin{aligned} z(k+1) &= (1 - \alpha)z(k) + \alpha \operatorname{refl}_g^\rho(2x(k) - z(k)), \\ x(k+1) &= \operatorname{prox}_f^\rho(z(k+1)) \end{aligned}$$

converges (**not necessarily at a linear rate**) to a fixed point z^* such that $x^* = \operatorname{prox}_f^\rho(z^*)$ is a minimizer of $f + g$ for any $\alpha \in (0, 1)$ and for any $\rho > 0$, i.e.,

$$\lim_{k \rightarrow \infty} z(k) = z^*, \quad \lim_{k \rightarrow \infty} x(k) = x^* \in \mathcal{X}^*, \quad \rho > 0, \quad \alpha \in (0, 1),$$

where $\mathcal{X}^*$ is the set of minimizers of $f + g$.

The Peaceman-Rachford algorithm: proof for convex functions

Proof (via Krasnoselskij, Theorem 5.15 in [R1]): Let $R := \text{refl}_g^\rho \circ \text{refl}_f^\rho$ and recall the relaxed PRS algorithm:

$$z(k+1) = (1-\alpha)z(k) + \alpha R(z(k)), \quad x(k+1) = \text{prox}_f^\rho(z(k)).$$

Step 1: fixed points and minimizers. By Fermat's rule and the qualification condition, for any minimizer $x^* \in \mathcal{X}^*$ it holds $0 \in (\partial f + \partial g)(x^*)$, and vice versa. Equivalently, there exists $u^* \in \partial f(x^*)$ such that $-u^* \in \partial g(x^*)$. **Recall that:** $x = \text{prox}_f^\rho(x + \rho u) \iff u \in \partial f(x)$.

Minimizers yield fixed points. Let $x^* \in \mathcal{X}^*$ and set $z^* := x^* + \rho u^*$. Then

$$\text{prox}_f^\rho(z^*) = \text{prox}_f^\rho(x^* + \rho u^*) = x^*, \quad \Rightarrow \quad \text{refl}_f^\rho(z^*) = 2 \text{prox}_f^\rho(z^*) - z^* = 2x^* - z^* = x^* - \rho u^*,$$

and thus

$$R(z^*) = \text{refl}_g^\rho(\text{refl}_f^\rho(z^*)) = \text{refl}_g^\rho(x^* - \rho u^*) = 2 \text{prox}_g^\rho(x^* - \rho u^*) - (x^* - \rho u^*) = x^* + \rho u^* = z^*.$$

Fixed points yield minimizers. Let $z^* \in \text{fix}(R)$ and set $x^* = \text{prox}_f^\rho(z^*)$ and $u^* = (z^* - x^*)/\rho$. Then

$$\begin{aligned} x^* = \text{prox}_f^\rho(z^*) &= \text{prox}_f^\rho(x^* + \rho u^*) && \iff u^* \in \partial f(x^*), \\ x^* = z^* - \rho u^* &= 2 \text{prox}_g^\rho(x^* - \rho u^*) - x^* && \iff -u^* \in \partial g(x^*), \end{aligned}$$

Thus, $0 = u^* - u^* \in (\partial f + \partial g)(x^*)$, i.e., $x^* \in \mathcal{X}^*$.

Step 2: R is nonexpansive. Since refl_f^ρ and refl_g^ρ are nonexpansive, their composition is nonexpansive.

Step 3: convergence $z(k) \rightarrow z^* \in \text{fix}(R)$ follows by Thm 5.15 in [R1]. Thus, $x(k) \rightarrow \text{prox}_f^\rho(z^*) \in \mathcal{X}^*$.

Additional assumptions: Assume that f is m -strongly convex and L -smooth.

Theorem 1 in [R9]

Let $f : \mathbb{R}^P \rightarrow \mathbb{R}$ be a proper, lsc, m -strongly convex and L -smooth function. Then, for any $\rho > 0$, the reflective operator refl_f^ρ is contractive with contraction factor $\gamma \in [0, 1)$ given by

$$\gamma := \max \left\{ \frac{\rho L - 1}{\rho L + 1}, \frac{1 - \rho m}{1 + \rho m} \right\}, \quad \left\| \text{refl}_f^\rho(x) - \text{refl}_f^\rho(y) \right\| \leq \gamma \|x - y\|, \quad \forall x, y \in \mathbb{R}^P.$$

[R9] Giselsson P and Boyd S (2017). Linear Convergence and Metric Selection for Douglas-Rachford Splitting and ADMM, IEEE Transactions on Automatic Control.

The Peaceman-Rachford algorithm: the main result for strongly convex and smooth functions

Theorem (Relaxed Peaceman-Rachford Algorithm)

Consider the unconstrained composite optimization

$$\min_{x \in \mathbb{R}^p} f(x) + g(x),$$

and assume that f and g are proper, lsc, convex. **Additionally, assume that f is m -strongly convex and L -smooth.** Then, the iteration

$$\begin{aligned} z(k+1) &= (1 - \alpha)z(k) + \alpha \operatorname{refl}_g^\rho(2x(k) - z(k)), \\ x(k+1) &= \operatorname{prox}_f^\rho(z(k+1)) \end{aligned}$$

converges **at a linear rate** to the unique fixed point z^* such that $x^* = \operatorname{prox}_f^\rho(z^*)$ is the unique minimizer of $f + g$ for any $\alpha \in (0, 1)$ and for any $\rho > 0$, i.e.,

$$\|z(k) - z^*\| \leq \ell^k \|z(0) - z^*\|, \quad \ell = 1 - \alpha(1 - \gamma), \quad \gamma = \max \left\{ \frac{\rho L - 1}{\rho L + 1}, \frac{1 - \rho m}{1 + \rho m} \right\} < 1.$$

Consequently, also $x(k)$ converges linearly to x^* . Moreover, the best value of this contraction bound is achieved for $\rho^* = 1/\sqrt{mL}$ and $\alpha^* = 1$, and it is equal to $\ell^* = (\sqrt{L/m} - 1)/(\sqrt{L/m} + 1)$ (Prop. 3 [R9]).

[R9] Giselsson P and Boyd S (2017). Linear Convergence and Metric Selection for Douglas-Rachford Splitting and ADMM, IEEE Transactions on Automatic Control.

PRS vs PG: bound on the convergence rate comparison

Consider the composite problem $\min_x f(x) + g(x)$ with f being m -strongly convex and L -smooth ($\kappa := L/m$).

Proximal Gradient (optimal step $\rho^* = 2/(m + L)$):

$$\ell_{\text{PG}}^* = \frac{L - m}{L + m} = \frac{\kappa - 1}{\kappa + 1}.$$

Peaceman–Rachford (optimal $\rho^* = 1/\sqrt{mL}$, $\alpha = 1$):

$$\ell_{\text{PRS}}^* = \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}.$$

Claim: $\ell_{\text{PRS}}^* < \ell_{\text{PG}}^*$ for all $\kappa > 1$.

Proof. Let $t = \sqrt{\kappa} > 1$. Then $\ell_{\text{PG}}^* = (t^2 - 1)/(t^2 + 1)$ and $\ell_{\text{PRS}}^* = (t - 1)/(t + 1)$.

$$\ell_{\text{PG}}^* - \ell_{\text{PRS}}^* = \frac{t^2 - 1}{t^2 + 1} - \frac{t - 1}{t + 1} = \frac{(t^2 - 1)(t + 1) - (t - 1)(t^2 + 1)}{(t^2 + 1)(t + 1)} = \frac{(t - 1)[(t + 1)^2 - (t^2 + 1)]}{(t^2 + 1)(t + 1)} = \frac{2t(t - 1)}{(t^2 + 1)(t + 1)} > 0.$$

Thus PRS achieves a **strictly better bound on the contraction rate** than PG.

Relaxed PRS vs PG: a numerical example

Minimize $f(x) + \lambda \|x\|_1$ with $\lambda = 1.5$:

$$f(x) = \frac{1}{2} \sum_i \sigma_i (x_i - c_i)^2$$

$\sigma = (2, 2, 1, 1)$, $c = (5, 3, 1, 0.5)$, $\kappa = 2$.

