Federated Learning in Open Multi-Agent Systems via Peaceman-Rachford Splitting ⋆

Diego Deplano * **Nicola Bastianello** ** **Mauro Franceschelli** *
Karl H. Johansson **

* *DIEE, University of Cagliari, 09123 Cagliari, Italy (e-mail:* {diego.deplano, mauro.franceschelli}@unica.it*).*
** *School of Electrical Engineering and Computer Science and Digital Futures, KTH Royal Institute of Technology, Stockholm, Sweden (e-mail:* {nicolba,kallej}@kth.se.*)*

Abstract: Modern AI systems are built on networks of agents that acquire data, perform local computations, and communicate with neighbors to cooperatively address optimization and learning tasks. This paper introduces *Open FedPRS*, a novel federated algorithm to address a broad class of these problems in *open networks*, where the number of participating agents may vary due to several factors, such as autonomous decisions, heterogeneous resource availability, or failures. Extending the current literature, the convergence analysis of the proposed algorithm is based on the *Theory of Open Operators*, which allows one to prove (1) linear convergence and (2) bounded asymptotic error of the distance between the trained model and the optimal model, rather than exploiting the commonly employed regret-based metrics that only describe cumulative performance over a finite-time horizon. As an illustrative example, the proposed algorithm is used to solve logistic learning problems.

Keywords: Federated learning, open mult-agent systems, decentralized learning, Peaceman-Rachford, operator theory.

1. INTRODUCTION

Federated Learning (FL) has revolutionized distributed machine learning by enabling multiple clients (e.g., devices, organizations or, in general, agents) to collaboratively train a shared model under the coordination of a central server. In a typical FL system, agents train local models based on their private data and share only the model updates with the server, thus preserving data privacy by design. The server then aggregates these updates to form a global model, which is redistributed to the agents for subsequent training iterations. The federated setting mitigates risks of centralized data breaches, and addresses challenges like regulatory compliance and communication efficiency. A common assumption is that the set of participating agents remains fixed throughout the training process (Omori and Kashima, 2023; Dal Fabbro et al., 2023; Chellapandi et al., 2023; Öksüz et al., 2024). However, modern applications in areas such as robotics (Yuan et al., 2024), healthcare (Arunan et al., 2023), finance (Zheng et al., 2023), mobile computing (Perifanis et al., 2023) involve scenarios in which agents may drop out of the training process due to, for instance, sudden unavailability or loss of interest; at the same time, it is often desirable to allow new agents to join in real-time without requiring a re-initialization of the learning process. An emerging framework to model and study stability of networks where agents can join or leave the network at any time due to resource constraints, network variability, or strategic considerations is that of *open multi-agent systems* (OMAS). The dynamic nature of agents participation introduces significant challenges for ensuring convergence and stability of the global model.

In the last few years, different algorithms have been proposed to solve optimization and learning problems over open networks of agents in fully distributed (Hsieh et al., 2021, 2022; Hayashi, 2023; Sawamura et al., 2024; Hendrickx and Rabbat, 2020; Deplano et al., 2026) and federated settings (Gu et al., 2021; Sun et al., 2023; Han et al., 2023; Wang and Xu, 2023; Chang et al., 2024; Sun et al., 2024). In the former setting, assuming that all cost functions are static, smooth, and strongly convex – with minimizers confined within a predetermined ball – Hendrickx et al. (Hendrickx and Rabbat, 2020) demonstrated that the Decentralized Gradient Descent (DGD) algorithm remains stable even when departing agents are instantly replaced by new ones. By relaxing the assumptions from

⋆ *Diego Deplano is the corresponding author.*

The work of Diego Deplano was supported by the project e.INS-Ecosystem of Innovation for Next Generation Sardinia (cod. ECS 00000038) funded by the Italian Ministry for Research and Education (MUR) under the National Recovery and Resilience Plan (NRRP) - MISSION 4 COMPONENT 2, "From research to business" INVESTMENT 1.5, "Creation and strengthening of Ecosystems of innovation" and construction of "Territorial R& D Leaders".
The work of N. Bastianello and K. H. Johansson was partially supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101070162, and partially by the Swedish Research Council Distinguished Professor Grant 2017-01078 Knut and Alice Wallenberg Foundation Wallenberg Scholar Grant.

© 2026 the authors. This work has been accepted to IFAC for publication under a Creative Commons Licence CC-BY-NC-ND.

