DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Nicolò Dal Fabbro

Pushing the Boundaries of Federated Learning:

Superlinear Convergence and Reinforcement Learning over Wireless

GTTI meeting

September, 2024

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Advisors:
Prof. Luca Schenato
Prof. Michele Rossi

Distributed Machine Learning

Massive data production

Artificial Intelligence = Data-Driven Algorithms

Distributed Machine Learning

Massive data production

Artificial Intelligence = Data-Driven Algorithms

Privacy

Decentralized datasets

Parallelization

Distributed Machine Learning

Massive data production

Artificial Intelligence = Data-Driven Algorithms

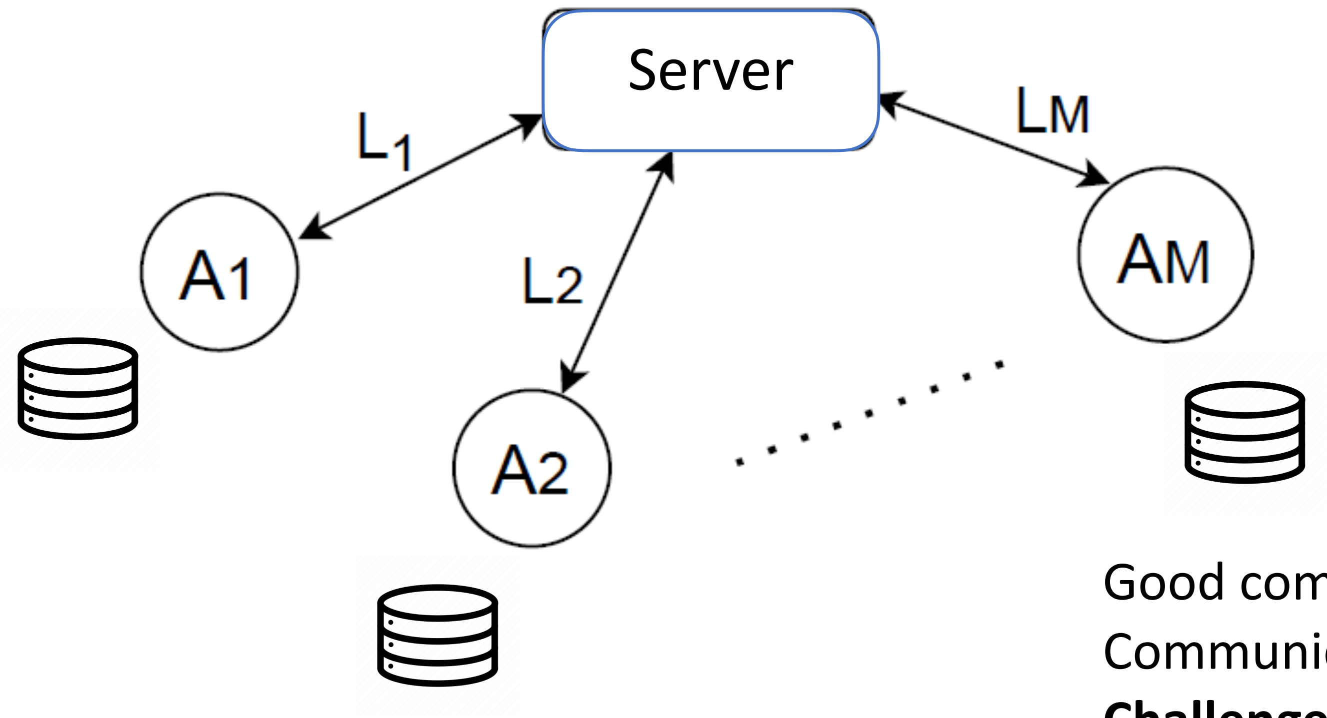
Privacy

Decentralized datasets

Parallelization

Federated Learning

Federated Learning



Decentralized datasets

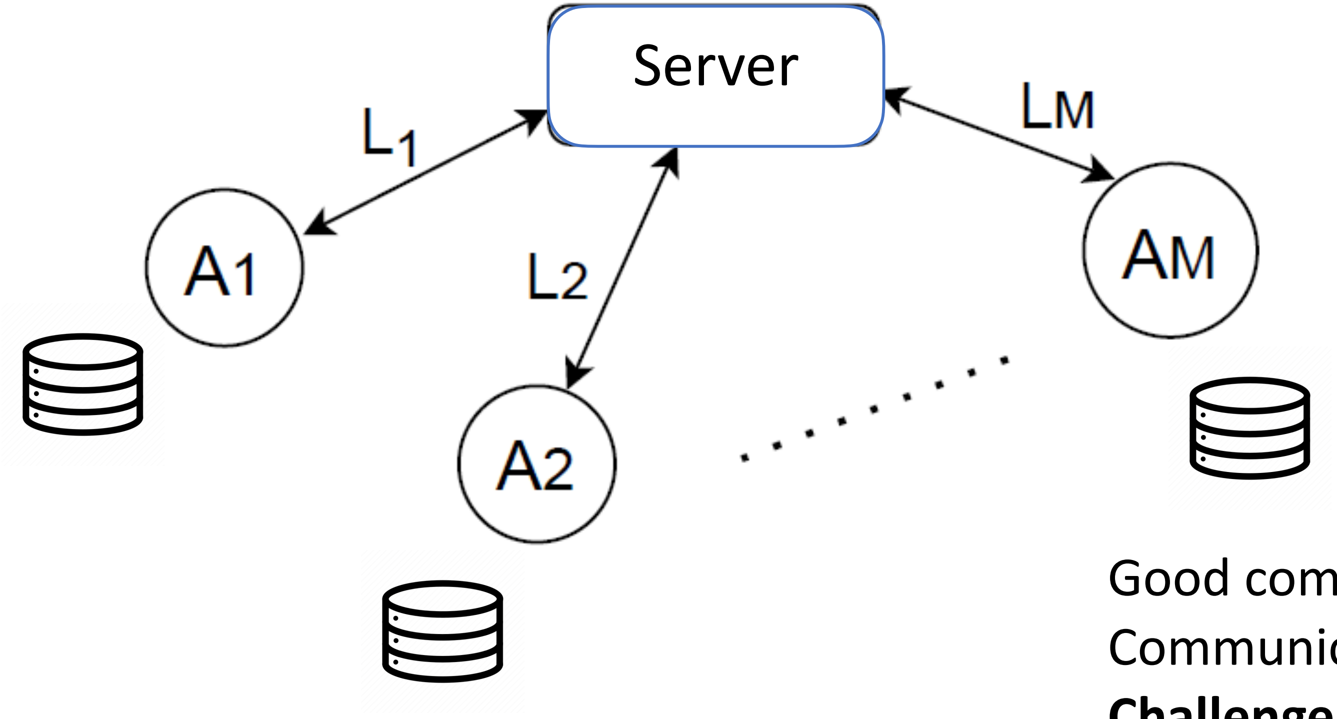
Parallelization

Privacy

Federated Learning

Good computational power at the edge devices
Communication is usually expensive
Challenge: reducing the amount of communication

Federated Learning



Decentralized datasets

Parallelization

Privacy

Federated Learning

Good computational power at the edge devices
Communication is usually expensive
Challenge: reducing the amount of communication

In this thesis:
- Design communication efficient algorithms
- Analyze the effect of communication constraints on algorithms’ convergence

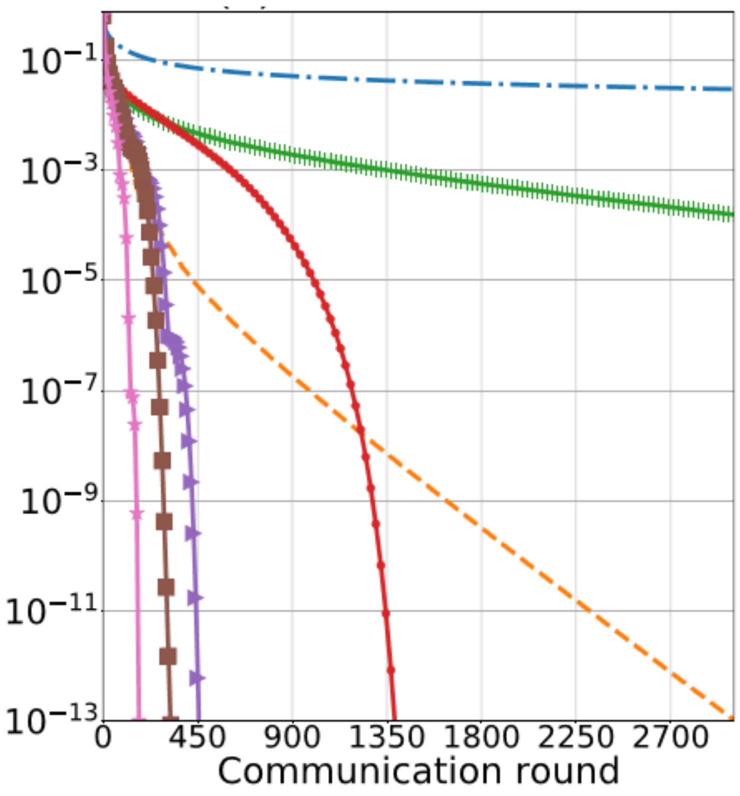
Research questions

Research questions

1. Can we design **communication-efficient** algorithms for federated learning with **superlinear convergence**?

Research questions

1. Can we design **communication-efficient** algorithms for federated learning with **superlinear convergence**?



Research questions

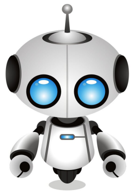
1. Can we design **communication-efficient** algorithms for federated learning with **superlinear convergence**?

2. Can we provide finite-time analyses of **federated reinforcement learning** algorithms under **communication constraints**?

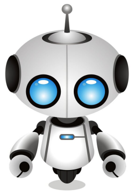
Research questions

1. Can we design **communication-efficient** algorithms for federated learning with **superlinear convergence**?

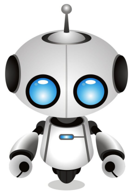
2. Can we provide finite-time analyses of **federated reinforcement learning** algorithms under **communication constraints**?



