FedCBO: Reaching Group Consensus in Clustered Federated Learning and Robustness to Backdoor Adversarial Attacks

Nicolás García Trillos
University of Wisconsin-Madison

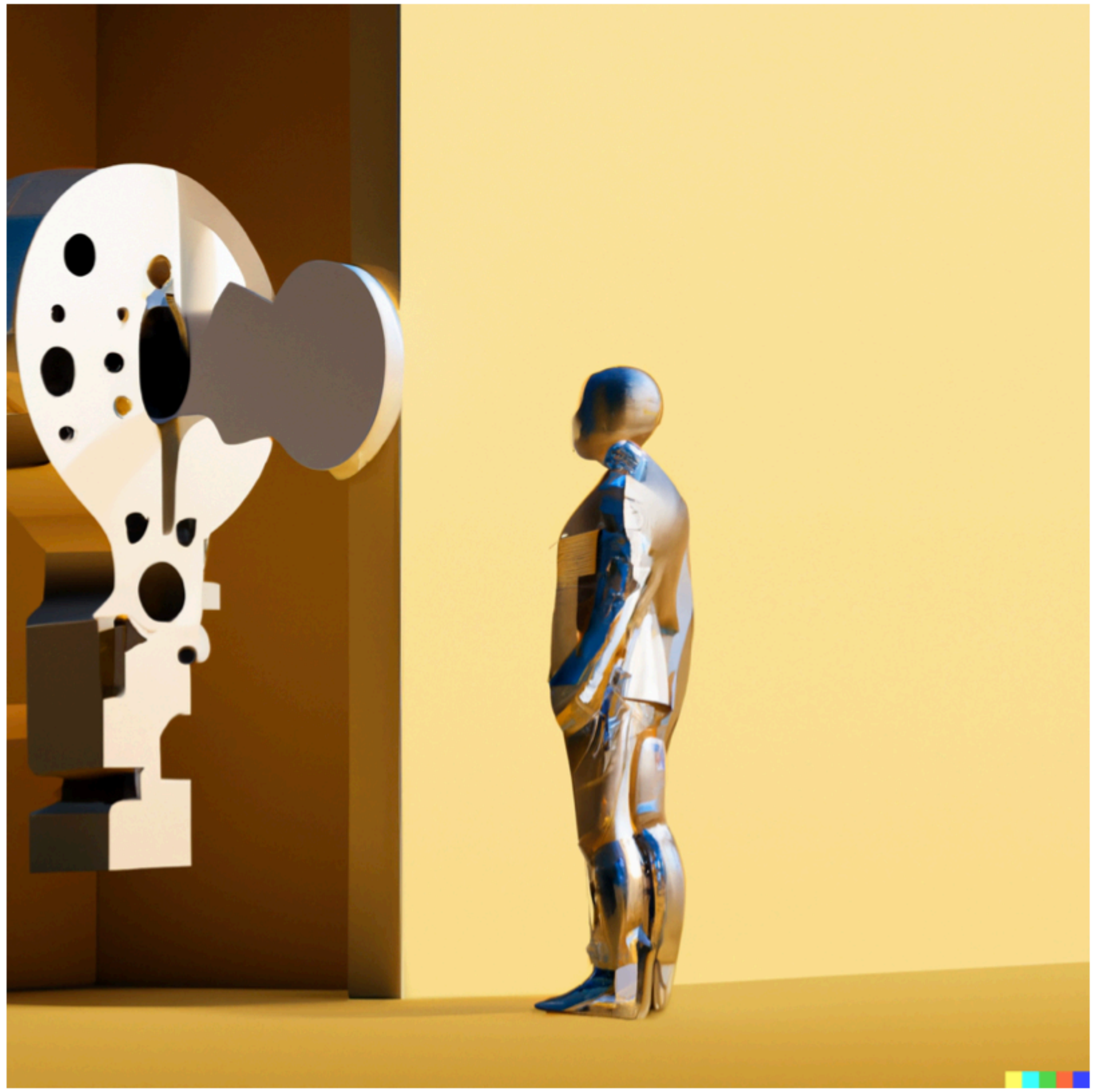
ICERM
May 2024

Based on

- J.A. Carrillo, **NGT**, S. Li, Y. Zhu "*FedCBO: Reaching Group Consensus in Clustered Federated Learning through Consensus-based Optimization*." https://arxiv.org/abs/2305.02894

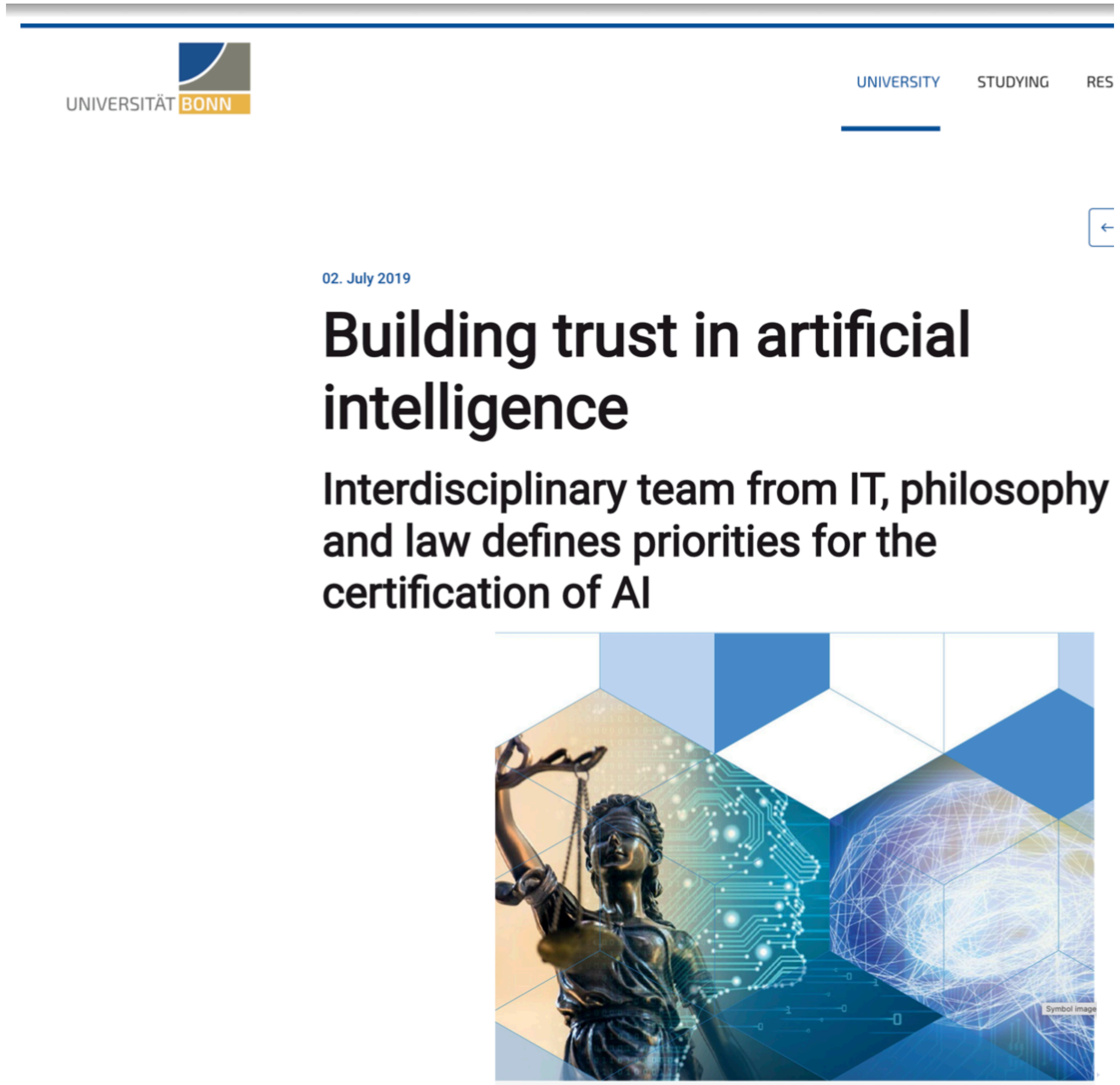
- **NGT,** S. Li, K. Riedl, Y. Zhu "*Bilevel Consensus-based Optimization and applications to backdoor attack robustness in Clustered Federated Learning*." (In preparation).

The success stories of AI



From Dall-e: generate an image representing the success stories of AI.

However…



Safety (e.g., systems should be robust to perturbations of data)

Security (e.g, systems should be robust to adversarial agents)

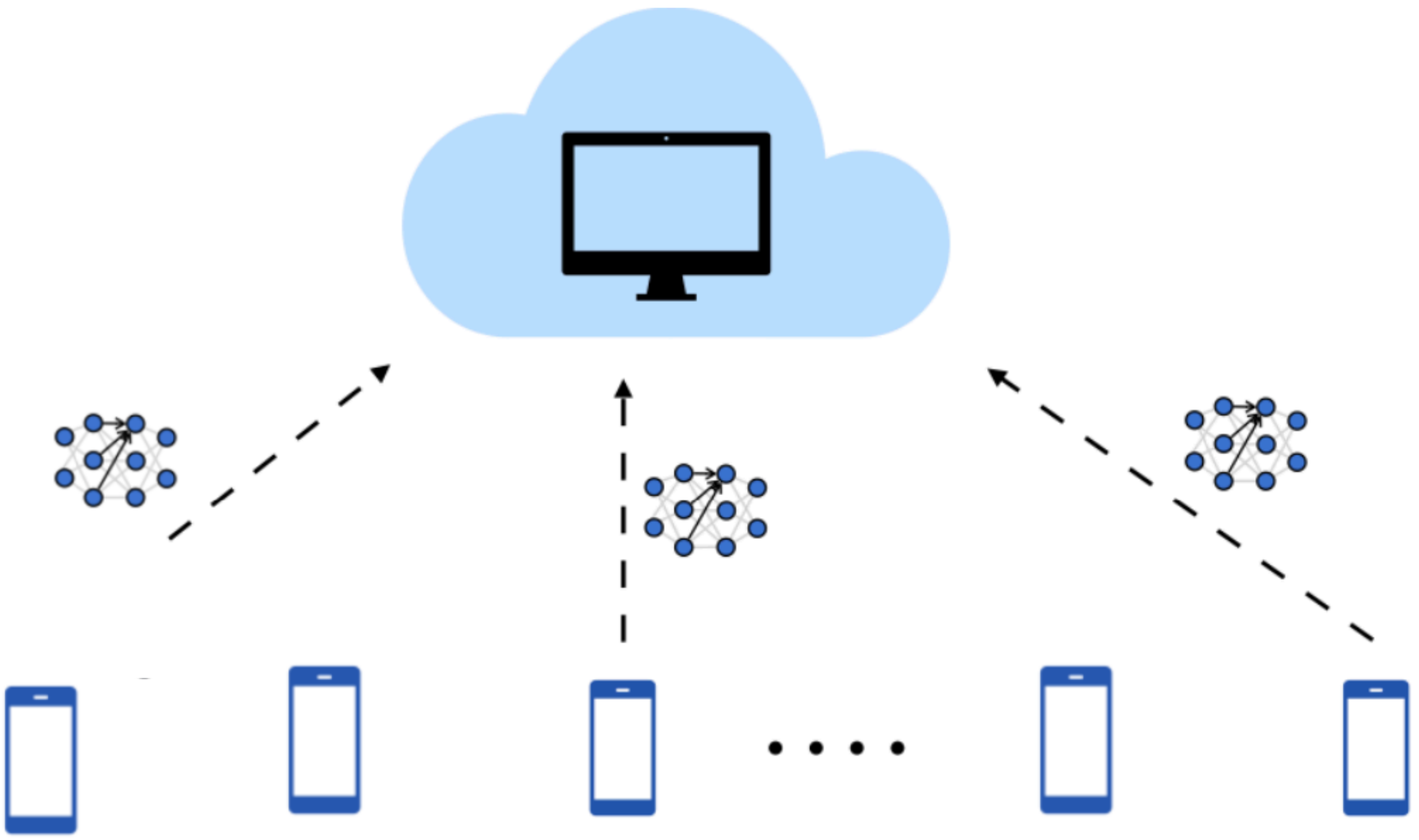
Trust (e.g., privacy, fairness, watermarks for generated images)

1.

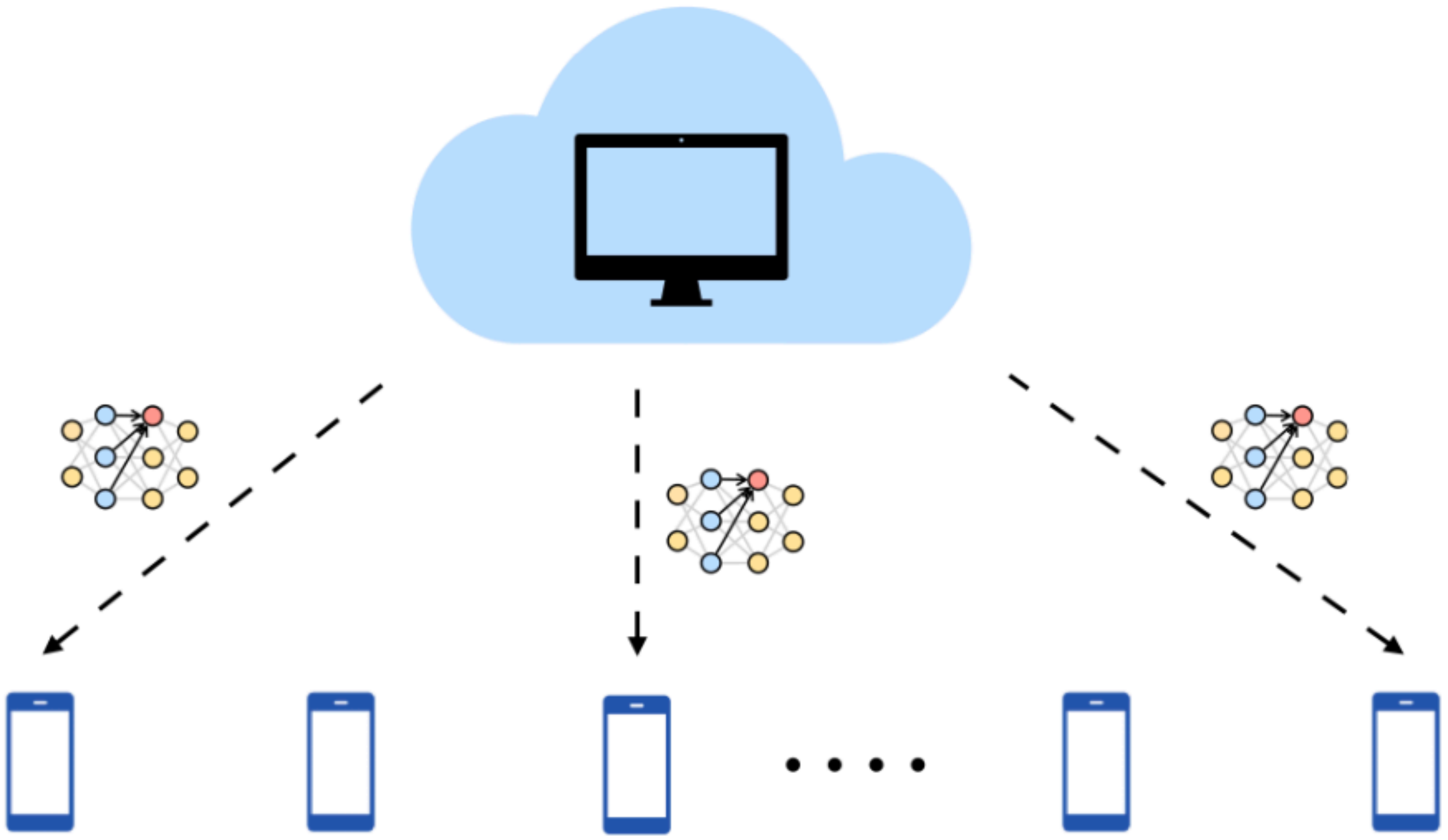
Outline

1. Introduction: federated learning and clustered federated learning
2. Federated learning through consensus based optimization (CBO)
3. Backdoor attacks.