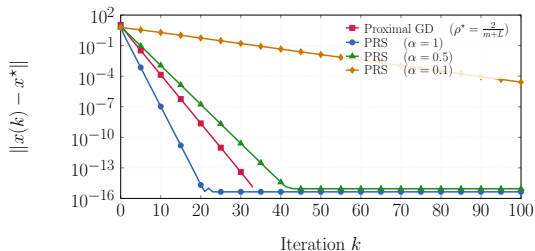
Proximal GD (optimal $\rho^* = \frac{2}{m+L} = \frac{2}{3}$):

$$\ell_{\text{PG}}^* = \frac{\kappa-1}{\kappa+1} = \frac{1}{3} \approx 0.33.$$

PRS (optimal $\rho^* = \frac{1}{\sqrt{mL}} = \frac{1}{\sqrt{2}}$), rate $\ell = |1-\alpha| + \alpha\gamma$:

- $\alpha = 1$: $\ell = \frac{\sqrt{2}-1}{\sqrt{2}+1} \approx 0.17$
- $\alpha = 0.5$: $\ell \approx 0.59$ (slower than PG)
- $\alpha = 0.1$: $\ell \approx 0.92$ (slower than PG)

All start from $x(0) = (10, 8, 6, 4)$.



Outline

- 1 Recap
- 2 The Peaceman-Rachford splitting (PRS) algorithm
- 3 The distributed PRS algorithm in OMAS (Open ADMM)**
- 4 The federated PRS algorithm in OMAS (Open FedPRS)
- 5 A numerical example on the EMNIST dataset
- 6 Conclusions, future directions, and debate

Consider a network of $n \in \mathbb{N}$ agents connected according to a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Each agent $i \in \mathcal{V}$ has a local cost function $f_i : \mathbb{R}^p \rightarrow \mathbb{R}$ and the network wants to minimize the sum of all cost functions:

$$\min_{x \in \mathbb{R}^p} \sum_{i \in \mathcal{V}} f_i(x).$$

We have seen that this problem can be equivalently written in a distributed way as follows:

$$\begin{aligned} \min_{x_1, \dots, x_n \in \mathbb{R}^p} \quad & \sum_{i \in \mathcal{V}} f_i(x_i) \\ \text{s.t.} \quad & x_i = x_j \quad \forall (i, j) \in \mathcal{E} \end{aligned}$$

where $x_i \in \mathbb{R}^p$ is the local copy of the i -th agent's estimate. Alternatively, instead of enforcing directly $x_i = x_j$, we can enforce the consensus constraints on auxiliary variable y_{ij} defined for each edge $(i, j) \in \mathcal{E}$, namely,

$$\begin{aligned} \min_{x_1, \dots, x_n \in \mathbb{R}^p} \quad & \sum_{i \in \mathcal{V}} f_i(x_i) \\ \text{s.t.} \quad & x_i = y_{ij}, \quad x_j = y_{ji}, \quad y_{ij} = y_{ji}, \quad \forall (i, j) \in \mathcal{E}. \end{aligned}$$

We can rewrite the consensus optimization problem as a separable composite problem with linear constraints as follows

$$\begin{aligned} \min_{x \in \mathbb{R}^{pn}, y \in \mathbb{R}^{p|\mathcal{E}|}} \quad & f(x) + g(y), \\ \text{s.t} \quad & Ax + By = c. \end{aligned}$$

where $x = [x_1^\top, \dots, x_n^\top]^\top$ is the vector stacking all the agents states, and where

$$f(x) = \sum_{i \in \mathcal{V}} f_i(x_i), \quad g(y) = \iota_{\mathcal{Y}}(y) := \begin{cases} 0 & \text{if } y \in \mathcal{Y} = \{y \in \mathbb{R}^{p|\mathcal{E}|} : y_{ij} = y_{ij}, \forall (i, j) \in \mathcal{E}\}, \\ +\infty & \text{otherwise,} \end{cases}$$

and where

$$A = \text{block-diag}(\mathbf{1}_{|\mathcal{N}_i|}), \quad B = -I, \quad c = 0.$$

Problem 1: This is a **constrained** problem, so the PRS algorithm cannot be applied to it directly.

Solution 1: We can apply the PRS algorithm to the **unconstrained dual problem** instead, yielding the so called **Relaxed Alternating Direction Method of Multipliers (R-ADMM)**. Indeed, by Proposition 19.21 in [R1], the dual problem is given by

$$\min_{w \in \mathbb{R}^q} \underbrace{f^*(A^\top w)}_{\hat{f}(w)} + \underbrace{g^*(C^\top w) - c^\top w}_{\hat{g}(w)}.$$

where $h^*(v) = \sup_{x \in \mathbb{R}^q} \{v^\top x - h(x)\}$ is the convex conjugate function of a function h .

Problem 2: The dual objective $\hat{f}(w) = f(Aw)$ is **not strongly convex** even if f is strongly convex, because A has surely not full row rank.

Consequence: We cannot enforce (global) linear convergence by simply asking that f is strongly convex (as we did so far). Indeed, strong convexity of f (with additional assumption) only allows to prove local convergence (close to the optimum).

Hint: The fundamental issue is the rank deficiency of A , which implies **the set of fixed points is not a singleton**, and thus the resulting operator cannot be contractive by construction.

Solution: We will enforce **paracontractivity** instead of contractivity.

Remark. The particular separable structure of $f(x)$, $g(y)$, A , B allows to specialize the PRS algorithm so that the update of the $x(k)$ and $z(k)$ variables are amenable of distributed implementations, as formalized next.

Distributed R-ADMM in Appendix A-A of [R10]

Consider the following optimization problem

$$\begin{aligned} \min_{x \in \mathbb{R}^{pn}, y \in \mathbb{R}^{p|\mathcal{E}|}} \quad & f(x) + g(y), \\ \text{s.t} \quad & Ax + By = c. \end{aligned}$$

where $x = [x_1^\top, \dots, x_n^\top]^\top$ is the vector stacking all the agents states, and where

$$f(x) = \sum_{i \in \mathcal{V}} f_i(x_i), \quad g(y) = \iota_{\mathcal{Y}}(y) := \begin{cases} 0 & \text{if } y \in \mathcal{Y} = \{y \in \mathbb{R}^{p|\mathcal{E}|} : y_{ij} = y_{ji}, \forall (i, j) \in \mathcal{E}\}, \\ +\infty & \text{otherwise,} \end{cases}$$

and where $A = \text{block-diag}(\mathbf{1}_{|\mathcal{N}_i|})$, $B = -I$, $c = 0$. Then, the PRS operator applied to the dual yields the so-called R-ADMM algorithm, ruled by the following set of updates (with $\eta_i = |\mathcal{N}_i|$)

$$\begin{aligned} z_{ij}(k+1) &= (1 - \alpha)z_{ij}(k) + \alpha(2\rho x_j(k) - z_{ji}(k)), \quad \forall i \in \mathcal{V}, j \in \mathcal{N}_i, \\ x_i(k+1) &= \text{prox}_{f_i}^{1/\rho\eta_i} \left(\frac{1}{\rho\eta_i} \sum_{j \in \mathcal{N}_i} z_{ij}(k+1) \right) \quad \forall i \in \mathcal{V}. \end{aligned}$$

[R10] Bastianello N, Carli R, Schenato L, Toderascato M (2021). Asynchronous Distributed Optimization Over Lossy Networks via Relaxed ADMM: Stability and Linear Convergence, IEEE Transactions on Automatic Control.

Theorem (Distributed Peaceman-Rachford Algorithm)

Consider the distributed optimization problem

$$\begin{aligned} \min_{\{x_i\}_{i \in \mathcal{V}}} \quad & \sum_{i \in \mathcal{V}} f_i(x_i) =: f(x) \\ \text{s.t.} \quad & x_i = x_j, \quad \forall (i, j) \in \mathcal{E} \end{aligned}$$

Assume that the graph is connected and that the functions $f_i : \mathbb{R}^p \rightarrow \mathbb{R}$ are proper, lsc, **convex**. Then, the distributed R-ADMM iteration

$$z_{ij}(k+1) = (1 - \alpha)z_{ij}(k) + \alpha(2\rho x_j(k) - z_{ji}(k)), \quad \forall i \in \mathcal{V}, j \in \mathcal{N}_i,$$

$$x_i(k+1) = \text{prox}_{f_i}^{1/\rho\eta_i} \left(\frac{1}{\rho\eta_i} \sum_{j \in \mathcal{N}_i} z_{ij}(k+1) \right) \quad \forall i \in \mathcal{V}.$$

converges (**not necessarily at a linear rate**) to a fixed point z^* such that $x^* = \text{prox}_{f_i}^{1/\rho\eta_i}(\frac{1}{\rho\eta_i} \sum_{j \in \mathcal{N}_i} z_{ij}^*)$ is a minimizer of $f + g$ for any $\alpha \in (0, 1)$ and for any $\rho > 0$, i.e.,

$$\lim_{k \rightarrow \infty} z(k) = z^*, \quad \lim_{k \rightarrow \infty} x(k) = \mathbf{1} \otimes x^* \text{ with } x^* \in \mathcal{X}^*, \quad \rho > 0, \alpha \in (0, 1),$$

where $\mathcal{X}^*$ is the set of minimizers of $f + g$.