strong convexity to convexity and from smoothness to Lipschitz continuity, Hsieh et al. (2021) developed a dual averaging method that guarantees sublinear convergence of the running loss, provided that the ratio of the quadratic mean to the average number of active agents is bounded, thereby accommodating scenarios of unbounded network growth. Under the more stringent condition that the network size is capped, Hayashi (2023) introduced a subgradient approach which also achieves sublinear convergence of the running loss while taking into account time-varying local loss functions. However, the works (Hsieh et al., 2021; Hayashi, 2023) require the network to be (jointly) vertex-connected and evaluate performance through a regret-based metric. In contrast, by extending distributed ADMM (Bastianello et al., 2025) to the open scenario, we have proposed and characterized *Open ADMM* in (Deplano et al., 2026) which only requires that the network is connected and focus on quantifying the precise deviation of the local estimates from the global optimum; this is done by establishing general convergence results exploiting the newly developed "Theory of Open Operators". In federated settings, all works (Sun et al., 2023; Han et al., 2023; Wang and Xu, 2023; Chang et al., 2024; Sun et al., 2024) considered strongly convex and smooth problems, except for (Gu et al., 2021) which focused on acceleration methods under arbitrary *dropouts*, i.e., temporary unavailability of an agent, thus making the set of agents static and not fully open. The same restrictive scenario is also considered by (Wang and Xu, 2023) and (Sun et al., 2024): Sun et al. introduce a method that combats agent dropouts by mimicking central updates (Sun et al., 2024), while Wang and Xu propose friend model substitution to maintain model consistency when certain agents temporarily become unavailable (Wang and Xu, 2023). On the other hand, the authors in (Han et al., 2023) do not consider the possibility of new agents joining the network, but they only examine the effects of agent *defections*, i.e., permanent leaving of agents that have reached a good local performance and, and propose measures to mitigate their impact (Han et al., 2023). The only works that consider an open scenario in the most general sense are (Chang et al., 2024) and (Sun et al., 2023): Chang et al. tackle dynamic agent arrival and departure by designing mechanisms for rapid adaptation via initial model construction (Chang et al., 2024); Sun et al. provide a general stability analysis that bounds the norm of local models in open federated learning systems with application to DGD and local ADAM (Sun et al., 2023).

This paper proposes a novel algorithm for federated optimization and learning over open networks, which we call Open FedPRS as it is based on the Peaceman-Rachford Splitting (PRS) (Ryu and Boyd, 2016, Section 7.3). The **main contributions** and **novelties with respect to the state-of-the-art** are:

(i) Open FedPRS enjoys a finite asymptotic bound on the distance of the local models from the optimal model in the general scenario of open networks;
(ii) Open FedPRS enjoys a linear convergence rate while reaching the bounded optimality region.

These results have been derived by exploiting the novel "Theory of Open Operators", recently proposed by us in (Deplano et al., 2026), whose roots lie in the pioneering work (Franceschelli and Frasca, 2020). Both the asymptotic bound on the training error and the convergence rate can be computed a-priori based on the algorithm parameters, assumptions on the network variability, and the class of optimization problems under study, i.e., with strongly convex and smooth costs as in (Sun et al., 2023, Assumption 1) and (Chang et al., 2024, Assumption 3). We note that neither (Sun et al., 2023) or (Chang et al., 2024) provide an asymptotic bound on the training error or describe the convergence rate for the proposed algorithms. Instead, (Sun et al., 2023) quantifies what the *radius of stability*, that is an upper bound on the squared norm of the vector local models (not their distance from the optimal model), while the authors in (Chang et al., 2024) quantify what they call *optimality gap*, that is the punctual distance between the local models and the optimal model.

2. PROBLEM OF INTEREST AND PROPOSED ALGORITHM: OPEN FEDPRS

We are interested in federated algorithms to be executed in open multi-agent systems (OMAS) to solve optimization and learning problems of the following kind

$$\min_{y\in\mathbb{R}^p} \sum_{i\in\mathcal{V}_k} f_i(y), \qquad \forall k \in \mathbb{N}, \tag{1}$$

where $p \in \mathbb{N}_+$, $f_i : \mathbb{R}^p \to \mathbb{R}$ denotes the local objective function of an agent $i \in \mathcal{V}_k$ in the network at time k, with $\mathcal{V}_k$ representing the time-varying set of agents. Most of the literature addresses this problem under the assumption that the set of agents in the network remains constant over time, i.e., $\mathcal{V} = \mathcal{V}_0 = \mathcal{V}_1 \cdots$, which enables the application of various results from operator theory to study the convergence of custom-designed algorithms. Instead, this work considers the problem without this assumption, allowing the agents to leave and join the network arbitrarily, which results in a time-varying number of agents $n_k = |\mathcal{V}_k| \in (0, \infty)$.

2.1 Preliminaries and working assumptions

Let $x_k^i \in \mathbb{R}^p$ denote the state of agent $i \in \mathcal{V}_k$ at time $k \in \mathbb{N}$ and let $\boldsymbol{x}_k = [{x_k^1}^\top, \cdots, {x_k^{n_k}}^\top]^\top \in \mathbb{R}^{pn_k}$ be the vector stacking all agents states. The sequence $\{\boldsymbol{x}_k \in \mathbb{R}^{pn_k} : k \in \mathbb{N}\}$ contains vectors of possibly different dimension, and thus is called an *open sequence*. We now introduce how the concept of convergence for open sequences in the light of (Deplano et al., 2026, Definition 2). As in this manuscript we measure distances between vectors by means of the standard Euclidean norm $d(x, y) := \|x - y\|_2$ then, comparing distances of points in spaces of different dimensions is unfair without normalization. Indeed, without normalization, open sequences containing vectors with an increasing number of components may diverge one from the other even if the distance between new components are bounded (cfr. (Deplano et al., 2026, Example 4). This is not always the case as, for instance, with the infinity norm (Deplano et al., 2023, 2024). Thus, our definition accounts for normalization of distances.

Definition 1. Two open sequences $\{\boldsymbol{x}_k \in \mathbb{R}^{m_k} : k \in \mathbb{N}\}$ and $\{\boldsymbol{y}_k \in \mathbb{R}^{m_k} : k \in \mathbb{N}\}$ are said to converge to one another within a radius $R \geq 0$ if

$$\limsup_{k\to\infty} \frac{d(\boldsymbol{x}_k, \boldsymbol{y}_k)}{\sqrt{m_k}} \leq R.$$

© 2026 the authors. Licensed under CC-BY-NC-ND.