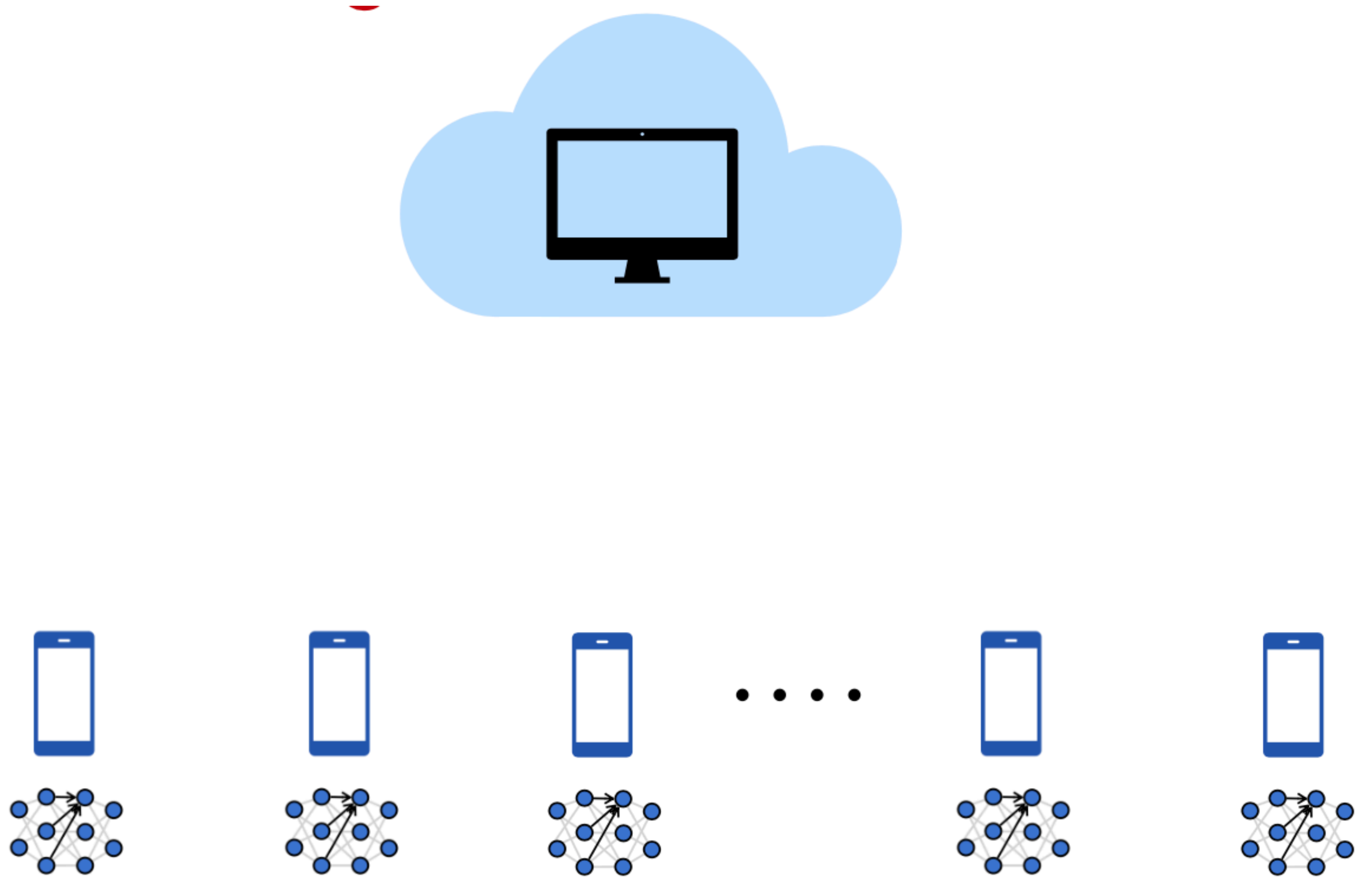
Federated Learning



Federated Learning



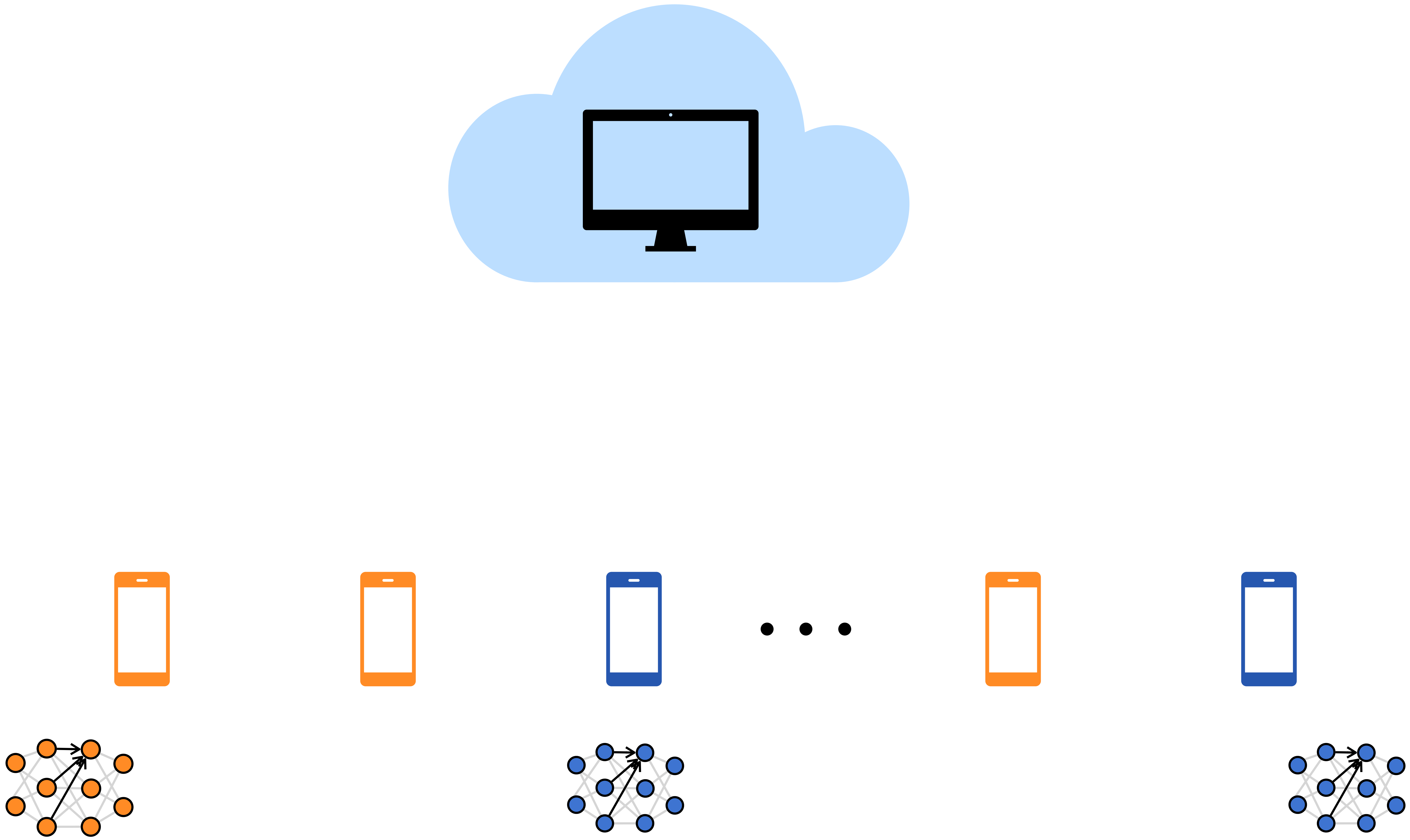
Federated Average

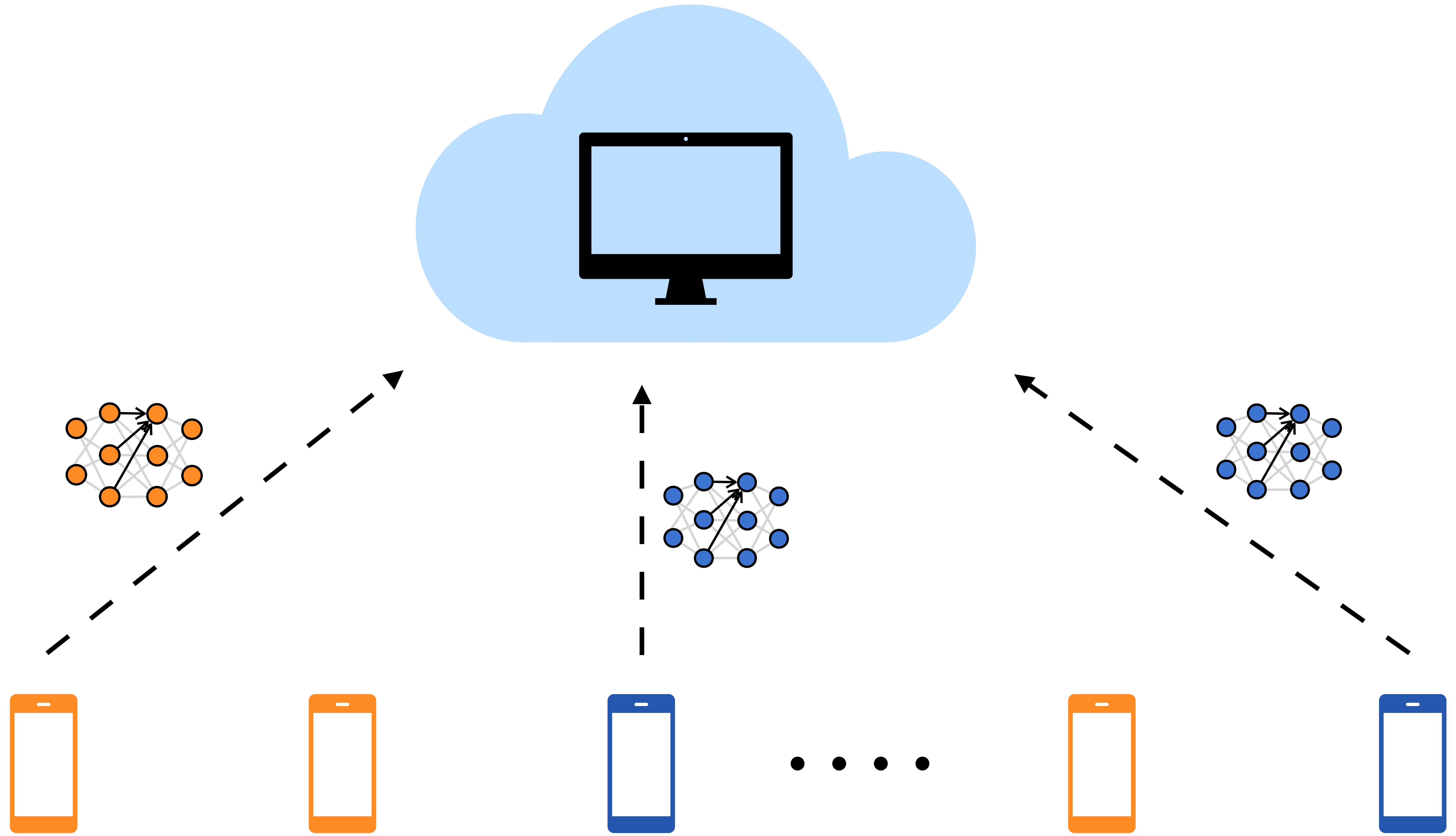


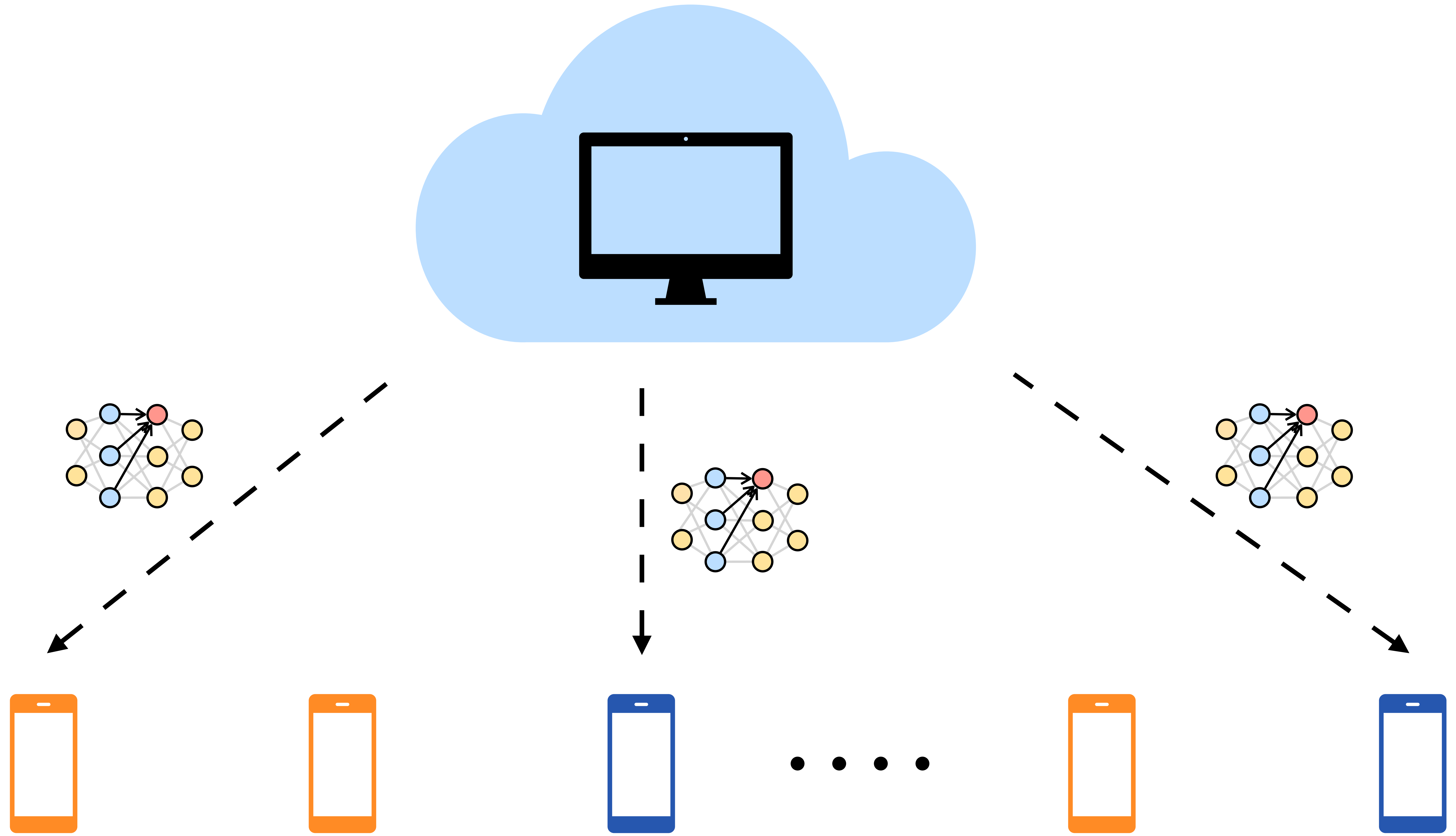
FedAvg [McMahan et al 16']:

$$\begin{cases} d\theta_t^i = -\nabla L_i(\overline{\theta}_t) + \text{Noise} , \quad i = 1, \ldots, N. \\ \overline{\theta}_t = \frac{1}{N} \sum_{i=1}^N \theta_t^i \end{cases}$$

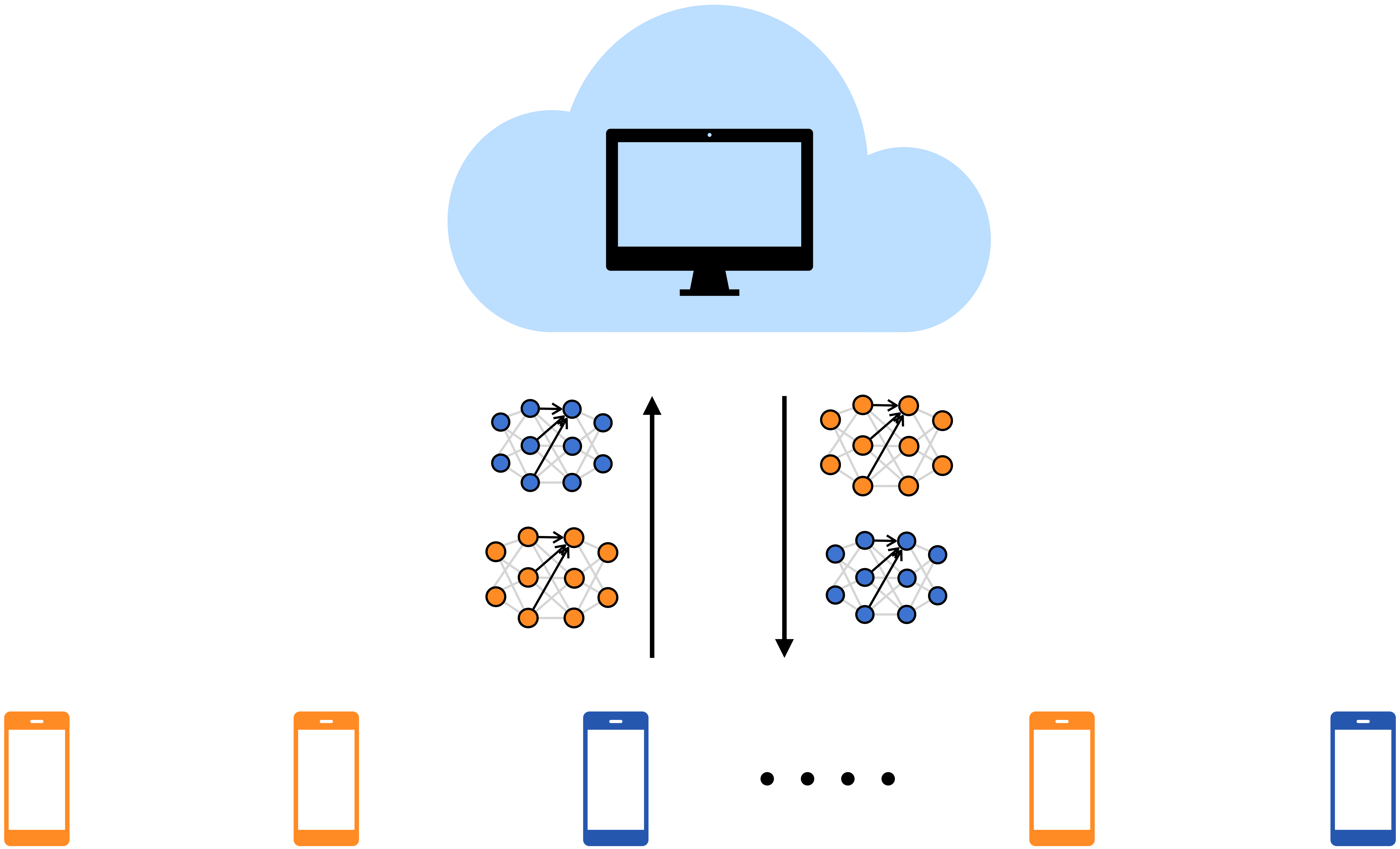
Federated Learning (heterogeneous setting)