Additional assumption: Assume that f is such that its proximal prox_f is an affine function.

Eq. (5) in [R11]

An operator $T : \mathcal{X} \rightarrow \mathcal{Y}$ with $\mathcal{X} \subseteq \mathcal{Y} \subseteq \mathbb{R}^p$ is (globally) “ μ -metric subregular” if, for some $\mu > 0$ and for all $x \in \mathcal{X}$, it satisfies

$$d(x, \text{fix}(T)) \leq \mu \cdot d(x, T(x)).$$

[R11] Robinson SM (2009). Some continuity properties of polyhedral multifunctions, Mathematical Programming at Oberwolfach.

Special case of Proposition 1 and the discussion within Pages 6-7 in [R11]

An affine operator $T : x \mapsto Mx + q$ with $x \in \mathbb{R}^p$ is (globally) “ μ -metric subregular” where μ is the smallest strictly positive singular value of the matrix $I - M$, namely,

$$\mu = \sigma_{\min}^+(I - M) = \min_{i=1, \dots, p} \{\sigma_i(I - M) : \sigma_i(I - M) > 0\}.$$

[R1] Robinson SM (2009). Some continuity properties of polyhedral multifunctions, Mathematical Programming at Oberwolfach.

Proposition 1 in [R2]

If an operator $T : \mathcal{X} \rightarrow \mathcal{Y}$ with $\mathcal{X} \subseteq \mathcal{Y} \subseteq \mathbb{R}^p$ is:

- α -averaged, i.e., there is $\alpha \in (0, 1)$ such that $\|T(x) - T(y)\|^2 \leq \|x - y\|^2 - \frac{1-\alpha}{\alpha} \|(x - y) - (T(x) - T(y))\|^2$;
- μ -metric subregular, i.e., $d(x, \text{fix}(T)) \leq \mu \cdot d(x, T(x))$;

then, T is γ -paracontractive, i.e., $d(T(x), \text{fix}(T)) \leq \gamma \cdot d(x, \text{fix}(T))$, with contraction factor given by

$$\gamma = \sqrt{1 - \frac{1-\alpha}{\alpha\mu^2}}.$$

[R2] D. Deplano, N. Bastianello, M. Franceschelli, K.H. Johansson (2026). Optimization and learning in Open Multi-Agent Systems, IEEE Transactions on Automatic Control.

Proof. Set $y = \text{proj}(x, \text{fix}(T))$, such that $y = T(y)$, then:

$$\begin{aligned} d^2(T(x), \text{fix}(T)) &\leq \|T(x) - T(y)\|^2 \leq \|x - y\|^2 - \frac{1-\alpha}{\alpha} \|(x - y) - (T(x) - y)\|^2 \\ &= d^2(x, \text{fix}(T)) - \frac{1-\alpha}{\alpha} d^2(x, T(x)) \leq d^2(x, \text{fix}(T)) - \frac{1-\alpha}{\alpha\mu^2} d^2(x, \text{fix}(T)) \\ &= \left(1 - \frac{1-\alpha}{\alpha\mu^2}\right) d^2(x, \text{fix}(T)) \end{aligned}$$

Theorem (Distributed Peaceman-Rachford Algorithm)

Consider the distributed optimization problem

$$\begin{aligned} \min_{\{x_i\}_{i \in \mathcal{V}}} \quad & \sum_{i \in \mathcal{V}} f_i(x_i) =: f(x) \\ \text{s.t.} \quad & x_i = x_j, \quad \forall (i, j) \in \mathcal{E} \end{aligned}$$

Assume that the graph is connected and that the functions $f_i : \mathbb{R}^p \rightarrow \mathbb{R}$ are proper, lsc, convex. **Additionally, assume the proximal of each local cost f_i is affine.** Then, the distributed R-ADMM iteration

$$\begin{aligned} x_i(k+1) &= \text{prox}_{f_i}^{1/\rho\eta_i} \left(\frac{1}{\rho\eta_i} \sum_{j \in \mathcal{N}_i} z_{ij}(k) \right) & \forall i \in \mathcal{V} \\ z_{ij}(k+1) &= (1 - \alpha)z_{ij}(k) + \alpha(2\rho x_j(k+1) - z_{ji}(k)), & \forall i \in \mathcal{V}, j \in \mathcal{N}_i. \end{aligned}$$

converges (**at a linear rate**) to a fixed point z^* such that $x^* = \text{prox}_{f_i}^{1/\rho\eta_i}(\frac{1}{\rho\eta_i} \sum_{j \in \mathcal{N}_i} z_{ij}^*)$ is the unique minimizer of $f + g$ for any $\alpha \in (0, 1)$ and for any $\rho > 0$, i.e.,

$$\|z(k) - z^*\| \leq \gamma^k \|z(0) - z^*\|, \quad \text{for some } \gamma = \sqrt{1 - \frac{1 - \alpha}{\alpha\mu^2}} \in (0, 1),$$

where $\mu > 0$ is the metric subregularity constant of the resulting operator. Consequently, also $x(k)$ converge linearly to the consensus state $\mathbf{1}_{n_k} \otimes x^*$ on the unique minimizer x^* .

Distributed PRS with affine proximal: proof of linear convergence

Proof. Define the PRS operator $T(z) := (1 - \alpha)z + \alpha \operatorname{refl}_g^\rho(\operatorname{refl}_f^\rho(z))$.

Step 1: T is α -averaged. Since $\operatorname{refl}_f^\rho$ and $\operatorname{refl}_g^\rho$ are nonexpansive, their composition $R := \operatorname{refl}_g^\rho \circ \operatorname{refl}_f^\rho$ is nonexpansive, and $T = (1 - \alpha)I + \alpha R$ is α -averaged by definition.

Step 2: T is affine. Since prox_f is affine by assumption, $\operatorname{refl}_f^\rho = 2\operatorname{prox}_f^\rho - I$ is affine. Since $g = \iota_Y$ is the indicator of a polyhedral set, $\operatorname{prox}_g^\rho = \operatorname{proj}_Y$ is affine, and so $\operatorname{refl}_g^\rho$ is affine. The composition $R = \operatorname{refl}_g^\rho \circ \operatorname{refl}_f^\rho$ is affine, and therefore $T = (1 - \alpha)I + \alpha R$ is affine.

Step 3: T is metrically subregular. Since T is affine, T is globally μ -metrically subregular for some $\mu > 0$ [R11].

Step 4: T is paracontractive. By Proposition 1 in [R2], averaged + metrically subregular $\Rightarrow \gamma$ -paracontractive with $\gamma = \sqrt{1 - \frac{1-\alpha}{\alpha\mu^2}} < 1$.

Step 5: linear convergence. Since T is paracontractive, $d(z(k), \operatorname{fix}(T)) \leq \gamma^k d(z(0), \operatorname{fix}(T)) \rightarrow 0$ at a linear rate. Since $\operatorname{prox}_f^\rho$ is Lipschitz, $x(k) = \operatorname{prox}_f^\rho(z(k))$ also converges linearly to a minimizer as it was proved earlier.

DGD vs Distributed ADMM: non-strongly convex quadratic costs with affine proximal

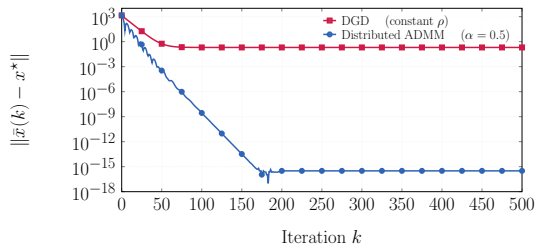
4 agents on a cycle graph in $\mathbb{R}^2$.

$$f_i(x) = (x - c_i)^\top Q_i (x - c_i), \quad Q_i \geq 0.$$

DGD (constant ρ): **sublinear** convergence (curve bends on log scale) + bias.

Distributed ADMM: still converges **linearly** to the **exact** x^* (straight line on log scale to machine precision) because prox_{f_i} is affine $\Rightarrow$ operator is affine $\Rightarrow$ metrically subregular $\Rightarrow$ paracontractive.

$x(0) = (1000, 1000)$ for all agents.