It is useful defining the sets of remaining, arriving, departing agents as follows

$$\mathcal{V}_k^{\mathrm{R}} = \mathcal{V}_k \cap V_{k-1},\ \ \mathcal{V}_k^{\mathrm{A}} = \mathcal{V}_k \setminus \mathcal{V}_{k-1},\ \ \mathcal{V}_k^{\mathrm{D}} = \mathcal{V}_k \setminus \mathcal{V}_{k+1},$$

The only assumption we make on the network is about the rate at which agents leave the network. In plain words, this assumption ensures that the network retains at least a fixed proportion of its agents from one time step to the next, which prevents the network from shrinking too rapidly or even becoming empty.

Assumption 2. It holds $|\mathcal{V}_k| \geq \beta |\mathcal{V}_{k-1}|$ for some $0 < \gamma \leq \beta < 1$.

A network of agents can actually solve problem (1) in a federated setting under suitable assumptions of the local objective functions and the corresponding local and global solutions. We formalize next our set of assumptions.

Assumption 3. The problem in (1) is such that, $\forall k \in \mathbb{N}$:

(i) the local cost functions f_i are: proper and closed, i.e., the epigraph is non-empty and closed (Bauschke and Combettes, 2017, Def. 1.4); L-smooth with finite $L > 0$ (Taylor et al., 2017, Def. 1(a)); μ-strongly convex with $\mu \in (0, L)$ (Taylor et al., 2017, Def. 1(b));
(ii) the distance between the current global solution $y_k^\star$ and the previous global solution $y_{k-1}^\star$ is upper bounded by $\sigma \geq 0$;
(iii) the distance between any local solution $y^{i,\star}$ and the global solution $y_k^\star$ is upper bounded by $\omega \geq 0$.

Assumption 3(i) is quite common in the context of distributed and federated learning, as it ensures that each local cost function has exactly one unique minimizer. Assumption 3(ii) ensures an upper bound on the distance between any solution at time k and any solution at time $k-1$. Due the heterogeneity of the local objective functions, the local minimizers in $y^{i,\star}$ can be arbitrarily far away from the global minimizers in $y_k^\star$. Thus, Assumption 3(iii) ensures that such distance is bounded by a constant at any time k.

2.2 Open FedPRS and convergence analysis

To solve the problem in (1) in a federated way over an open network of agents, we propose the open version of the Peaceman-Rachford Splitting (PRS), which we call *Open FedPRS*, whose implementation is detailed in Algorithm 1 provided in the next page. Open FedPRS requires each agent $i \in \mathcal{V}_k$ to update/initialize two state variables $x_k^i, y_k^i \in \mathbb{R}^p$ and requires the coordinator to update only a state variable $x_k^c \in \mathbb{R}^p$. To prove the convergence of Open FedPRS, we resort to the main convergence result in (Deplano et al., 2026, Theorem 1) for open operators.

Theorem 4. Consider an OMAS executing Open FedPRS to solve a problem as in eq. (1) under Assumptions 3-2. Then, the open sequence $\{\boldsymbol{y}_k = [y_k^1, \cdots, y_k^{n_k}] \in \mathbb{R}^{pn_k} : k \in \mathbb{N}\}$ of agents estimates converges with linear rate

$$\theta = \frac{\gamma}{\beta}, \quad \text{where} \quad \gamma := \max\left\{\frac{\rho L - 1}{\rho L + 1}, \frac{1 - \rho\mu}{\rho\mu + 1}\right\}, \tag{2}$$

to consensus $\mathcal{C}_k^\star = \{\mathbb{1}_{n_k} \otimes y_k^\star \in \mathbb{R}^{pn_k} : k \in \mathbb{N}\}$ on the optimal solution within a distance bounded by R given next

$$\limsup_{k\to\infty} \frac{d(\boldsymbol{y}_k, \mathcal{C}_k^\star)}{\sqrt{pn_k}} = \frac{(1+\rho L)(\sigma+\omega)}{\sqrt{p}(1-\theta)} =: R \tag{3}$$

Algorithm 1 Open FedPRS

Input: The relaxation $\alpha \in (0,1)$ and the penalty $\rho > 0$.
Output: Each agent returns y_k^i that is an (approximated) solution to the optimization problem in (1).

for $k = 0, 1, 2, \ldots$**:**
  for each agent $i \in \mathcal{V}_k$:
    if $i \in \mathcal{V}_k^{\mathrm{A}}$ **is an arriving agent:**
      initializes the state variable at the local optimum

$$x_k^i = y^{i,\star} \tag{6a}$$

    else if $i \in \mathcal{V}_k^{\mathrm{R}}$ **is a remaining agent:**
      receives y_{k-1}^c from the coordinator
      updates the state variable according to

$$y_k^i = \mathrm{prox}_{f_i}^{\rho}\left(2y_{k-1}^c - x_{k-1}^i\right) \tag{6b}$$

$$x_k^i = x_{k-1}^i + 2(y_k^i - y_{k-1}^c) \tag{6c}$$

    end if
    transmits x_k^i to the coordinator
  end for
  The coordinator averages the information x_k^i received from all agents $i \in \mathcal{V}_k$ and sends it back to them:

$$y_k^c = \frac{1}{n_k} \sum_{i \in \mathcal{V}_k} x_k^i$$

end for

Proof. The first part of the proof consists in showing that there exists R such that

$$\limsup_{k\to\infty} \frac{d(\boldsymbol{x}_k, \hat{\boldsymbol{x}}_k)}{\sqrt{pn_k}} \leq R \tag{4}$$

i.e., the distance between the open sequence $\{\boldsymbol{x}_k : k \in \mathbb{N}\}$ generated by Open FedPRS converges toward the fixed point $\hat{\boldsymbol{x}}$ with linear rate $\theta \in (0,1)$ within a radius R, where θ, R are given in eqs. (2)-(3). We do so by exploiting [1] (Deplano et al., 2026, Theorem 1).