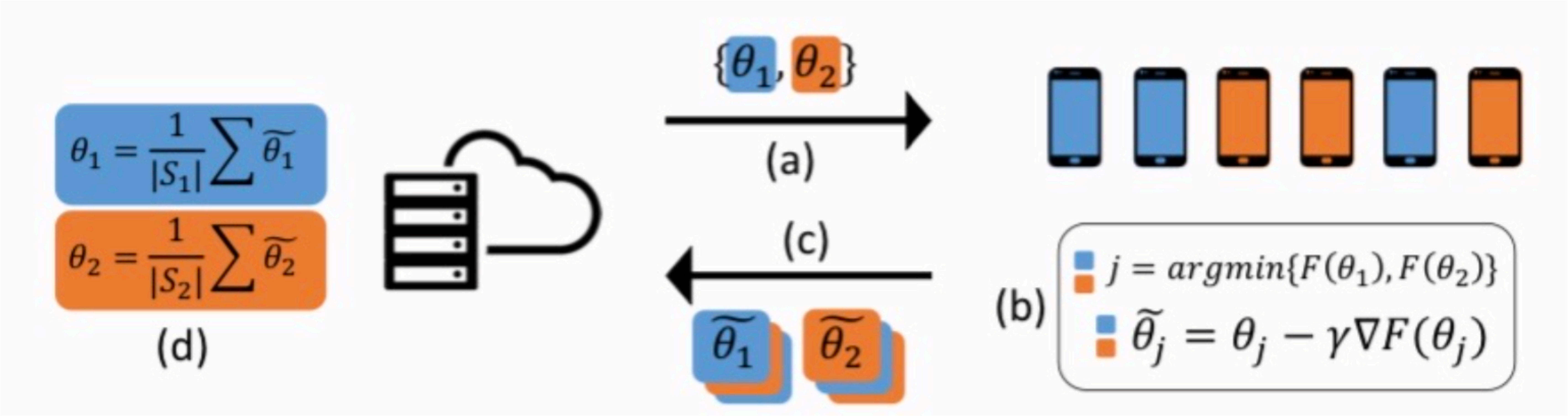


Clustered Federated Learning



Iterated Federated Clustering Algorithm

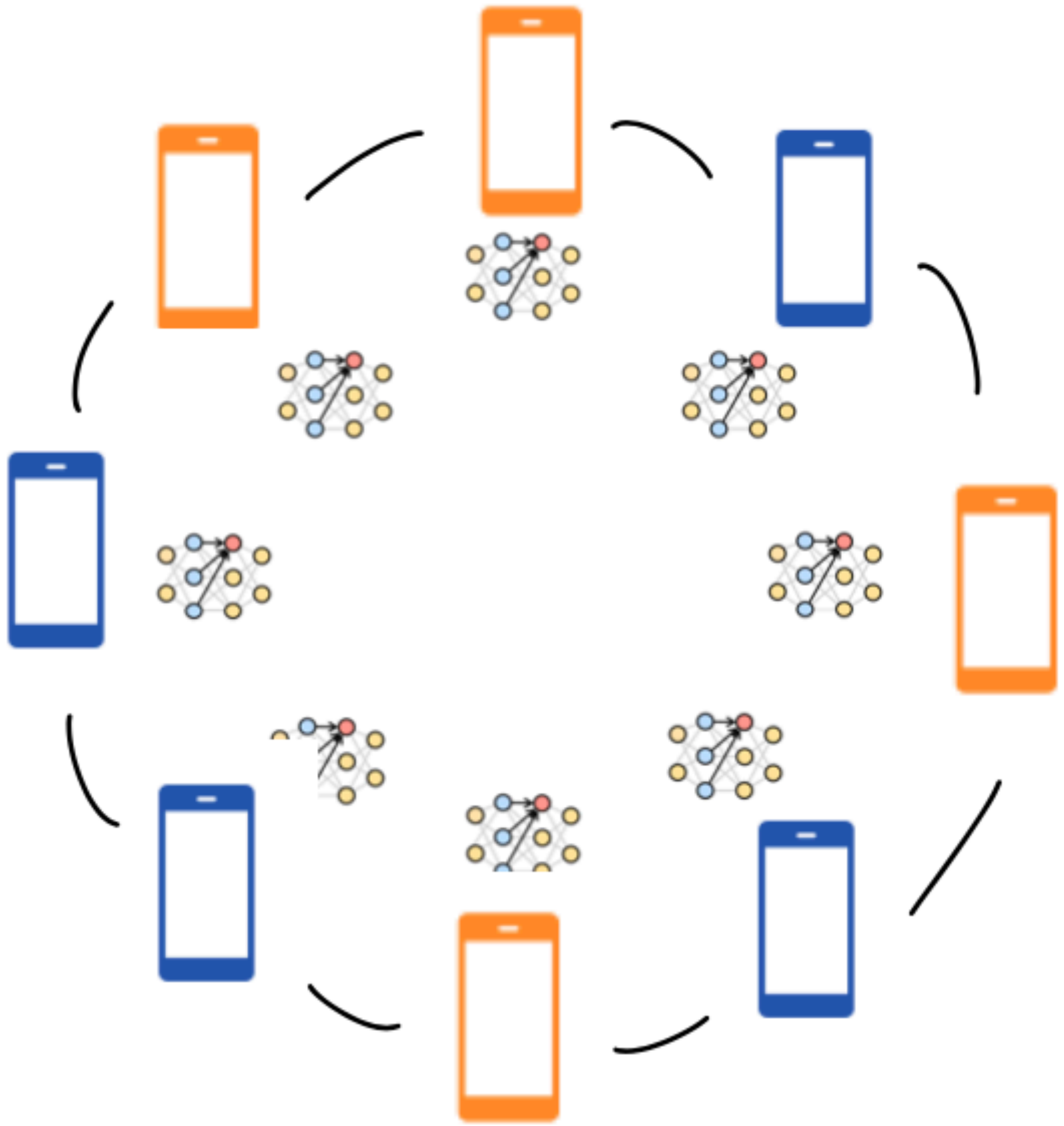
IFCA [Ghosh et al 20']:



Some References

- McMahan et al. *Communication-Efficient Learning of Deep Networks from Decentralized Data*, Proceedings of the 20th AISTATS, 2017.
- Karimireddy et al. *Scaffold: Stochastic controlled averaging for federated learning*, PMLR, 2020.
- Li et al. *Federated optimization in heterogeneous networks*, Proceedings of Machine learning and systems, 2020.
- M. Mohri, et al. *Agnostic federated learning*. ICML, 2019.
- A. Ghosh et al. *An efficient framework for clustered federated learning*, Neurips, 2020.
- G. Long et al. *Multi-center federated learning: clients clustering for better personalization*, World Wide Web, 2023.
- Y. Ruan and C. Joe-Wong. *Fedsoft: Soft clustered federated learning with proximal local updating*, Proceedings of the AAAI Conference on Artificial Intelligence, 2022.

Today: decentralized Clustered Federated Learning based on CBO



Setting for Clustered Federated Learning

Number of agents = N

Number of clusters = K

$$L_k(\theta) := \mathbb{E}_{(x,y)\sim\mathcal{D}_k}\left[l\left(f(x;\theta), y\right)\right], \qquad k = 1, 2, \ldots, K.$$

$$\theta_k^* := \underset{\theta\in\mathbb{R}^d}{\arg\min}\; L_k(\theta)$$

Part 1

Clustered Federated Learning through CBO

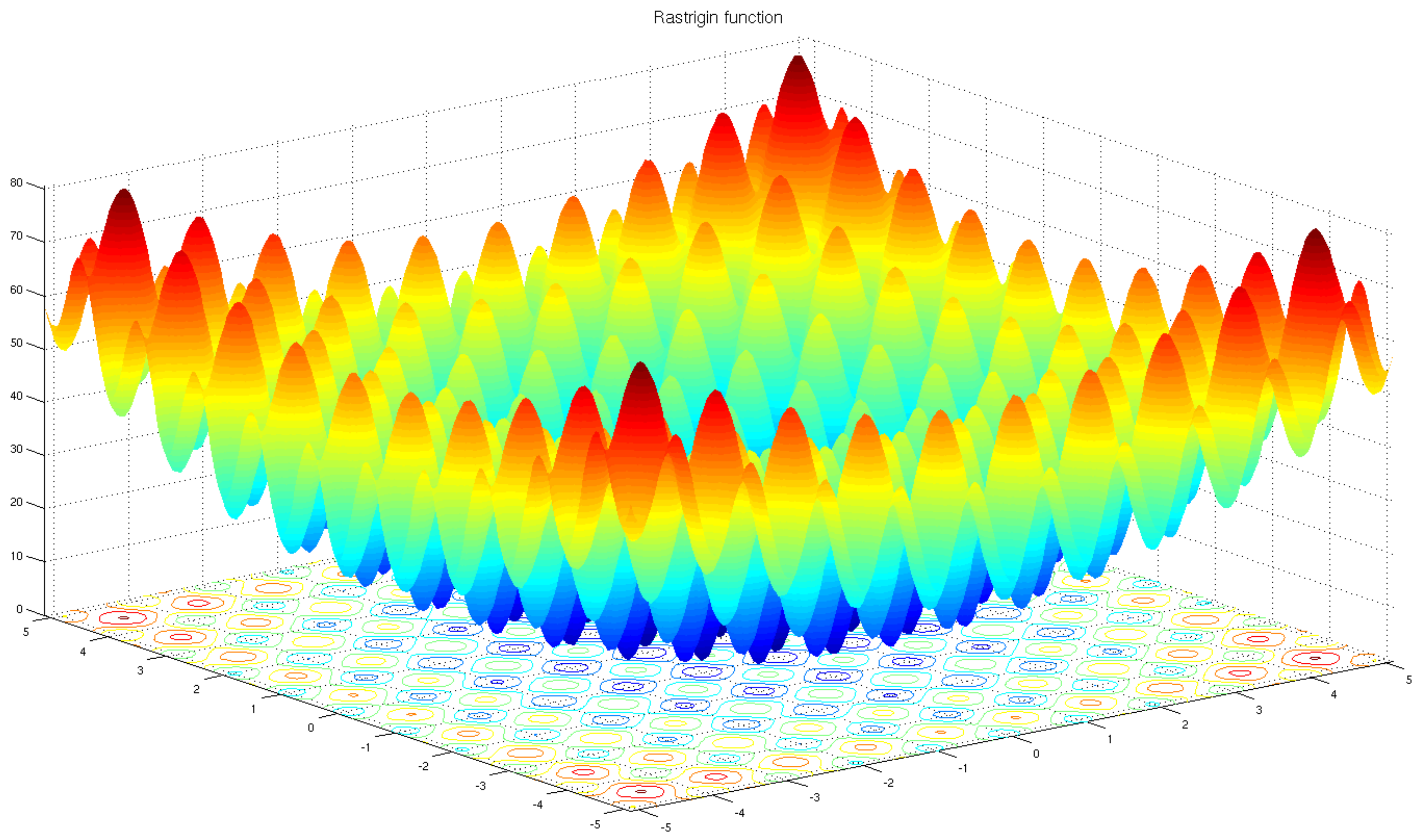
Consensus-based Optimization (CBO)

Optimization problem:

$$\min_{\theta \in \mathbb{R}^d} \; L(\theta)$$

Assumptions:

- L has unique global min θ^*.



Consensus-based Optimization (CBO)

Interacting particle system:

$$d\theta_t^i = -\lambda\left(\theta_t^i - m_L^\alpha[\rho_t^N]\right) dt + \sigma \left|\theta_t^i - m_L^\alpha[\rho_t^N]\right| dB_t^i, \qquad i = 1, 2, \ldots, N,$$

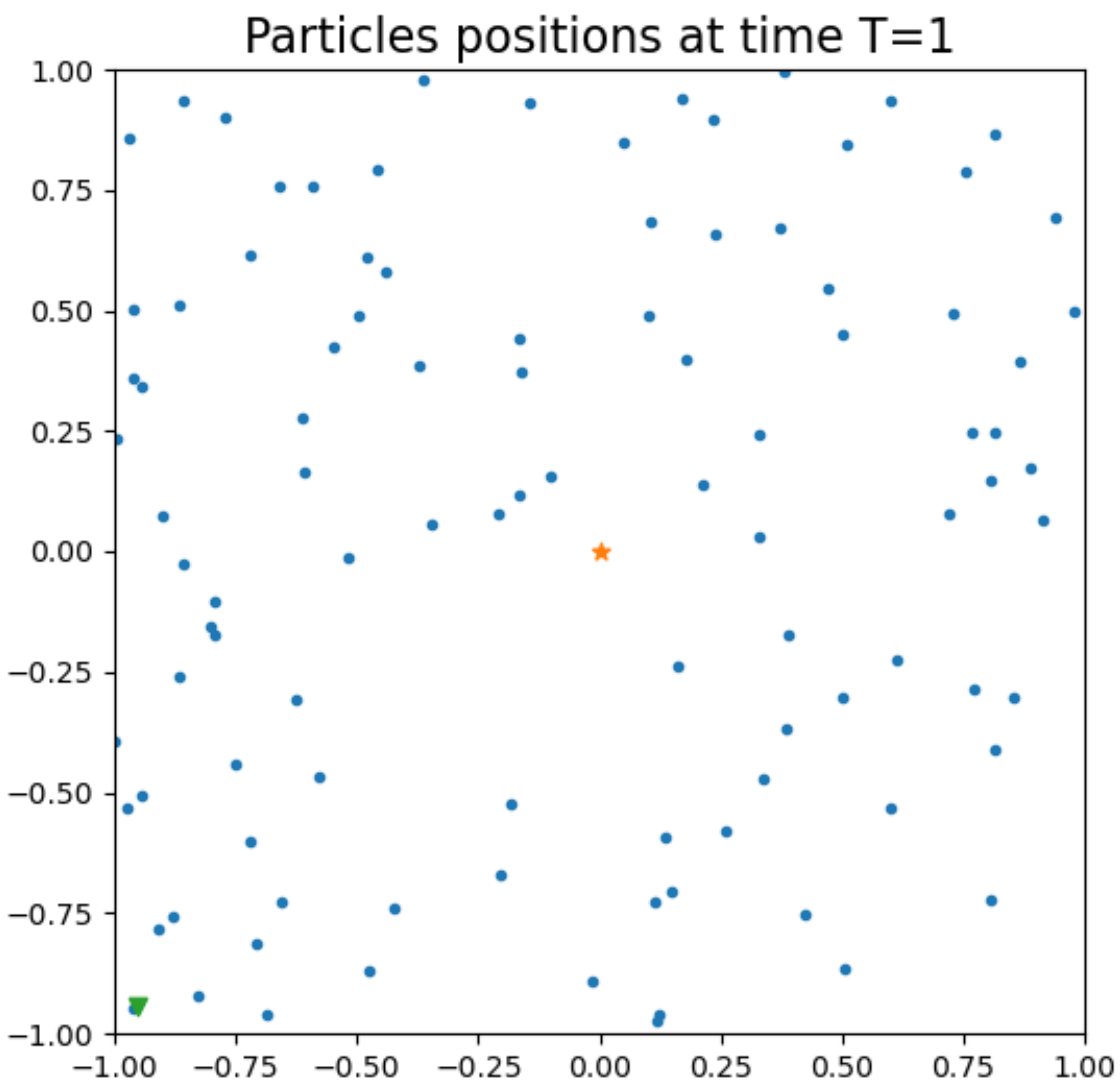
where

$$\rho_t^N := \frac{1}{N}\sum_{i=1}^{N} \delta_{\theta_t^i}$$

$$m_L^\alpha\left[\rho_t^N\right] := \int \theta \frac{\exp\left(-\alpha L(\theta)\right)\rho_t^N}{\int \exp\left(-\alpha L(\theta)\right)\rho_t^N \, d\theta} d\theta = \sum_{i=1}^{N} w_L^i \theta_t^i, \qquad \text{with } w_L^i := \frac{\exp\left(-\alpha L(\theta_t^i)\right)}{\sum_{j=1}^{N} \exp\left(-\alpha L(\theta_t^j)\right)}$$

Carrillo, José A., et al. "An analytical framework for consensus-based global optimization method." *Mathematical Models and Methods in Applied Sciences* 28.06 (2018): 1037-1066.