DGD vs Distributed ADMM: non-strongly convex quadratic costs with non-affine proximal

4 agents on a cycle graph in $\mathbb{R}^2$.

$$f_i(x) = \sum_j (x_j - c_{i,j})^4$$

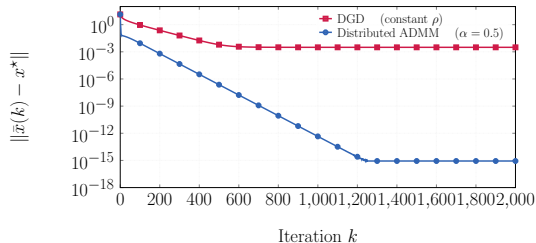
Smooth, convex, **not strongly convex**.

DGD (constant ρ): **sublinear** convergence (curve bends on log scale) + bias.

Distributed ADMM: still converges **linearly** to the **exact** x^* (straight line on log scale to machine precision).

Remark. The ADMM operator is not affine here, but it is smooth and **locally** metrically subregular near the fixed point.

$x(0) = (10, 10)$ for all agents.



Consider an **open network** of $n_k \in \mathbb{N}$ agents, with $n_k = |\mathcal{V}_k|$, interacting according to a **time-varying graph** $\mathcal{G}_k = (\mathcal{V}_k, \mathcal{E}_k)$, connected at every k . Each agent $i \in \mathcal{V}_k$ has a local cost function $f_i : \mathbb{R}^p \rightarrow \mathbb{R}$, and the goal is to minimize the sum of all local cost functions:

$$\begin{aligned} \min_{\{x_i\}_{i \in \mathcal{V}_k}} \quad & \sum_{i \in \mathcal{V}_k} f_i(x_i) \\ \text{s.t.} \quad & x_i = x_j, \quad \forall (i, j) \in \mathcal{E}_k \text{ (consensus constraints)} \end{aligned}$$

where $x_i \in \mathbb{R}^p$ is agent i 's local copy of the decision variable.

Since to execute the PRS algorithm the agents need to store and updated per-edge variables, we need to distinguish between remaining, arriving, and departing neighbors:

- **Remaining neighbors** $\mathcal{R}_{i,k+1} = \mathcal{N}_{i,k+1} \cap \mathcal{N}_{i,k}$: neighbors present both at time k and time $k+1$;
- **Arriving neighbors** $\mathcal{A}_{i,k+1} = \mathcal{N}_{i,k+1} \setminus \mathcal{N}_{i,k}$: neighbors present at time $k+1$ but not at time k ;
- **Departing neighbors** $\mathcal{D}_{i,k} = \mathcal{N}_{i,k} \setminus \mathcal{N}_{i,k+1}$: neighbors present at time k but not at time $k+1$.

We formalize our assumptions, which use the following notation for the set of solutions to the distributed optimization problem:

$$\mathcal{X}_k^* = \left\{ x_k^* \in \mathbb{R}^P : \sum_{i \in \mathcal{V}_k} f_i(x_k^*) = \min_{x \in \mathbb{R}^P} \sum_{i \in \mathcal{V}_k} f_i(x) \right\},$$

and for the minimizers of each local cost function,

$$\mathcal{X}_{i,k}^* = \left\{ x_{i,k}^* \in \mathbb{R}^P : f_i(x_{i,k}^*) = \min_{x \in \mathbb{R}^P} f_i(x) \right\}, \quad \forall i \in \mathcal{V}_k. \tag{2}$$

Main assumptions. For any $k \in \mathbb{N}$:

- (i) the local cost functions f_i are proper, lsc, and convex for all $i \in \mathcal{V}_k$, and $\mathcal{X}_{i,k}^*, \mathcal{X}_k^*$ are not empty;
- (ii) the distance between two consecutive global solutions $x_k^* \in \mathcal{X}_k^*$ and $x_{k+1}^* \in \mathcal{X}_{k+1}^*$ is upper bounded by a constant $\sigma \geq 0$;
- (iii) the distance between each local solution $x_{i,k}^* \in \mathcal{X}_{i,k}^*$ and the global solution $x_k^* \in \mathcal{X}_k^*$ is upper bounded by $\omega \geq 0$.

The distributed R-ADMM in open multi-agent systems (called **Open ADMM**) can be written as follows:

$$z_{ij}(k+1) = \begin{cases} (1-\alpha)z_{ij}(k) + \alpha(2\rho x_j(k) - z_{ji}(k)) & \text{if } i \in \mathcal{V}_{k+1}^R \text{ and } j \in \mathcal{R}_{i,k+1}, \\ z_{ij,k+1}^A & \text{if } i \in \mathcal{V}_{k+1}^A \text{ and } j \in \mathcal{A}_{i,k+1}, \end{cases}$$

$$x_i(k+1) = \text{prox}_{f_i}^{1/\rho\eta_i} \left(\frac{1}{\rho\eta_i} \sum_{j \in \mathcal{N}_{i,k+1}} z_{ij}(k+1) \right).$$

Design: Several choices are possible for initializing the variable $z_{ij,k+1}^A$ of arriving channels:

- Initialization to (penalized) local minimizer: $z_{ij,k+1}^A = \rho x_{i,k+1}^*$ (our choice);
- Blind initialization: $z_{ij,k+1}^A = \mathbf{0}$;
- Initialization to the average of neighboring states $z_{ij,k+1}^A = \frac{1}{|\mathcal{N}_{i,k}|} \sum_{j \in \mathcal{N}_{i,k}} x_j(k)$ (this requires that the arriving agents can communicate with neighbors before joining the network);
- Initialization within a bounded set $z_{ij,k+1}^A \in \overline{\mathcal{X}}$ (e.g., in the case of the projected DGD);
- Other choices.

[R2] D. Deplano, N. Bastianello, M. Franceschelli, K.H. Johansson (2026). Optimization and learning in Open Multi-Agent Systems, IEEE Transactions on Automatic Control.

The standard iteration of Open ADMM ruled by the operator $F_k = [\cdots, F_{ij,k}, \cdots]^\top$ is

$$z(k+1) = F_k(z(k)) = [(1-\alpha)I - \alpha P_k] z(k) + 2\alpha\rho P_k A_k x(k)$$

$$x(k+1) = P_k(z(k)) = \text{prox}_f^{1/\rho\eta_k}(D_k A_k^\top z(k+1))$$

where:

- the operator $\text{prox}_f^{1/\rho\eta_k} : \mathbb{R}^{\mathcal{J}_k} \rightarrow \mathbb{R}^{\mathcal{J}_k}$ with $\mathcal{J}_k = \{(i, \ell) : i \in \mathcal{V}_k \text{ and } \ell = 1, \dots, p\}$ applies block-wise the proximal of the local costs f_i ;
- the matrix $A_k = \Lambda \otimes I_p \in \{0, 1\}^{\mathcal{I}_k \times \mathcal{J}_k}$ is given by $\Lambda \in \{0, 1\}^{\mathcal{E}_k \times \mathcal{V}_k}$ defined block-wise for $i \in \mathcal{V}_k$ and $j \in \mathcal{N}_{i,k}$ by the matrices $\Lambda_{ij} = [e_i, e_j]^\top$ for $j > i$ where $e_\ell \in \{0, 1\}^{\mathcal{V}_k}$ denotes a canonical vector;
- the matrix $D_k \in \mathbb{R}^{\mathcal{J}_k \times \mathcal{J}_k}$ is given by $D_k = \text{blkdiag}\{(\rho\eta_{i,k})^{-1} I_p\}_{i=1}^n$;
- the matrix $P_k \in \{0, 1\}^{\mathcal{I}_k \times \mathcal{I}_k}$ is a squared block-diagonal matrix given by $P_k = I_{\xi_k/2} \otimes (\Pi \otimes I_p)$ with $\Pi = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

Special case of Theorem 1 in [R2]

Consider an OMAS and assume that

- (a) the standard iteration is paracontractive with $\gamma \in (0, 1)$;
- (b) the TSI has bounded variation $B \geq 0$, i.e., $d_{\text{SH}}(\hat{\mathcal{Z}}_{k+1}, \hat{\mathcal{Z}}_k) / \sqrt{2p|\mathcal{E}_{k+1}|} \leq B$;
- (c) the arrival process is bounded with $H \geq 0$, i.e., $\exists \hat{z}_k \in \text{proj}(z_k^{\text{R}}, \hat{\mathcal{Z}}_k)$ s.t. $d(z_k^{\text{A}}, \hat{z}_k) / \sqrt{2p|\mathcal{E}_k|} \leq H$;
- (d) the departure process is bounded with $\beta \in (\gamma, 1)$, i.e., $\sqrt{|\mathcal{E}_{k+1}|} / \sqrt{|\mathcal{E}_k|} \geq \beta$.