Agent 1



Agent 2



Agent N

PhD thesis

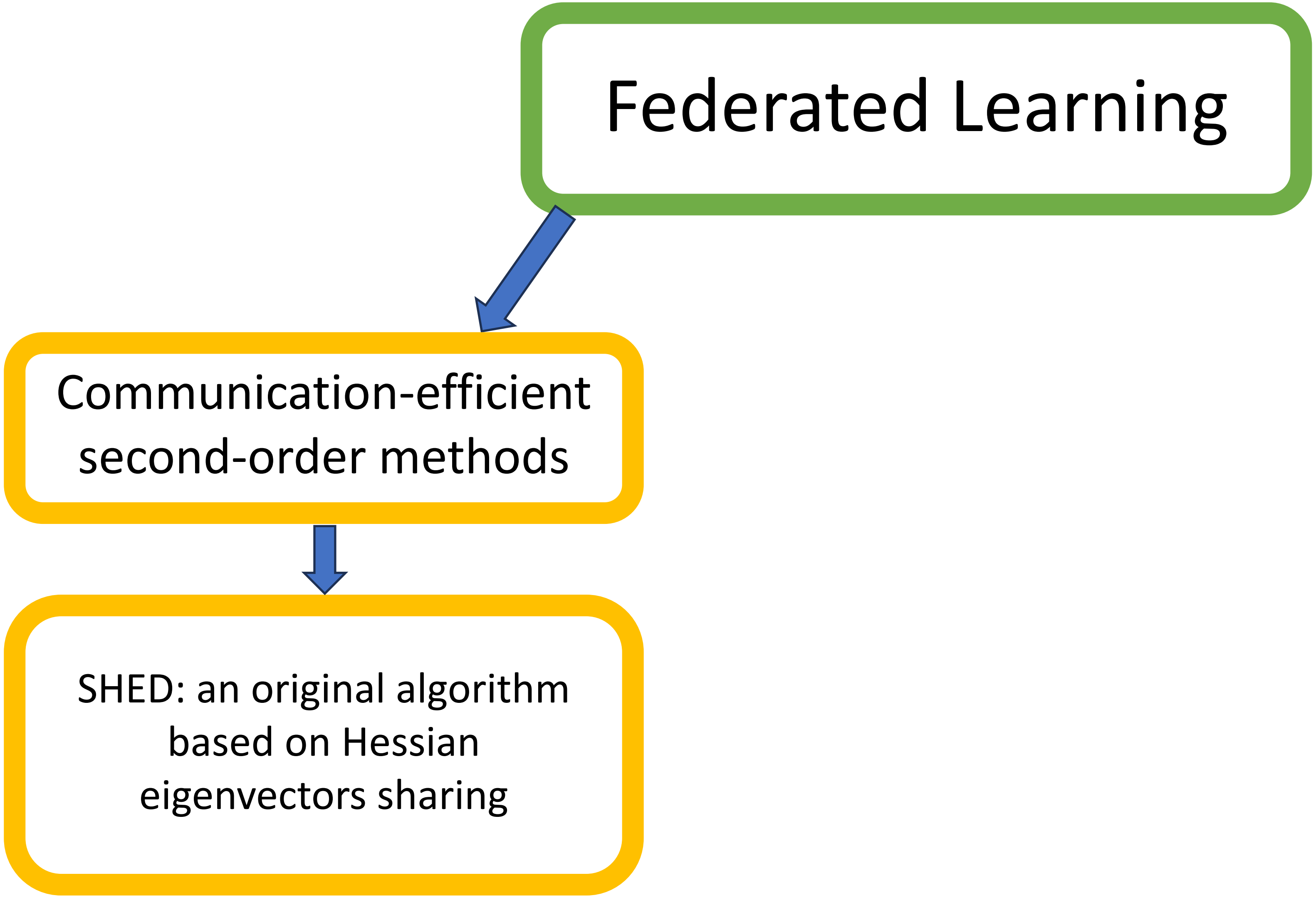
Federated Learning

PhD thesis

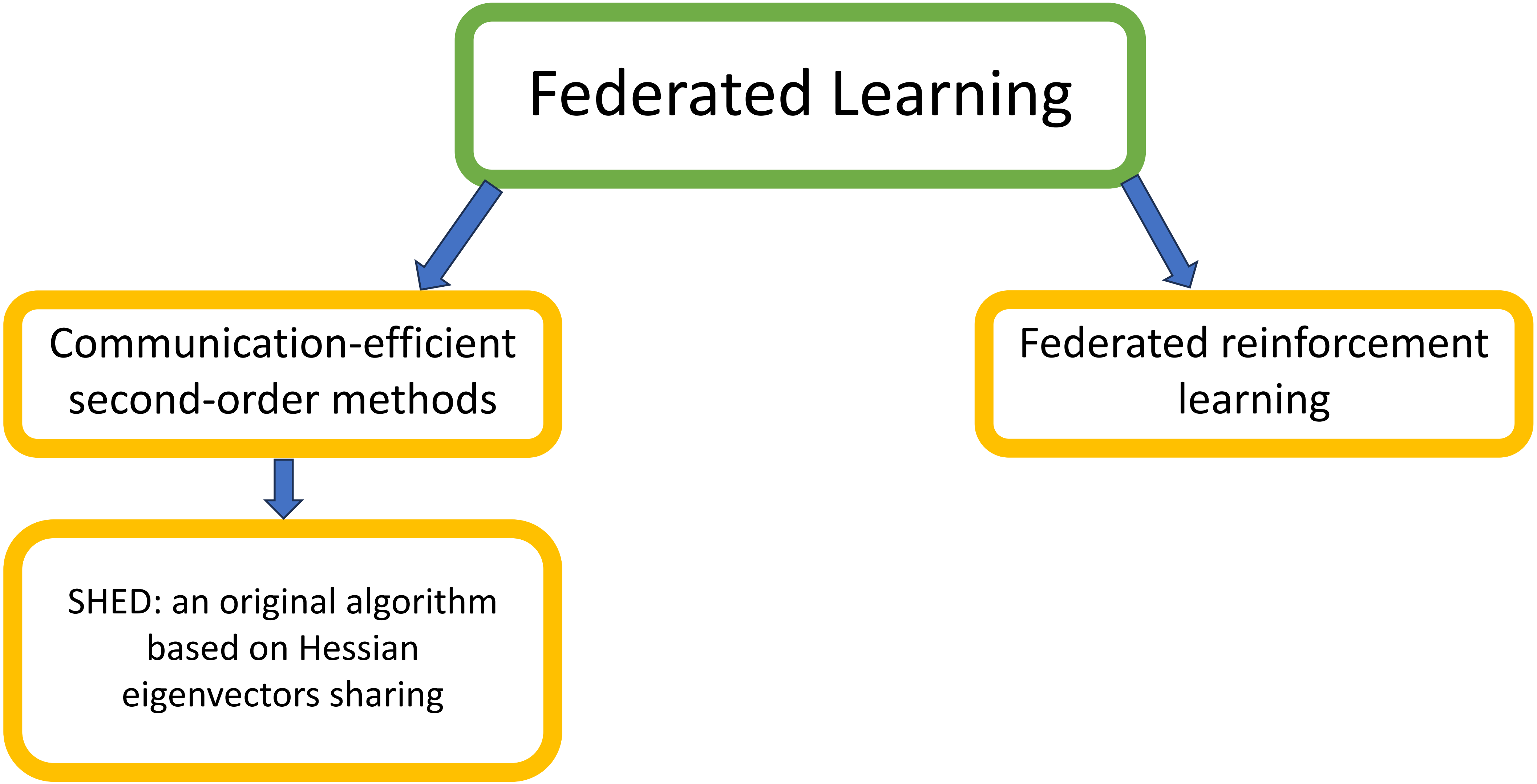
Federated Learning

Communication-efficient second-order methods

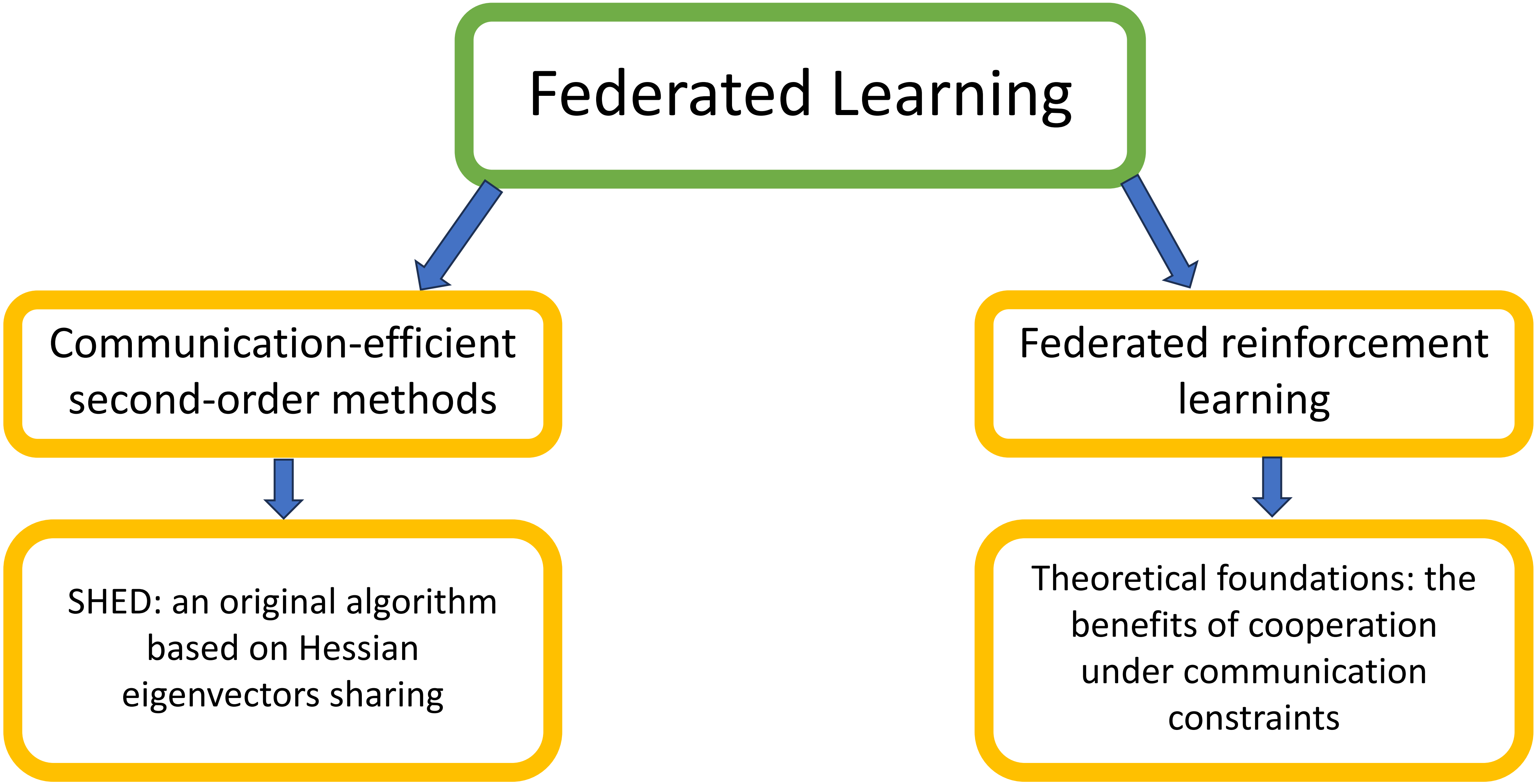
PhD thesis


Federated Learning
Communication-efficient second-order methods
SHED: an original algorithm based on Hessian eigenvectors sharing

PhD thesis

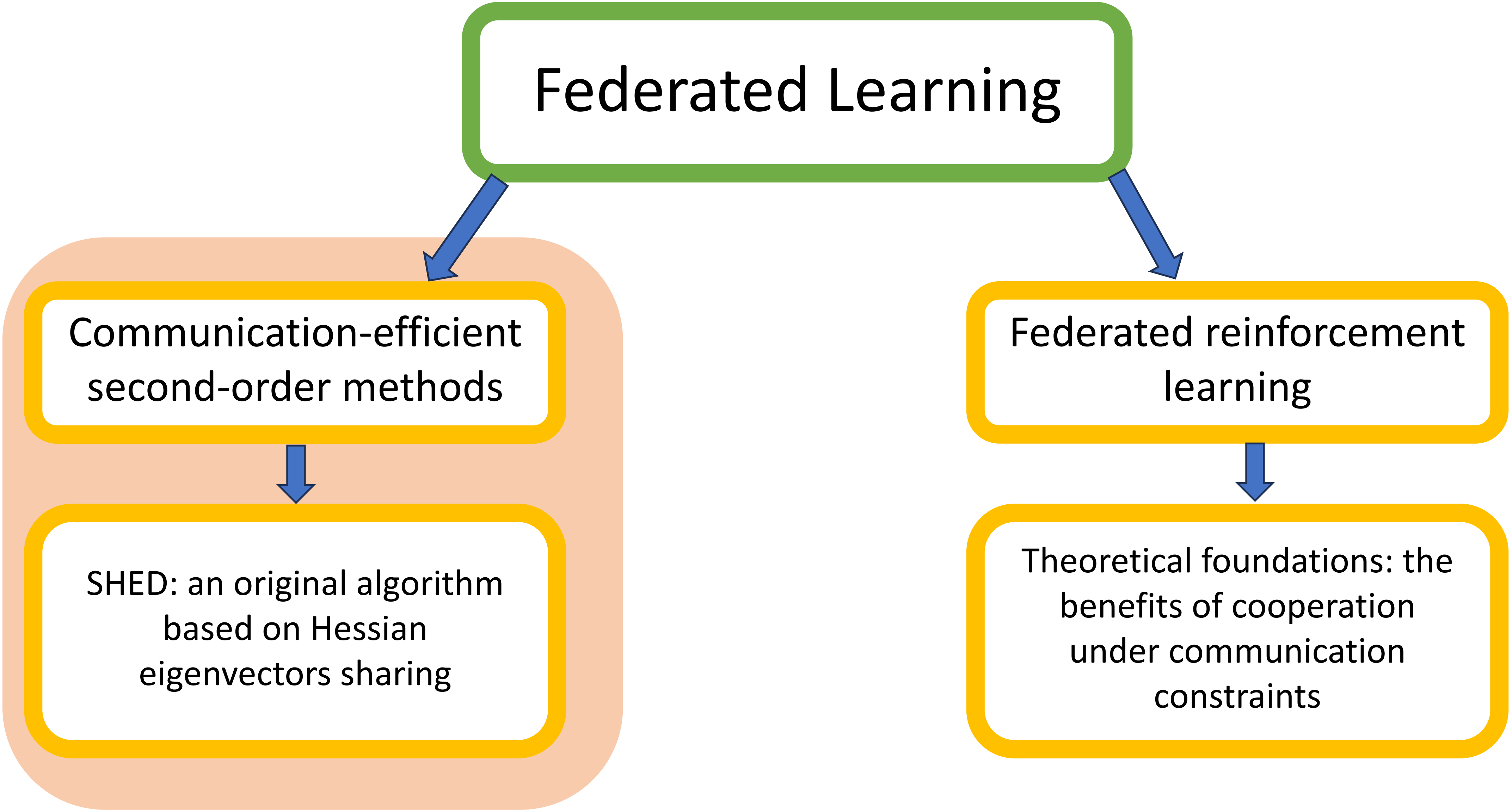

Federated Learning
Communication-efficient second-order methods
Federated reinforcement learning
SHED: an original algorithm based on Hessian eigenvectors sharing

PhD thesis



Federated Learning
Communication-efficient second-order methods
SHED: an original algorithm based on Hessian eigenvectors sharing
Federated reinforcement learning
Theoretical foundations: the benefits of cooperation under communication constraints

PhD thesis


Federated Learning
Communication-efficient second-order methods
SHED: an original algorithm based on Hessian eigenvectors sharing
Federated reinforcement learning
Theoretical foundations: the benefits of cooperation under communication constraints

Luca Schenato
Unipd

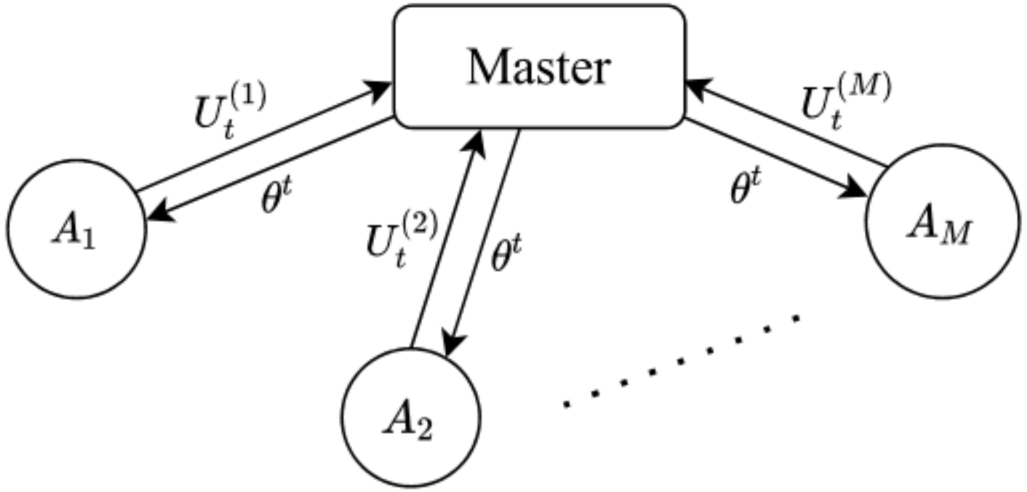
Michele Rossi
Unipd

Subhrakanti Dey,
Uppsala University

Superlinear federated learning: notation

M agents with datasets $\{\mathcal{D}_1, \ldots, \mathcal{D}_M\}$ aim to iteratively minimize a cost function $f(\boldsymbol{\theta}) = \frac{1}{M}\sum_{i=1}^{M} f_i(\boldsymbol{\theta})$

- $\boldsymbol{\theta}^t \in \mathbb{R}^n$ is the n-dimensional global parameter at iteration t,
- $f_i(\boldsymbol{\theta})$ is the local cost of agent i,
- $U_t^{(i)}$ is the optimization set shared by agent i at iteration t.