Consensus-based Optimization (CBO)



Clustered Federated Learning

Optimization problem:

$$\min_{\theta \in \mathbb{R}^d} L_1(\theta) \quad \text{and} \quad \min_{\theta \in \mathbb{R}^d} L_2(\theta)$$

Number of cluster 1 agents = N_1

Number of cluster 2 agents = N_2

Total number of agents = N

FedCBO System

$$d\theta_t^{1,i} = -\lambda_1 \left(\theta_t^{1,i} - m_{L_1}^{\alpha}[\rho_t^N]\right) dt - \lambda_2 \nabla L_1(\theta_t^{1,i}) dt + \sigma_1 \left|\theta_t^{1,i} - m_{L_1}^{\alpha}[\rho_t^N]\right| dB_t^{1,i} + \sigma_2 \left|\nabla L_1(\theta_t^{1,i})\right| d\widetilde{B}_t^{1,i}$$

$$d\theta_t^{2,j} = -\lambda_1 \left(\theta_t^{2,j} - m_{L_2}^{\alpha}[\rho_t^N]\right) dt - \lambda_2 \nabla L_2(\theta_t^{2,j}) dt + \sigma_1 \left|\theta_t^{2,j} - m_{L_2}^{\alpha}[\rho_t^N]\right| dB_t^{2,j} + \sigma_2 \left|\nabla L_2(\theta_t^{2,j})\right| d\widetilde{B}_t^{2,j}$$

where

$$\rho_t^{1,N} := \frac{1}{N_1} \sum_{i=1}^{N_1} \delta_{\theta_t^{1,i}}, \qquad \rho_t^{2,N} := \frac{1}{N_2} \sum_{j=1}^{N_2} \delta_{\theta_t^{2,j}}, \qquad \rho_t^N := \frac{N_1}{N} \rho_t^{1,N} + \frac{N_2}{N} \rho_t^{2,N}.$$

FedCBO System

$$d\theta_t^{1,i} = -\lambda_1 \left(\theta_t^{1,i} - m_{L_1}^{\alpha}[\rho_t^N]\right) dt - \lambda_2 \nabla L_1(\theta_t^{1,i}) dt + \sigma_1 \left|\theta_t^{1,i} - m_{L_1}^{\alpha}[\rho_t^N]\right| dB_t^{1,i} + \sigma_2 \left|\nabla L_1(\theta_t^{1,i})\right| d\widetilde{B}_t^{1,i}$$

$$d\theta_t^{2,j} = -\lambda_1 \left(\theta_t^{2,j} - m_{L_2}^{\alpha}[\rho_t^N]\right) dt - \lambda_2 \nabla L_2(\theta_t^{2,j}) dt + \sigma_1 \left|\theta_t^{2,j} - m_{L_2}^{\alpha}[\rho_t^N]\right| dB_t^{2,j} + \sigma_2 \left|\nabla L_2(\theta_t^{2,j})\right| d\widetilde{B}_t^{2,j}$$

where

$$m_{L_1}^{\alpha}[\rho_t^N] := \int \theta \frac{\exp\left(-\alpha L_1(\theta)\right) \rho_t^N}{\int \exp\left(-\alpha L_1(\theta)\right) \rho_t^N} d\theta = \sum_{i=1}^{N_1} w_{L_1}^{1,i} \theta_t^{1,i} + \sum_{j=1}^{N_2} w_{L_1}^{2,j} \theta_t^{2,j},$$

$$w_{L_1}^{1,i} := \frac{\exp\left(-\alpha L_1\left(\theta_t^{1,i}\right)\right)}{Z_{L_1}}, \qquad w_{L_1}^{2,j} := \frac{\exp\left(-\alpha L_1\left(\theta_t^{2,j}\right)\right)}{Z_{L_1}}$$

FedCBO System

$$d\theta_t^{1,i} = -\lambda_1 \left(\theta_t^{1,i} - m_{L_1}^{\alpha}[\rho_t^N]\right) dt - \lambda_2 \nabla L_1(\theta_t^{1,i}) dt + \sigma_1 \left|\theta_t^{1,i} - m_{L_1}^{\alpha}[\rho_t^N]\right| dB_t^{1,i} + \sigma_2 \left|\nabla L_1(\theta_t^{1,i})\right| d\widetilde{B}_t^{1,i}$$

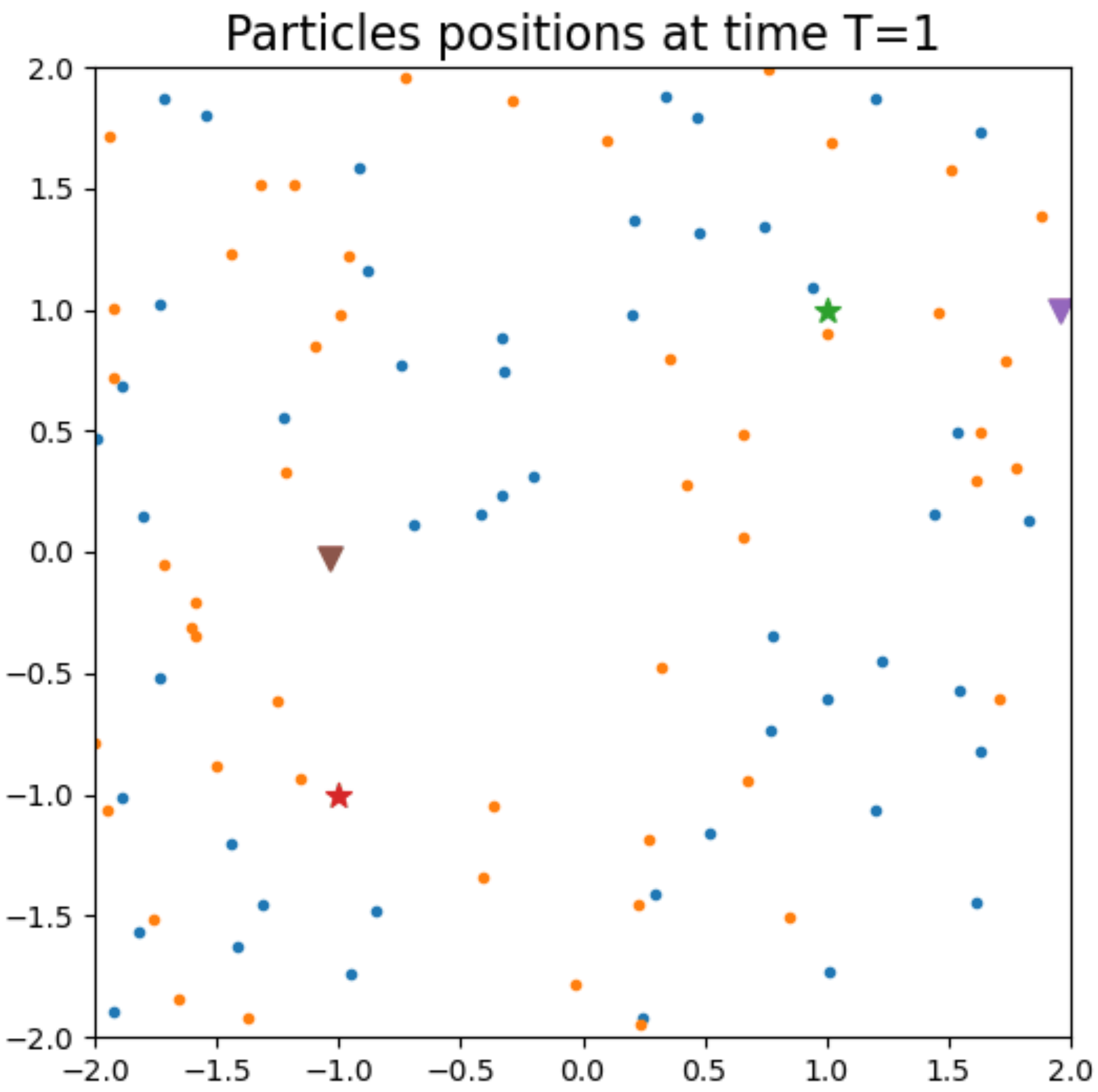
$$d\theta_t^{2,j} = -\lambda_1 \left(\theta_t^{2,j} - m_{L_2}^{\alpha}[\rho_t^N]\right) dt - \lambda_2 \nabla L_2(\theta_t^{2,j}) dt + \sigma_1 \left|\theta_t^{2,j} - m_{L_2}^{\alpha}[\rho_t^N]\right| dB_t^{2,j} + \sigma_2 \left|\nabla L_2(\theta_t^{2,j})\right| d\widetilde{B}_t^{2,j}$$

where

$$m_{L_2}^{\alpha}[\rho_t^N] := \int \theta \frac{\exp\left(-\alpha L_2(\theta)\right) \rho_t^N}{\int \exp\left(-\alpha L_2(\theta)\right) \rho_t^N} d\theta = \sum_{i=1}^{N_1} w_{L_2}^{1,i} \theta_t^{1,i} + \sum_{j=1}^{N_2} w_{L_2}^{2,j} \theta_t^{2,j},$$

$$w_{L_2}^{1,i} := \frac{\exp\left(-\alpha L_2\left(\theta_t^{1,i}\right)\right)}{Z_{L_2}}, \qquad w_{L_2}^{2,j} := \frac{\exp\left(-\alpha L_2\left(\theta_t^{2,j}\right)\right)}{Z_{L_2}}$$

FedCBO System



FedCBO System

$$d\theta_t^{1,i} = -\lambda_1 \left(\theta_t^{1,i} - m_{L_1}^{\alpha}[\rho_t^N]\right) dt - \lambda_2 \nabla L_1(\theta_t^{1,i}) dt + \sigma_1 \left|\theta_t^{1,i} - m_{L_1}^{\alpha}[\rho_t^N]\right| dB_t^{1,i} + \sigma_2 \left|\nabla L_1(\theta_t^{1,i})\right| d\widetilde{B}_t^{1,i}$$

$$d\theta_t^{2,j} = -\lambda_1 \left(\theta_t^{2,j} - m_{L_2}^{\alpha}[\rho_t^N]\right) dt - \lambda_2 \nabla L_2(\theta_t^{2,j}) dt + \sigma_1 \left|\theta_t^{2,j} - m_{L_2}^{\alpha}[\rho_t^N]\right| dB_t^{2,j} + \sigma_2 \left|\nabla L_2(\theta_t^{2,j})\right| d\widetilde{B}_t^{2,j}$$

$$\Big\downarrow \quad \text{As N goes to } \infty$$

$$d\theta_t^{1} = -\lambda_1 \left(\theta_t^{1} - m_{L_1}^{\alpha}[\rho_t]\right) dt - \lambda_2 \nabla L_1(\theta_t^{1}) dt + \sigma_1 \left|\theta_t^{1} - m_{L_1}^{\alpha}[\rho_t]\right| dB_t^{1} + \sigma_2 \left|\nabla L_1(\theta_t^{1})\right| d\widetilde{B}_t^{1}$$

$$d\theta_t^{2} = -\lambda_1 \left(\theta_t^{2} - m_{L_2}^{\alpha}[\rho_t]\right) dt - \lambda_2 \nabla L_2(\theta_t^{2}) dt + \sigma_1 \left|\theta_t^{2} - m_{L_2}^{\alpha}[\rho_t]\right| dB_t^{2} + \sigma_2 \left|\nabla L_2(\theta_t^{2})\right| d\widetilde{B}_t^{2}$$

FedCBO System

$$d\theta_t^{1,i} = -\lambda_1 \left(\theta_t^{1,i} - m^{\alpha}_{L_1}[\rho_t^N]\right) dt - \lambda_2 \nabla L_1(\theta_t^{1,i}) dt + \sigma_1 \left|\theta_t^{1,i} - m^{\alpha}_{L_1}[\rho_t^N]\right| dB_t^{1,i} + \sigma_2 \left|\nabla L_1(\theta_t^{1,i})\right| d\widetilde{B}_t^{1,i}$$

$$d\theta_t^{2,j} = -\lambda_1 \left(\theta_t^{2,j} - m^{\alpha}_{L_2}[\rho_t^N]\right) dt - \lambda_2 \nabla L_2(\theta_t^{2,j}) dt + \sigma_1 \left|\theta_t^{2,j} - m^{\alpha}_{L_2}[\rho_t^N]\right| dB_t^{2,j} + \sigma_2 \left|\nabla L_2(\theta_t^{2,j})\right| d\widetilde{B}_t^{2,j}$$

$$\Big\downarrow \text{ As N goes to } \infty$$

$$d\theta_t^{1} = -\lambda_1 \left(\theta_t^{1} - m^{\alpha}_{L_1}[\rho_t]\right) dt - \lambda_2 \nabla L_1(\theta_t^{1}) dt + \sigma_1 \left|\theta_t^{1} - m^{\alpha}_{L_1}[\rho_t]\right| dB_t^{1} + \sigma_2 \left|\nabla L_1(\theta_t^{1})\right| d\widetilde{B}_t^{1}$$

$$d\theta_t^{2} = -\lambda_1 \left(\theta_t^{2} - m^{\alpha}_{L_2}[\rho_t]\right) dt - \lambda_2 \nabla L_2(\theta_t^{2}) dt + \sigma_1 \left|\theta_t^{2} - m^{\alpha}_{L_2}[\rho_t]\right| dB_t^{2} + \sigma_2 \left|\nabla L_2(\theta_t^{2})\right| d\widetilde{B}_t^{2}$$

Consensus-based Optimization (CBO)

Theorem (Mean-field limit; Carrillo, **NGT**, Li, Zhu, 23'):

Suppose $\theta_0^i \sim \rho_0^1$ and $\theta_0^j \sim \rho_0^2$. Also, suppose $N \to \infty$ and

$$\frac{N_1}{N} \to w_1, \quad \frac{N_2}{N} \to w_2.$$

Then $\rho^{N,1} \to \rho^1$ and $\rho^{N,2} \to \rho^2$.

$$\partial_t \rho_t^1 := \Delta(\kappa_t^1 \rho_t^1) + \nabla \cdot (\mu_t^1 \rho_t^1), \qquad \lim_{t\to 0} \rho_t^1 = \rho_0^1$$

$$\partial_t \rho_t^2 := \Delta(\kappa_t^2 \rho_t^2) + \nabla \cdot (\mu_t^2 \rho_t^2), \qquad \lim_{t\to 0} \rho_t^2 = \rho_0^2,$$

$$\rho = w_1 \rho^1 + w_2 \rho^2$$

$$\mu_t^k := \lambda_1 \left(\theta - m_{L_k}^\alpha[\rho_t]\right) + \lambda_2 \nabla L_k(\theta), \quad \kappa_t^k := \frac{\sigma_1^2}{2}\left|\theta - m_{L_k}^\alpha[\rho_t]\right|^2 + \frac{\sigma_2^2}{2}\left|\nabla L_k(\theta)\right|^2, \quad \text{for } k = 1, 2.$$

Consensus-based Optimization (CBO)

Theorem (Long-time behavior mean field; Carrillo, **NGT**, Li, Zhu, 23'):

Let ρ_0^k give positive mass around θ_k^* (global minimizer of L_k) for each $k = 1, 2$. Let (ρ_t^1, ρ_t^2) be solution of mean field PDE. Let $\varepsilon > 0$.

Provided parameters λ, σ, α are chosen appropriately, we have, for some T^*,

$$\mathcal{V}(\rho_t^1) + \mathcal{V}(\rho_t^2) \leq \exp(-ct)(\mathcal{V}(\rho_0^1) + \mathcal{V}(\rho_0^2)), \quad \forall t \in [0, T^*]$$

and

$$\min_{t\in[0,T^*]} \mathcal{V}(\rho_t^1) + \mathcal{V}(\rho_t^2) \leq \varepsilon,$$

where

$$\mathcal{V}(\rho_t^k) := \int |\theta - \theta_k^*|^2 d\rho_t^k(\theta).$$

Consensus-based Optimization (CBO)

Theorem (Long-time behavior mean field; Carrillo, **NGT**, Li, Zhu, 23'):

Let ρ_0^k give positive mass around θ_k^* (global minimizer of L_k) for each $k = 1, 2$. Let (ρ_t^1, ρ_t^2) be solution of mean field PDE. Let $\varepsilon > 0$.

Provided parameters λ, σ, α are chosen appropriately, we have, for some T^*,

$$\boxed{T^* := \frac{1}{(1-\vartheta)(2\lambda_1 - 2\lambda_2 M - d\sigma_1^2 - d\sigma_2^2 M^2)} \log\left(\frac{\mathcal{V}(\rho_0^1) + \mathcal{V}(\rho_0^2)}{\varepsilon}\right)}$$

$$\mathcal{V}(\rho_t^1) + \mathcal{V}(\rho_t^2) \leq \exp(-ct)(\mathcal{V}(\rho_0^1) + \mathcal{V}(\rho_0^2)), \quad \forall t \in [0, T^*]$$

and

$$\min_{t\in[0,T^*]} \mathcal{V}(\rho_t^1) + \mathcal{V}(\rho_t^2) \leq \varepsilon,$$

where

$$\mathcal{V}(\rho_t^k) := \int |\theta - \theta_k^*|^2 d\rho_t^k(\theta).$$

Some References

- R. Pinnau, C. Totzeck, O. Tse, and S. Martin. *A consensus-based model for global optimization and its mean-field limit*. Mathematical Models and Methods in Applied Sciences, 27(01):183–204, 2017.

- J. A. Carrillo, Y.-P. Choi, C. Totzeck, and O. Tse. *An analytical framework for consensus-based global optimization method*. Mathematical Models and Methods in Applied Sciences, 28(06):1037–1066, 2018.

- M. Fornasier, T. Klock, and K. Riedl. *Consensus-based optimization methods converge globally in mean-field law*. arXiv preprint arXiv:2103.15130, 2021.

- Huang H, Qiu J. *On the mean-field limit for the consensus-based optimization*. Math Meth Appl Sci. 2022; 45(12): 7814-7831. doi:10.1002/mma.8279

- Riedl K. *Leveraging memory effects and gradient information in consensus-based optimisation: On global convergence in mean-field law*. EJAM. Published online 2023:1-32. doi:10.1017/S0956792523000293

Some References

- J. A. Carrillo, C. Totzeck, and U. Vaes. *Consensus-based optimization and ensemble kalman inversion for global optimization problems with constraints*. In Modeling and Simulation for Collective Dynamics, Lecture Notes Series, Institute for Mathematical Sciences, NUS, volume 40, 2023.

- J. A. Carrillo, S. Jin, L. Li, and Y. Zhu. *A consensus-based global optimization method for high dimensional machine learning problems*. ESAIM: Control, Optimisation and Calculus of Variations, 27:S5, 2021.