Then, the open sequence $\{z(k)\}_{k \in \mathbb{N}}$ converges linearly with rate $\theta = \gamma/\beta \in (0, 1)$ to the TPI within a radius

$$R = \frac{B + H}{1 - \theta}, \quad (3)$$

according to the following pointwise upper bound

$$\frac{d(z(k), \hat{z}_k)}{\sqrt{p|\mathcal{V}_k|}} \leq \theta^k \frac{d(z(0), \hat{z}_0)}{\sqrt{p|\mathcal{V}_0|}} + \frac{1 - \theta^k}{1 - \theta} (B + H).$$

Consequently, also $x(k)$ converges linearly to the consensus state $\mathbf{1}_{n_k} \otimes x^*$ on the unique minimizer x^* .

[R2] D. Deplano, N. Bastianello, M. Franceschelli, K.H. Johansson (2026). Optimization and learning in Open Multi-Agent Systems, IEEE Transactions on Automatic Control.

Proof of condition (a) (the standard iteration is paracontractive with $\gamma \in (0, 1)$).

The standard iteration is simply the PRS algorithm, which we have already proved to be paracontractive with

constant $\gamma_k = \sqrt{1 - \frac{1-\alpha}{\alpha\mu_k^2}}.$

Remark. Obtaining a uniform upper bound $\gamma < 1$ on the contraction factor γ_k would require to know the metric subregularity constant μ_k , which depends on:

- the local costs f_i for $i \in \mathcal{V}_k$ (through the specific form of the proximal associated with);
- the graph $\mathcal{G}_k$ (through the matrices P_k and A_k).

Conclusion: We assume that there exists such a uniform upper bound $\gamma \geq \gamma_k$ for all $k \in \mathbb{N}$.

Proof of condition (b) – Part 1 (the TPI has bounded variation $B \geq 0$, i.e., $d_{\text{SH}}(\hat{\mathcal{Z}}_k, \hat{\mathcal{Z}}_{k-1})/\sqrt{2p|\mathcal{E}_k|} \leq B$).

The standard iteration of Open ADMM ruled by the operator $F_k = [\cdots, F_{ij,k}, \cdots]^\top$ is

$$\begin{aligned} z(k+1) &= F_k(z(k)) = [(1-\alpha)I - \alpha P_k] z(k) + 2\alpha\rho P_k A_k x(k) \\ x(k+1) &= P_k(z(k)) = \text{prox}_f^{1/\rho\eta_k}(D_k A_k^\top z(k+1)) \end{aligned}$$

At a fixed point $\hat{z}_k$ of the standard iteration it holds $(\mathbf{1}_{n_k} \otimes x_k^*) = \text{prox}_f^{1/\rho\eta_k}(D_k A_k^\top \hat{z}_k)$, yielding

$$(I + P_k)\hat{z}_k = 2\rho P_k A_k (\mathbf{1}_{n_k} \otimes x_k^*).$$

Since $I + P_k = I_{\xi_k/2} \otimes (J \otimes I_p)$ with $J = \Pi + I_2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and pseudoinverse $J^\dagger = \frac{1}{4}J$, the projection of any $z \in \mathbb{R}^{\mathcal{I}_k}$ onto $\hat{\mathcal{Z}}_k$ decomposes into edge blocks $\hat{z}_{ij,k} \in \mathbb{R}^{2p}$ as

$$\hat{z}_{ij,k} = z_{ij} - \frac{1}{2}(J \otimes I_p)(z_{ij} - \rho(\mathbf{1}_2 \otimes x_k^*)).$$

Similarly, $\hat{z}_{ij,k-1} = z_{ij} - \frac{1}{2}(J \otimes I_p)(z_{ij} - \rho(\mathbf{1}_2 \otimes x_{k-1}^*))$, hence,

$$\hat{z}_{ij,k} - \hat{z}_{ij,k-1} = \frac{\rho}{2}(J \otimes I_p)(\mathbf{1}_2 \otimes (x_k^* - x_{k-1}^*)).$$

Proof of condition (b) – Part 2.

$$\begin{aligned}
 d(\text{proj}(z, \hat{Z}_k), \text{proj}(z, \hat{Z}_{k-1})) &= \sqrt{\sum_{i \in \mathcal{V}_k^R} \sum_{i < j \in \mathcal{R}_{i,k}} \|\hat{z}_{ij,k} - \hat{z}_{ij,k-1}\|^2} \\
 &= \sqrt{\sum_{i \in \mathcal{V}_k^R} \sum_{i < j \in \mathcal{R}_{i,k}} \frac{\rho^2}{4} \|(J \otimes I_p)(\mathbf{1}_2 \otimes (x_k^* - x_{k-1}^*))\|^2} \\
 &\stackrel{(i)}{\leq} \sqrt{\sum_{i \in \mathcal{V}_k^R} \sum_{i < j \in \mathcal{R}_{i,k}} \frac{\rho^2}{4} \|J \otimes I_p\|^2 \|\mathbf{1}_2 \otimes (x_k^* - x_{k-1}^*)\|^2} \\
 &\stackrel{(ii)}{\leq} \sqrt{\sum_{i \in \mathcal{V}_k^R} \sum_{i < j \in \mathcal{R}_{i,k}} \frac{\rho^2 \sigma^2}{2} \|J\|^2} \stackrel{(iii)}{\leq} \sqrt{\sum_{i \in \mathcal{V}_k^R} \sum_{i < j \in \mathcal{R}_{i,k}} \frac{\rho^2 \sigma^2}{2} \|J\|_1 \|J\|_\infty} \\
 &\stackrel{(iv)}{=} \sqrt{\sum_{i \in \mathcal{V}_k^R} \sum_{i < j \in \mathcal{R}_{i,k}} 2\rho^2 \sigma^2} = \rho\sigma \sqrt{\sum_{i \in \mathcal{V}_k^R} \sum_{i < j \in \mathcal{R}_{i,k}} 2} \stackrel{(v)}{\leq} \rho\sigma \sqrt{|\mathcal{R}_k|},
 \end{aligned}$$

where (i) sub-multiplicativity; (ii) $\|J \otimes I_p\| = \|J\| \|I_p\| = \|J\|$ and $\|\mathbf{1}_2 \otimes (x_k^* - x_{k-1}^*)\| = \sqrt{2} \|x_k^* - x_{k-1}^*\| \leq \sqrt{2}\sigma$; (iii) Riesz-Thorin; (iv) $\|J\|_1 = \|J\|_\infty = 2$; (v) $|\mathcal{R}_k|$ is the number of remaining channels. Taking the sup over z yields $B = \rho\sigma$.

Proof of condition (c) (the arrival process is bounded with $H \geq 0$, i.e., $d(z_k^A, \hat{z}_k)/\sqrt{2p|\mathcal{E}_k|} \leq H$).

Under initialization $z_{ij}(k) = \rho x_{i,k}^*$ for arriving channels, decomposing into edge blocks:

$$\begin{aligned}
 d(z_k^A, \hat{z}_k) &= \sqrt{\sum_{i \in \mathcal{V}_k} \sum_{i < j \in \mathcal{A}_{i,k}} \left\| \rho \begin{bmatrix} x_{i,k}^* \\ x_{j,k}^* \end{bmatrix} - \begin{bmatrix} \hat{z}_{ij,k} \\ \hat{z}_{ji,k} \end{bmatrix} \right\|^2} \\
 &\stackrel{(i)}{\leq} \sqrt{\sum_{i \in \mathcal{V}_k} \sum_{i < j \in \mathcal{A}_{i,k}} \left\| \frac{\rho}{2} (J \otimes I_p) \left(\begin{bmatrix} x_{i,k}^* \\ x_{j,k}^* \end{bmatrix} - \mathbf{1}_2 \otimes x_k^* \right) \right\|^2} \\
 &\stackrel{(ii)}{\leq} \sqrt{\frac{\rho^2}{4} \|J \otimes I_p\|^2 \sum_{i \in \mathcal{V}_k} \sum_{i < j \in \mathcal{A}_{i,k}} \left\| \begin{bmatrix} x_{i,k}^* - x_k^* \\ x_{j,k}^* - x_k^* \end{bmatrix} \right\|^2} \\
 &\stackrel{(iii)}{\leq} \sqrt{\rho^2 \sum_{i \in \mathcal{V}_k} \sum_{i < j \in \mathcal{A}_{i,k}} 2\omega^2} = \rho\omega \sqrt{\sum_{i \in \mathcal{V}_k} \sum_{i < j \in \mathcal{A}_{i,k}} 2} \stackrel{(iv)}{\leq} \rho\omega \sqrt{|\mathcal{A}_k|},
 \end{aligned}$$

where (i) uses the projection formula from the proof of condition (b); (ii) sub-multiplicativity; (iii) $\|J\| \leq 2$ as before and $\|x_{i,k}^* - x_k^*\| \leq \omega$ by assumption; (iv) $|\mathcal{A}_k|$ is the number of arriving channels. Therefore, $H = \rho\omega$.

