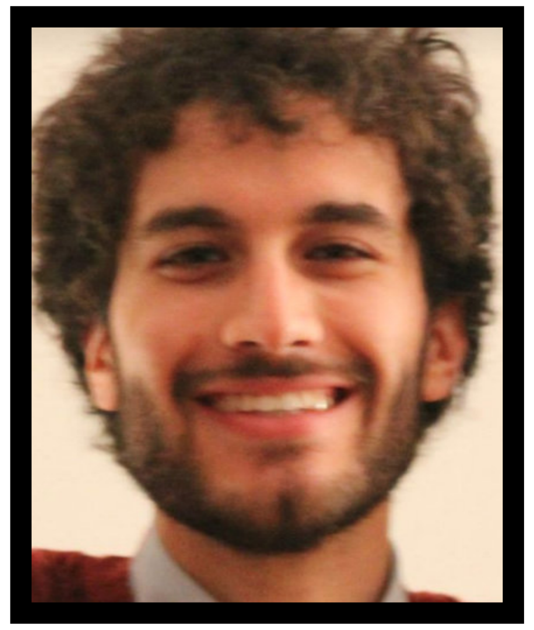
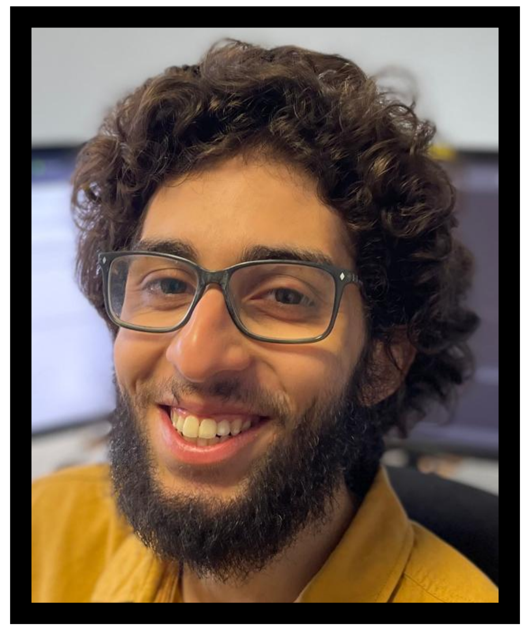
FL under heterogeneous and correlated client availability