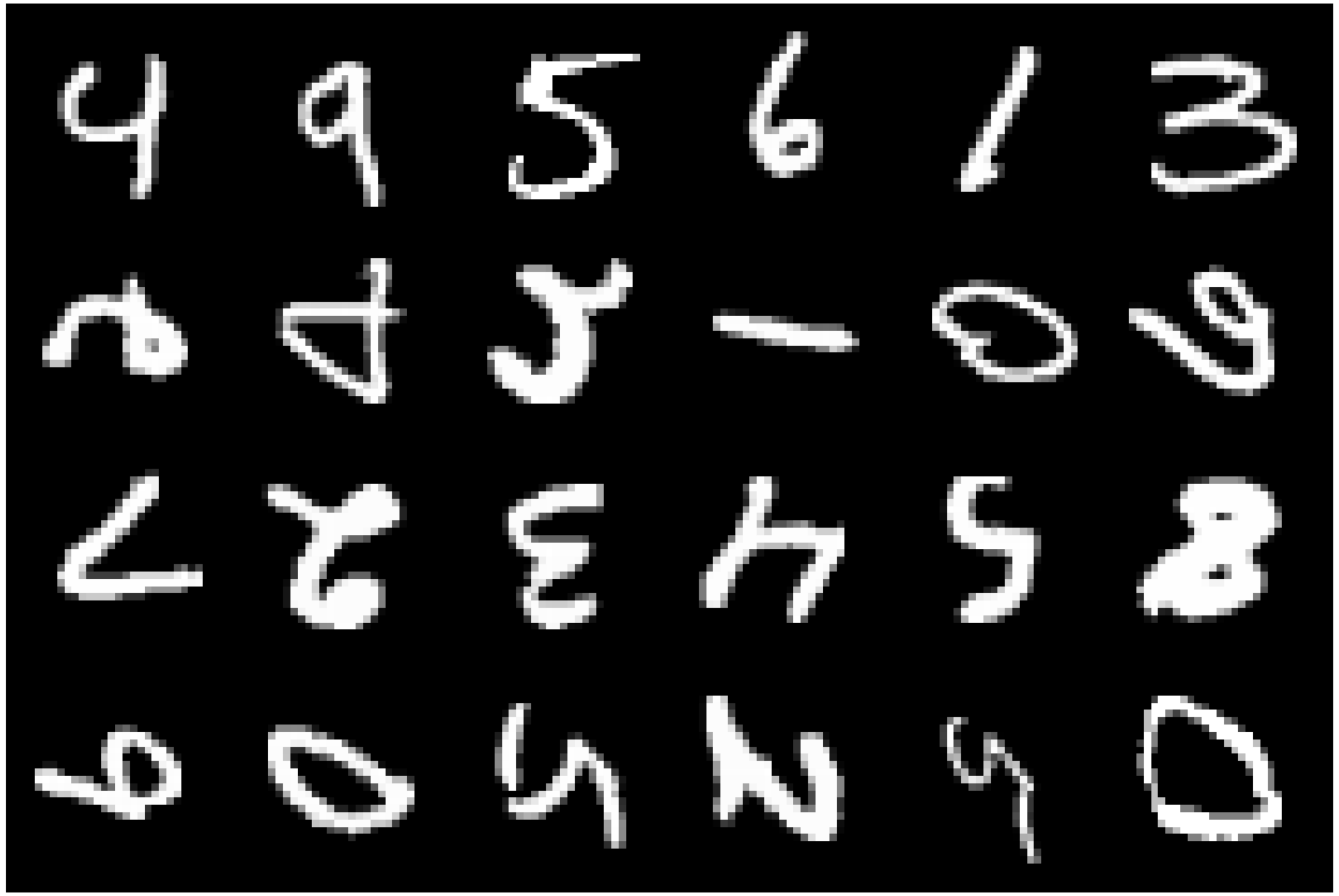
- L. Bungert, P. Wacker, and T. Roith. *Polarized consensus-based dynamics for optimization and sampling*. arXiv preprint arXiv:2211.05238, 2022.

- J. A. Carrillo, F. Hoffmann, A. M. Stuart, and U. Vaes. *Consensus-based sampling*. Studies in Applied Mathematics, 148(3):1069–1140, 2022

- Borghi, G., Herty, M. & Pareschi, L. *An Adaptive Consensus Based Method for Multi-objective Optimization with Uniform Pareto Front Approximation*. Appl Math Optim **88**, 58 (2023). https://doi.org/10.1007/s00245-023-10036-y

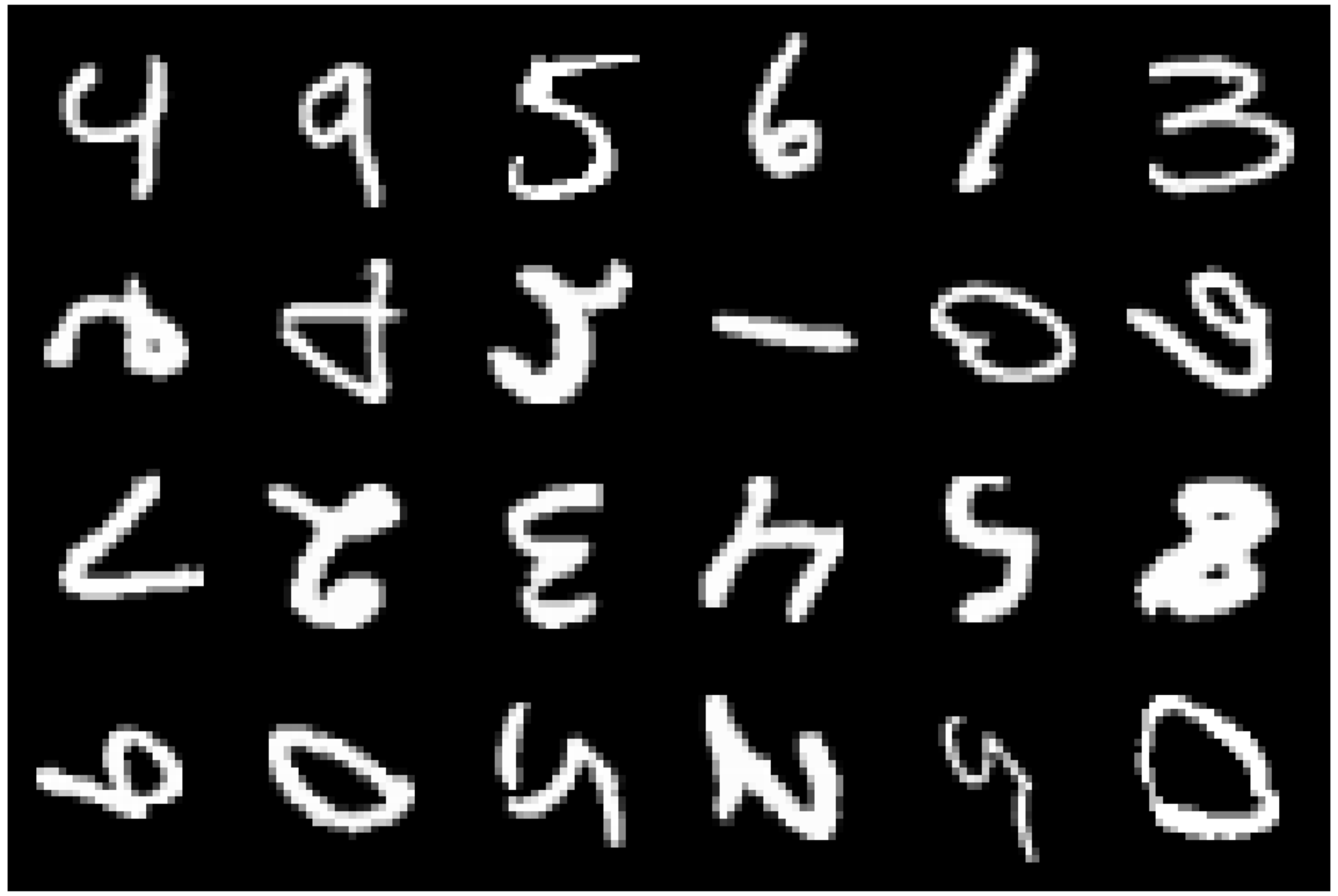
Experiments

Rotated MNIST:



90 degrees rotation

Experiments

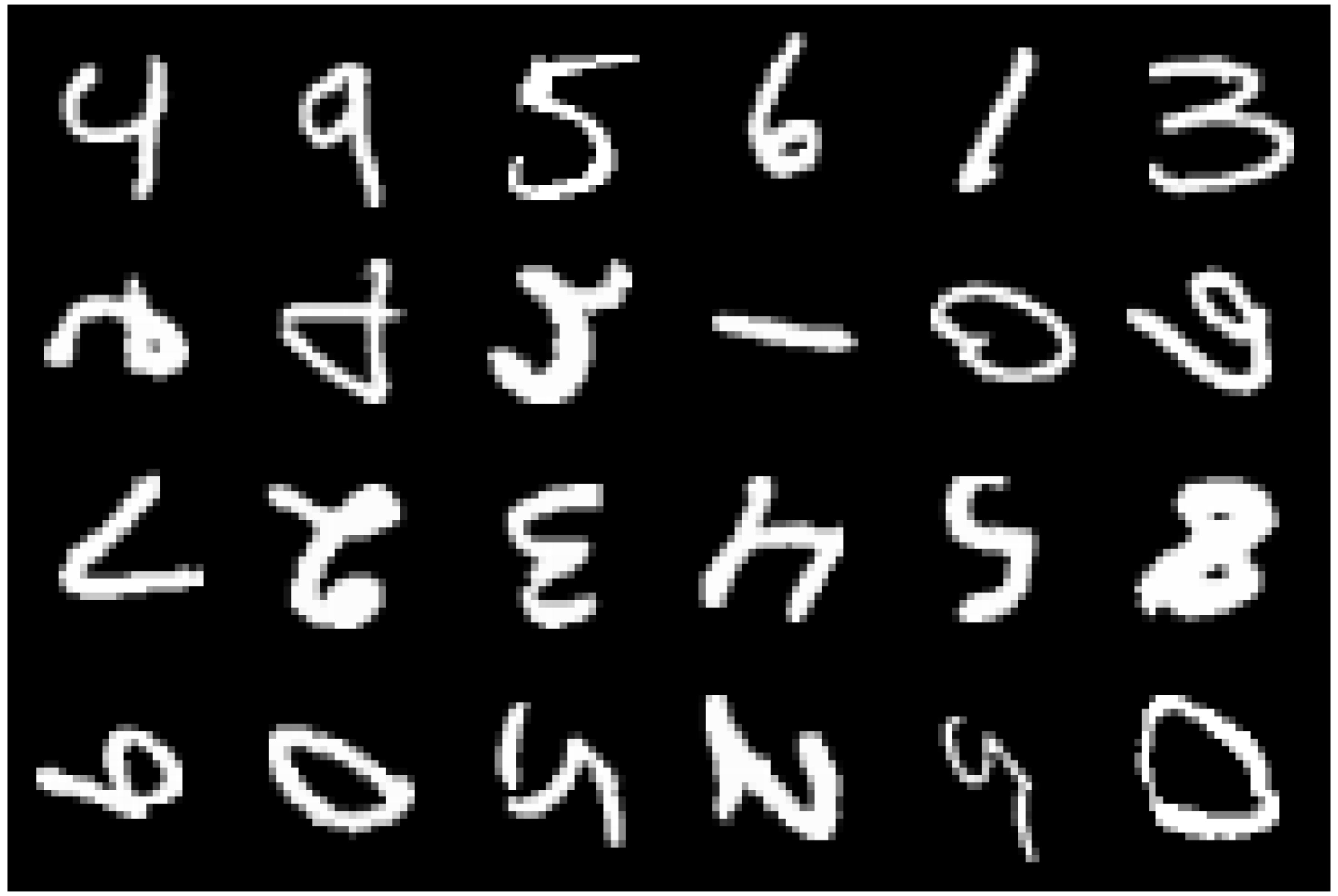


Number of clusters = 4

Number of agents in each cluster = 300

Number of data points in each agent = 200

Experiments

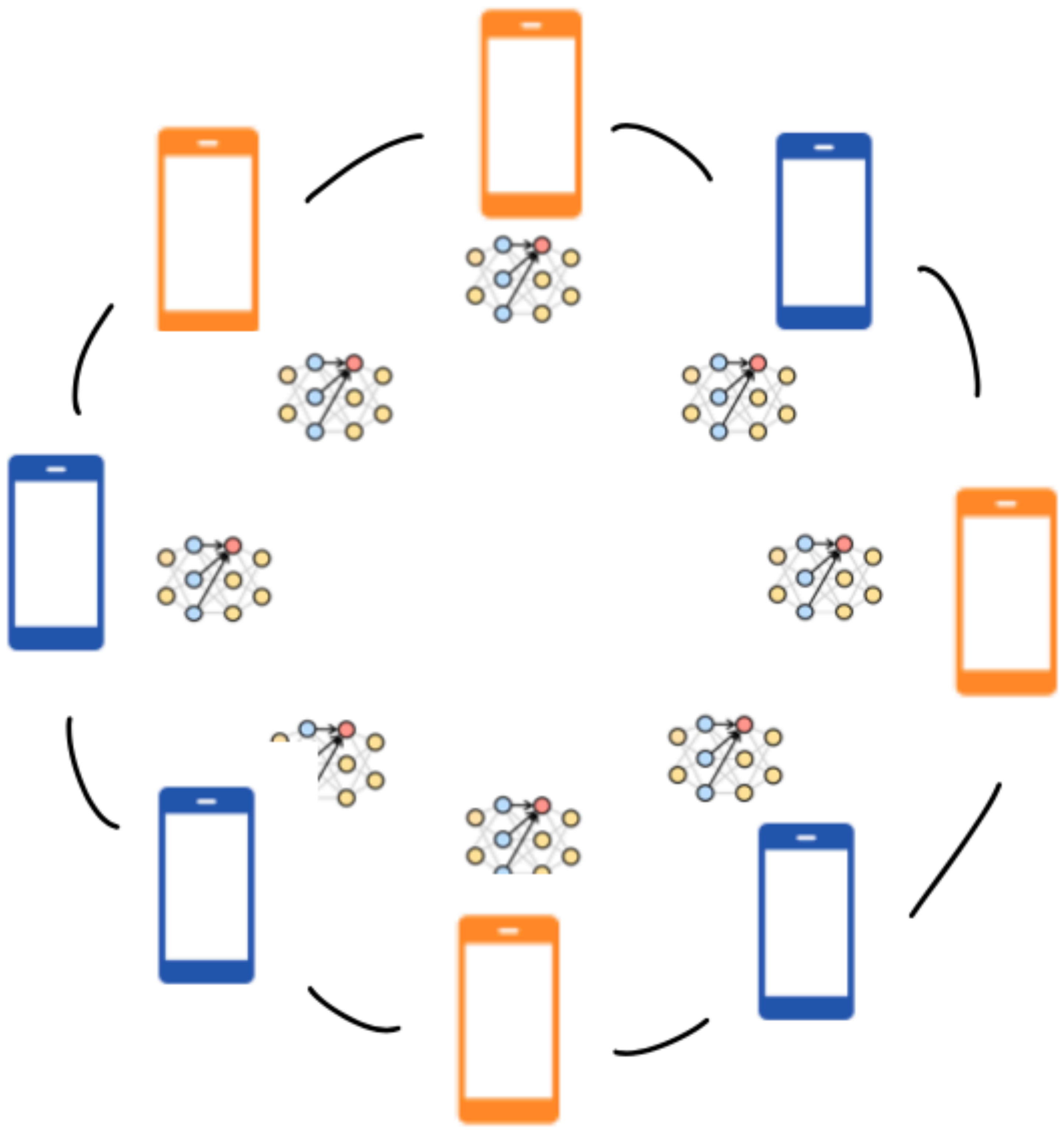


FedCBO	IFCA	FedAvg	Local
96.51 ± 0.04	94.44 ± 0.01	85.50 ± 0.19	81.27 ± 0.02

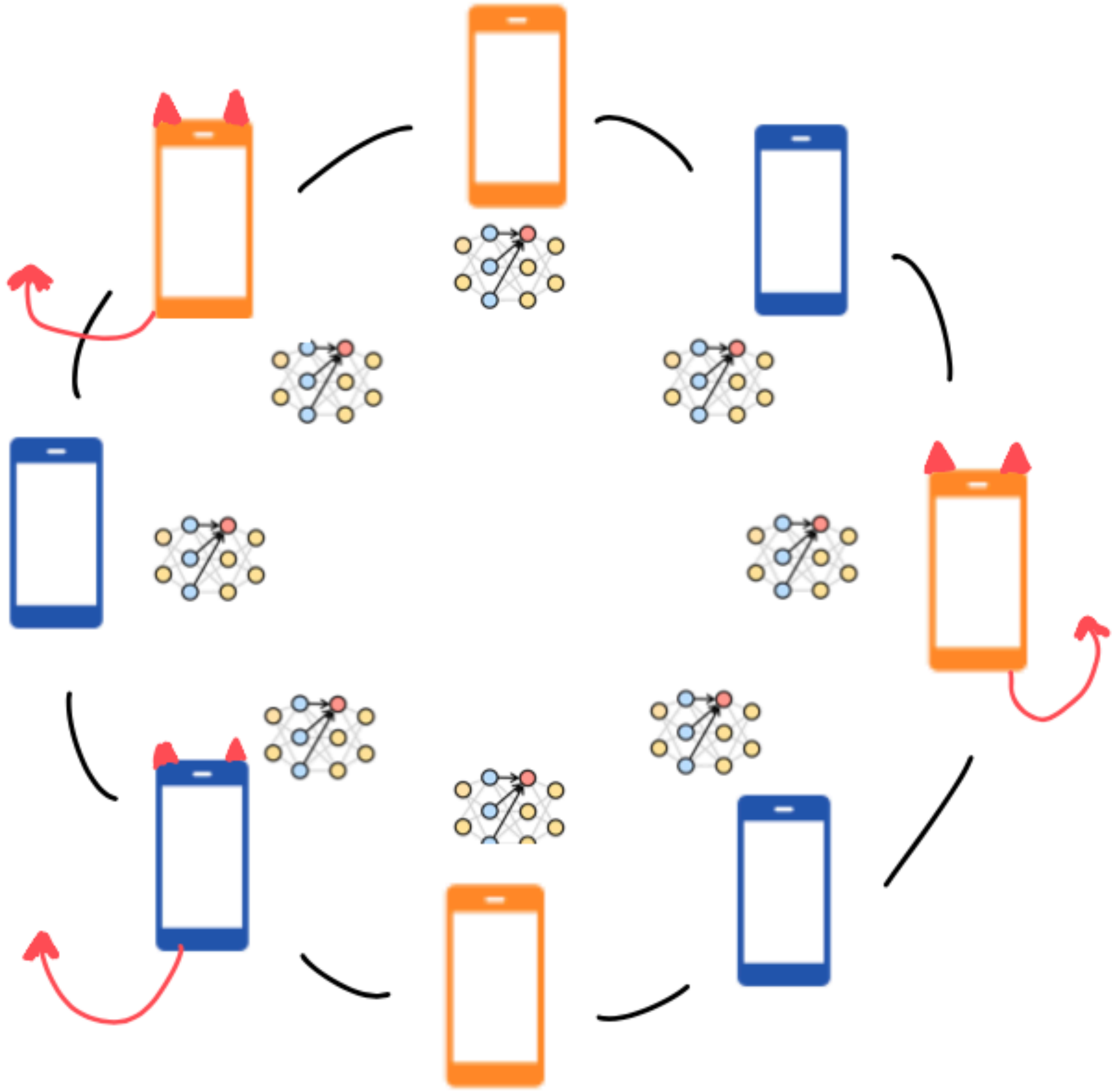
Part 2

Backdoor attacks

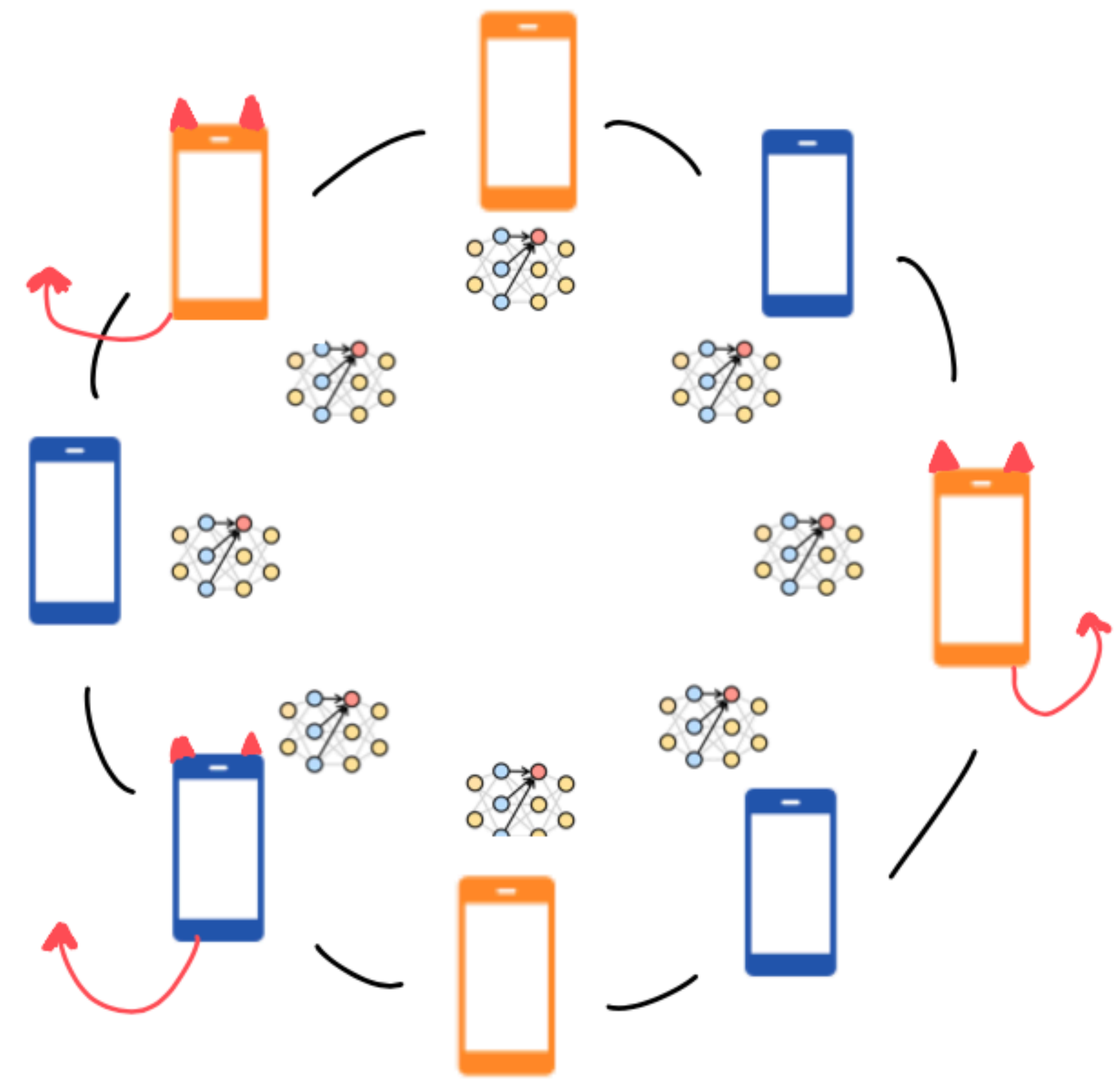
Backdoor attacks



Backdoor attacks



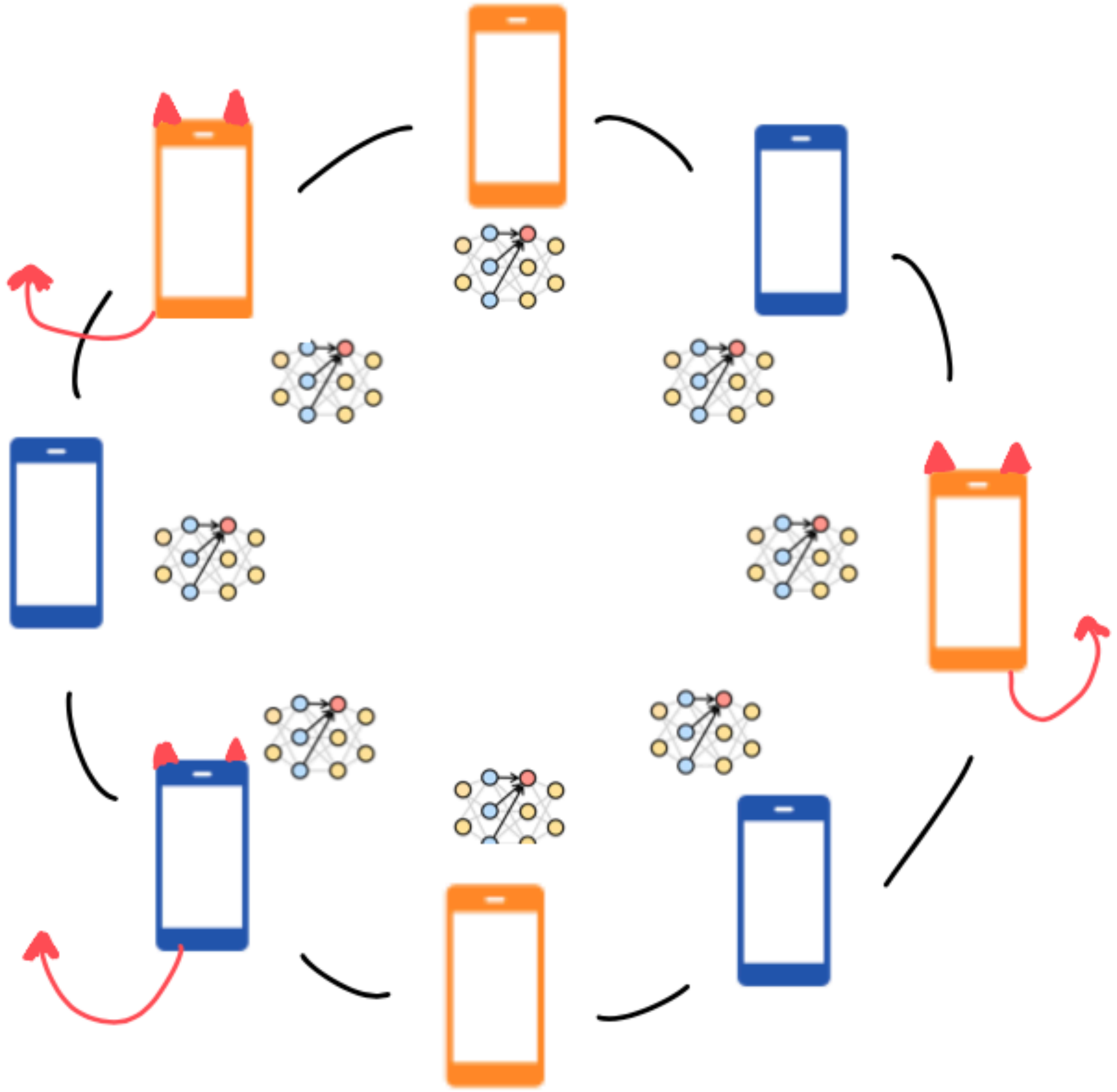
Backdoor attacks



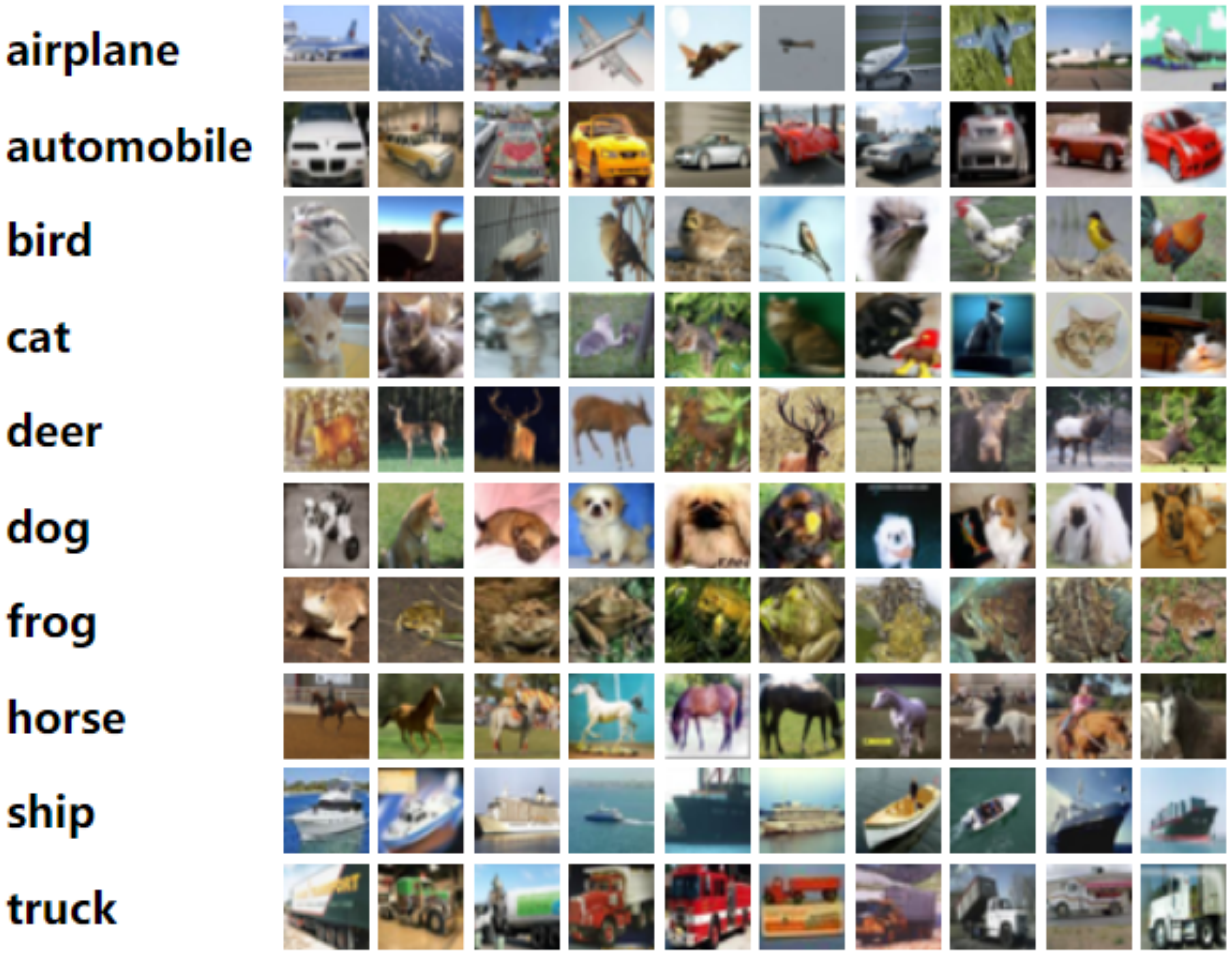
$$d\theta_t^{1,i} = -\lambda_1 \left(\theta_t^{1,i} - m_{L_1}^{\alpha}[\rho_t^N]\right) dt - \lambda_2 \nabla L_1(\theta_t^{1,i}) dt + \sigma_1 \left|\theta_t^{1,i} - m_{L_1}^{\alpha}[\rho_t^N]\right| dB_t^{1,i} + \sigma_2 \left|\nabla L_1(\theta_t^{1,i})\right| d\widetilde{B}_t^{1,i}$$

$$d\theta_t^{2,j} = -\lambda_1 \left(\theta_t^{2,j} - m_{L_2}^{\alpha}[\rho_t^N]\right) dt - \lambda_2 \nabla L_2(\theta_t^{2,j}) dt + \sigma_1 \left|\theta_t^{2,j} - m_{L_2}^{\alpha}[\rho_t^N]\right| dB_t^{2,j} + \sigma_2 \left|\nabla L_2(\theta_t^{2,j})\right| d\widetilde{B}_t^{2,j}$$

Backdoor attacks



Malicious agents' goal: make other agents predict points of class c_S as class c_T.


airplane
automobile
bird
cat
deer
dog
frog
horse
ship
truck

Backdoor attacks via label flipping

Instead of aiming to optimize

$$L_k(\theta) = \mathbb{E}_{(x,y)\sim\mathcal{D}_k}[l(f(x;\theta),y)] = \sum_{c=1}^{C} w_c \mathbb{E}_{x|y=c}[l(f(x;\theta),c)]$$

a malicious agent picks parameters to optimize:

$$L_k^{\text{mal}}(\theta) := \sum_{c \neq c_S}^{C} w_c \mathbb{E}_{x|y=c}[l(f(x;\theta),c)] + w_{c_S} \mathbb{E}_{x|y=c_S}[l(f(x;\theta),c_T)]$$

Backdoor attacks via label flipping

Instead of aiming to optimize

$$L_k(\theta) = \mathbb{E}_{(x,y)\sim\mathcal{D}_k}[l(f(x;\theta),y)] = \sum_{c=1}^{C} w_c \mathbb{E}_{x|y=c}[l(f(x;\theta),c)]$$

a malicious agent picks parameters to optimize:

$$L_k^{\mathsf{mal}}(\theta) := \sum_{c \neq c_S}^{C} w_c \mathbb{E}_{x|y=c}[l(f(x;\theta),c)] + w_{c_S} \mathbb{E}_{x|y=c_S}[l(f(x;\theta),c_T)]$$

Benign agents: introduce additional robustness criterion to protect against these attacks.

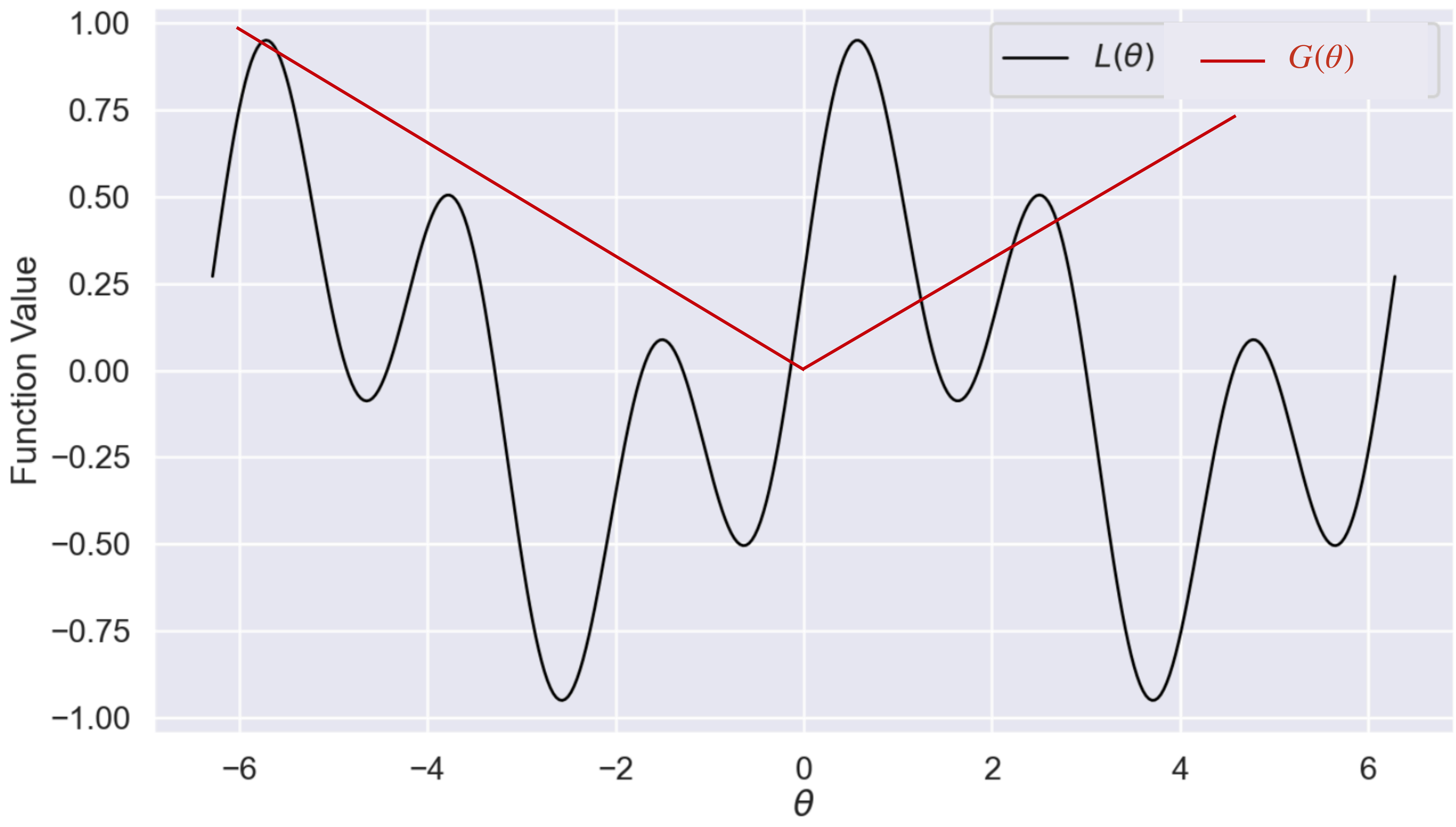
Bi-Level Optimization

Optimization problem:

$$\min_{\theta \in \Theta} \quad G(\theta)$$
$$\text{s.t.} \quad \theta \in \arg\min L$$