Theorem (Open ADMM Algorithm in OMAS)

Consider the distributed optimization problem over an open network

$$\begin{aligned} \min_{\{x_i\}_{i \in \mathcal{V}_k}} \quad & \sum_{i \in \mathcal{V}_k} f_i(x_i) \\ \text{s.t.} \quad & x_i = x_j, \quad \forall (i, j) \in \mathcal{E}_k \end{aligned}$$

Assume that each graph $\mathcal{G}_k$ is connected, that $f_i : \mathbb{R}^p \rightarrow \mathbb{R}$ are proper, lsc, and convex, and that the standard iteration is uniformly paracontractive with $\gamma \in (0, 1)$. Consider the Open ADMM iteration

$$\begin{aligned} z_{ij}(k+1) &= \begin{cases} (1-\alpha)z_{ij}(k) + \alpha(2\rho x_j(k) - z_{ji}(k)) & i \in \mathcal{V}_{k+1}^R, j \in \mathcal{R}_{i,k+1}, \\ \rho x_{i,k+1}^* & i \in \mathcal{V}_{k+1}^A, j \in \mathcal{A}_{i,k+1}, \end{cases} \\ x_i(k+1) &= \text{prox}_{f_i}^{1/\rho\eta_i} \left(\frac{1}{\rho\eta_i} \sum_{j \in \mathcal{N}_{i,k+1}} z_{ij}(k+1) \right). \end{aligned}$$

Assume also that $\sqrt{|\mathcal{E}_{k+1}|}/\sqrt{|\mathcal{E}_k|} \geq \beta$ with $\beta \in (\gamma, 1)$ for all $k \in \mathbb{N}$. Then, the open sequence $\{z(k)\}_{k \in \mathbb{N}}$ converges at a linear rate $\theta = \gamma/\beta$ to the TPI within radius R ,

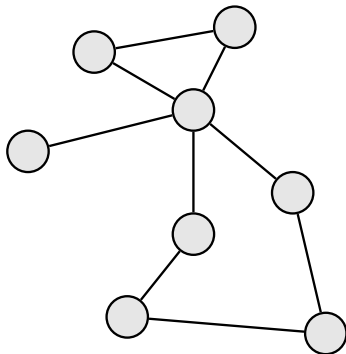
$$\frac{d(z(k), \hat{z}_k)}{\sqrt{2p|\mathcal{E}_k|}} \leq \theta^k \frac{d(z(0), \hat{z}_0)}{\sqrt{2p|\mathcal{E}_0|}} + \underbrace{\frac{B+H}{1-\theta}}_R, \quad \text{with} \quad B = \rho\sigma, \quad H = \rho\omega.$$

Consequently, $x(k)$ converges linearly to the consensus state $\mathbf{1}_{n_k} \otimes x_k^*$.

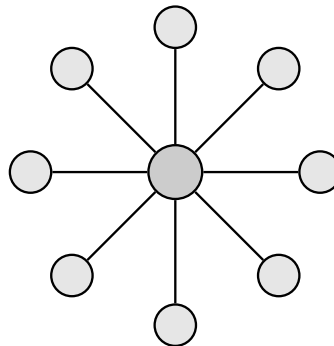
Outline

- 1 Recap
- 2 The Peaceman-Rachford splitting (PRS) algorithm
- 3 The distributed PRS algorithm in OMAS (Open ADMM)
- 4 The federated PRS algorithm in OMAS (Open FedPRS)**
- 5 A numerical example on the EMNIST dataset
- 6 Conclusions, future directions, and debate

Distributed VS Federated



peer-to-peer



federated

Consider a network of $n \in \mathbb{N}$ agents connected according to a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Each agent $i \in \mathcal{V}$ has a local cost function $f_i : \mathbb{R}^p \rightarrow \mathbb{R}$ and the network wants to minimize the sum of all cost functions:

$$\min_{x \in \mathbb{R}^p} \sum_{i \in \mathcal{V}} f_i(x).$$

We have seen that this problem can be equivalently written in a distributed way as follows:

$$\begin{aligned} \min_{x_1, \dots, x_n \in \mathbb{R}^p} \quad & \sum_{i \in \mathcal{V}} f_i(x_i) \\ \text{s.t.} \quad & x_i = x_j \quad \forall (i, j) \in \mathcal{E} \end{aligned}$$

where $x_i \in \mathbb{R}^p$ is the local copy of the i -th agent's estimate. **Alternatively, instead of enforcing directly $x_i = x_j$, we can enforce the consensus constraints on auxiliary variable z_i defined for each node $i \in \mathcal{V}$, namely,**

$$\begin{aligned} \min_{x_1, \dots, x_n \in \mathbb{R}^p} \quad & \sum_{i \in \mathcal{V}} f_i(x_i) \\ \text{s.t.} \quad & x_i = z_i, \quad \forall i \in \mathcal{V} \\ & z_i = z_j, \quad (i, j) \in \mathcal{E}. \end{aligned}$$

We can rewrite the consensus optimization problem as a separable composite problem with linear constraints as follows

$$\begin{aligned} \min_{x \in \mathbb{R}^{pn}, y \in \mathbb{R}^{p|\mathcal{E}|}} \quad & f(x) + g(y), \\ \text{s.t.} \quad & Ax + Bz = c. \end{aligned}$$

where $x = [x_1^\top, \dots, x_n^\top]^\top$ is the vector stacking all the agents states, and where

$$f(x) = \sum_{i \in \mathcal{V}} f_i(x_i), \quad g(y) = \iota_{\mathcal{Z}}(y) := \begin{cases} 0 & \text{if } y \in \mathcal{Y} = \{y \in \mathbb{R}^{p|\mathcal{E}|} : y_i = y_j, \forall (i, j) \in \mathcal{E}\}, \\ +\infty & \text{otherwise,} \end{cases}$$

and where

$$A = I, \quad B = -I, \quad c = 0.$$

Therefore, in principle, we could apply the (non-relaxed) PRS algorithm to solve the problem:

$$\begin{aligned}\xi(k+1) &= \text{refl}_g^\rho(2x(k) - \xi(k)), \\ x(k+1) &= \text{prox}_f^\rho(\xi(k+1))\end{aligned}$$

Apply the change of variable $z(k) = 2x(k) - \xi(k)$ and obtain

$$\begin{aligned}\xi(k+1) &= 2 \text{prox}_g^\rho(z(k)) - z(k), \\ x(k+1) &= \text{prox}_f^\rho(\xi(k+1))\end{aligned}$$

Let $z_c(k) = \text{prox}_g^\rho(z(k))$ be the computation carried out by the coordination. Since g is the indicator function over the consensus set $\mathcal{C}$, such that $\text{prox}_g^\rho(x) = \text{proj}_{\mathcal{C}}(x) = \frac{1}{n_k} (\mathbf{1}_{n_k} \mathbf{1}_{n_k}^\top \otimes I_p) x$, then $\xi(k+1) = 2z_c(k) - z(k)$, yielding

$$\begin{aligned}x(k+1) &= \text{prox}_f^\rho(2z_c(k) - z(k)), \\ z(k+1) &= z(k) + 2(x(k+1) - z_c(k)), \\ z_c(k+1) &= \frac{1}{n_k} (\mathbf{1}_{n_k} \mathbf{1}_{n_k}^\top \otimes I_p) z(k+1),\end{aligned}$$

which, component-wise, can be written as follows

$$\begin{aligned}x_i(k+1) &= \text{prox}_{f_i}^\rho(2z_c(k) - z_i(k)) \\ z_i(k+1) &= z_i(k) + 2(x_i(k+1) - z_c(k))\end{aligned}$$