Let us first formulate the operator F_k ruling the so-called *standard dynamics*, that is the operator ruling the state transition under the assumption that the set of agents remains unchanged in one step (Deplano et al., 2024, 2026). By denoting $\boldsymbol{y}_k^c = \mathbb{1}_{n_k} \otimes y_k^c$, the distributed updates (6) can be written in compact form as follows (cfr. (Bastianello et al., 2024, Section IV.A)):

$$\boldsymbol{y}_k = \mathrm{prox}_{f_k}^{\rho}(2\boldsymbol{y}_{k-1}^c - \boldsymbol{x}_{k-1}), \tag{5a}$$

$$\boldsymbol{x}_k = \boldsymbol{x}_{k-1} + 2(\boldsymbol{y}_k - \boldsymbol{y}_{k-1}^c), \tag{5b}$$

$$\boldsymbol{y}_k^c = n_k^{-1}(\mathbb{1}_{n_k} \otimes \mathbb{1}_{n_k}^\top) \otimes I_p \boldsymbol{x}_k =: J_k \boldsymbol{x}_k \tag{5c}$$

By plugging $\boldsymbol{y}_k$, $\boldsymbol{y}_k^c$ into the iteration of $\boldsymbol{x}_k$ we get the operator $\mathsf{F}_k : \mathbb{R}^{n_{k-1}} \to \mathbb{R}^{n_{k-1}}$ ruling the standard dynamics

$$\mathsf{F}_k(\boldsymbol{x}_{k-1}) := \mathrm{refl}_{f_k}^{\rho}((2J_{k-1} - I)\boldsymbol{x}_{k-1}), \tag{7}$$

which is the Peaceman-Rachford Splitting applied to

$$\min f_k(\boldsymbol{y}) + \iota_{\mathcal{C}}(\boldsymbol{y}) \tag{8}$$

where $f_k(\boldsymbol{y}) = \sum_{i \in \mathcal{V}_k} f_i(y^i)$ and $\iota_{\mathcal{C}}$ is the indicator function of the consensus set. We can now proceed showing

[1] In the scenario of this manuscript, (Deplano et al., 2026, Theorem 1) differs from (Franceschelli and Frasca, 2020, Theorem 3.8) in the definition of *open distance* between vectors with different labeled components because it only considers common components, as previously proposed in (Deplano et al., 2024) and (Vizuete et al., 2024, Section 10.4.2)

© 2026 the authors. Licensed under CC-BY-NC-ND.

that all conditions of (Deplano et al., 2026, Theorem 1) hold by going through them one by one.

Condition (a) – To satisfy this condition is sufficient to show that the operator F_k is contractive. First of all, we note that F_k is the composition of the operator refl_f^ρ with a linear operator $\mathcal{J}_{k-1} : \boldsymbol{x} \mapsto (2J_{k-1} - I)\boldsymbol{x}$. The operator $\mathcal{J}_{k-1}$ is nonexpansive because all eigenvalues of $2J_{k-1} - I$ are within the unit circle (Belgioioso et al., 2018, Lemma 3): the matrix $n_{k-1}^{-1}\mathbb{1}_{k-1}\otimes\mathbb{1}_{k-1}^\top$ 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. The reflected operator refl_f^ρ is γ-contractive because f is μ-strongly convex and L-smooth by Assumption 3(i), and the contraction rate γ is given in (2) (Giselsson and Boyd, 2016, Theorem 1). Since F_k is the composition of a nonexpansive operator with a γ-contractive operator, than F_k is γ-contractive.

Condition (b) – To satisfy this condition is sufficient to show that the normalized *open distance* (Deplano et al., 2026, Definition 3) between (unique) fixed points $\hat{\boldsymbol{x}}_{k-1} = \mathsf{F}_k(\hat{\boldsymbol{x}}_{k-1})$ at consecutive steps is bounded by some constant. The *open distance* refers to the distance between vectors with different labeled components, i.e., the distance of the vectors obtained by only considering the remaining agents, namely:

$$d(\hat{\boldsymbol{x}}_k, \hat{\boldsymbol{x}}_{k-1}) := \sqrt{\textstyle\sum_{i\in\mathcal{V}_k^{\mathrm{R}}} \left\|\hat{x}_k^i - \hat{x}_{k-1}^i\right\|^2}).$$

According to eq. (7), fixed points are such that

$$J_{k-1}\hat{\boldsymbol{x}}_{k-1} = \text{prox}_{f_k}^\rho\left((2J_{k-1} - I)\hat{\boldsymbol{x}}_{k-1}\right) \tag{9}$$

Since F_k is the PRS operator, the coordinator's state variable $\hat{y}_k^c$ at a fixed point is equal to the unique solution $y_k^\star$ of problem. (1), cfr. (Bauschke and Combettes, 2017, Proposition 25.1), namely, $\hat{x}_k^c = y_k^\star$, and thus

$$\hat{\boldsymbol{x}}_k^c = J_k\hat{\boldsymbol{x}}_k = \mathbb{1}_{n_k} \otimes y_k^\star. \tag{10}$$