Distributed gradient descent

Let $\mathbf{g}_t = \nabla f(\boldsymbol{\theta}^t) \in \mathbb{R}^n$ and $\mathbf{H}_t = \nabla^2 f(\boldsymbol{\theta}^t) \in \mathbb{R}^{n \times n}$ denote the gradient and the Hessian matrix of the cost function.

Distributed gradient descent consists of iteratively performing:

$$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t - \eta_t \mathbf{g}_t$$

where η_t is the step size. We have assumed w.l.o.g. that all agents have the same number of data samples.

Distributed gradient descent

Let $\mathbf{g}_t = \nabla f(\boldsymbol{\theta}^t) \in \mathbb{R}^n$ and $\mathbf{H}_t = \nabla^2 f(\boldsymbol{\theta}^t) \in \mathbb{R}^{n\times n}$ denote the gradient and the Hessian matrix of the cost function.

Distributed gradient descent consists of iteratively performing:

$$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t - \eta_t \mathbf{g}_t = \boldsymbol{\theta}^t - \eta_t\left(\frac{1}{M}\sum_{i=1}^{M}\mathbf{g}_t^{(i)}\right),$$

where η_t is the step size. We have assumed w.l.o.g. that all agents have the same number of data samples.

Pros: scales well

Cons: convergence rate heavily impacted by the condition number

Distributed Newton method

Let $\mathbf{g}_t = \nabla f(\boldsymbol{\theta}^t) \in \mathbb{R}^n$ and $\mathbf{H}_t = \nabla^2 f(\boldsymbol{\theta}^t) \in \mathbb{R}^{n \times n}$ denote the gradient and the Hessian matrix of the cost function.

The Newton method consists of iteratively performing the Newton update:

$$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t - \eta_t \mathbf{H}_t{}^{-1} \mathbf{g}_t$$

where η_t is the step size. We have assumed w.l.o.g. that all agents have the same number of data samples.

Distributed Newton method

Let $\mathbf{g}_t = \nabla f(\boldsymbol{\theta}^t) \in \mathbb{R}^n$ and $\mathbf{H}_t = \nabla^2 f(\boldsymbol{\theta}^t) \in \mathbb{R}^{n \times n}$ denote the gradient and the Hessian matrix of the cost function.

The Newton method consists of iteratively performing the Newton update:

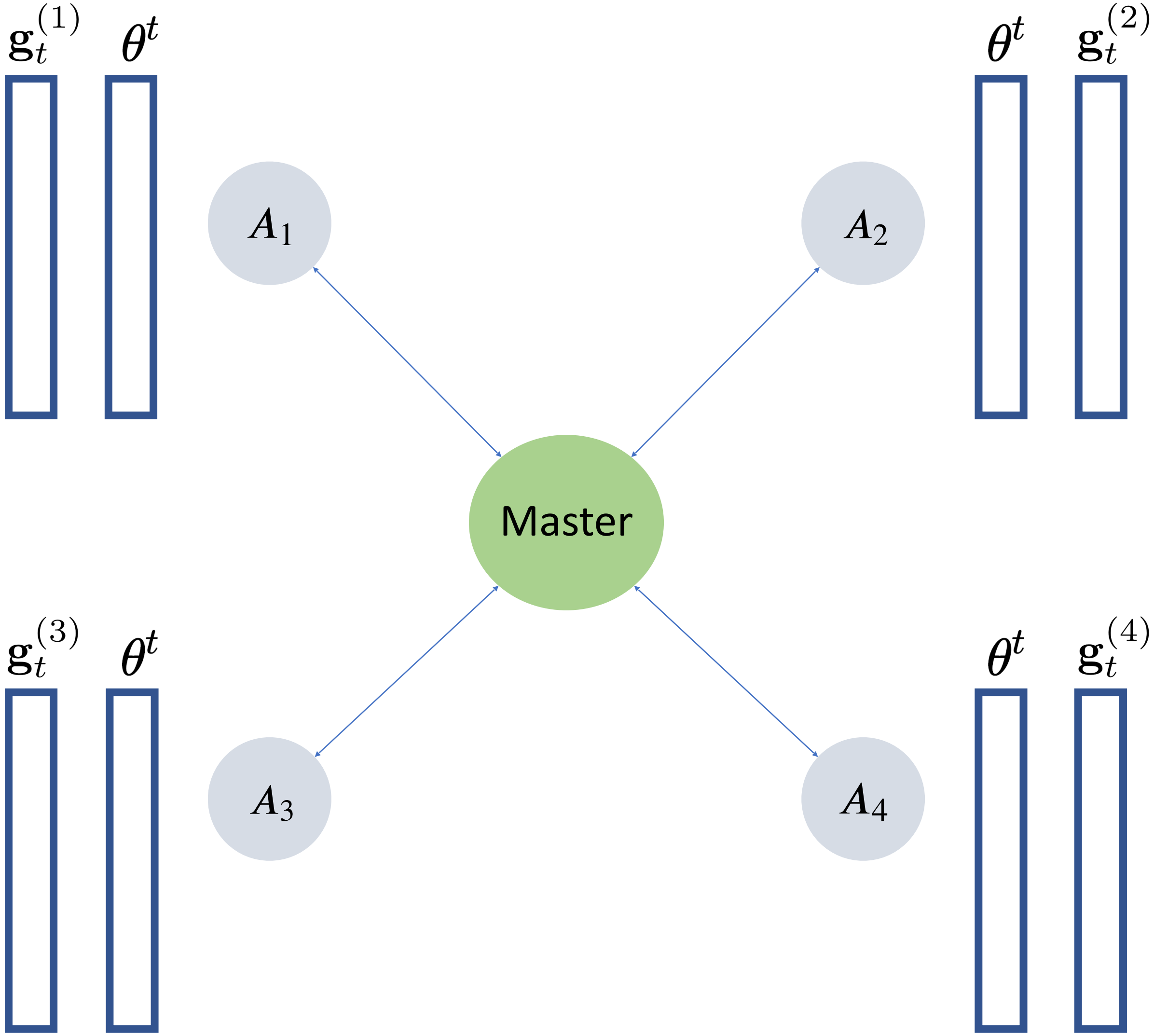
$$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t - \eta_t \mathbf{H}_t^{-1} \mathbf{g}_t = \boldsymbol{\theta}^t - \eta_t \Big(\frac{1}{M} \sum_{i=1}^{M} \mathbf{H}_t^{(i)}\Big)^{-1} \Big(\frac{1}{M} \sum_{i=1}^{M} \mathbf{g}_t^{(i)}\Big),$$

where η_t is the step size. We have assumed w.l.o.g. that all agents have the same number of data samples.

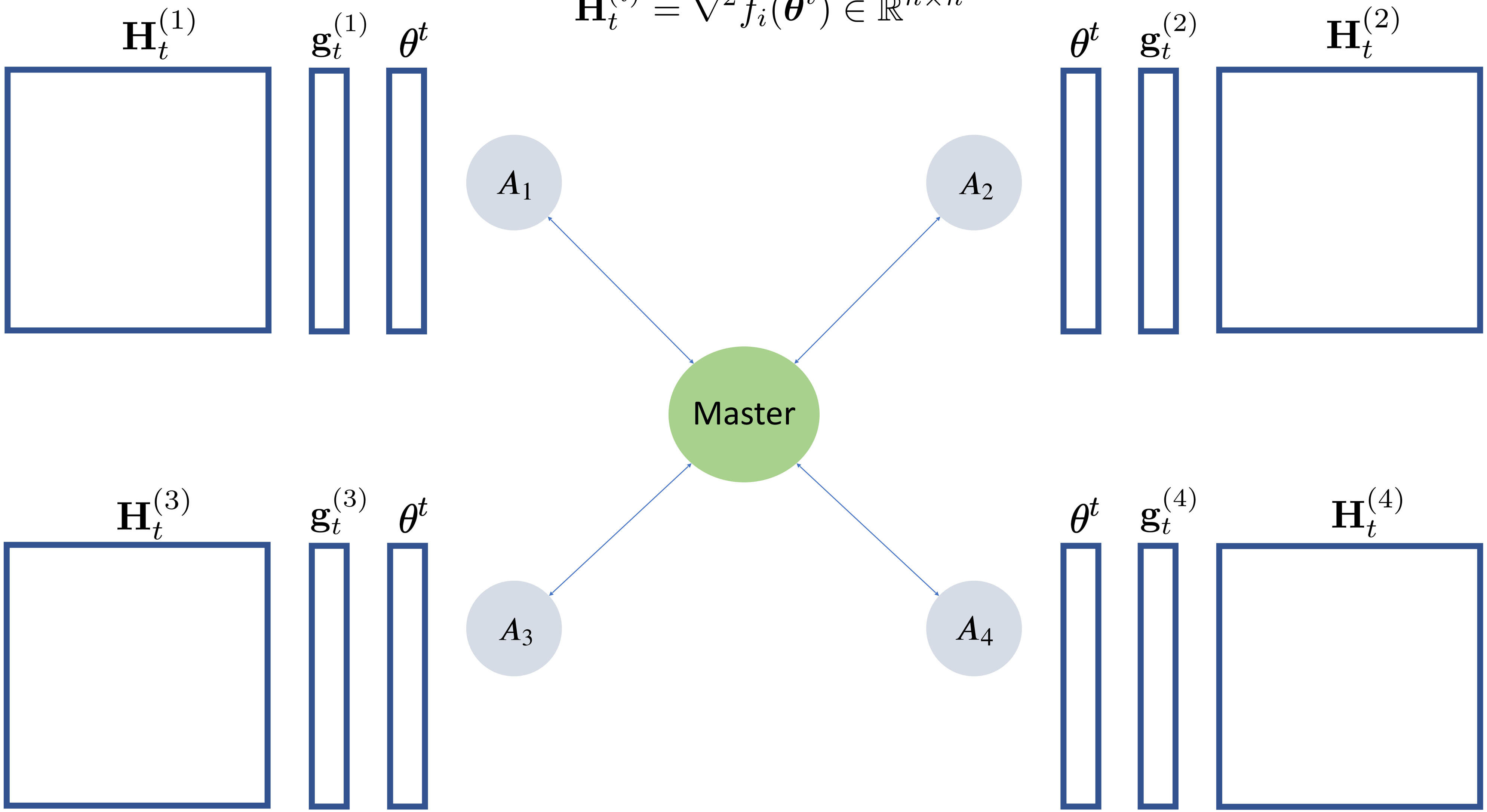
Pros: superlinear convergence speed independent of the condition number

Cons: significantly more demanding from a **computation** and **communication** point of view

$$\mathbf{g}_t^{(i)} = \nabla f_i(\boldsymbol{\theta}^t) \in \mathbb{R}^n$$



$$\mathbf{g}_t^{(i)} = \nabla f_i(\boldsymbol{\theta}^t) \in \mathbb{R}^n$$
$$\mathbf{H}_t^{(i)} = \nabla^2 f_i(\boldsymbol{\theta}^t) \in \mathbb{R}^{n \times n}$$



Approximate Newton method

Approximate Newton-type methods use approximations of the Hessian, $\hat{\mathbf{H}}_t$, so a Newton-type parameter update is

$$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t - \eta_t \hat{\mathbf{H}}_t^{-1} \mathbf{g}_t,$$

where $\hat{\mathbf{H}}_t$ is an approximation of the Hessian matrix.

Question: can we provide superlinear convergence in a communication-efficient way?

N. Dal Fabbro, S. Dey, M. Rossi and L. Schenato, "SHED: A Newton-type algorithm for federated learning based on incremental Hessian eigenvector sharing", 2024, *Automatica*

N. Dal Fabbro, M. Rossi, L. Schenato, S. Dey "Q-SHED: Distributed Optimization at the Edge via Hessian Eigenvectors Quantization", *IEEE International Conference on Communications,* Rome 2023

State of the art

Wang, Shusen, et al. 'GIANT: Globally improved approximate newton method for distributed optimization.' *Advances in Neural Information Processing Systems* 31 (2018).

Rixon Crane and Fred Roosta. 'DINGO: Distributed Newton-type method for gradient-norm optimization'. *Advances in Neural Information Processing Systems* 32 (2019).

Safaryan, Mher, et al. 'FedNL: Making Newton-type methods applicable to federated learning.' *International Conference on Machine Learning* 39 (2022).

Agafonov, Artem, et al. 'FLECS: A Federated Learning Second-Order Framework via Compression and Sketching.' arXiv preprint arXiv:2206.02009 (2022)

Elgabli, Anis, et al. 'FedNew: A Communication-Efficient and Privacy-Preserving Newton-Type Method for Federated Learning.' *International Conference on Machine Learning* 39 (2022).

State of the art

Wang, Shusen, et al. 'GIANT: Globally improved approximate newton method for distributed optimization.' *Advances in Neural Information Processing Systems* 31 (2018).