The federated PRS in open multi-agent systems (called **Open FedPRS**) can be written as follows:

$$\begin{aligned}
 x_i(k+1) &= \text{prox}_{f_i}^{\rho} (2z_c(k) - z_i(k)) && \text{if } i \in \mathcal{V}_{k+1}^R \\
 z_i(k+1) &= \begin{cases} z_i(k) + 2(x_i(k+1) - z_c(k)) & \text{if } i \in \mathcal{V}_{k+1}^R, \\ x_{i,k+1}^* & \text{if } i \in \mathcal{V}_{k+1}^A, \end{cases}
 \end{aligned}$$

where the local coordinator computes and transmit at each step the following aggregate information

$$z_c(k) = \frac{1}{|\mathcal{V}_k|} \sum_{i \in \mathcal{V}_k} z_i(k)$$

Letting $J_k = |\mathcal{V}_k|^{-1} (\mathbf{1}_{|\mathcal{V}_k|} \mathbf{1}_{|\mathcal{V}_k|}^{\top} \otimes I_p)$, the standard iteration of Open FedPRS ruled by the operator $F_k = [\dots, F_{i,k}, \dots]^{\top}$ is

$$\begin{aligned}
 z(k+1) &= F_k(z(k)) = z(k) + 2(x(k+1) - z_c(k)) = 2x(k+1) - (2z_c(k) - z(k)) \\
 &= 2 \text{prox}_{f_i}^{\rho} (2z_c(k) - z_i(k)) - (2z_c(k) - z(k)) \\
 &= 2 \text{refl}_{f_i}^{\rho} (2z_c(k) - z_i(k)) = 2 \text{refl}_{f_i}^{\rho} (2J_k z(k) - z_i(k)) \\
 &= 2 \text{refl}_{f_i}^{\rho} ((2J_k - I)z_i(k))
 \end{aligned}$$

Special case of Theorem 1 in [R2]

Consider an OMAS and assume that

- (a) the standard iteration is paracontractive with $\gamma \in (0, 1)$;
- (b) the TSI has bounded variation $B \geq 0$, i.e., $d_{\text{SH}}(\hat{\mathcal{Z}}_{k+1}, \hat{\mathcal{Z}}_k) / \sqrt{p|\mathcal{V}_{k+1}|} \leq B$;
- (c) the arrival process is bounded with $H \geq 0$, i.e., $\exists \hat{z}_k \in \text{proj}(z_k^R, \hat{\mathcal{Z}}_k)$ s.t. $d(z_k^A, \hat{z}_k) / \sqrt{p|\mathcal{V}_{k+1}|} \leq H$;
- (d) the departure process is bounded with $\beta \in (\gamma, 1)$, i.e., $\sqrt{p|\mathcal{V}_{k+1}|} / \sqrt{p|\mathcal{V}_k|} \geq \beta$.

Then, the open sequence $\{z(k)\}_{k \in \mathbb{N}}$ converges linearly with rate $\theta = \gamma/\beta \in (0, 1)$ to the TPI within a radius

$$R = \frac{B + H}{1 - \theta},$$

according to the following pointwise upper bound

$$\frac{d(z(k), \hat{z}_k)}{\sqrt{p|\mathcal{V}_k|}} \leq \theta^k \frac{d(z(0), \hat{z}_0)}{\sqrt{p|\mathcal{V}_0|}} + \frac{1 - \theta^k}{1 - \theta} (B + H).$$

Consequently, also $x(k)$ converge linearly to the consensus state $\mathbf{1}_{n_k} \otimes x^*$ on the unique minimizer x^* .

[R2] D. Deplano, N. Bastianello, M. Franceschelli, K.H. Johansson (2026). Optimization and learning in Open Multi-Agent Systems, IEEE Transactions on Automatic Control.

Proof of condition (a) (the standard iteration is contractive with $\ell \in (0, 1)$).

To satisfy this condition is sufficient to show that the operator F_k is contractive.

- 1 F_k is the composition of the operator refl_f^ρ with a linear operator $\mathcal{J}_{k-1} : x \mapsto (2J_{k-1} - I)x$;
- 2 The operator $\mathcal{J}_{k-1}$ is nonexpansive because all eigenvalues of $2J_{k-1} - I$ are within the unit circle (see Lemma 3 in [R12]): the matrix $|\mathcal{V}_k|^{-1}(\mathbf{1}_{|\mathcal{V}_k|}\mathbf{1}_{|\mathcal{V}_k|}^\top \otimes I_p)$ has all eigenvalues equal to 0 but one that is equal to 1 and the Kronecker product with the identity matrix I_p leaves the eigenvalues unchanged;
- 3 The reflected operator refl_f^ρ is ℓ -contractive when f is μ -strongly convex and L -smooth, and the contraction rate $\ell := \max\left\{\frac{\rho L - 1}{\rho L + 1}, \frac{1 - \rho m}{1 + \rho m}\right\}$ holds by Theorem 1 in [R9];
- 4 Since F_k is the composition of a nonexpansive operator with a ℓ -contractive operator, then F_k is ℓ -contractive.

Remark. The constant ℓ already holds at any time $k \in \mathbb{N}$ and it does not depend on the network topology due to the presence of the coordinator.

Conclusion: The contraction factor ℓ holds uniformly.

[R9] Giselsson P and Boyd S (2017). Linear Convergence and Metric Selection for Douglas-Rachford Splitting and ADMM, IEEE Transactions on Automatic Control.

[R12] Belgioioso G, Fabiani F, Blanchini F, and Grammatico S (2018). On the Convergence of Discrete-Time Linear Systems: A Linear Time-Varying Mann Iteration Converges IFF Its Operator Is Strictly Pseudocontractive, IEEE Control Systems Letters.

Proof of condition (b) (the TPI has bounded variation $B \geq 0$, i.e., $d_{\text{SH}}(\hat{z}_k, \hat{z}_{k-1})/\sqrt{p|\mathcal{V}_k|} \leq B$).

Step 1: characterize the fixed points. At a fixed point $\hat{z}_k$ of the standard iteration, the coordinator's output satisfies $J_k \hat{z}_k = \mathbf{1}_{n_k} \otimes x_k^*$ and, by the update rule, it holds $J_k \hat{z}_k = \text{prox}_f^\rho((2J_k - I)\hat{z}_k)$. Since the proximal operator is characterized by $a = \text{prox}_f^\rho(b) \Leftrightarrow b - a \in \rho \partial f(a)$, then:

$$\begin{aligned} (2J_k - I)\hat{z}_k - J_k \hat{z}_k &= \rho \nabla f_k(J_k \hat{z}_k) \\ \hat{z}_k &= J_k \hat{z}_k - \rho \nabla f_k(J_k \hat{z}_k) \\ \hat{z}_k &= \mathbf{1}_{n_k} \otimes x_k^* - \rho \nabla f_k(\mathbf{1}_{n_k} \otimes x_k^*). \end{aligned}$$

Component-wise: $\hat{z}_{i,k} = x_k^* - \rho \nabla f_i(x_k^*)$, $\forall i \in \mathcal{V}_k$.

Step 2: bound for remaining agents. For $i \in \mathcal{V}_k^{\text{R}}$:

$$\begin{aligned} \|\hat{z}_{i,k} - \hat{z}_{i,k-1}\| &= \|x_k^* - x_{k-1}^* - \rho(\nabla f_i(x_k^*) - \nabla f_i(x_{k-1}^*))\| \\ &\stackrel{(i)}{\leq} \|x_k^* - x_{k-1}^*\| + \rho \|\nabla f_i(x_k^*) - \nabla f_i(x_{k-1}^*)\| \\ &\stackrel{(ii)}{\leq} (1 + \rho L) \|x_k^* - x_{k-1}^*\| \stackrel{(iii)}{\leq} (1 + \rho L) \sigma, \end{aligned}$$

where (i) triangle inequality; (ii) L -smoothness of f_i ; (iii) assumption (ii).

Step 3: normalize. Since $|\mathcal{V}_k^{\text{R}}| \leq n_k$:

$$\frac{d(\hat{z}_k, \hat{z}_{k-1})}{\sqrt{p|\mathcal{V}_k|}} \leq \frac{1}{\sqrt{p|\mathcal{V}_k|}} \sqrt{\sum_{i \in \mathcal{V}_k} \|\hat{z}_{i,k} - \hat{z}_{i,k-1}\|^2} \leq \frac{(1 + \rho L)\sigma}{\sqrt{p}} =: B.$$

Proof of condition (c) (the arrival process is bounded with $H \geq 0$, i.e., $d(z_k^A, \hat{z}_k)/\sqrt{p|\mathcal{V}_k|} \leq H$).

Open FedPRS initializes $z_{i,k} = x_{i,k}^*$ for arriving agents $i \in \mathcal{V}_k^A$. From the fixed-point characterization:
 $\hat{z}_{i,k} = x_k^* - \rho \nabla f_i(x_k^*)$.