Angelo Rodio

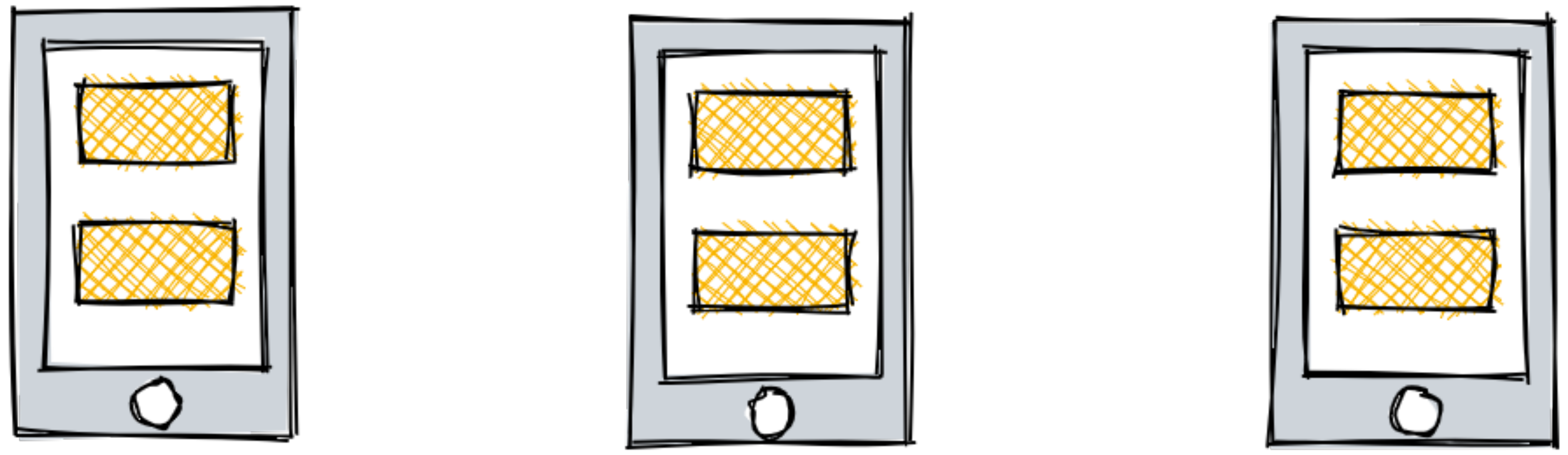
Francescomaria Faticanti

Othmane Marfoq

Giovanni Neglia

Emilio Leonardi

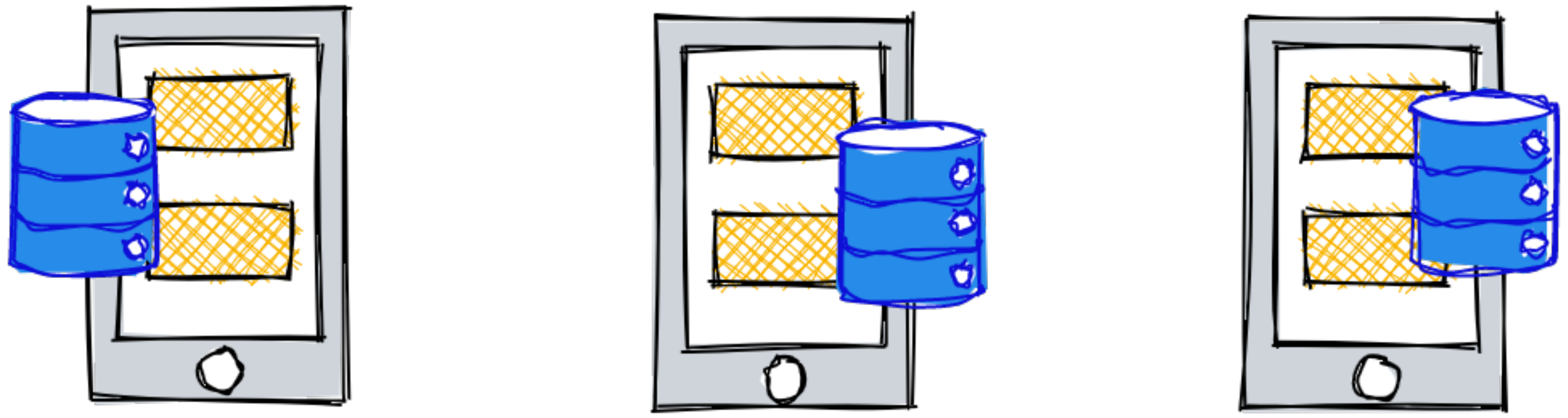
Federated Learning



Clients $k = 1, \ldots, K$

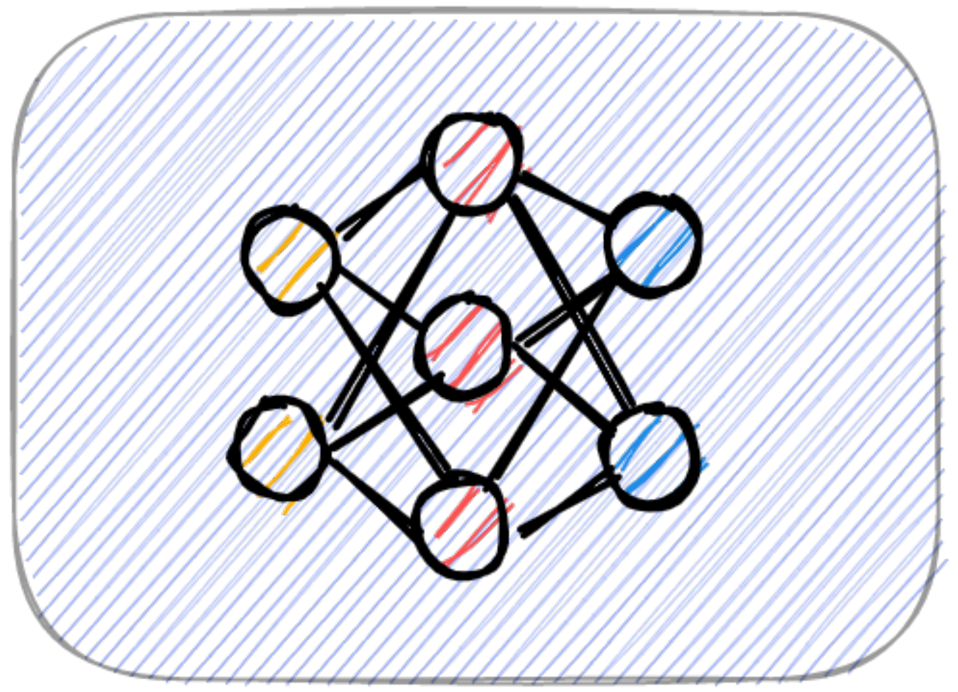
Federated Learning

Dataset $D_k = \{\xi_{kl}\}_{l=1}^{n_k}$



Clients $k = 1, \ldots, K$

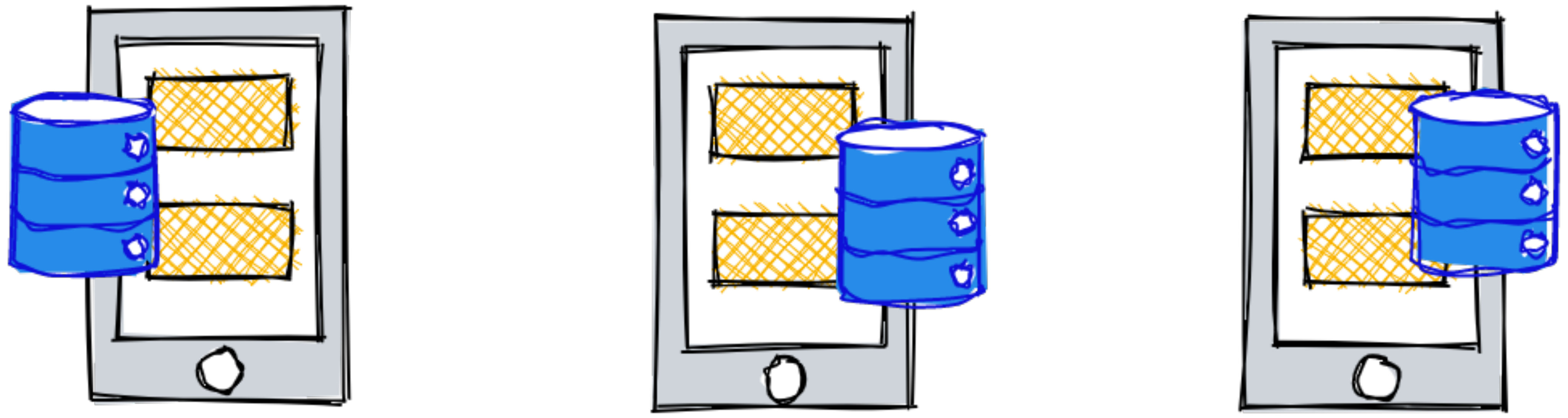
Federated Learning



Global model