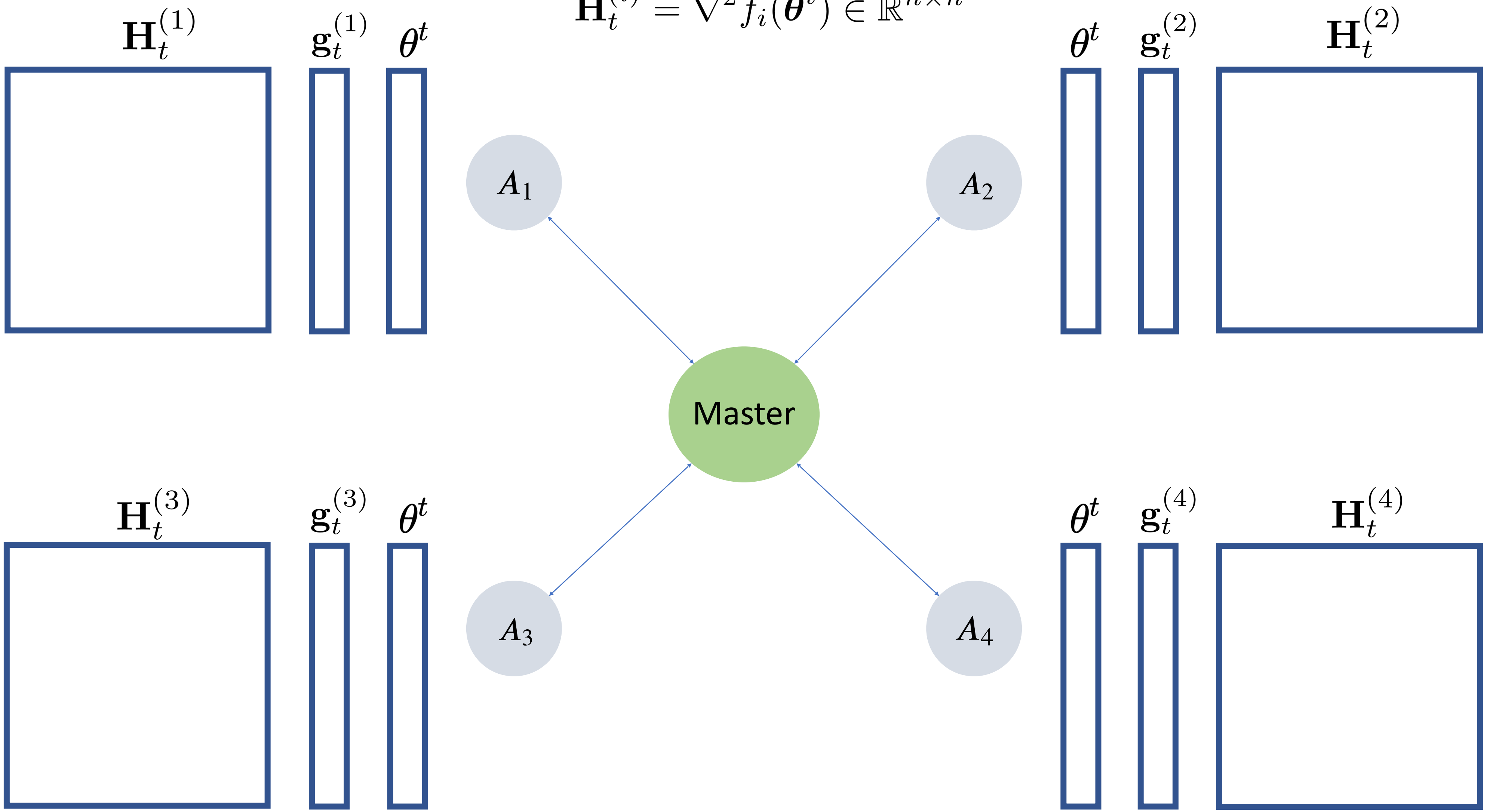
Rixon Crane and Fred Roosta. 'DINGO: Distributed Newton-type method for gradient-norm optimization'. *Advances in Neural Information Processing Systems* 32 (2019).

Safaryan, Mher, et al. 'FedNL: Making Newton-type methods applicable to federated learning.' *International Conference on Machine Learning* 39 (2022).

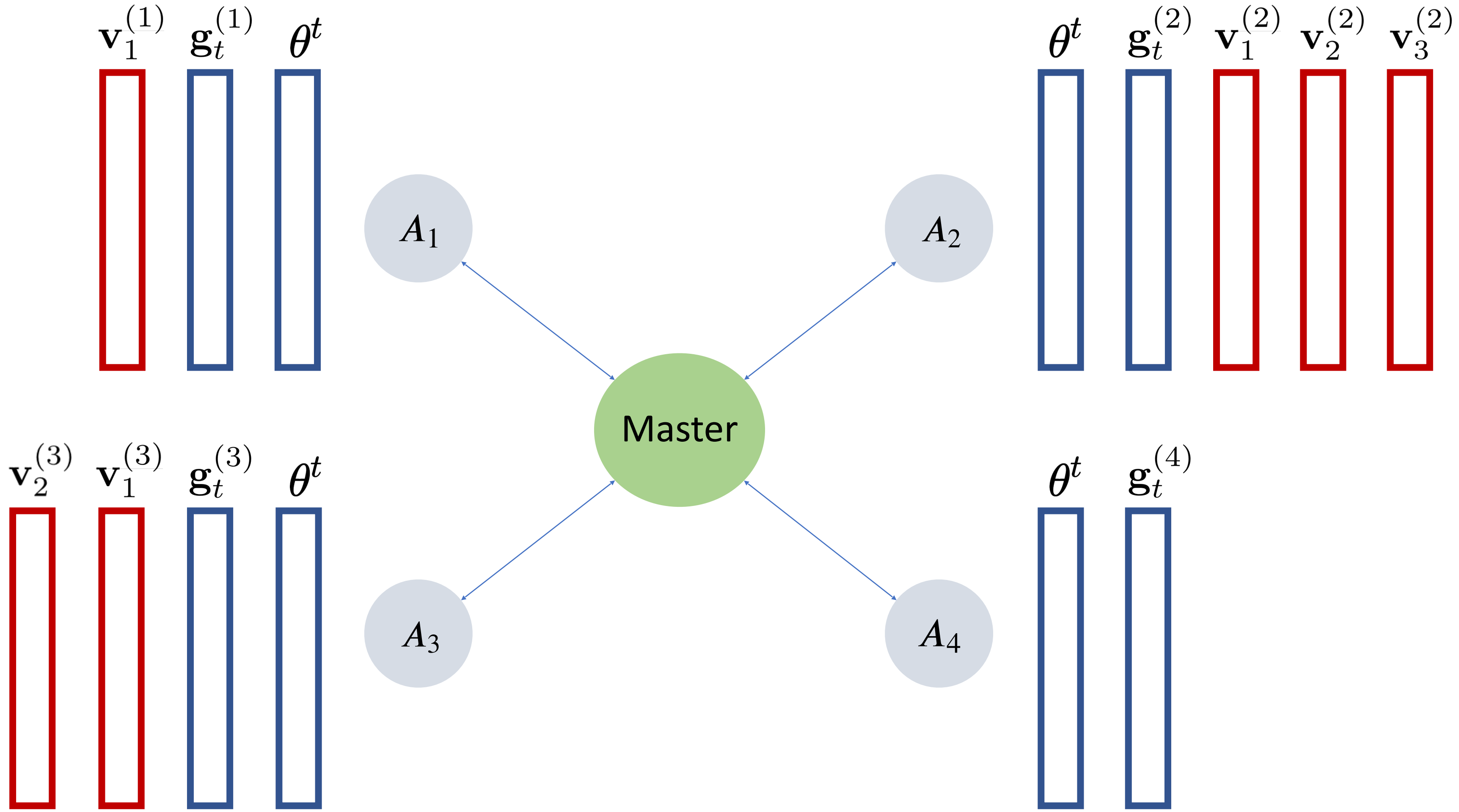
Agafonov, Artem, et al. 'FLECS: A Federated Learning Second-Order Framework via Compression and Sketching.' arXiv preprint arXiv:2206.02009 (2022)

Elgabli, Anis, et al. 'FedNew: A Communication-Efficient and Privacy-Preserving Newton-Type Method for Federated Learning.' *International Conference on Machine Learning* 39 (2022).

$$\mathbf{g}_t^{(i)} = \nabla f_i(\boldsymbol{\theta}^t) \in \mathbb{R}^n$$
$$\mathbf{H}_t^{(i)} = \nabla^2 f_i(\boldsymbol{\theta}^t) \in \mathbb{R}^{n\times n}$$



SHED: a Newton-type algorithm for FL based on eigendecomposition



$$\{\{\mathbf{v}_j^{(i)}, \lambda_j^{(i)}\}_{j=1}^{q_t^{(i)}}, \rho_t^{(i)}\}$$

$$\rho_t^{(i)} \geq (\lambda_{q_t^{(i)}+1}^{(i)} + \lambda_n^{(i)})/2$$

$$\|\boldsymbol{\theta}^{t+1} - \boldsymbol{\theta}^*\| \leq \left(1 - \frac{\bar{\lambda}_n}{\bar{\rho}_t}\right)\|\boldsymbol{\theta}^t - \boldsymbol{\theta}^*\|$$

Strengths of SHED

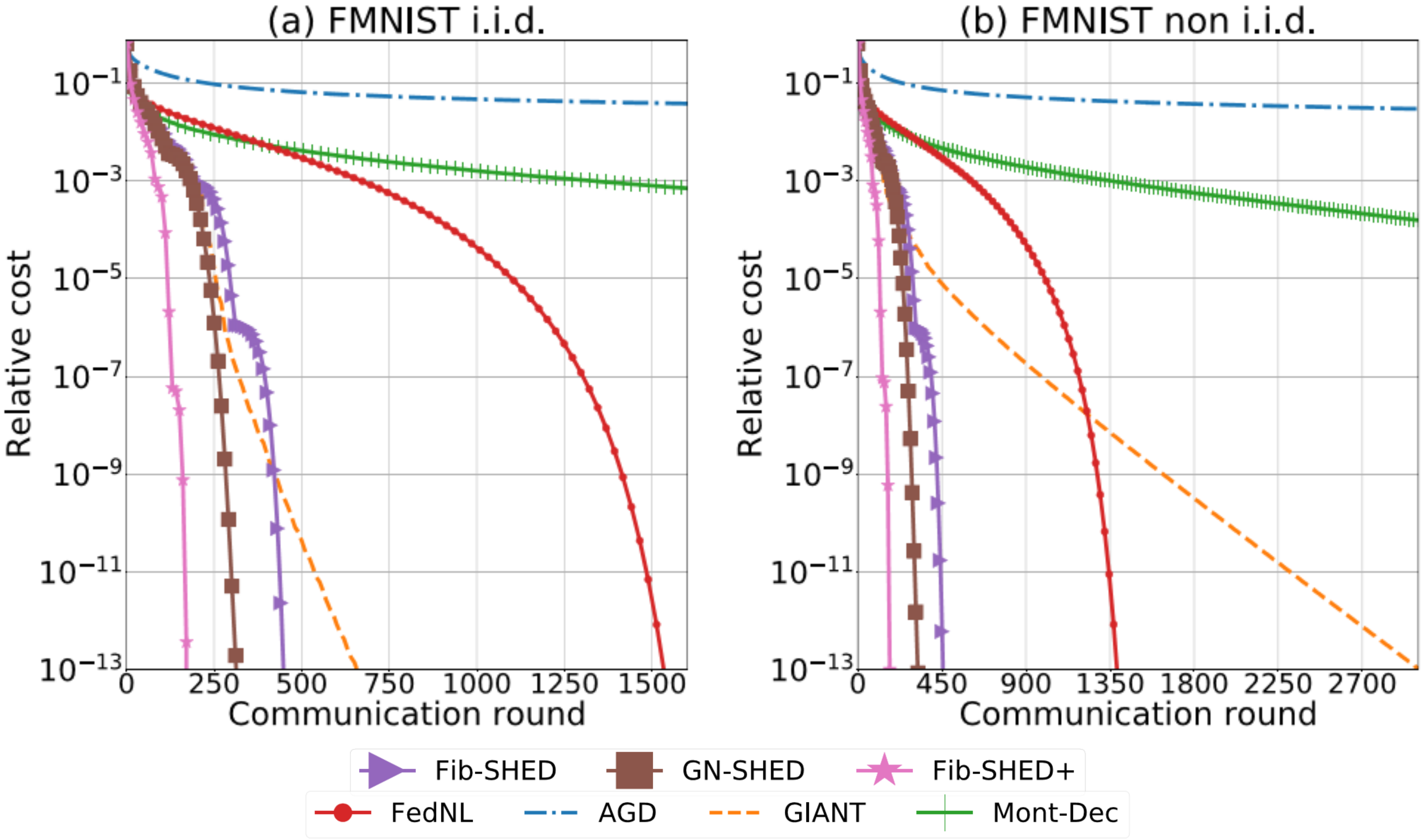
- Global convergence with asymptotic **superlinear rate**

- **Versatility** - each agent can share a number of eigenvectors based on their communication resources

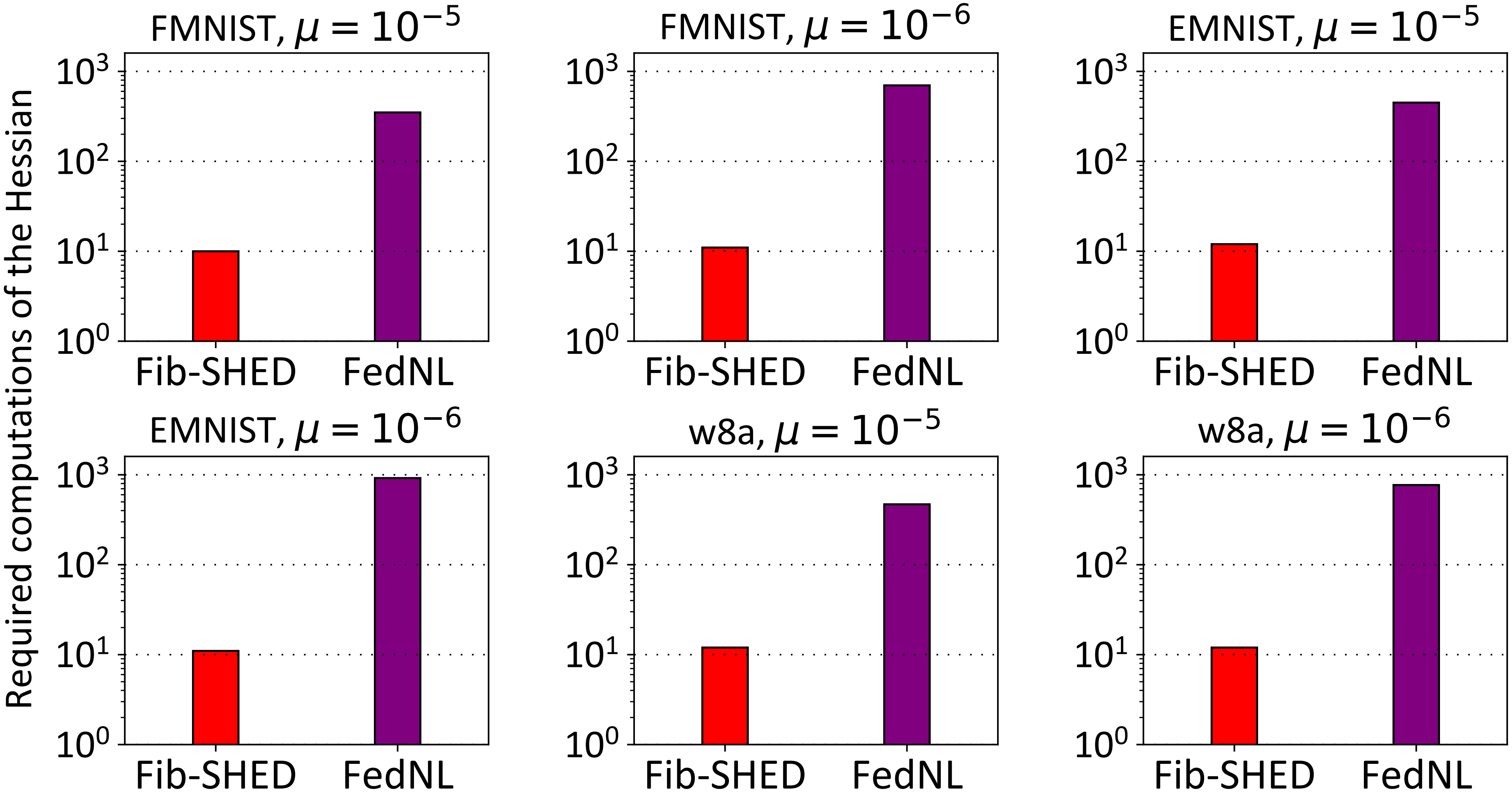
- Only **sporadic** Hessian computations required