Since the proximal operator is characterized by the inclusion $a = \text{prox}_f(b) \Leftrightarrow b - a = \rho \cdot \nabla f(a)$, we conclude that

$$\begin{aligned}
(2J_{k-1}-I)\hat{\boldsymbol{x}}_{k-1} - J_{k-1}\hat{\boldsymbol{x}}_{k-1} &= \rho\cdot\nabla f_k(J_{k-1}\hat{\boldsymbol{x}}_{k-1})\\
J_{k-1}\hat{\boldsymbol{x}}_{k-1} - \hat{\boldsymbol{x}}_{k-1} &= \rho\cdot\nabla f_k(J_{k-1}\hat{\boldsymbol{x}}_{k-1})\\
\hat{\boldsymbol{x}}_{k-1} &= J_{k-1}\hat{\boldsymbol{x}}_{k-1} - \rho\cdot\nabla f_k(J_{k-1}\hat{\boldsymbol{x}}_{k-1})\\
\hat{\boldsymbol{x}}_{k-1} &= \mathbb{1}_{n_k}\otimes y_k^\star - \rho\cdot\nabla f_k(\mathbb{1}_{n_k}\otimes y_k^\star)
\end{aligned}$$

which, block-wise, corresponds to $\hat{x}_k^i = y_k^\star - \rho \cdot \nabla f_i(y_k^\star)$, $\forall i \in \mathcal{V}_k$. Therefore, for $i \in \mathcal{V}_k^{\mathrm{R}}$ it holds that:

$$\begin{aligned}
\left\|\hat{x}_k^i - \hat{x}_{k-1}^i\right\| &= \left\|y_k^\star - y_{k-1}^\star - \rho(\nabla f_i(y_k^\star) - \nabla f_i(y_{k-1}^\star))\right\|\\
&\overset{(i)}{\leq} \left\|y_k^\star - y_{k-1}^\star\right\| + \rho\left\|\nabla f_i(y_k^\star) - \nabla f_i(y_{k-1}^\star)\right\|\\
&\overset{(ii)}{\leq} \left\|y_k^\star - y_{k-1}^\star\right\| + \rho L\left\|y_k^\star - y_{k-1}^\star\right\|\\
&= (1+\rho L)\left\|y_k^\star - y_{k-1}^\star\right\| \overset{(iii)}{\leq} (1+\rho L)\sigma
\end{aligned}$$

where (i) holds by triangle inequality; (ii) holds because the local costs f_i are L-smoth according to Assumption 3(i); (iii) holds because the distance between two consecutive solutions are bounded by σ according to Assumption 3(i). Then, by the fact that $\left|\mathcal{V}_k^{\mathrm{R}}\right| \leq n_k$, we prove condition (b) as follows

$$\frac{d(\hat{\boldsymbol{x}}_k, \hat{\boldsymbol{x}}_{k-1})}{\sqrt{pn_k}} \leq \frac{(1+\rho L)\sigma}{\sqrt{pn_k}}\sqrt{\left|\mathcal{V}_k^{\mathrm{R}}\right|} \leq \frac{(1+\rho L)\sigma}{\sqrt{p}}.$$

Condition (c) – To satisfy this condition is sufficient to show that the normalized distance between arriving agents' state and fixed points is bounded. Letting $\boldsymbol{x}_k^{\mathrm{A}}$ be the vector stacking the state variables of all arriving agents, the open distance with respect to the fixed points corresponds to:

$$d(\boldsymbol{x}_k^{\mathrm{A}}, \hat{\boldsymbol{x}}_k) = \sqrt{\textstyle\sum_{i\in\mathcal{V}_k^{\mathrm{A}}} \left\|\hat{x}_k^i - \hat{x}_{k-1}^i\right\|}).$$

Since Open FedPRS requires that the state variables x_k^i of arriving agents $i \in \mathcal{V}_k^{\mathrm{A}}$ are initialized near the local optimum: $x_k^i = y^{i,\star} - \rho \cdot \nabla f_i(y^{i,\star})$, $\forall i \in \mathcal{V}_k^{\mathrm{A}}$, then, by similar steps as for condition (b), for $i \in \mathcal{V}_k^{\mathrm{A}}$ it holds

$$\begin{aligned}
\left\|x_k^i - \hat{x}_k^i\right\| &= \left\|y^{i,\star} - (y_k^\star - \rho\cdot\nabla f_i(y_k^\star))\right\|\\
&\overset{(i)}{=} \left\|y^{i,\star} - y_k^\star - \rho(\nabla f_i(y^{i,\star}) - \nabla f_i(y_k^\star))\right\|\\
&\leq \left\|y^{i,\star} - y_k^\star\right\| + \rho\left\|\nabla f_i(y^{i,\star}) - \nabla f_i(y_k^\star)\right\|\\
&\leq \left\|y^{i,\star} - y_k^\star\right\| + \rho L\left\|y^{i,\star} - y_k^\star\right\|\\
&= (1+\rho L)\left\|y^{i,\star} - y_k^\star\right\| \leq (1+\rho L)w,
\end{aligned}$$

where (i) holds because $\nabla f_i(y^{i,\star}) = 0$. Thus, by the fact that $\left|\mathcal{V}_k^{\mathrm{A}}\right| \leq n_k$ we prove condition (c) as follows

$$\frac{d(\boldsymbol{x}_k^{\mathrm{A}}, \hat{\boldsymbol{x}}_k)}{\sqrt{pn_k}} \leq \frac{(1+\rho L)w}{\sqrt{pn_k}}\sqrt{\left|\mathcal{V}_k^{\mathrm{A}}\right|} \leq \frac{(1+\rho L)w}{\sqrt{p}}.$$

Condition (d) – This condition holds by Assumption 2.