Step 1: bound per-agent distance. Since $\nabla f_i(x_{i,k}^*) = 0$ (local minimizer), for $i \in \mathcal{V}_k^A$:

$$\begin{aligned}
 \|z_{i,k} - \hat{z}_{i,k}\| &= \|x_{i,k}^* - x_k^* + \rho \nabla f_i(x_k^*)\| \\
 &\stackrel{(i)}{=} \|x_{i,k}^* - x_k^* - \rho(\nabla f_i(x_{i,k}^*) - \nabla f_i(x_k^*))\| \\
 &\stackrel{(ii)}{\leq} \|x_{i,k}^* - x_k^*\| + \rho \|\nabla f_i(x_{i,k}^*) - \nabla f_i(x_k^*)\| \\
 &\stackrel{(iii)}{\leq} (1 + \rho L) \|x_{i,k}^* - x_k^*\| \stackrel{(iv)}{\leq} (1 + \rho L) \omega,
 \end{aligned}$$

where (i) $\nabla f_i(x_{i,k}^*) = 0$; (ii) triangle inequality; (iii) L -smoothness of f_i ; (iv) assumption (iii).

Step 2: normalize. Since $|\mathcal{V}_k^A| \leq n_k$:

$$\frac{d(z_k^A, \hat{z}_k)}{\sqrt{p|\mathcal{V}_k|}} \leq \frac{(1 + \rho L) \omega}{\sqrt{p|\mathcal{V}_k|}} \sqrt{|\mathcal{V}_k^A|} \leq \frac{(1 + \rho L) \omega}{\sqrt{p}} =: H.$$

Theorem (Open FedPRS Algorithm in OMAS)

Consider the distributed optimization problem over an open network

min_{\{x_i\}_{i \in \mathcal{V}_k}} \sum_{i \in \mathcal{V}_k} f_i(x_i)

s.t. x_i = x_j, \quad \forall (i, j) \in \mathcal{E}_k

Assume that each $f_i : \mathbb{R}^p \rightarrow \mathbb{R}$ is proper, lsc, **m -strongly convex**, and L -smooth. Consider the Open FedPRS iteration

$$x_i(k+1) = \text{prox}_{f_i}^\rho(2z_c(k) - z_i(k)), \qquad i \in \mathcal{V}_{k+1}^R,$$
$$z_i(k+1) = \begin{cases} z_i(k) + 2(x_i(k+1) - z_c(k)) & i \in \mathcal{V}_{k+1}^R, \\ x_{i,k+1}^* & i \in \mathcal{V}_{k+1}^A, \end{cases}$$
$$z_c(k+1) = \frac{1}{|\mathcal{V}_{k+1}|} \sum_{i \in \mathcal{V}_{k+1}} z_i(k+1).$$

Assume also the departure process is bounded with $\beta \in (\gamma, 1)$ for all $k \in \mathbb{N}$, where $\gamma = \{(\rho L - 1)(\rho L + 1), (1 - \rho m)(1 + \rho m)\} < 1$. Then, the open sequence $\{z(k)\}_{k \in \mathbb{N}}$ converges **at a linear rate** $\theta = \gamma/\beta$ to the TPI within radius R ,

$$\frac{d(z(k), \hat{z}_k)}{\sqrt{pn_k}} \leq \theta^k \frac{d(z(0), \hat{z}_0)}{\sqrt{pn_0}} + \underbrace{\frac{B + H}{1 - \theta}}_R, \quad \text{with} \quad \begin{aligned} B &= \frac{(1 + \rho L)\sigma}{\sqrt{p}}, \\ H &= \frac{(1 + \rho L)\omega}{\sqrt{p}}. \end{aligned}$$

Consequently, $x(k)$ converges linearly to the consensus state $\mathbf{1}_{n_k} \otimes x_k^*$.

[R13] D. Deplano, N. Bastianello, M. Franceschelli, K.H. Johansson (2026). Federated Learning in Open Multi-Agent Systems via Peaceman-Rachford Splitting, IFAC WC.

Outline

- 1 Recap
- 2 The Peaceman-Rachford splitting (PRS) algorithm
- 3 The distributed PRS algorithm in OMAS (Open ADMM)
- 4 The federated PRS algorithm in OMAS (Open FedPRS)
- 5 A numerical example on the EMNIST dataset**
- 6 Conclusions, future directions, and debate

Distributed binary classification

Dataset. The EMNIST dataset contains 70 000+ handwritten digit images from 3 400+ writers. Each image is a 28×28 grayscale scan, stored as a vector $a_{ij} \in [-1, 1]^p$ with $p = 784$.

Agent structure. Each agent $i \in \mathcal{V}_k$ holds the data of one writer:

$$\{(a_{ij}, b_{ij})\}_{j=1}^{n_i}, \quad b_{ij} = \begin{cases} +1 & \text{if digit} = \delta, \\ -1 & \text{otherwise.} \end{cases}$$



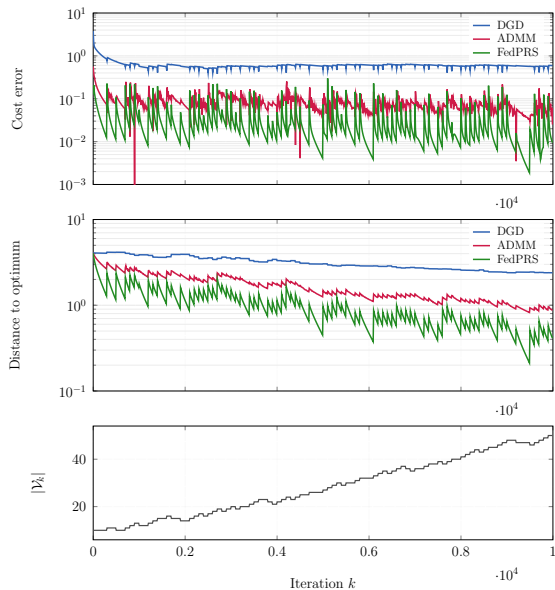
28×28 digit example

Stack rows: $A_i \in [-1, 1]^{n_i \times p}$, $b_i \in \{-1, +1\}^{n_i}$.

Goal. The agents collaboratively minimize the global penalized least-squares cost

$$f_k(x) := \sum_{i \in \mathcal{V}_k} f_i(x), \quad f_i(x) = \frac{1}{2n_i} \|A_i x - b_i\|^2 + \frac{\lambda}{2} \|x\|^2,$$

obtaining a linear classifier $x_k^* \in \mathbb{R}^p$ without sharing raw data.

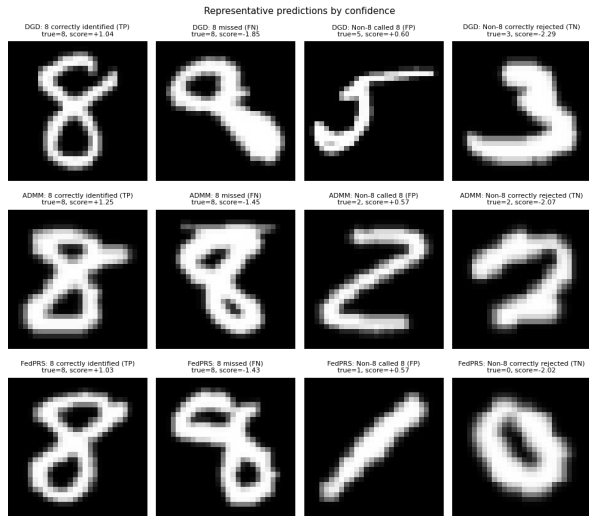


Classification accuracy (final agents)

Model evaluated on the joint dataset of all agents present at $k = T$.

Metric	Open DGD	Open ADMM	Open FedPRS
Overall accuracy	0.9295	0.9422	0.9426
Balanced accuracy	0.7566	0.8012	0.8031
Sensitivity (digit 8)	0.5292	0.6158	0.6196
Specificity (non-8)	0.9841	0.9866	0.9866

Example predictions



Outline

- 1 Recap
- 2 The Peaceman-Rachford splitting (PRS) algorithm
- 3 The distributed PRS algorithm in OMAS (Open ADMM)
- 4 The federated PRS algorithm in OMAS (Open FedPRS)
- 5 A numerical example on the EMNIST dataset
- 6 Conclusions, future directions, and debate

Q.A.



UNIVERSITÀ DEGLI STUDI
DI **CAGLIARI**



Introduction to Open Multi-Agent Systems – Part IV

Splitting algorithms via operator theory

Thank you for your attention!

Diego Deplano

Email: diego.deplano@unica.it

Webpage: <https://diegodeplano.github.io/>