N. Dal Fabbro, S. Dey, M. Rossi and L. Schenato, "SHED: A Newton-type algorithm for federated learning based on incremental Hessianeigenvector sharing", *Automatica,* 2024

$\mu = 10^{-6}$

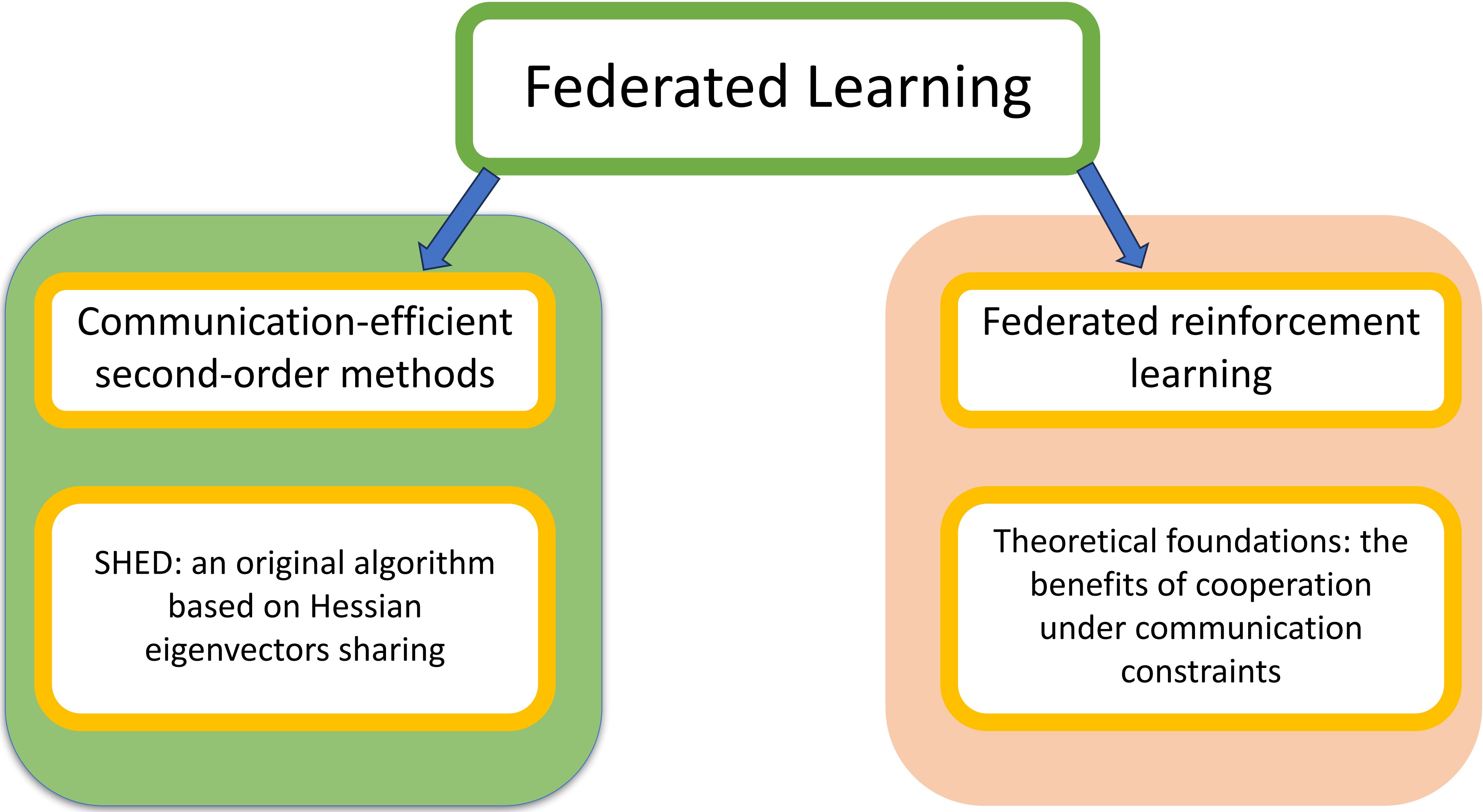


(a) FMNIST i.i.d.
(b) FMNIST non i.i.d.
Relative cost
Communication round
10−1
10−3
10−5
10−7
10−9
10−11
10−13
0
250
500
750
1000
1250
1500
450
900
1350
1800
2250
2700
Fib-SHED
GN-SHED
Fib-SHED+
FedNL
AGD
GIANT
Mont-Dec



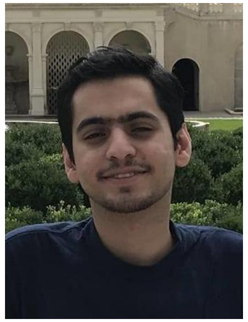
Required computations of the Hessian
FMNIST, $\mu = 10^{-5}$
FMNIST, $\mu = 10^{-6}$
EMNIST, $\mu = 10^{-5}$
EMNIST, $\mu = 10^{-6}$
w8a, $\mu = 10^{-5}$
w8a, $\mu = 10^{-6}$
10^0
10^1
10^2
10^3
Fib-SHED
FedNL

This thesis

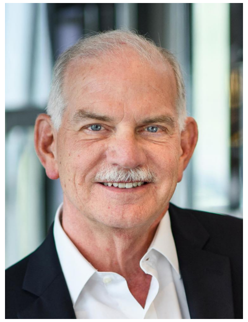

Federated Learning
Communication-efficient second-order methods
SHED: an original algorithm based on Hessian eigenvectors sharing
Federated reinforcement learning
Theoretical foundations: the benefits of cooperation under communication constraints

Aritra Mitra
NCState

George Pappas
Upenn



Arman Adibi
Princeton

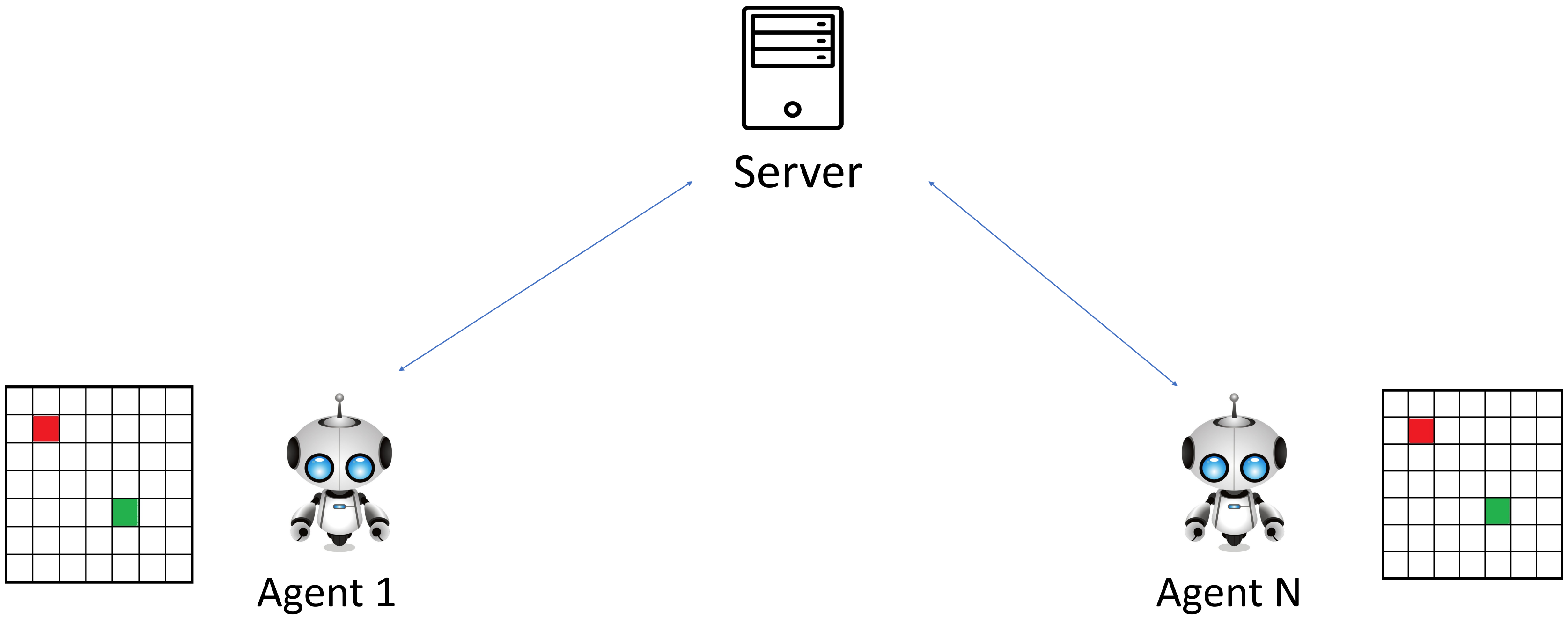


Vince Poor
Princeton

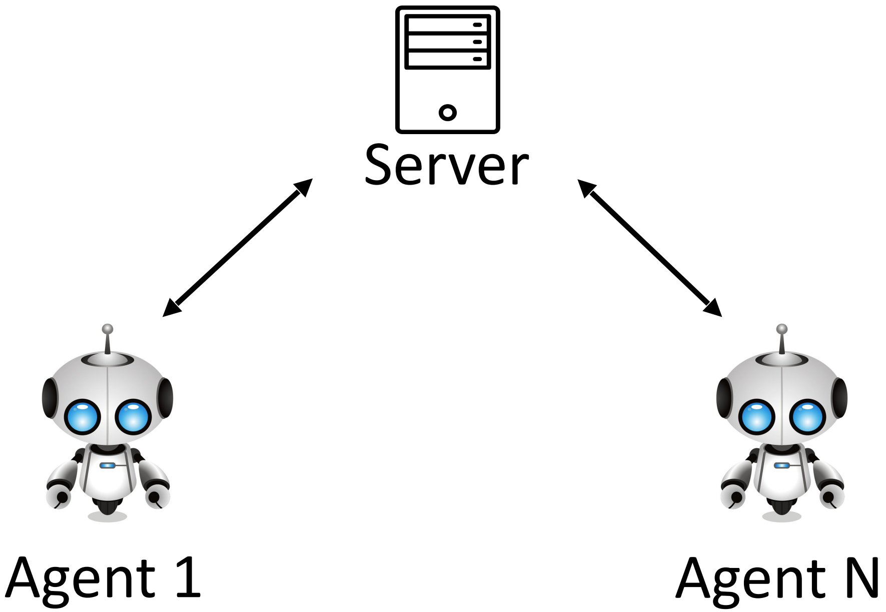
Hamed Hassani
Upenn

Sanjeev Kulkarni
Princeton

Federated reinforcement learning



Is it possible to provide finite-sample analysis for federated reinforcement learning under communication constraints?

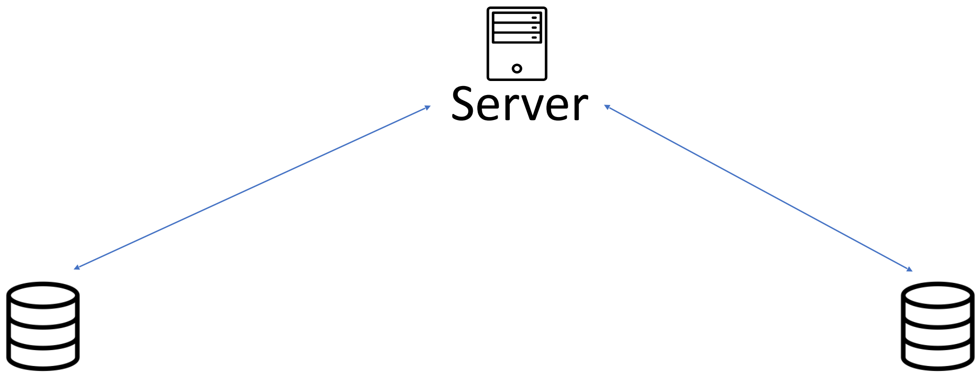


- **Goal**
  - Finite-sample convergence guarantees
  - Achieve a linear convergence speedup w.r.t. the number of agents N
- **Challenges**
  - Markovian sampling
  - Communication constraints

  (e.g., wireless networks)

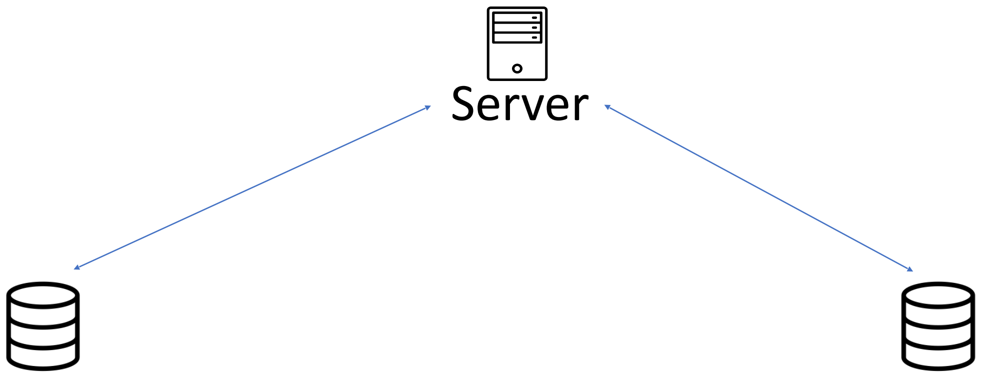
Related work and novelty

Federated Learning



Related work and novelty

Federated Learning



Distributed optimization under communication constraints

Li, T., Sahu, A. K., Talwalkar, A., and Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine*, *37*(3), 50-60.