Conclusion of the proof: We now show that that the distance of the estimation vector $\boldsymbol{y}_k = [y_k^i, \cdots, y_k^{n_k}] \in \mathbb{R}^{n_k}$ of the whole network from the consensus state on the solution $\mathcal{C}_k^\star = \{\mathbb{1}_{n_k} \otimes y_k^\star\}$ is upper bounded by $d(\boldsymbol{x}_k, \hat{\boldsymbol{x}}_k)$ as follows: $d(\boldsymbol{y}_k, \mathcal{C}_k^\star) = \|\boldsymbol{y}_k - (\mathbb{1}_{n_k} \otimes y_k^\star)\| =$

$$\overset{(i)}{=} \left\|\text{prox}_{f_k}^\rho((2J_k - I)\boldsymbol{x}_{k-1}) - \text{prox}_{f_k}^\rho((2J_k - I)\hat{\boldsymbol{x}}_{k-1})\right\|$$

$$\overset{(ii)}{\leq} \|(2J_k - I)(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k)\| \overset{(iii)}{\leq} \|(2J_k - I)\|d(\boldsymbol{x}_k, \hat{\boldsymbol{x}}_k) \overset{(iv)}{=} d(\boldsymbol{x}_k, \hat{\boldsymbol{x}}_k)$$

where (i) holds by the specific update rule in eq. (5a) of Open FedPRS and by the relation between the fixed point and the optimal solution in eqs. (9)-(10); (ii) follows by the non-expansiveness of the proximal; (iii) holds by the sum-multiplicative property; (iv) holds because matrix $2J_k - I$ is doubly stochastic. We complete the proof exploiting (4):

$$\limsup_{k\to\infty} \frac{d(\boldsymbol{y}_k, \mathcal{C}_k^\star)}{\sqrt{pn_k}} \leq \limsup_{k\to\infty} \frac{d(\boldsymbol{x}_k, \hat{\boldsymbol{x}}_k)}{\sqrt{pn_k}} \leq R. \qquad \square$$

3. NUMERICAL SIMULATIONS

In this section we apply Open FedPRS to classification problems characterized by the local costs

$$f_i(x) = \tfrac{1}{m_i}\textstyle\sum_{j=1}^{m_i} \log\left(1 + \exp\left(-b_{i,j}a_{i,j}^\top x\right)\right) + \frac{\epsilon}{2}\|x\|^2$$

where $a_{i,j} \in \mathbb{R}^p$ and $b_{i,j} \in \{-1, 1\}$ are pairs of feature vector and label, randomly generated with `scikit-learn`. In particular, $p = 5$, $m_i = 50$, and $\epsilon = 0.05$. Notice that the local costs do not have a closed form prox, hence the agents approximate it via accelerated gradient descent, up to precision 10^{-10}. The results will be presented in terms of the normalized distance e_k between the local models and the consensus set on the optimal model as in (3) and in terms of the normalized distance ε_k between the coordinator model and the optimal model:

$$e_k := \frac{d(\boldsymbol{y}_k, \mathcal{C}_k^\star)}{\sqrt{pn_k}}, \qquad \varepsilon_k = \frac{d(y_k^c, y_k^\star)}{\sqrt{p}}. \tag{11}$$

© 2026 the authors. Licensed under CC-BY-NC-ND.

3.1 Comparison with the state of the art

We benchmark Open FedPRS against two state-of-the-art algorithms in federated learning over open networks, namely FedAVG (McMahan et al., 2017), one of the pioneering methods that introduced the federated learning paradigm, and FedACG (Kim et al., 2024), arguably one of the most competitive approach currently available in the literature. Since these algorithms were originally designed for closed (static) networks, we adapt them to the open-network setting by incorporating the mechanisms proposed in (Sun et al., 2024) and (Chang et al., 2024), respectively. The chosen local optimizers are accelerated gradient descent (AGD) for Open FedPRS, stochastic gradient descent (SGD) and Adam for FedAVG, and regularized SGD (rSGD) for FedACG. All algorithms are carefully hand-tuned to ensure optimal performance in the considered experimental scenario.

We model arrival and departure events as Poisson r.v.s $\mathrm{Pois}(\lambda^{\mathrm{join}})$ and $\mathrm{Pois}(\lambda^{\mathrm{leave}})$, and assume that at least 5 iterations occur between events. We simulate four modes: 1) same arrivals and departures $\lambda^{\mathrm{join}} = \lambda^{\mathrm{leave}} = 1$; 2) more arrivals $\lambda^{\mathrm{join}} = 2\lambda^{\mathrm{leave}} = 1$; 3) more departures $\lambda^{\mathrm{leave}} = 2\lambda^{\mathrm{join}} = 1$, 4 fixed network $\lambda^{\mathrm{leave}} = \lambda^{\mathrm{join}} = 0$.