$\boldsymbol{w} \in \mathbb{R}^d$

Dataset $D_k = \{\xi_{kl}\}_{l=1}^{n_k}$

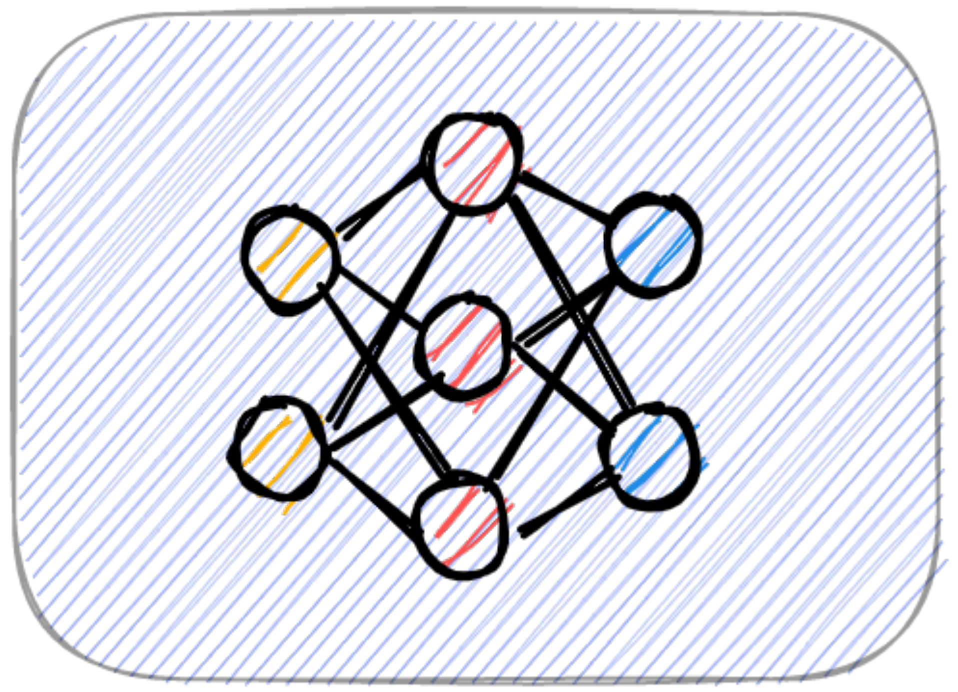


Clients $k = 1, \ldots, K$

Federated Learning

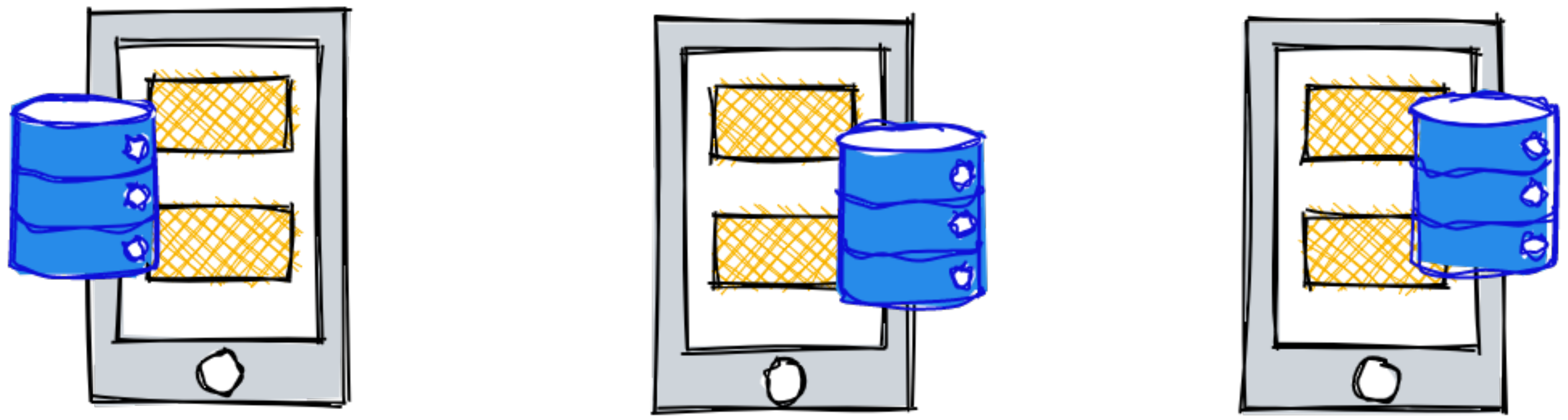
Solve the optimization problem

$$\min_{\boldsymbol{w}} F(\boldsymbol{w}) = \sum_{k=1}^{K} \alpha_k \, F_k(\boldsymbol{w})$$



where

$$F_k(\boldsymbol{w}) = \frac{1}{n_k} \sum_{l=1}^{n_k} \ell(\boldsymbol{w}, \xi_{kl})$$