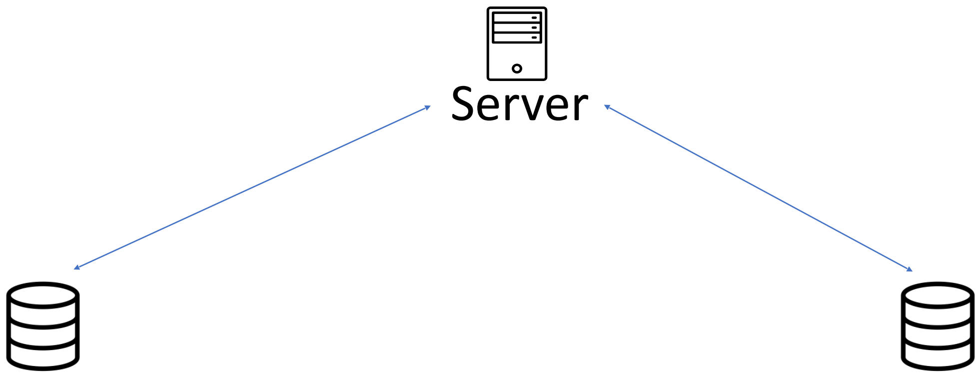
Amiri, Mohammad Mohammadi, and Deniz Gündüz. "Federated learning over wireless fading channels." *IEEE Transactions on Wireless Communications* 19.5 (2020): 3546-3557.

Konečný, Jakub, et al. "Federated learning: Strategies for improving communication efficiency." *arXiv preprint arXiv:1610.05492* (2016).

Chen, Mingzhe, et al. "A joint learning and communications framework for federated learning over wireless networks." *IEEE Transactions on Wireless Communications* (2020)

Related work and novelty

Federated Learning



Distributed optimization under communication constraints

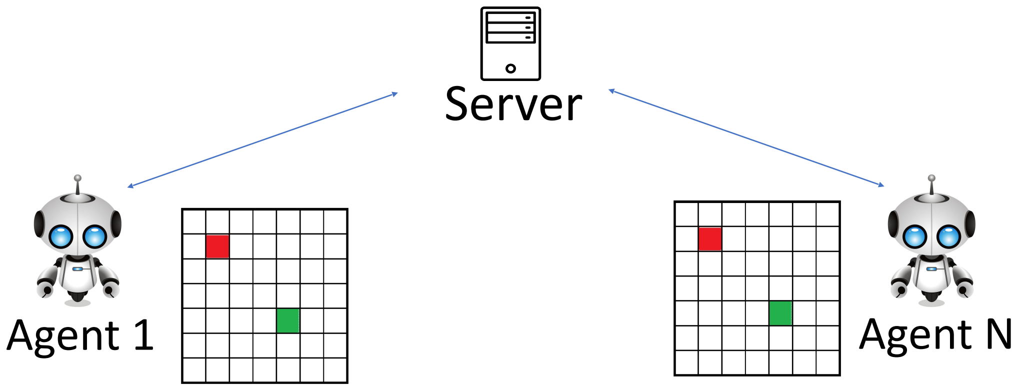
Li, T., Sahu, A. K., Talwalkar, A., and Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine*, *37*(3), 50-60.

Amiri, Mohammad Mohammadi, and Deniz Gündüz. "Federated learning over wireless fading channels." *IEEE Transactions on Wireless Communications* 19.5 (2020): 3546-3557.

Konečný, Jakub, et al. "Federated learning: Strategies for improving communication efficiency." *arXiv preprint arXiv:1610.05492* (2016).

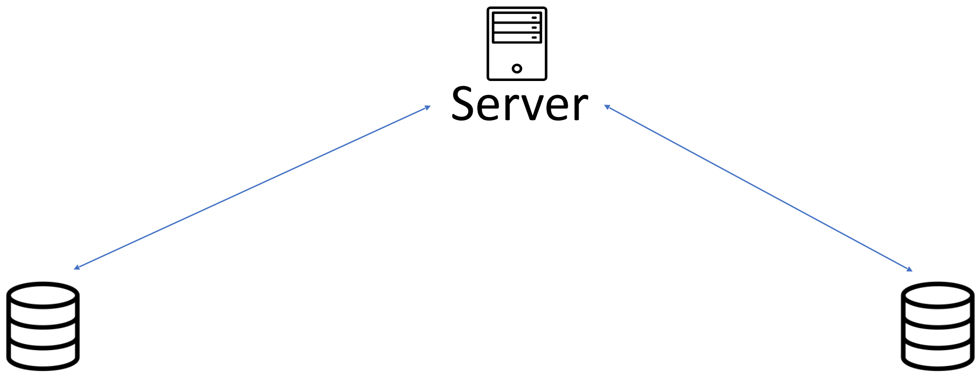
Chen, Mingzhe, et al. "A joint learning and communications framework for federated learning over wireless networks." *IEEE Transactions on Wireless Communications* (2020)

Federated Reinforcement Learning



Related work and novelty

Federated Learning



Distributed optimization under communication constraints

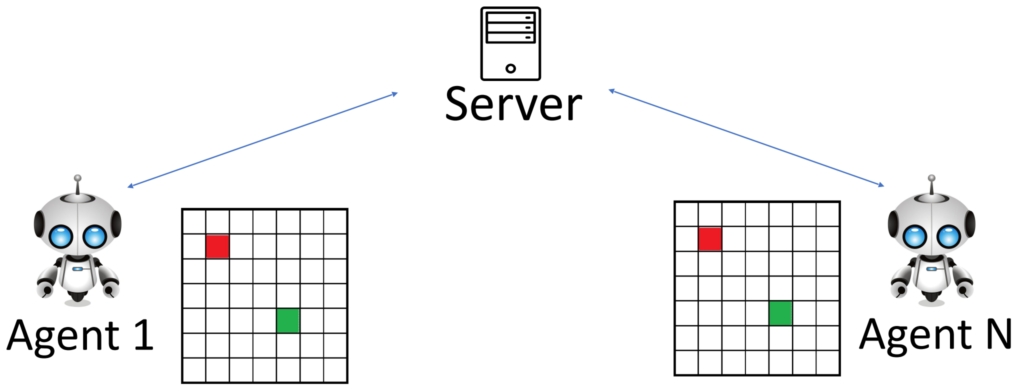
Li, T., Sahu, A. K., Talwalkar, A., and Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine*, *37*(3), 50-60.

Amiri, Mohammad Mohammadi, and Deniz Gündüz. "Federated learning over wireless fading channels." *IEEE Transactions on Wireless Communications* 19.5 (2020): 3546-3557.

Konečný, Jakub, et al. "Federated learning: Strategies for improving communication efficiency." *arXiv preprint arXiv:1610.05492* (2016).

Chen, Mingzhe, et al. "A joint learning and communications framework for federated learning over wireless networks." *IEEE Transactions on Wireless Communications* (2020)

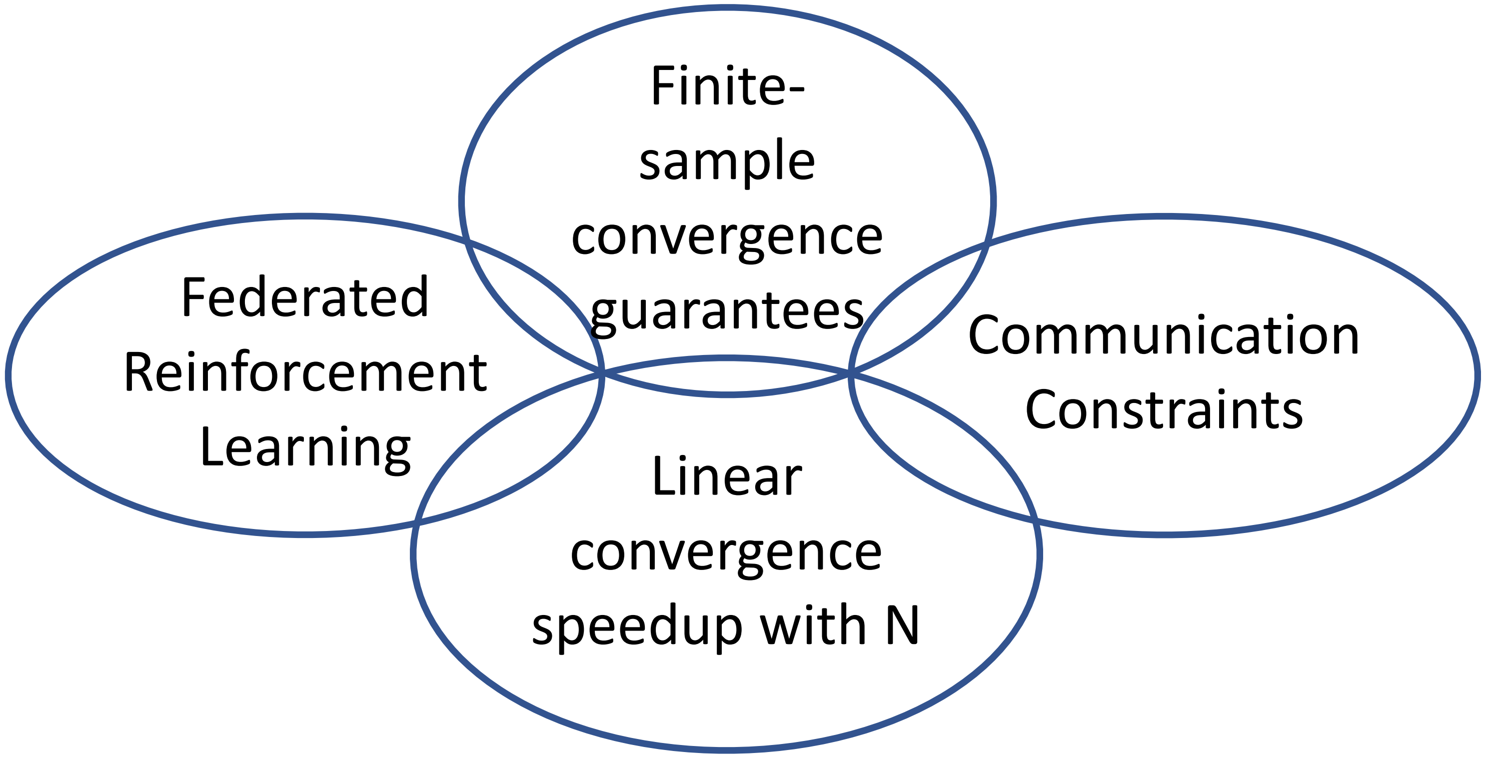
Federated Reinforcement Learning



?

Contributions

We provide the first finite-sample convergence analysis for federated reinforcement learning under communication constraints, establishing a linear convergence speedup with the number of agents



Temporal difference (TD) learning

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \alpha \mathbf{g}_k(\boldsymbol{\theta}_k, o_k)$$

Finite-sample analysis of this update rule under Markovian sampling has been recently established and provides

$O\left(\frac{1}{T}\right)$ approximation error after T iterations

Bhandari, Jalaj, Daniel Russo, and Raghav Singal. "A finite time analysis of temporal difference learning with linear function approximation." *Conference on learning theory*, 2018

Srikant, Rayadurgam, and Lei Ying. "Finite-time error bounds for linear stochastic approximation andtd learning." *Conference on Learning Theory*, 2019

Temporal difference (TD) learning

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \alpha \mathbf{g}_k(\boldsymbol{\theta}_k, o_k)$$

Finite-sample analysis of this update rule under Markovian sampling has been recently established and provides

$O\left(\frac{1}{T}\right)$ approximation error after T iterations

In federated RL with N agents, can we obtain

$O\left(\frac{1}{NT}\right)$ error in FRL under **communication constraints?**

Bhandari, Jalaj, Daniel Russo, and Raghav Singal. "A finite time analysis of temporal difference learning with linear function approximation." *Conference on learning theory*, 2018

Srikant, Rayadurgam, and Lei Ying. "Finite-time error bounds for linear stochastic approximation andtd learning." *Conference on Learning Theory*, 2019

Temporal difference (TD) learning

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \alpha \mathbf{g}_k(\boldsymbol{\theta}_k, o_k)$$

Finite-sample analysis of this update rule under Markovian sampling has been recently established and provides