Figure 1 illustrates the evolution of the training errors e_k and ε_k for the coordinator and the agents, respectively, as defined in (11), across all algorithms and different numbers of agents. According to Theorem 4, after an initial transient phase in which convergence occurs at a linear rate, the agents' training error ε_k under Open FedPRS eventually becomes upper bounded by a constant that does not scale with the network size. For the other algorithms, however, the metric ε_k is less informative since, by design, the agents' local variables do not converge exactly to consensus. To provide a fairer comparison, we also evaluated the algorithms in terms of the coordinator's estimation error e_k. Even under this metric, Open FedPRS outperforms the other methods, achieving both a tighter bound on the tracking error and a faster convergence rate.

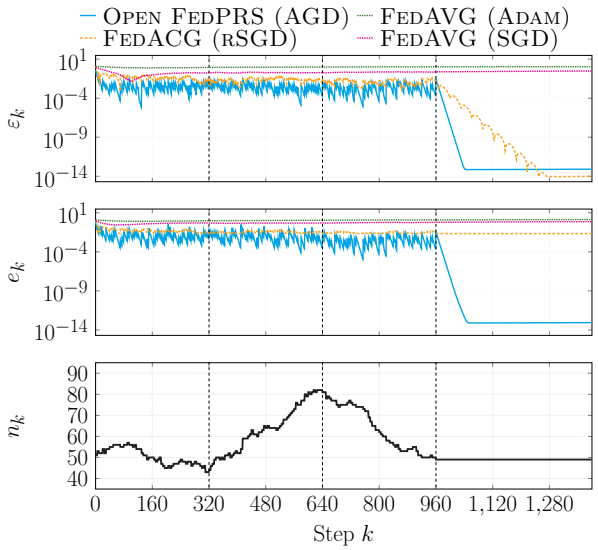


Fig. 1. Comparison of Open FedPRS performance with the state of the art in open networks.

This advantage persists even in scenarios with unbalanced agent arrivals and departures, i.e., when the network size varies over time—highlighting the superior resilience of the proposed Open FedPRS. Finally, when the network size becomes fixed, only Open FedPRS and FedACG converge asymptotically to the global solution.

3.2 Different initializations

Open FedPRS is designed to initialize the local state x_k^i of an arriving agent at its local optimum model $y^{i,\star}$. However, in practice this choice might not be available, as it assumes pre-computation of the local optimum, which might be prohibitively time-consuming. Therefore, in this section we test empirically two alternative initializations: 1) the coordinator model, i.e., $x_k^i = y_{k-1}^c$, which only requires one communication from the coordinator to each arriving agent; the zero vector, i.e., $x_k^i = \mathbb{0}$. which is simple and cheap from computational and communication perspectives. To test these initializations, we apply Open FedPRS on a network with Poisson arrivals/departures with mean $\lambda^{\mathrm{join}} = \lambda^{\mathrm{leave}} = 1$. Table 1 reports the results, presented in terms of the minimum, maximum, mean and standard deviation of ε_k over the last 4/5 of the simulation (to ensure that steady-state behavior is reached). The choice $y^{i,\star}$ is the one that gives the best minimum and mean training error ε_k, but this comes to the cost of a larger standard deviation and maximum value. The other choices y^c and $\mathbb{0}$ instead have quite similar, worse performance, but have the advantage of being much simpler and cheap. A future direction is to investigate this aspect from a theoretical perspective.

Table 1. ε_k of Open FedPRS for different initializations of arriving agents.

Init.	Min	Mean ± Std	Max
$y^{i,\star}$	6.013×10^{-6}	$(4.646 \pm 12.880) \times 10^{-4}$	1.179×10^{-2}
y^c	6.546×10^{-5}	$(8.409 \pm 9.141) \times 10^{-4}$	5.854×10^{-3}
0	1.905×10^{-5}	$(7.619 \pm 8.298) \times 10^{-4}$	5.145×10^{-3}

4. CONCLUDING REMARKS AND FUTURE WORKS

We have introduced Open FedPRS, a novel federated learning algorithm tailored for open multi-agent systems where agents may join or leave freely during the learning process. By leveraging the Theory of Open Operators, we established a rigorous convergence analysis that guarantees a finite asymptotic bound on the local models' distance from the optimal model and achieves a linear convergence rate. Simulation results confirm the theoretical robustness of the method across various initialization strategies. Future improvements could include the timing and manner in which agents leave the network and new agents join, for instance by allowing the central coordinator to control these events (Bistritz and Bambos, 2024), and overcoming the need of the current PRS formulation to perform a proximal operation at each time step, for instance by replacing this step with approximated solutions as outlined in (Bastianello et al., 2024).

REFERENCES

Arunan, A., Qin, Y., Li, X., and Yuen, C. (2023). A federated learning-based industrial health prognostics

© 2026 the authors. Licensed under CC-BY-NC-ND.

for heterogeneous edge devices using matched feature extraction. *IEEE Transactions on Automation Science and Engineering*.

Bastianello, N., Deplano, D., Franceschelli, M., and Johansson, K.H. (2025). Robust online learning over networks. *IEEE Transactions on Automatic Control*, 70(2), 933–946. doi:10.1109/TAC.2024.3441723.

Bastianello, N., Liu, C., and Johansson, K.H. (2024). Enhancing privacy in federated learning through local training. *arXiv preprint arXiv:2403.17572*.

Bauschke, H.H. and Combettes, P.L. (2017). *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, Cham, Switzerland, 2 edition.

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*, 2(3), 453–458.

Bistritz, I. and Bambos, N. (2024). Gamekeeper: Online learning for admission control of networked open multiagent systems. *IEEE Transactions on Automatic Control*.