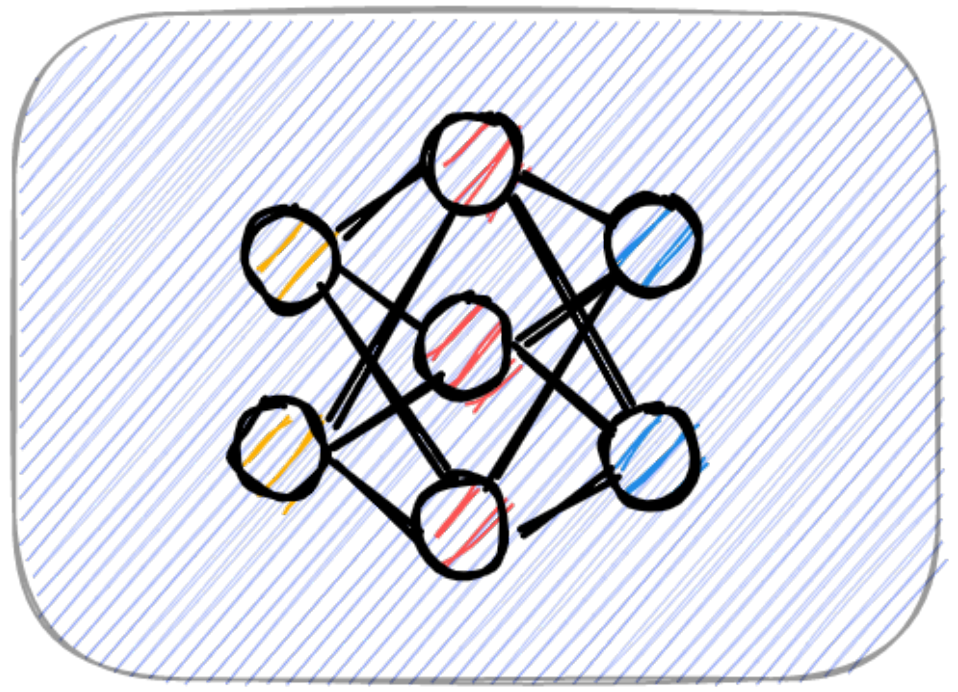


Federated Learning

Solve the optimization problem

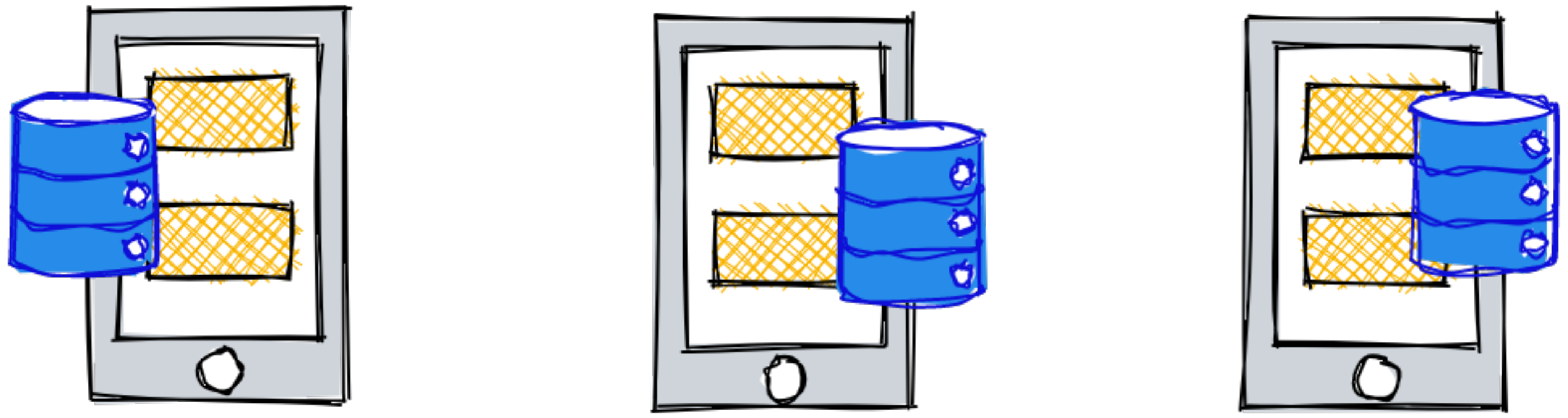
$$\min_{\boldsymbol{w}} F(\boldsymbol{w}) = \sum_{k=1}^{K} \alpha_k \, F_k(\boldsymbol{w})$$

$\boldsymbol{\alpha}$: target importance



where

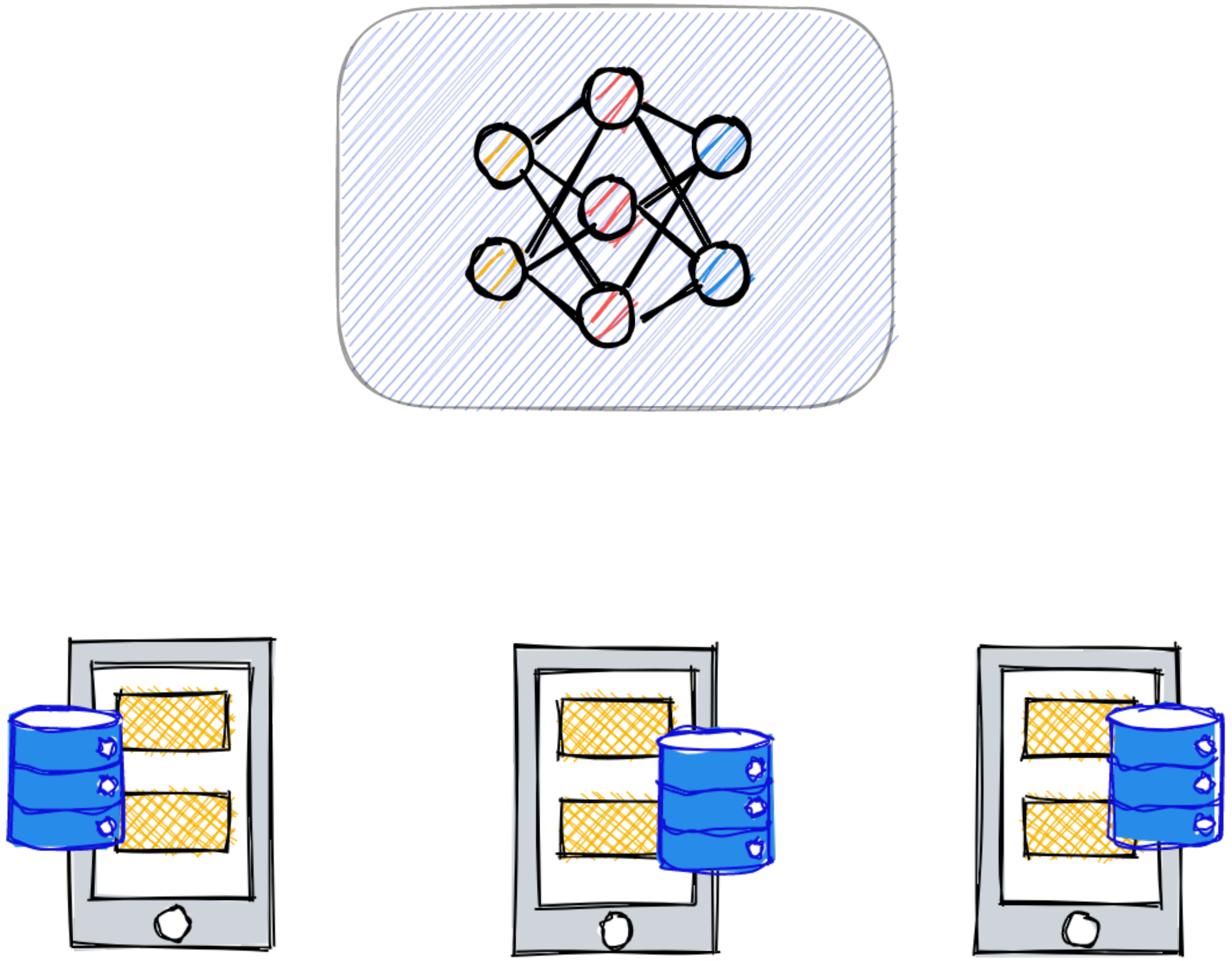
$$F_k(\boldsymbol{w}) = \frac{1}{n_k} \sum_{l=1}^{n_k} \ell(\boldsymbol{w}, \xi_{kl})$$