$O\left(\frac{1}{T}\right)$ approximation error after T iterations

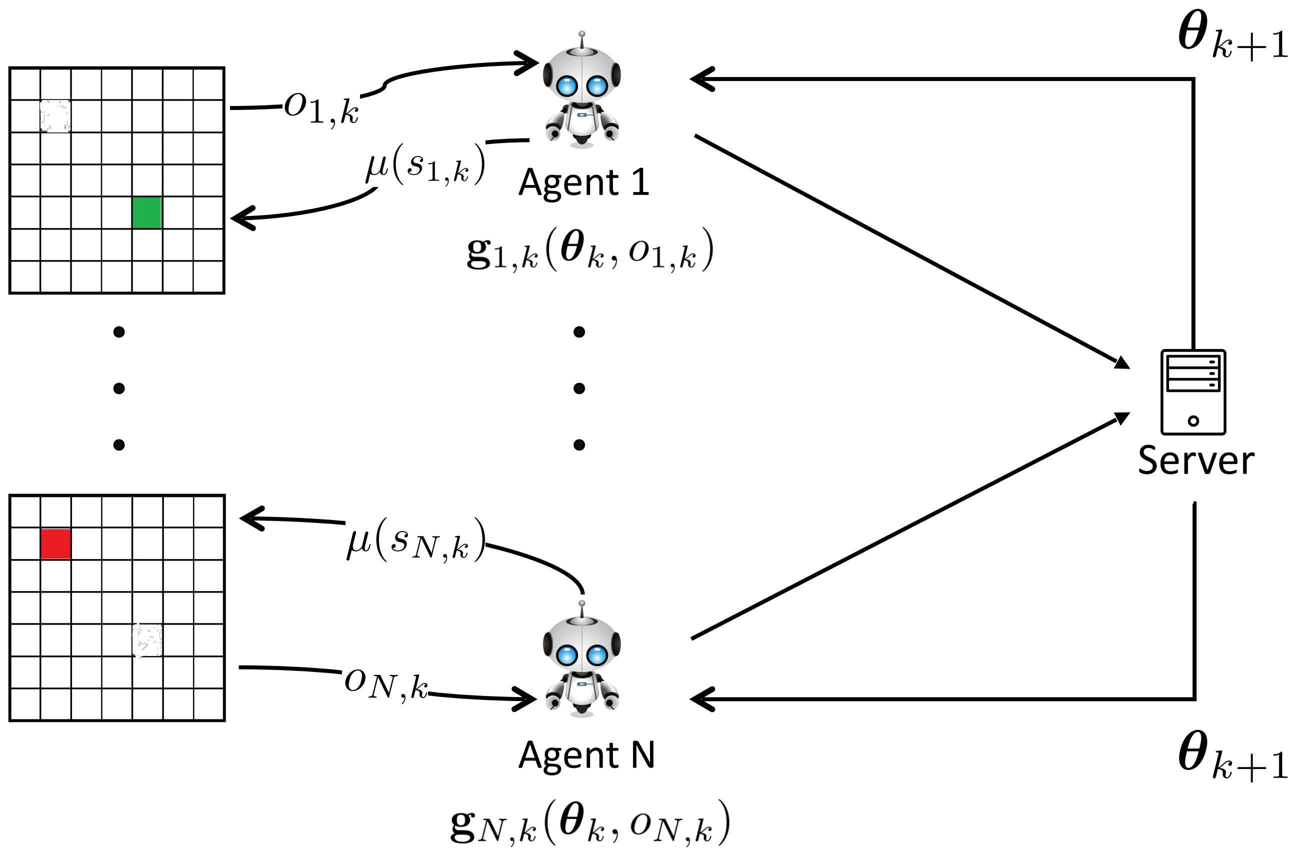
In federated RL with N agents, can we obtain

$O\left(\frac{1}{NT}\right)$ error in FRL under **communication constraints?**

Bhandari, Jalaj, Daniel Russo, and Raghav Singal. "A finite time analysis of temporal difference learning with linear function approximation." *Conference on learning theory*, 2018

Srikant, Rayadurgam, and Lei Ying. "Finite-time error bounds for linear stochastic approximation andtd learning." *Conference on Learning Theory*, 2019

Federated TD learning



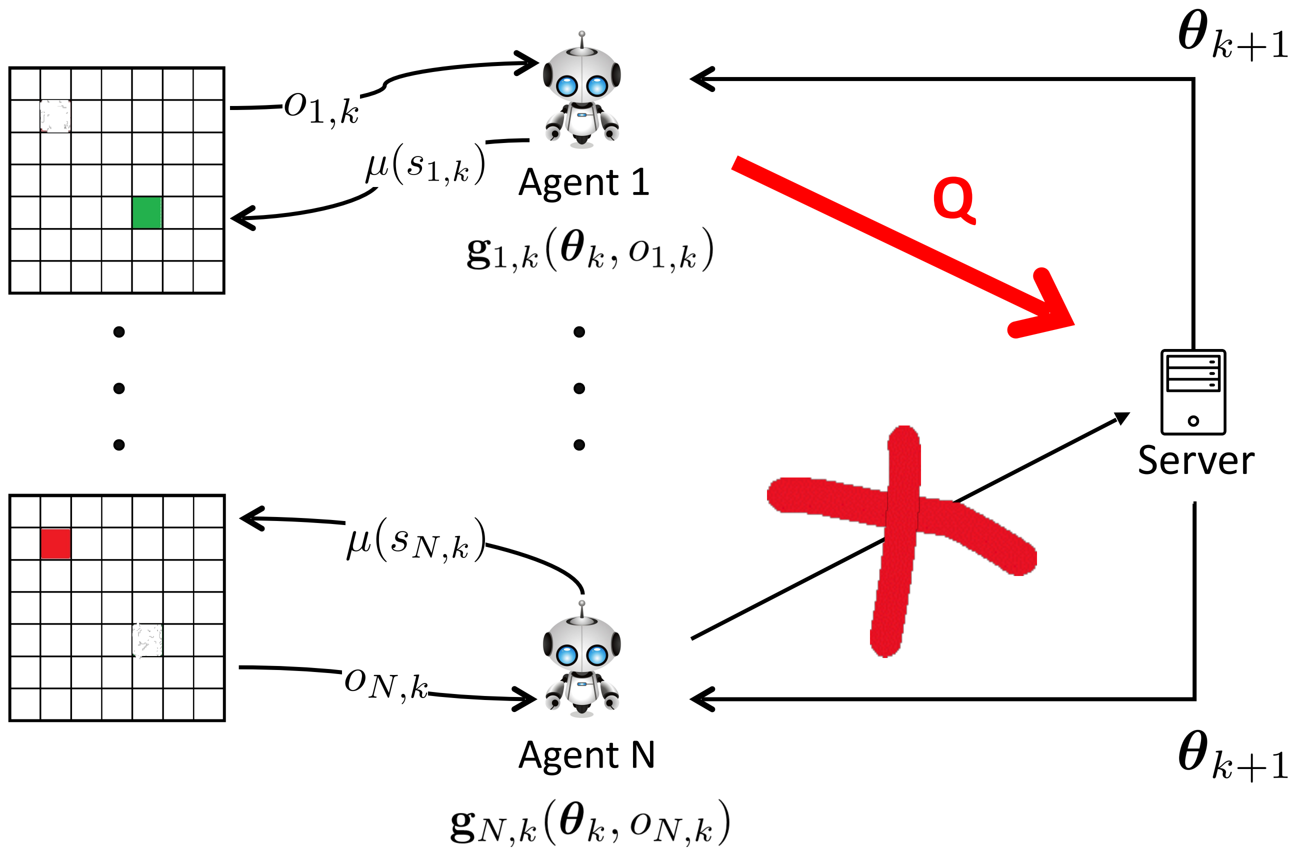
$\mathbf{g}_{i,k}(\boldsymbol{\theta}_k, o_{i,k})$

TD update direction of agent i at iteration k

$o_{i,k} = (s_{i,k}, r_{i,k}, s_{i,k+1})$

Observation of agent i at iteration k

Federated TD learning over quantized communication: QFedTD



$\mathbf{g}_{i,k}(\boldsymbol{\theta}_k, o_{i,k})$

TD update direction of agent i at iteration k

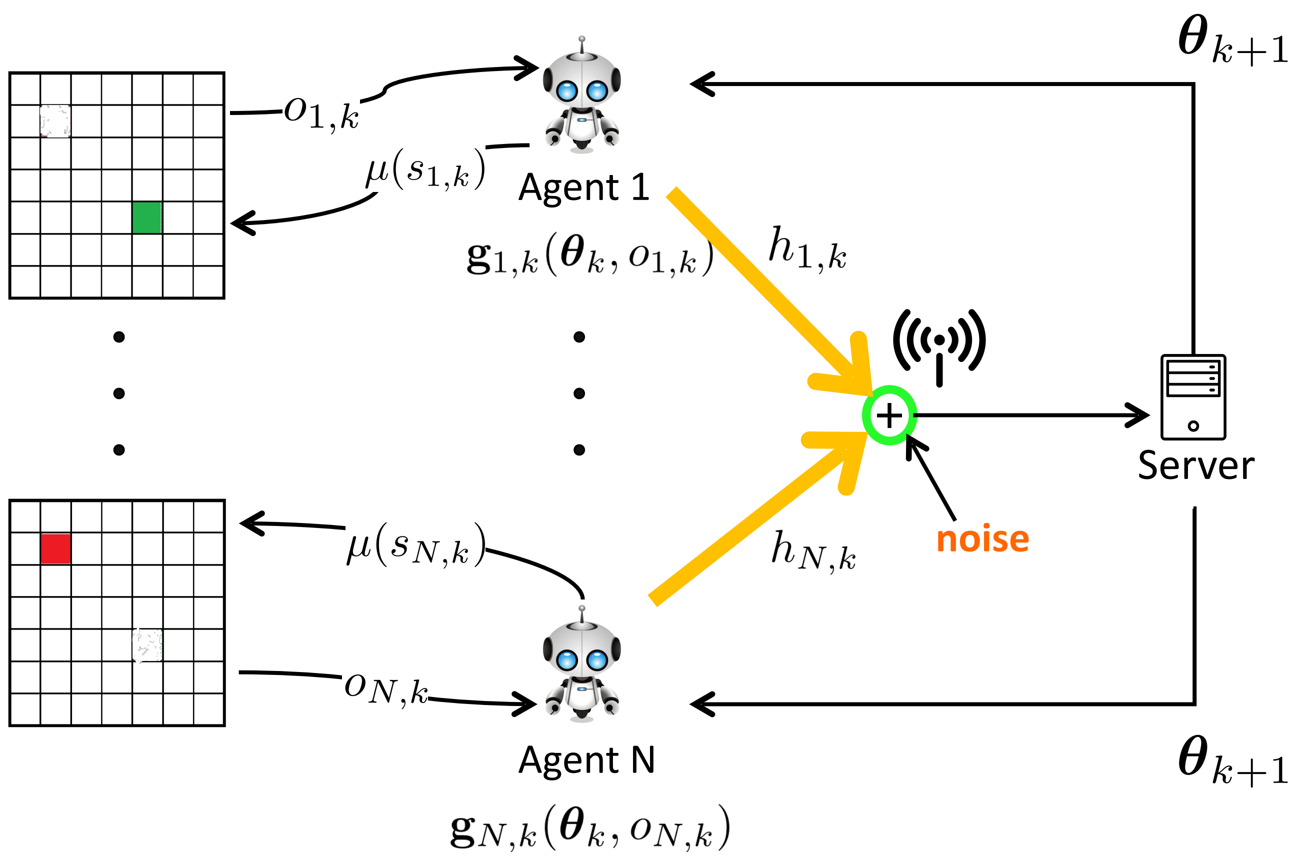
$o_{i,k} = (s_{i,k}, r_{i,k}, s_{i,k+1})$

Observation of agent i at iteration k

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \alpha \mathbf{v}_k \qquad \mathbf{v}_k = \frac{1}{N}\sum_{i=1}^{N} b_{i,k}\mathcal{Q}\left(\mathbf{g}_{i,k}(\boldsymbol{\theta}_k, o_{i,k})\right)$$

N. Dal Fabbro, A. Mitra, G. J. Pappas, "Federated TD Learning over Finite-Rate Erasure Channels: Linear Speedup under Markovian Sampling". IEEE Control Systems Letters, 2023

Over-the-air federated TD learning: OAC-FedTD



$\mathbf{g}_{i,k}(\boldsymbol{\theta}_k, o_{i,k})$

TD update direction of agent i at iteration k

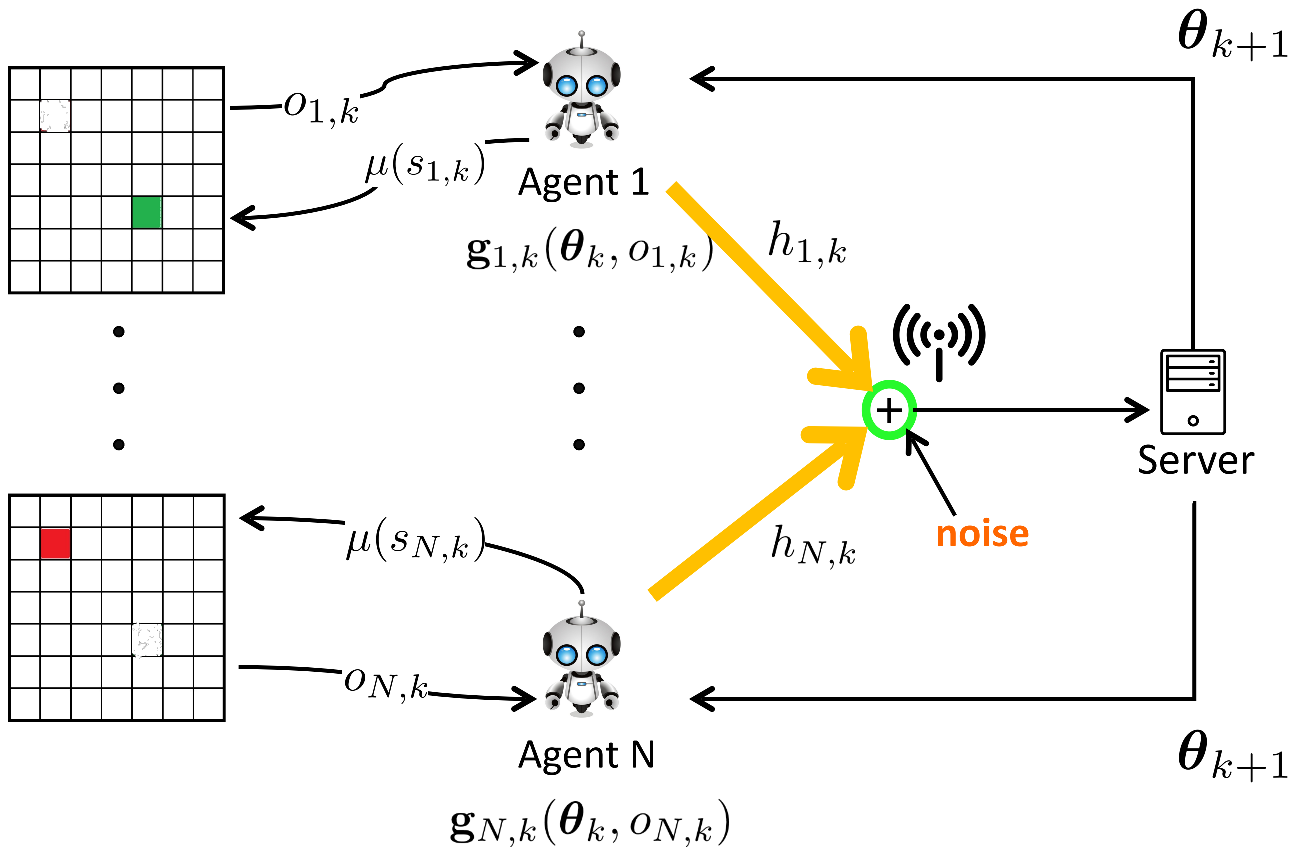
$o_{i,k} = (s_{i,k}, r_{i,k}, s_{i,k+1})$