Chang, Z.L., Han, D.J., Parasnis, R., Hosseinalipour, S., and Brinton, C.G. (2024). Federated Learning with Dynamic Client Arrival and Departure: Convergence and Rapid Adaptation via Initial Model Construction.

Chellapandi, V.P., Upadhyay, A., Hashemi, A., and Żak, S.H. (2023). On the convergence of decentralized federated learning under imperfect information sharing. *IEEE Control Systems Letters*, 7, 2982–2987.

Dal Fabbro, N., Mitra, A., and Pappas, G.J. (2023). Federated td learning over finite-rate erasure channels: Linear speedup under markovian sampling. *IEEE Control Systems Letters*, 7, 2461–2466.

Deplano, D., Franceschelli, M., and Giua, A. (2024). Stability of paracontractive open multi-agent systems. In *2024 IEEE 63rd Conference on Decision and Control (CDC)*, 3031–3036.

Deplano, D., Bastianello, N., Franceschelli, M., and Johansson, K.H. (2026). Optimization and learning in open multi-agent systems. *IEEE Transactions on Automatic Control*.

Deplano, D., Franceschelli, M., and Giua, A. (2023). Novel stability conditions for nonlinear monotone systems and consensus in multiagent networks. *IEEE Transactions on Automatic Control*, 68(12), 7028–7040.

Franceschelli, M. and Frasca, P. (2020). Stability of open multiagent systems and applications to dynamic consensus. *IEEE Transactions on Automatic Control*, 66(5), 2326–2331.

Giselsson, P. and Boyd, S. (2016). Linear convergence and metric selection for douglas-rachford splitting and admm. *IEEE Transactions on Automatic Control*, 62(2), 532–544.

Gu, X., Huang, K., Zhang, J., and Huang, L. (2021). Fast federated learning in the presence of arbitrary device unavailability. *Advances in Neural Information Processing Systems*, 34, 12052–12064.

Han, M., Patel, K.K., Shao, H., and Wang, L. (2023). On the Effect of Defections in Federated Learning and How to Prevent Them.

Hayashi, N. (2023). Distributed Subgradient Method in Open Multiagent Systems. *IEEE Transactions on Automatic Control*, 68(10), 6192–6199.

Hendrickx, J.M. and Rabbat, M.G. (2020). Stability of decentralized gradient descent in open multi-agent systems. In *59th IEEE Conference on Decision and Control*, 4885–4890. Jeju Island, Korea (South).

Hsieh, Y.G., Iutzeler, F., Malick, J., and Mertikopoulos, P. (2021). Optimization in Open Networks via Dual Averaging. In *2021 60th IEEE Conference on Decision and Control (CDC)*, 514–520. Austin, TX, USA.

Hsieh, Y.G., Iutzeler, F., Malick, J., and Mertikopoulos, P. (2022). Multi-agent online optimization with delays: Asynchronicity, adaptivity, and optimism. *Journal of Machine Learning Research*, 23(78), 1–49.

Kim, G., Kim, J., and Han, B. (2024). Communication-efficient federated learning with accelerated client gradient. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12385–12394.

McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B.A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, 1273–1282. PMLR.

Öksüz, H.Y., Molinari, F., Sprekeler, H., and Raisch, J. (2024). Boosting fairness and robustness in over-the-air federated learning. *IEEE Control Systems Letters*.

Omori, T. and Kashima, K. (2023). Combinatorial optimization approach to client scheduling for federated learning. *IEEE Control Systems Letters*, 7, 2203–2208.

Perifanis, V., Pavlidis, N., Koutsiamanis, R.A., and Efraimidis, P.S. (2023). Federated learning for 5g base station traffic forecasting. *Computer Networks*, 235, 109950.

Ryu, E.K. and Boyd, S. (2016). Primer on monotone operator methods. *Appl. comput. math*, 15(1), 3–43.

Sawamura, R., Hayashi, N., and Inuiguchi, M. (2024). A distributed primal-dual push-sum algorithm on open multiagent networks. *IEEE Transactions on Automatic Control*.

Sun, Y., Fernando, H., Chen, T., and Shahrampour, S. (2023). On the Stability Analysis of Open Federated Learning Systems. In *2023 American Control Conference (ACC)*, 867–872. San Diego, CA, USA.

Sun, Y., Mao, Y., and Zhang, J. (2024). MimiC: Combating Client Dropouts in Federated Learning by Mimicking Central Updates. *IEEE Transactions on Mobile Computing*.

Taylor, A.B., Hendrickx, J.M., and Glineur, F. (2017). Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Mathematical Programming*, 161, 307–345.

Vizuete, R., Monnoyer de Galland, C., Frasca, P., Panteley, E., and Hendrickx, J.M. (2024). Trends and questions in open multi-agent systems. In *Hybrid and Networked Dynamical Systems: Modeling, Analysis and Control*, 219–252. Springer.

Wang, H. and Xu, J. (2023). Combating Client Dropout in Federated Learning via Friend Model Substitution.

Yuan, Z., Xu, S., and Zhu, M. (2024). Federated reinforcement learning for robot motion planning with zero-shot generalization. *Automatica*, 166, 111709.

Zheng, G., Kong, L., and Brintrup, A. (2023). Federated machine learning for privacy preserving, collective supply chain risk prediction. *International Journal of Production Research*, 61(23), 8115–8132.

© 2026 the authors. Licensed under CC-BY-NC-ND.