General FL Algorithm

for $t \in \{0, \ldots, T-1\}$ **do**:

A_t: set of active clients at time t

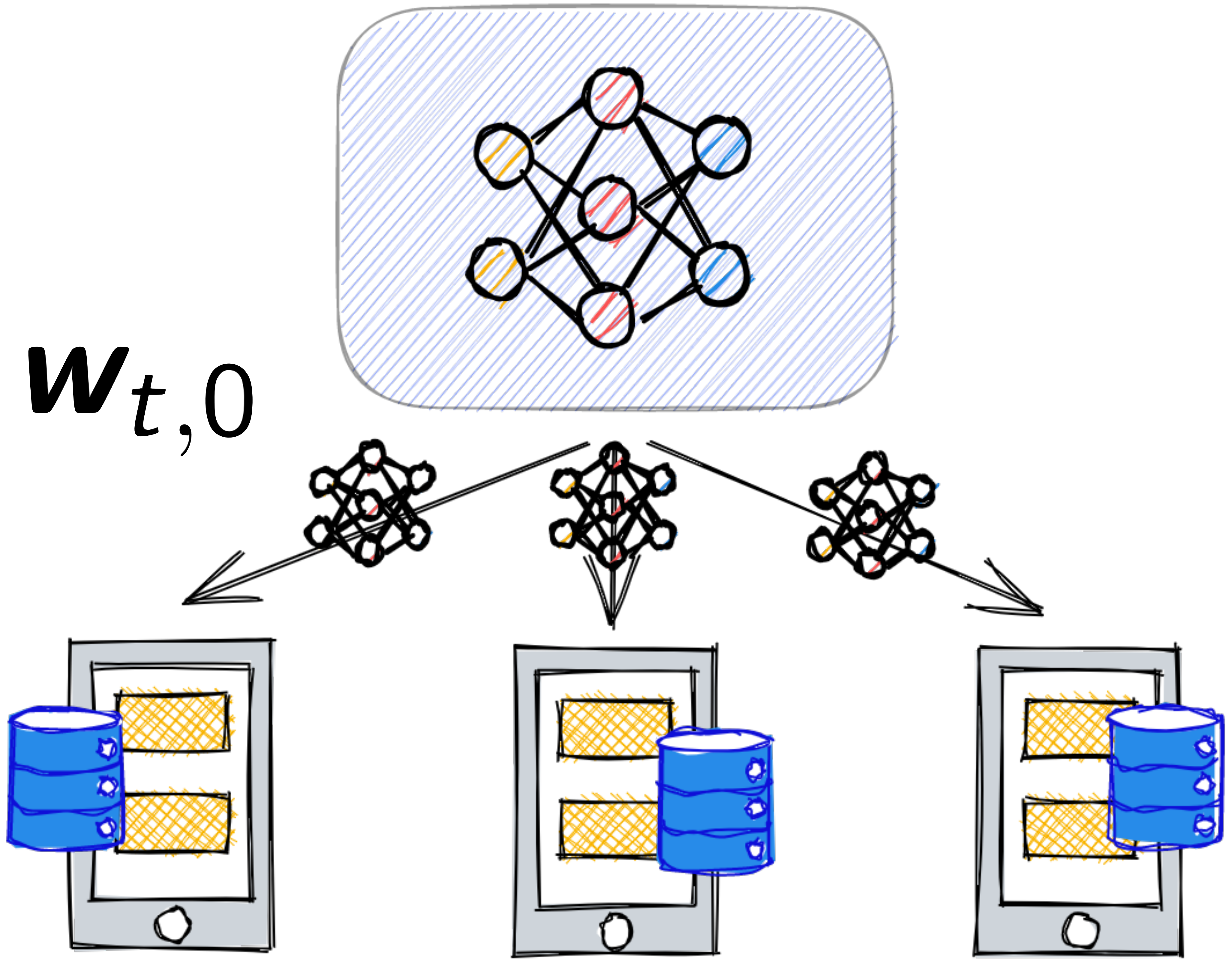


General FL Algorithm

for $t \in \{0, \ldots, T-1\}$ **do**:

A_t: set of active clients at time t

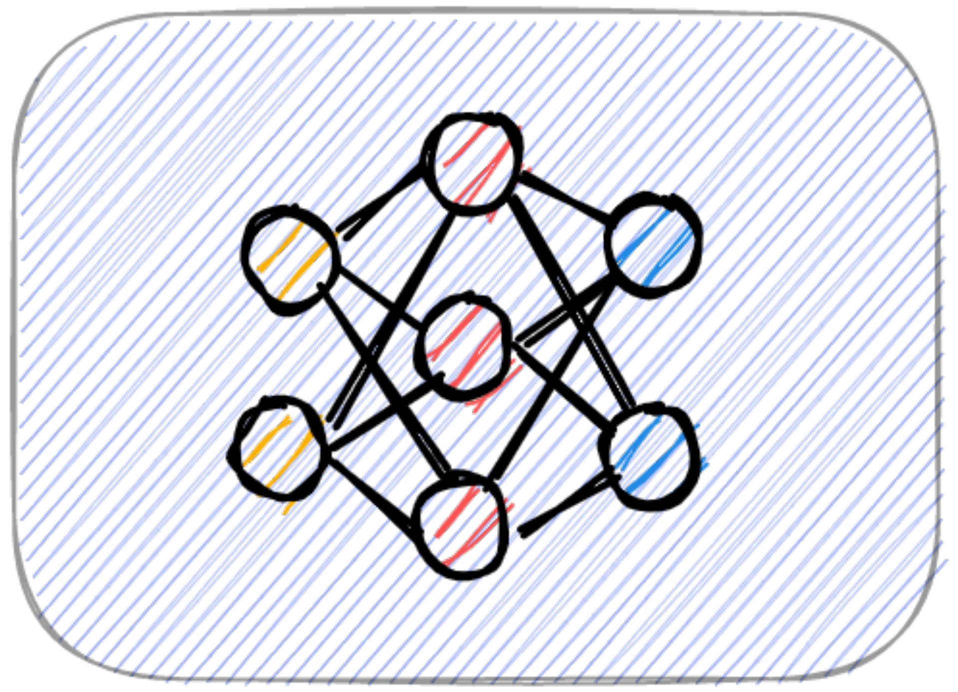
(1)