Assumptions:

- Unique solution θ^*_{good}



Bilevel CBO

Interacting particle system:

$$d\theta_t^i = -\lambda(\theta_t^i - m^{\alpha,\beta}[\rho_t^N])dt + \sigma|\theta_t^i - m^{\alpha,\beta}[\rho_t^N]|dB_t^i, \quad i = 1, \dots N.$$

where

$$\rho_t^N := \frac{1}{N}\sum_{i=1}^N \delta_{\theta_t^i}$$

$$m^{\alpha,\beta}[\rho_t^N] := \int \theta \frac{\exp(-\alpha G(\theta))}{\int \exp(-\alpha G(\theta)) dI_\beta[\rho_t^N](\theta)} dI_\beta[\rho_t^N](\theta)$$

$$I_\beta[\rho_t^N] := \rho_t^N(\cdot \cap Q_\beta[\rho_t^N]) \qquad Q_\beta[\rho_t^N] := \{\theta \text{ s.t. } L(\theta) \leq q_\beta[\rho_t^N]\} \qquad q_\beta[\rho_t^N] := \inf\{q \text{ s.t. } \rho_t^N(\{L(\theta) \leq q\}) \geq \beta\}$$

Bilevel CBO

Interacting particle system:

$$d\theta_t^i = -\lambda(\theta_t^i - m^{\alpha,\beta}[\rho_t^N])dt + \sigma|\theta_t^i - m^{\alpha,\beta}[\rho_t^N]|dB_t^i, \quad i = 1, \dots N.$$

where

$$\rho_t^N := \frac{1}{N}\sum_{i=1}^N \delta_{\theta_t^i}$$

$$m^{\alpha,\beta}[\rho_t^N] := \int \theta \frac{\exp(-\alpha G(\theta))}{\int \exp(-\alpha G(\theta)) dI_\beta[\rho_t^N](\theta)} dI_\beta[\rho_t^N](\theta)$$

$$I_\beta[\rho_t^N] := \rho_t^N(\cdot \cap Q_\beta[\rho_t^N]) \qquad Q_\beta[\rho_t^N] := \{\theta \text{ s.t. } L(\theta) \leq q_\beta[\rho_t^N]\} \qquad q_\beta[\rho_t^N] := \inf\{q \text{ s.t. } \rho_t^N(\{L(\theta) \leq q\}) \geq \beta\}$$

Bilevel CBO

Interacting particle system:

$$d\theta_t^i = -\lambda(\theta_t^i - m^{\alpha,\beta}[\rho_t^N])dt + \sigma|\theta_t^i - m^{\alpha,\beta}[\rho_t^N]|dB_t^i, \quad i = 1, \dots N.$$

where

$$\rho_t^N := \frac{1}{N}\sum_{i=1}^{N} \delta_{\theta_t^i}$$

$$m^{\alpha,\beta}[\rho_t^N] := \int \theta \frac{\exp(-\alpha G(\theta))}{\int \exp(-\alpha G(\theta)) dI_\beta[\rho_t^N](\theta)} dI_\beta[\rho_t^N](\theta)$$

$$I_\beta[\rho_t^N] := \rho_t^N(\cdot \cap Q_\beta[\rho_t^N]) \qquad Q_\beta[\rho_t^N] := \{\theta \text{ s.t. } L(\theta) \leq q_\beta[\rho_t^N]\} \qquad q_\beta[\rho_t^N] := \inf\{q \text{ s.t. } \rho_t^N(\{L(\theta) \leq q\}) \geq \beta\}$$

Bilevel CBO

Interacting particle system:

$$d\theta_t^i = -\lambda(\theta_t^i - m^{\alpha,\beta}[\rho_t^N])dt + \sigma|\theta_t^i - m^{\alpha,\beta}[\rho_t^N]|dB_t^i, \quad i = 1, \dots N.$$

where

$$\rho_t^N := \frac{1}{N}\sum_{i=1}^{N} \delta_{\theta_t^i}$$

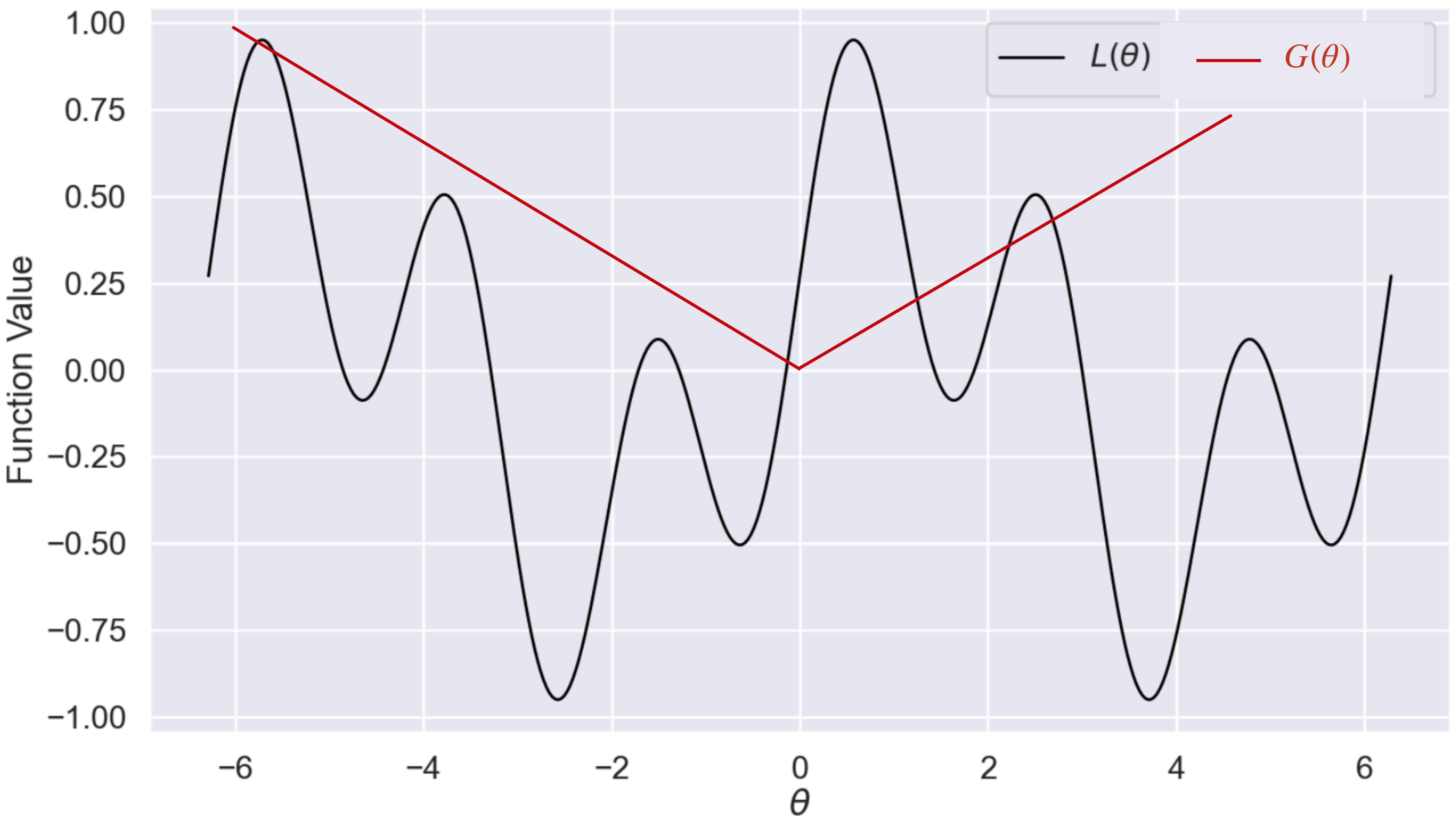
Within top $\beta \times (100)\,\%$, largest L

$$m^{\alpha,\beta}[\rho_t^N] := \int \theta \frac{\exp(-\alpha G(\theta))}{\int \exp(-\alpha G(\theta)) dI_\beta[\rho_t^N](\theta)} dI_\beta[\rho_t^N](\theta)$$

$$I_\beta[\rho_t^N] := \rho_t^N(\cdot \cap Q_\beta[\rho_t^N]) \qquad Q_\beta[\rho_t^N] := \{\theta \text{ s.t. } L(\theta) \le q_\beta[\rho_t^N]\} \qquad q_\beta[\rho_t^N] := \inf\{q \text{ s.t. } \rho_t^N(\{L(\theta) \le q\}) \ge \beta\}$$

Example:

$$\min_{\theta \in \Theta} \quad G(\theta)$$
$$\text{s.t.} \quad \theta \in \arg\min L$$



Experiments: $\theta_0^i \sim \text{Uniform}[-10, 10]$

$\beta = 0.1$	$\beta = 0.5$	$\beta = 0.8$	$\beta = 0.9$
10/10 (T=50)	10/10 (T=600)	10/10 (T=6000)	0/10 (T=20000)

Example: Constrained optimization via Bilevel CBO

$$\min_{\theta \in \mathbb{R}^3} \quad G(\theta)$$
$$\text{s.t.} \quad \theta \in \mathcal{C} := \partial B_1(0)$$

where

$$G(\theta) := -20 \exp\left(-0.2\sqrt{\frac{1}{3}\sum_{l=1}^{3}(\theta_l - p_l)}\right) + \exp\left(\frac{1}{3}\sum_{l=1}^{3}\cos(2\pi(\theta_l - p_l))\right)$$

$$p = (0.4, 0.4, 0.4)$$

Constrained optimization via Bilevel CBO

$$\min_{\theta \in \mathbb{R}^3} \quad G(\theta)$$
$$\text{s.t.} \quad \theta \in \mathcal{C} := \partial B_1(0)$$

$$\longrightarrow$$

$$\min_{\theta \in \Theta} \quad G(\theta)$$
$$\text{s.t.} \quad \theta \in \arg\min L$$

where

$$G(\theta) := -20 \exp\left(-0.2\sqrt{\frac{1}{3}\sum_{l=1}^{3}(\theta_l - p_l)}\right) + \exp\left(\frac{1}{3}\sum_{l=1}^{3}\cos(2\pi(\theta_l - p_l))\right)$$

$$p = (0.4, 0.4, 0.4)$$

$$L(\theta) = (1 - |\theta|)^2$$

Constrained optimization via Bilevel CBO

$$\min_{\theta \in \Theta} \quad G(\theta)$$

$$\text{s.t.} \quad \theta \in \arg\min L$$

Experiments: $\theta_0^i \sim \text{Uniform}[-10, 10]^3$

$\beta = 0.02$	$\beta = 0.05$	$\beta = 0.1$	$\beta = 0.2$	$\beta = 0.3$	$\beta = 0.5$	$\beta = 0.8$
1/10 (T=1000)	10/10 (T=200)	10/10 (T=200)	10/10 (T=200)	10/10 (T=1000)	10/10 (T=2000)	0/10 (T=2000)

BiLevel FedCBO

Optimization problems: for $k = 1, \ldots, K$

$$\begin{aligned} \min_{\theta \in \Theta} \quad & G_k(\theta) \\ \text{s.t.} \quad & \theta \in \arg\min L_k \end{aligned}$$

where $L_k(\theta)$ and $G_k(\theta)$ are, for example,

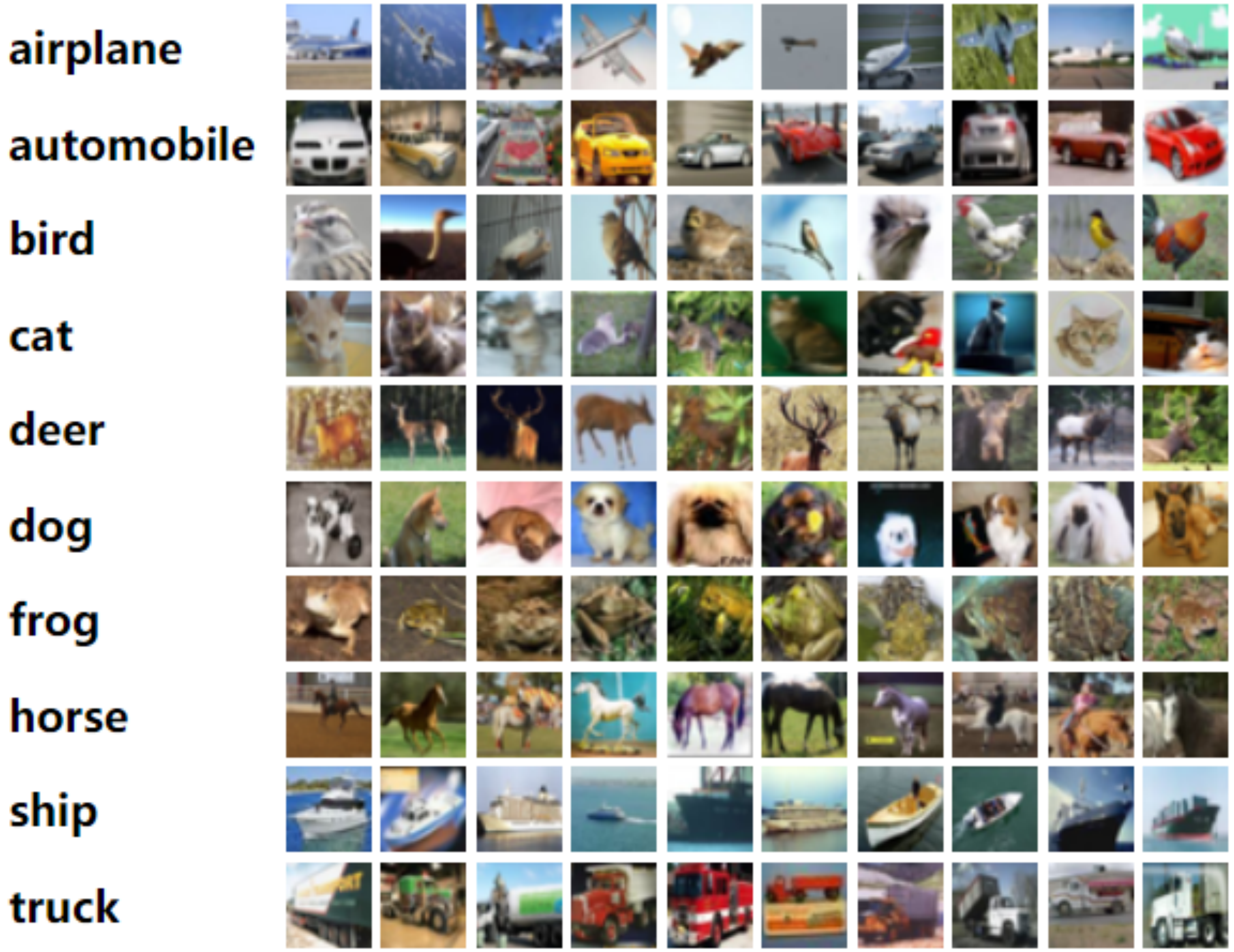
$$L_k(\theta) = \sum_{c=1}^{C} w_{k,c} L_{k,c}(\theta) \qquad G_k(\theta) = -\sum_{c=1}^{C} w_{k,c} \log\left(\frac{L_{k,c}(\theta)}{L_k(\theta)}\right)$$

BiLevel FedCBO

$$
\begin{aligned}
d\theta_t^{1,i} = &- \lambda_1(\theta_t^{1,i} - m^{\alpha,\beta}_{L_1,G_1}[\rho_t^N])dt - \lambda_2 \nabla L_1(\theta_t^{1,i})dt \\
&+ \sigma_1|\theta_t^{1,i} - m^{\alpha,\beta}_{L_1,G_1}[\rho_t^N]|dB_t^{1,i} + \sigma_2|\nabla L_1(\theta_t^{1,i})|d\tilde{B}_t^{1,i}
\end{aligned}
$$

$$
\begin{aligned}
d\theta_t^{2,j} = &- \lambda_1(\theta_t^{2,j} - m^{\alpha,\beta}_{L_2,G_2}[\rho_t^N])dt - \lambda_2 \nabla L_2(\theta_t^{2,j})dt \\
&+ \sigma_1|\theta_t^{2,j} - m^{\alpha,\beta}_{L_2,G_2}[\rho_t^N]|dB_t^{2,j} + \sigma_2|\nabla L_2(\theta_t^{2,j})|d\tilde{B}_t^{2,j}
\end{aligned}
$$

Experiments on CIFAR10



Experiments

Experimental setting 1 (CIFAR10 homogeneous case):

-Total number of agents N = 10;
-Num of benign agents = 7; Num of malicious agents = 3;
-Num of data for each benign agent = 500;
-Num of data for each malicious agent = 1200;

Attacks:

Source class: class 0 (images of planes)
Target class: class 2 (images of birds)
Label flipping: $0 \longrightarrow 2$.