Observation of agent i at iteration k

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \alpha \mathbf{v}_k \qquad \mathbf{v}_k = \frac{1}{N}\sum_{i=1}^{N} h_{i,k}\mathbf{g}_{i,k}(\boldsymbol{\theta}_k, o_{i,k}) + \mathbf{w}_k$$

N. Dal Fabbro, A. Mitra, R. W. Heath, L. Schenato, G. J. Pappas, "Over-the-Air Federated TD Learning" MLSys 2023 Workshop on Resource-Constrained Learning in Wireless Networks, Miami, Florida, 2023

Over-the-air federated TD learning: OAC-FedTD



$\mathbf{g}_{i,k}(\boldsymbol{\theta}_k, o_{i,k})$

TD update direction of agent i at iteration k

$o_{i,k} = (s_{i,k}, r_{i,k}, s_{i,k+1})$

Observation of agent i at iteration k

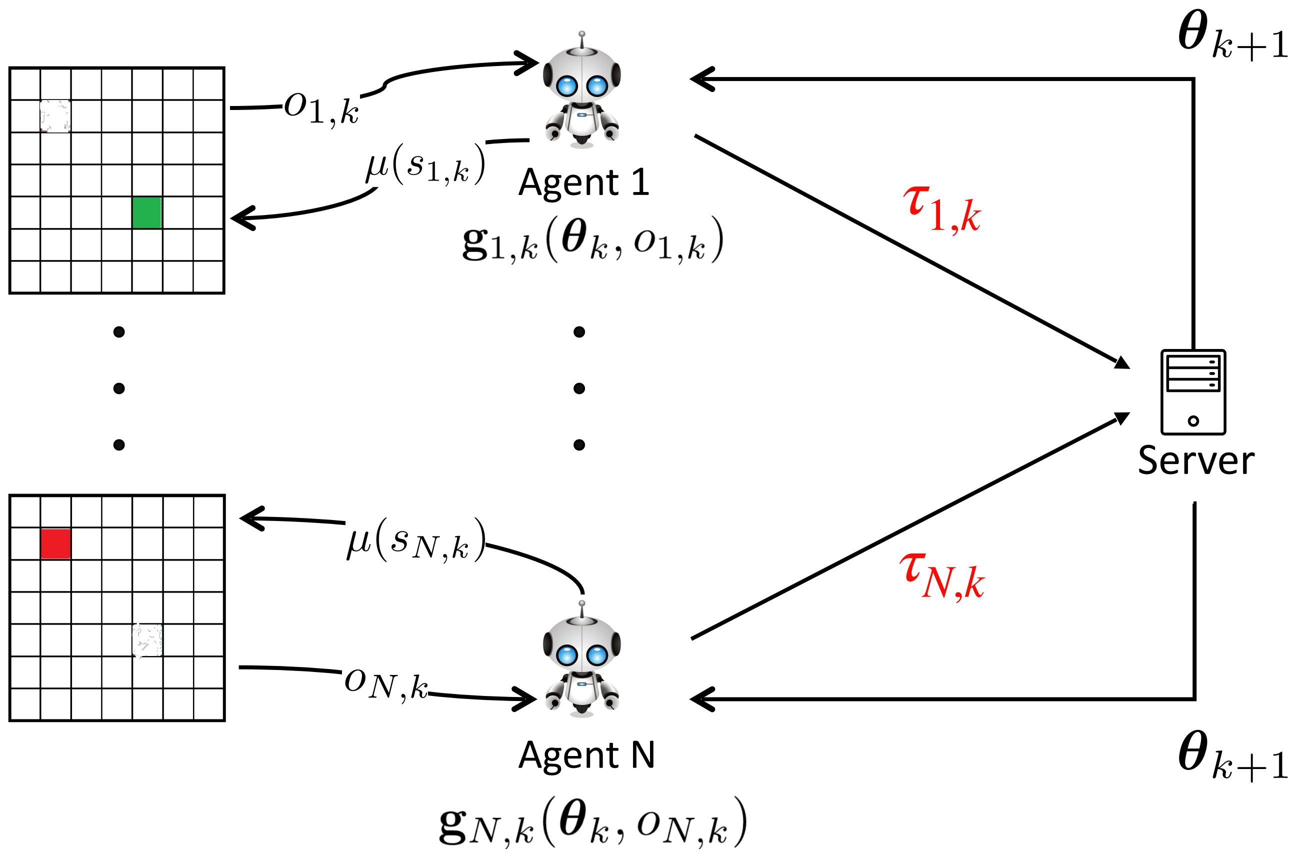
Channel distortion for agent i at iteration k

Measurement noise at the receiver

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \alpha \mathbf{v}_k \qquad \mathbf{v}_k = \frac{1}{N}\sum_{i=1}^{N} h_{i,k}\mathbf{g}_{i,k}(\boldsymbol{\theta}_k, o_{i,k}) + \mathbf{w}_k$$

N. Dal Fabbro, A. Mitra, R. W. Heath, L. Schenato, G. J. Pappas, "Over-the-Air Federated TD Learning" MLSys 2023 Workshop on Resource-Constrained Learning in Wireless Networks, Miami, Florida, 2023

Asynchronous multi-agent TD learning: AsyncMATD



$\mathbf{g}_{i,k}(\boldsymbol{\theta}_k, o_{i,k})$ TD update direction of agent i at iteration k

$o_{i,k} = (s_{i,k}, r_{i,k}, s_{i,k+1})$ observation of agent i at iteration k

delayed updates and observations

At the Server: $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \alpha \mathbf{v}_k$

$$\mathbf{v}_k = \frac{1}{N}\sum_{i=1}^{N} \mathbf{g}_{i,k}\boxed{\left(\boldsymbol{\theta}_{k-\tau_{i,k}}, o_{i,k-\tau_{i,k}}\right)}$$

Main Result

Theorem (convergence analysis of QFedTD)

If $\alpha > 0$ small enough, then the iterates of QFedTD are such that

$$\mathbb{E}\left[\|\boldsymbol{\theta}^* - \boldsymbol{\theta}_T\|^2\right] \leq \underbrace{C_1(1-\alpha p C_0)^T}_{\text{bias term}} + \underbrace{O\left(\frac{\zeta\alpha\tau\sigma^2}{N}\right)}_{\text{statistical terms}} + \underbrace{(\ldots)}_{\text{negligible terms}}$$

Main takeaways: we show the impact of the **channel effects** ζ and p on the convergence. We establish a linear convergence speedup with the number of agents N. We obtain a linear dependence on the mixing time of the Markov chain, τ.

Main Result

Theorem (convergence analysis of QFedTD)

If $\alpha > 0$ small enough, then the iterates of QFedTD are such that

$$\mathbb{E}\left[\|\boldsymbol{\theta}^* - \boldsymbol{\theta}_T\|^2\right] \leq \underbrace{C_1(1-\alpha p C_0)^T}_{\text{bias term}} + \underbrace{O\left(\frac{\zeta\alpha\tau\sigma^2}{N}\right)}_{\text{statistical terms}} + \underbrace{(\ldots)}_{\text{negligible terms}}$$

Main takeaways: we show the impact of the **channel effects** ζ and p on the convergence. We establish a linear convergence speedup with the number of agents N. We obtain a linear dependence on the mixing time of the Markov chain, τ .

Main Result

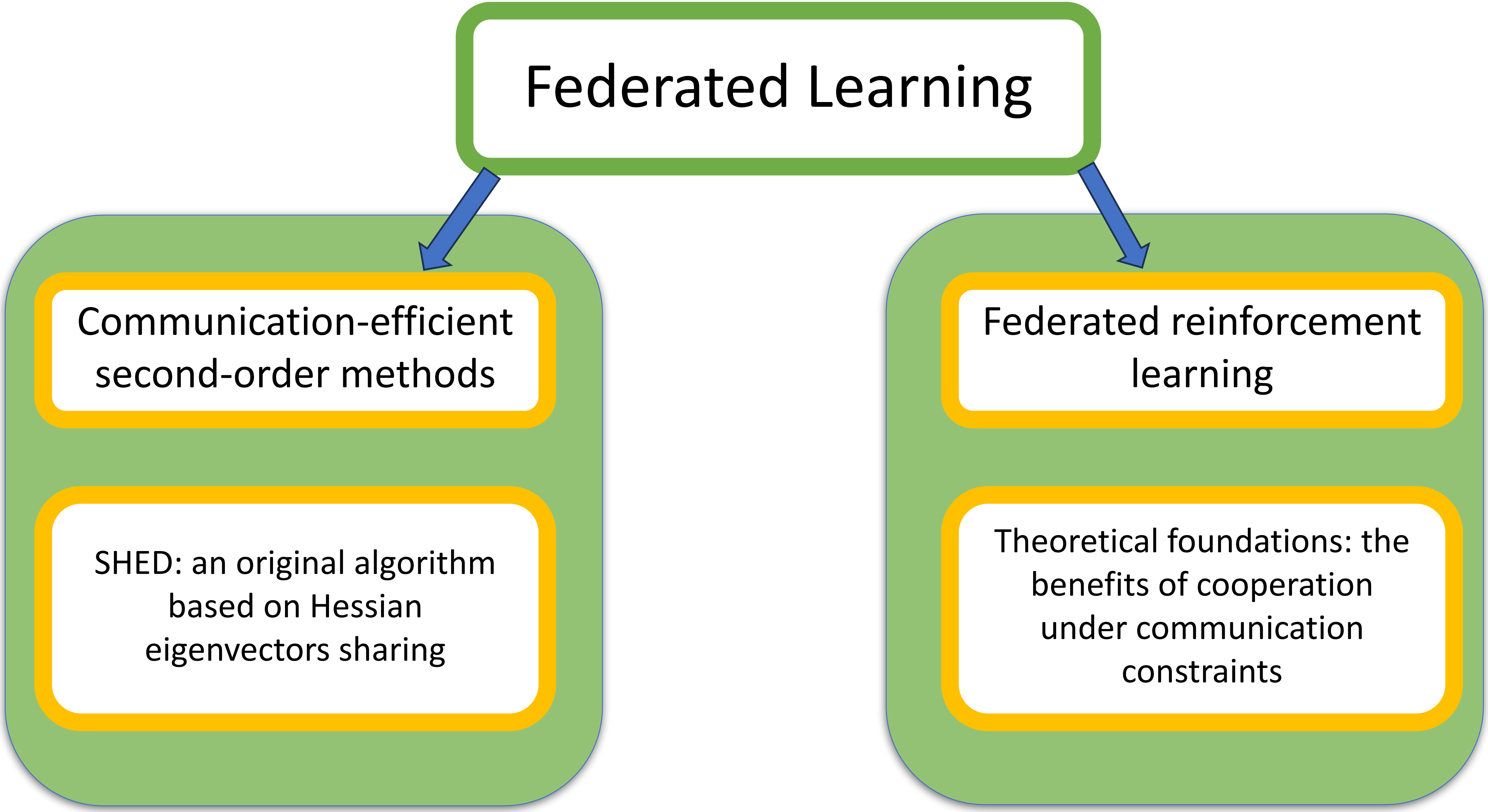
Theorem (convergence analysis of QFedTD)

If $\alpha > 0$ small enough, then the iterates of QFedTD are such that

$$\mathbb{E}\left[\|\boldsymbol{\theta}^* - \boldsymbol{\theta}_T\|^2\right] \leq \underbrace{C_1(1-\alpha p C_0)^T}_{\text{bias term}} + \underbrace{O\left(\frac{\zeta\alpha\tau\sigma^2}{N}\right)}_{\text{statistical terms}} + \underbrace{(\ldots)}_{\text{negligible terms}}$$

Main takeaways: we show the impact of the **channel effects** ζ and p on the convergence. We establish a linear convergence speedup with the number of agents N. We obtain a linear dependence on the mixing time of the Markov chain, τ.

This thesis


Federated Learning
Communication-efficient second-order methods
SHED: an original algorithm based on Hessian eigenvectors sharing
Federated reinforcement learning
Theoretical foundations: the benefits of cooperation under communication constraints

Future directions

Superlinear FL:

- Scaling up existing algorithms from a computational point of view

FRL:

- Heterogeneity/personalization, local optimization
    - Federated Multi-Agent reinforcement learning
        - sample complexity and communications **even more critical**

PhD thesis publications

Journals

N. Dal Fabbro, S. Dey, M. Rossi and L. Schenato, “SHED: A Newton-type algorithm for federated learning based on incremental Hessian eigenvector sharing”*, Automatica,* 2024

N. Dal Fabbro, A. Mitra, G. J. Pappas, "Federated TD Learning over Finite-Rate Erasure Channels: Linear Speedup under Markovian Sampling". *IEEE Control Systems Letters*, 2023

Conferences

N. Dal Fabbro, M. Rossi, L. Schenato, S. Dey “Q-SHED: Distributed Optimization at the Edge via Hessian Eigenvectors Quantization”*, IEEE International Conference on Communications,* Rome 2023

N. Dal Fabbro, A. Mitra, R. W. Heath, L. Schenato, G. J. Pappas, "Over-the-Air Federated TD Learning" *MLSys 2023* Workshop on Resource-Constrained Learning in Wireless Networks, Miami, Florida, 2023

A.Adibi, **N. Dal Fabbro**, L. Schenato, S. Kulkarni, H. V. Poor, G. J. Pappas, H. Hassani and A. Mitra. Stochastic Approximation with Delayed Updates: Finite-Time Rates under Markovian Sampling, *AISTATS*, 2024

N. Dal Fabbro, Arman Adibi, Aritra Mitra, George J. Pappas. Finite-Time Analysis of Asynchronous Multi-Agent TD Learning, *The 2024 American Control Conference (ACC)*, 2024

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Nicolò Dal Fabbro

Pushing the Boundaries of Federated Learning:

Superlinear Convergence and Reinforcement Learning over Wireless

GTTI meeting

September, 2024

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Advisors:
Prof. Luca Schenato
Prof. Michele Rossi

Thank you!