Experiments

	With backdoor attack (FedCBO $\alpha = 1$)	With backdoor attack (FedCBO $\alpha = 10$)	Without backdoor attack (FedCBO $\alpha = 1$)	Without malicious agents (FedCBO $\alpha = 1$)	With backdoor attack (Bilevel FedCBO $\alpha = 20,\ \beta = 1.0$)
Avg overall acc	63.85 ± 0.18 %	61.69 ± 1.21 %	**65.30 ± 0.35 %**	60.76 ± 0.29 %	61.06 ± 0.21 %
Acc on class 0	29.86 ± 1.79 %	44.38 ± 3.36 %	41.86 ± 5.22 %	**61.10 ± 3.49 %**	**58.62 ± 3.72 %**
Benign agents' models predict images of class 0 as label 2	34.84 ± 3.70 % (Attack success rate)	22.50 ± 1.86 % (Attack success rate)	11.38 ± 1.18 %	**6.90 ± 1.29 %**	**9.84 ± 1.43 %** (Attack success rate)

Experiments

	With backdoor attack (FedCBO $\alpha = 1$)	With backdoor attack (FedCBO $\alpha = 10$)	Without backdoor attack (FedCBO $\alpha = 1$)	Without malicious agents (FedCBO $\alpha = 1$)	With backdoor attack (Bilevel FedCBO $\alpha = 20,\ \beta = 1.0$)
Avg overall acc	63.85 ± 0.18 %	61.69 ± 1.21 %	**65.30 ± 0.35 %**	60.76 ± 0.29 %	61.06 ± 0.21 %
Acc on class 0	29.86 ± 1.79 %	44.38 ± 3.36 %	41.86 ± 5.22 %	**61.10 ± 3.49 %**	**58.62 ± 3.72 %**
Benign agents' models predict images of class 0 as label 2	34.84 ± 3.70 % (Attack success rate)	22.50 ± 1.86 % (Attack success rate)	11.38 ± 1.18 %	**6.90 ± 1.29 %**	**9.84 ± 1.43 %** (Attack success rate)

With Backdoor Attack:
Total number of class 0 images (with correct labels) from the benign agents = 284;
Total number of class 0 images (with wrong labels) from the malicious agents = 356;
(i.e. in the entire dataset, about 45% class 0 images have correct labels and 55% of them have wrong labels)

Experiments

	With backdoor attack (FedCBO $\alpha = 1$)	With backdoor attack (FedCBO $\alpha = 10$)	Without backdoor attack (FedCBO $\alpha = 1$)	Without malicious agents (FedCBO $\alpha = 1$)	With backdoor attack (Bilevel FedCBO $\alpha = 20,\ \beta = 1.0$)
Avg overall acc	63.85 ± 0.18 %	61.69 ± 1.21 %	**65.30 ± 0.35 %**	60.76 ± 0.29 %	61.06 ± 0.21 %
Acc on class 0	29.86 ± 1.79 %	44.38 ± 3.36 %	41.86 ± 5.22 %	**61.10 ± 3.49 %**	**58.62 ± 3.72 %**
Benign agents' models predict images of class 0 as label 2	34.84 ± 3.70 % (Attack success rate)	22.50 ± 1.86 % (Attack success rate)	11.38 ± 1.18 %	**6.90 ± 1.29 %**	**9.84 ± 1.43 %** (Attack success rate)

Without Backdoor Attack:
Remove all the class 0 images contained in malicious agents.

Experiments

	With backdoor attack (FedCBO $\alpha = 1$)	With backdoor attack (FedCBO $\alpha = 10$)	Without backdoor attack (FedCBO $\alpha = 1$)	Without malicious agents (FedCBO $\alpha = 1$)	With backdoor attack (Bilevel FedCBO $\alpha = 20, \beta = 1.0$)
Avg overall acc	63.85 ± 0.18 %	61.69 ± 1.21 %	**65.30 ± 0.35 %**	60.76 ± 0.29 %	61.06 ± 0.21 %
Acc on class 0	29.86 ± 1.79 %	44.38 ± 3.36 %	41.86 ± 5.22 %	**61.10 ± 3.49 %**	**58.62 ± 3.72 %**
Benign agents' models predict images of class 0 as label 2	34.84 ± 3.70 % (Attack success rate)	22.50 ± 1.86 % (Attack success rate)	11.38 ± 1.18 %	**6.90 ± 1.29 %**	**9.84 ± 1.43 %** (Attack success rate)

Without Malicious Agents:
Remove all the malicious agents.

Experiments

	With backdoor attack (FedCBO $\alpha = 1$)	With backdoor attack (FedCBO $\alpha = 10$)	Without backdoor attack (FedCBO $\alpha = 1$)	Without malicious agents (FedCBO $\alpha = 1$)	With backdoor attack (Bilevel FedCBO $\alpha = 20, \beta = 1.0$)
Avg overall acc	63.85 ± 0.18 %	61.69 ± 1.21 %	**65.30 ± 0.35 %**	60.76 ± 0.29 %	61.06 ± 0.21 %
Acc on class 0	29.86 ± 1.79 %	44.38 ± 3.36 %	41.86 ± 5.22 %	**61.10 ± 3.49 %**	**58.62 ± 3.72 %**
Benign agents' models predict images of class 0 as label 2	34.84 ± 3.70 % (Attack success rate)	22.50 ± 1.86 % (Attack success rate)	11.38 ± 1.18 %	**6.90 ± 1.29 %**	**9.84 ± 1.43 %** (Attack success rate)

Experiments

	With backdoor attack (FedCBO $\alpha = 1$)	With backdoor attack (FedCBO $\alpha = 10$)	Without backdoor attack (FedCBO $\alpha = 1$)	Without malicious agents (FedCBO $\alpha = 1$)	With backdoor attack (Bilevel FedCBO $\alpha = 20, \beta = 1.0$)
Avg overall acc	63.85 ± 0.18 %	61.69 ± 1.21 %	**65.30 ± 0.35 %**	60.76 ± 0.29 %	61.06 ± 0.21 %
Acc on class 0	29.86 ± 1.79 %	44.38 ± 3.36 %	41.86 ± 5.22 %	**61.10 ± 3.49 %**	**58.62 ± 3.72 %**
Benign agents' models predict images of class 0 as label 2	34.84 ± 3.70 % (Attack success rate)	22.50 ± 1.86 % (Attack success rate)	11.38 ± 1.18 %	**6.90 ± 1.29 %**	**9.84 ± 1.43 %** (Attack success rate)

Experiments

Experimental setting 2 (Rotated CIFAR10):

- Total number of agents N = 20;
- Num of clusters k = 2;
- Num of benign agents per cluster = 7; Num of malicious agents per cluster = 3;
- Num of data for each benign agent = 500;
- Num of data for each malicious agent = 1200;

Experiments

Experimental setting 2 (Rotated CIFAR10):

- Total number of agents N = 20;
- Num of clusters k = 2;
- Num of benign agents per cluster = 7; Num of malicious agents per cluster = 3;
- Num of data for each benign agent = 500;
- Num of data for each malicious agent = 1200;

	With backdoor attack (FedCBO $\alpha = 10$)	With backdoor attack (Bilevel FedCBO $\alpha = 20,\ \beta = 0.5$)	With backdoor attack (Bilevel FedCBO $\alpha = 10,\ \beta = 0.5$)
Avg overall acc	64.44 ± 0.80 %	62.96 ± 0.27 %	**65.57 ± 0.14** %
Acc on class 0	55.41 ± 3.07 %	62.52 ± 2.47 %	**63.88 ± 2.15** %
Benign agents' models predict images of class 0 as label 2	14.96 ± 2.87 %	**7.38 ± 1.62** %	9.01 ± 1.91 %

Future Works

1. Batched interactions.

2. Analysis of adaptive tuning of parameters.

3. Theoretical analysis of dynamics in low communication regime.

Thank you for your attention!

Special thanks to:

-NSF Grants: DMS-2005797 and DMS-2236447
-All my collaborators.

Discretized FedCBO System

FedCBO system:

$$d\theta_t^{1,i} = -\lambda_1 \left(\theta_t^{1,i} - m_{L_1}^{\alpha}[\rho_t^N]\right) dt - \lambda_2 \nabla L_1(\theta_t^{1,i}) dt + \sigma_1 \left|\theta_t^{1,i} - m_{L_1}^{\alpha}[\rho_t^N]\right| dB_t^{1,i} + \sigma_2 \left|\nabla L_1(\theta_t^{1,i})\right| d\widetilde{B}_t^{1,i}$$

$$d\theta_t^{2,j} = -\lambda_1 \left(\theta_t^{2,j} - m_{L_2}^{\alpha}[\rho_t^N]\right) dt - \lambda_2 \nabla L_2(\theta_t^{2,j}) dt + \sigma_1 \left|\theta_t^{2,j} - m_{L_2}^{\alpha}[\rho_t^N]\right| dB_t^{2,j} + \sigma_2 \left|\nabla L_2(\theta_t^{2,j})\right| d\widetilde{B}_t^{2,j}$$

Euler discretization:

$$\theta_{n+1}^{1,i} \leftarrow \theta_n^{1,i} - \lambda_1 \gamma \left(\theta_n^{1,i} - m_n^1\right) - \lambda_2 \gamma \nabla L_1(\theta_n^{1,i}) + \sigma_1 \sqrt{\gamma} \left|\theta_n^{1,i} - m_n^1\right| z_n^{1,i} + \sigma_2 \sqrt{\gamma} \left|\nabla L_1(\theta_n^{1,i})\right| \widetilde{z}_n^{1,i}$$

$$\theta_{n+1}^{2,j} \leftarrow \theta_n^{2,j} - \lambda_1 \gamma \left(\theta_n^{2,j} - m_n^2\right) - \lambda_2 \gamma \nabla L_2(\theta_n^{2,j}) + \sigma_1 \sqrt{\gamma} \left|\theta_n^{2,j} - m_n^2\right| z_n^{2,j} + \sigma_2 \sqrt{\gamma} \left|\nabla L_2(\theta_n^{2,j})\right| \widetilde{z}_n^{2,j}$$

Discretized FedCBO System

$$\theta_{n+1}^{1,i} \leftarrow \theta_n^{1,i} - \lambda_1 \gamma \left(\theta_n^{1,i} - m_n^1\right) - \lambda_2 \gamma \nabla L_1(\theta_n^{1,i}) + \sigma_1 \sqrt{\gamma} \left|\theta_n^{1,i} - m_n^1\right| z_n^{1,i} + \sigma_2 \sqrt{\gamma} \left|\nabla L_1(\theta_n^{1,i})\right| \tilde{z}_n^{1,i}$$

$$\theta_{n+1}^{2,j} \leftarrow \theta_n^{2,j} - \lambda_1 \gamma \left(\theta_n^{2,j} - m_n^2\right) - \lambda_2 \gamma \nabla L_2(\theta_n^{2,j}) + \sigma_1 \sqrt{\gamma} \left|\theta_n^{2,j} - m_n^2\right| z_n^{2,j} + \sigma_2 \sqrt{\gamma} \left|\nabla L_2(\theta_n^{2,j})\right| \tilde{z}_n^{2,j}$$

Remove noise terms

$$\theta_{n+1}^{1,i} \leftarrow \theta_n^{1,i} - \lambda_1 \gamma \left(\theta_n^{1,i} - m_n^1\right) - \lambda_2 \gamma \nabla L_1(\theta_n^{1,i})$$

$$\theta_{n+1}^{2,j} \leftarrow \theta_n^{2,j} - \lambda_1 \gamma \left(\theta_n^{2,j} - m_n^2\right) - \lambda_2 \gamma \nabla L_2(\theta_n^{2,j})$$

Discretized FedCBO System

$$\theta^{1,i}_{n+1} \leftarrow \theta^{1,i}_{n} - \lambda_1\gamma\left(\theta^{1,i}_{n} - m^1_n\right) - \lambda_2\gamma\nabla L_1(\theta^{1,i}_{n})$$

$$\theta^{2,j}_{n+1} \leftarrow \theta^{2,j}_{n} - \lambda_1\gamma\left(\theta^{2,j}_{n} - m^2_n\right) - \lambda_2\gamma\nabla L_2(\theta^{2,j}_{n})$$

$$\Big\downarrow \text{ Sum over } \tau \text{ times}$$

$$\theta^{1,i}_{(n+1)\tau} \leftarrow \theta^{1,i}_{n\tau} - \lambda_1\gamma\sum_{q=0}^{\tau-1}\left(\theta^{1,i}_{n\tau+q} - m^1_{n\tau+q}\right) - \lambda_2\gamma\sum_{q=0}^{\tau-1}\nabla L_1(\theta^{1,i}_{n\tau+q})$$

$$\theta^{2,j}_{(n+1)\tau} \leftarrow \theta^{2,j}_{n\tau} - \lambda_1\gamma\sum_{q=0}^{\tau-1}\left(\theta^{2,j}_{n\tau+q} - m^1_{n\tau+q}\right) - \lambda_2\gamma\sum_{q=0}^{\tau-1}\nabla L_2(\theta^{2,j}_{n\tau+q})$$

Discretized FedCBO System

$$\theta^{1,i}_{(n+1)\tau} \leftarrow \theta^{1,i}_{n\tau} - \lambda_1\gamma \sum_{q=0}^{\tau-1} \left(\theta^{1,i}_{n\tau+q} - m^{1}_{n\tau+q}\right) - \lambda_2\gamma \sum_{q=0}^{\tau-1} \nabla L_1(\theta^{1,i}_{n\tau+q})$$

$$\theta^{2,j}_{(n+1)\tau} \leftarrow \theta^{2,j}_{n\tau} - \lambda_1\gamma \sum_{q=0}^{\tau-1} \left(\theta^{2,j}_{n\tau+q} - m^{1}_{n\tau+q}\right) - \lambda_2\gamma \sum_{q=0}^{\tau-1} \nabla L_2(\theta^{2,j}_{n\tau+q})$$

Splitting Scheme

Step 1:

$$\widehat{\theta}^{1,i}_{n\tau} \leftarrow \theta^{1,i}_{n\tau}, \qquad \widehat{\theta}^{2,j}_{n\tau} \leftarrow \theta^{2,j}_{n\tau}$$

Step 2:

$$\widehat{\theta}^{1,i}_{n\tau+q+1} \leftarrow \widehat{\theta}^{1,i}_{n\tau+q} - \lambda_2 \gamma \nabla L_1(\widehat{\theta}^{1,i}_{n\tau+q}), \qquad \widehat{\theta}^{2,j}_{n\tau+q+1} \leftarrow \widehat{\theta}^{2,j}_{n\tau+q} - \lambda_2 \gamma \nabla L_2(\widehat{\theta}^{2,j}_{n\tau+q}) \qquad \text{for } q = 0, \ldots, \tau - 1.$$

Step 3:

$$\theta^{1,i}_{(n+1)\tau} \leftarrow \widehat{\theta}^{1,i}_{(n+1)\tau} - \lambda_1 \gamma \left(\widehat{\theta}^{1,i}_{(n+1)\tau} - m^1_{(n+1)\tau} \right), \qquad \theta^{2,j}_{(n+1)\tau} \leftarrow \widehat{\theta}^{2,j}_{(n+1)\tau} - \lambda_1 \gamma \left(\widehat{\theta}^{2,j}_{(n+1)\tau} - m^2_{(n+1)\tau} \right)$$

FedCBO Algorithm

Algorithm 1 FedCBO

Input: Initialized model $\theta_0^j \in \mathbb{R}^d, j \in [N]$; Number of iterations T; Number of local gradient steps τ; Number of models downloaded M; CBO system hyperparameters $\lambda_1, \lambda_2, \alpha$; Discretization step size γ; Initialized sampling likelihood $P_0 \in \mathbb{R}^{N\times(N-1)}$;

1: **for** $n = 0, \cdots, T-1$ **do**
2: $\quad G_n \leftarrow$ random subset of agents (participating devices);
3: $\quad$ **LocalUpdate**$(\theta_n^j, \tau, \lambda_2, \gamma)$ for $j \in G_n$;
4: $\quad$ **LocalAggregation**(agent j) for $j \in G_n$;
5: **end for**

Output: θ_T^j for $j \in [N]$.

LocalUpdate$(\widehat{\theta}_0, \tau, \lambda_2, \gamma)$ at j-th agent

6: **for** $q = 0, \cdots, \tau - 1$ **do**
7: $\quad$ (stochastic) gradient descent $\widehat{\theta}_{q+1} \leftarrow \widehat{\theta}_q - \lambda_2\gamma\nabla L_j(\widehat{\theta}_q)$;
8: **end for**
9: **return** $\widehat{\theta}_\tau$;

FedCBO Algorithm

Algorithm 2 LocalAggregation(agent j)

Input: Agent j's model $\theta_n^j \in \mathbb{R}^d$; Participating devices at n iteration G_n; Sampling likelihood $P_n^j \in \mathbb{R}^{N-1}$;
CBO system hyperparameters λ_1, α; Discretization step size γ; Random sample proportion $\varepsilon \in (0,1)$;
Number of models downloaded M;

1: $A_n \leftarrow$ **ε-greedySampling**(P_n^j, G_n, M);

2: Agent j downloads models θ_n^i for $i \in A_n$;

3: Evaluate models θ_n^i on agent j's data set respectively and denote the corresponding loss as L_j^i;

4: Calculate consensus point m_j by

$$m_j \leftarrow \frac{1}{\sum_{i \in A_n} \mu_j^i} \sum_{i \in A_n} \theta_n^i \mu_j^i, \qquad \text{with } \mu_j^i = \exp(-\alpha L_j^i) \tag{19}$$

5: Update agent j's model by

$$\theta_{n+1}^j \leftarrow \theta_n^j - \lambda_1 \gamma (\theta_n^j - m_j), \tag{20}$$

6: Update sampling likelihood P_n^j by

$$P_{n+1}^{j,i} \leftarrow P_n^{j,i} + (L_j^j - L_j^i), \qquad \text{for } i \in A_n \tag{21}$$

Output: $\theta_{n+1}^j, P_{n+1}^j$

<u>**ε-greedySampling**</u>(P_n^j, G_n, M)

7: Randomly sample $\varepsilon * M$ number of agents from G_n, denoted as A_n^1;

8: Select $(1-\varepsilon) * M$ numbers of agents in $G_n \backslash A_n^1$ with top value $P_j^{j,i}, i \in G_n \backslash A_n^1$, denoted as A_n^2;

9: **return** $A_n = A_n^1 \cup A_n^2$