General FL Algorithm

for $t \in \{0, \ldots, T-1\}$ **do**:

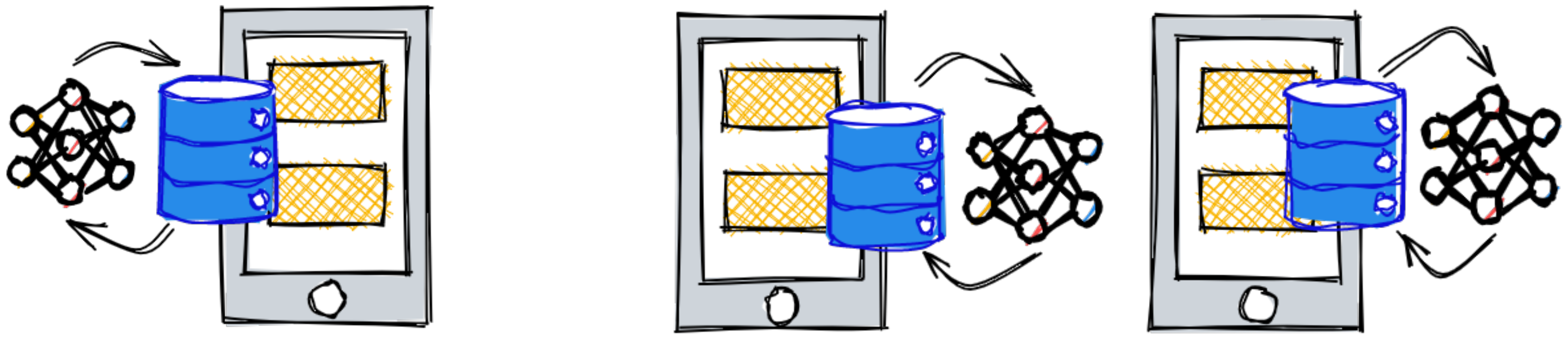
A_t: set of active clients at time t



(2) **For** $j = 0, \ldots, E-1$ **do** :

$$\boldsymbol{w}_{t,j+1}^{k} = \boldsymbol{w}_{t,j}^{k} - \eta_t \nabla F_k(\boldsymbol{w}_{t,j}^{k}, \mathcal{B}_{t,j}^{k})$$

$\boldsymbol{w}_{t,E}^{k}$

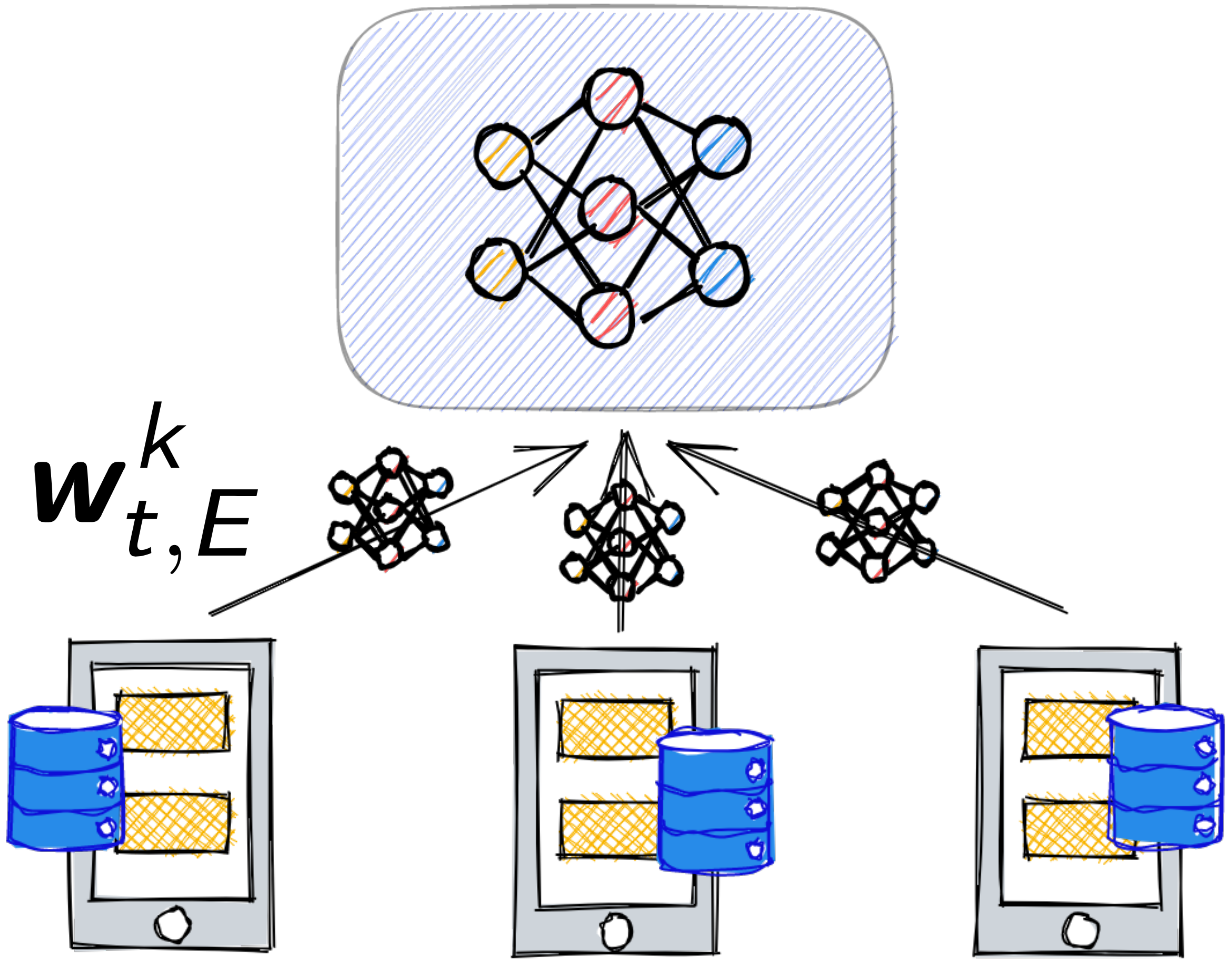


General FL Algorithm

for $t \in \{0, \ldots, T-1\}$ **do**:

A_t: set of active clients at time t

(3)



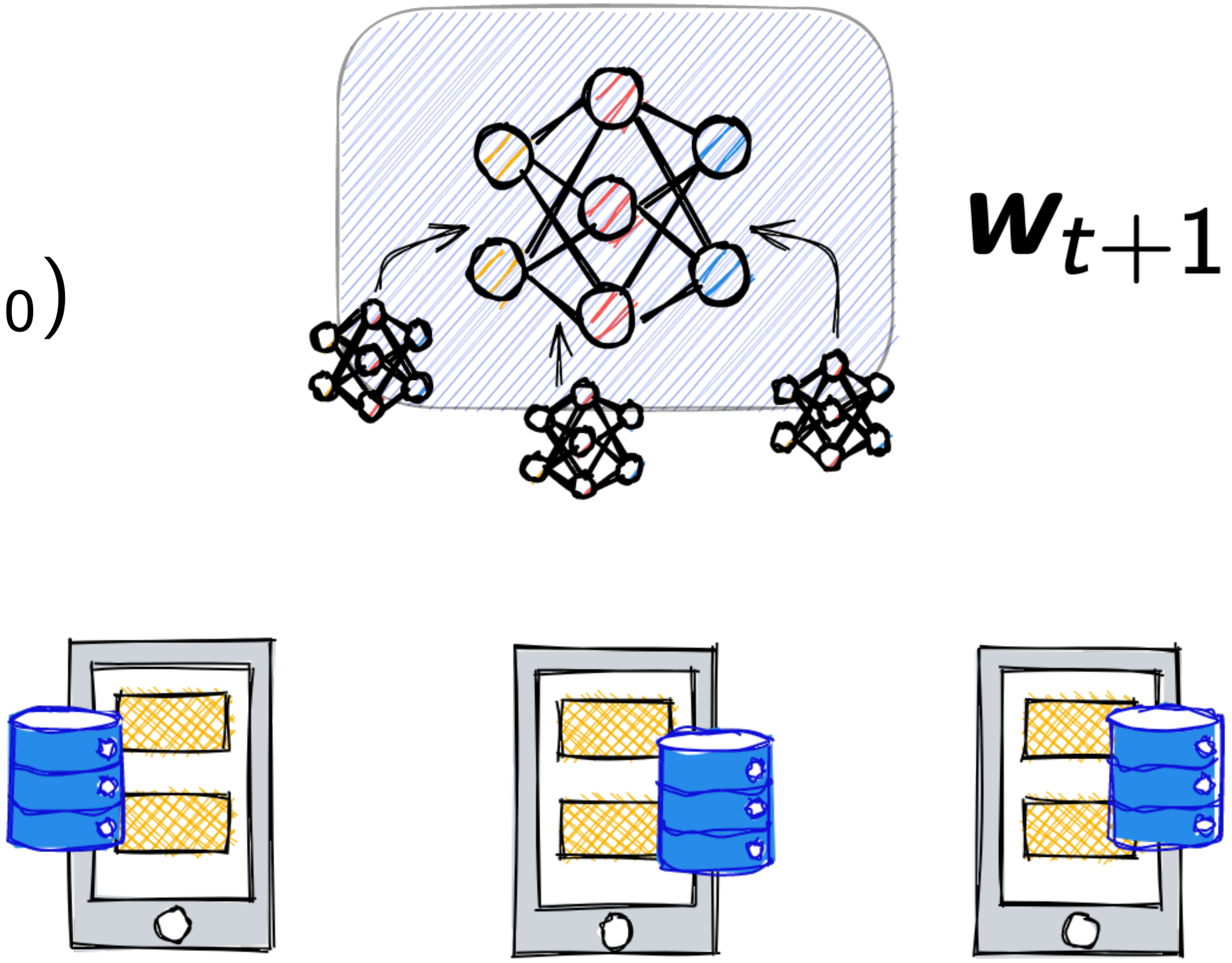
General FL Algorithm

for $t \in \{0, \ldots, T-1\}$ **do**:

A_t: set of active clients at time t

(4) $\boldsymbol{w}_{t+1,0} = \boldsymbol{w}_{t,0} + \sum_{k \in A_t} q_k (\boldsymbol{w}_{t,E}^k - \boldsymbol{w}_{t,0})$

$\boldsymbol{q}$: aggregation weights



$\boldsymbol{w}_{t+1,0}$

Client Availability in Federated Learning

Previous work in FL commonly assumed:

- Clients are always active or have uniform availability

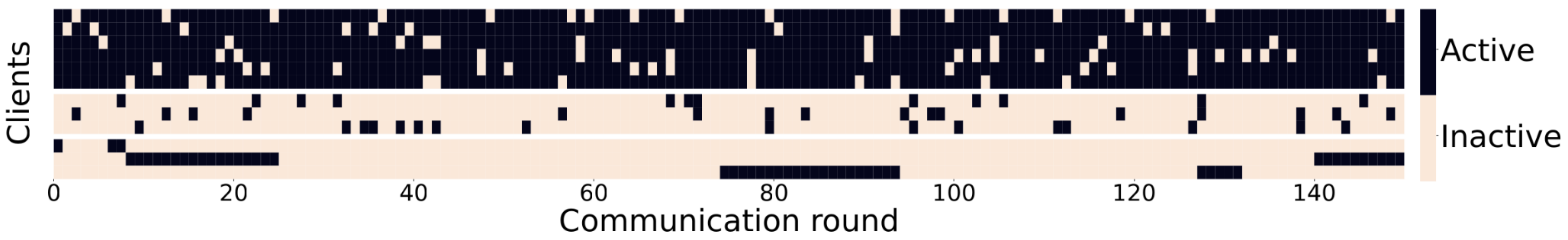
We take into account the Heterogeneity and Correlation in Client Availability

Client Availability in Federated Learning

Previous work in FL commonly assumed:

- Clients are always active or have uniform availability

We take into account the Heterogeneity and Correlation in Client Availability

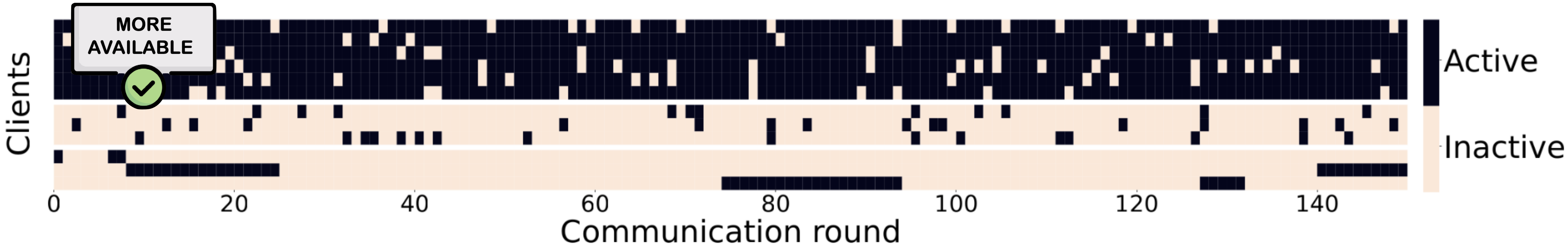


Client Availability in Federated Learning

Previous work in FL commonly assumed:

- Clients are always active or have uniform availability

We take into account the Heterogeneity and Correlation in Client Availability

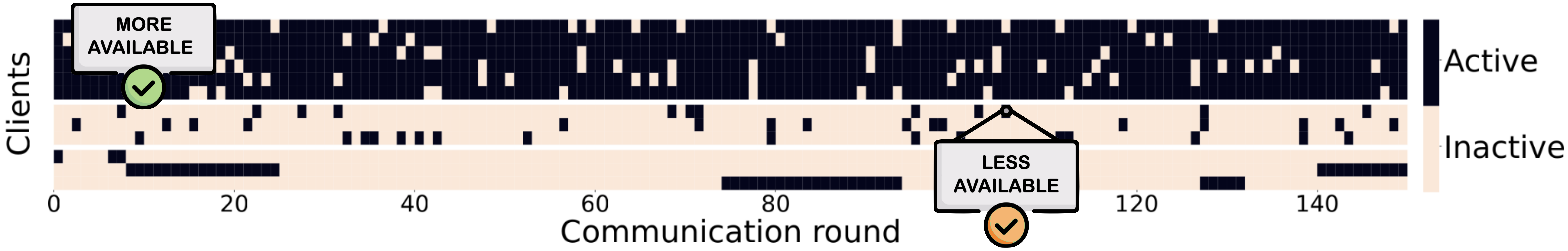


Client Availability in Federated Learning

Previous work in FL commonly assumed:

- Clients are always active or have uniform availability

We take into account the Heterogeneity and Correlation in Client Availability

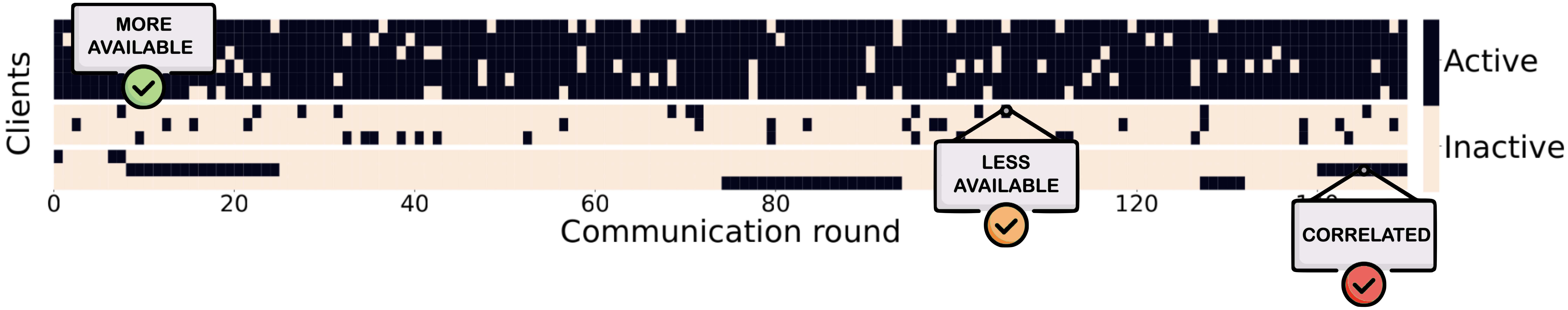


Client Availability in Federated Learning

Previous work in FL commonly assumed:

- Clients are always active or have uniform availability

We take into account the Heterogeneity and Correlation in Client Availability

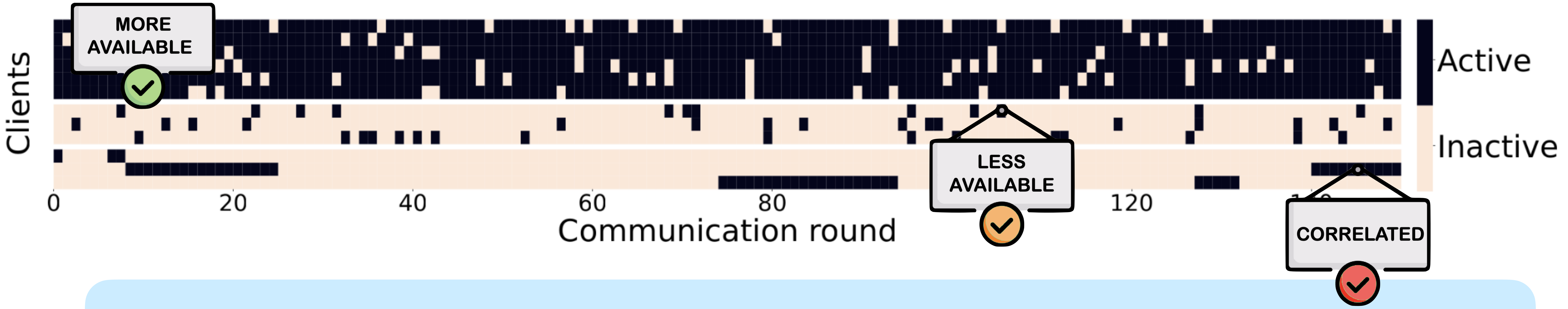


Client Availability in Federated Learning

Previous work in FL commonly assumed:

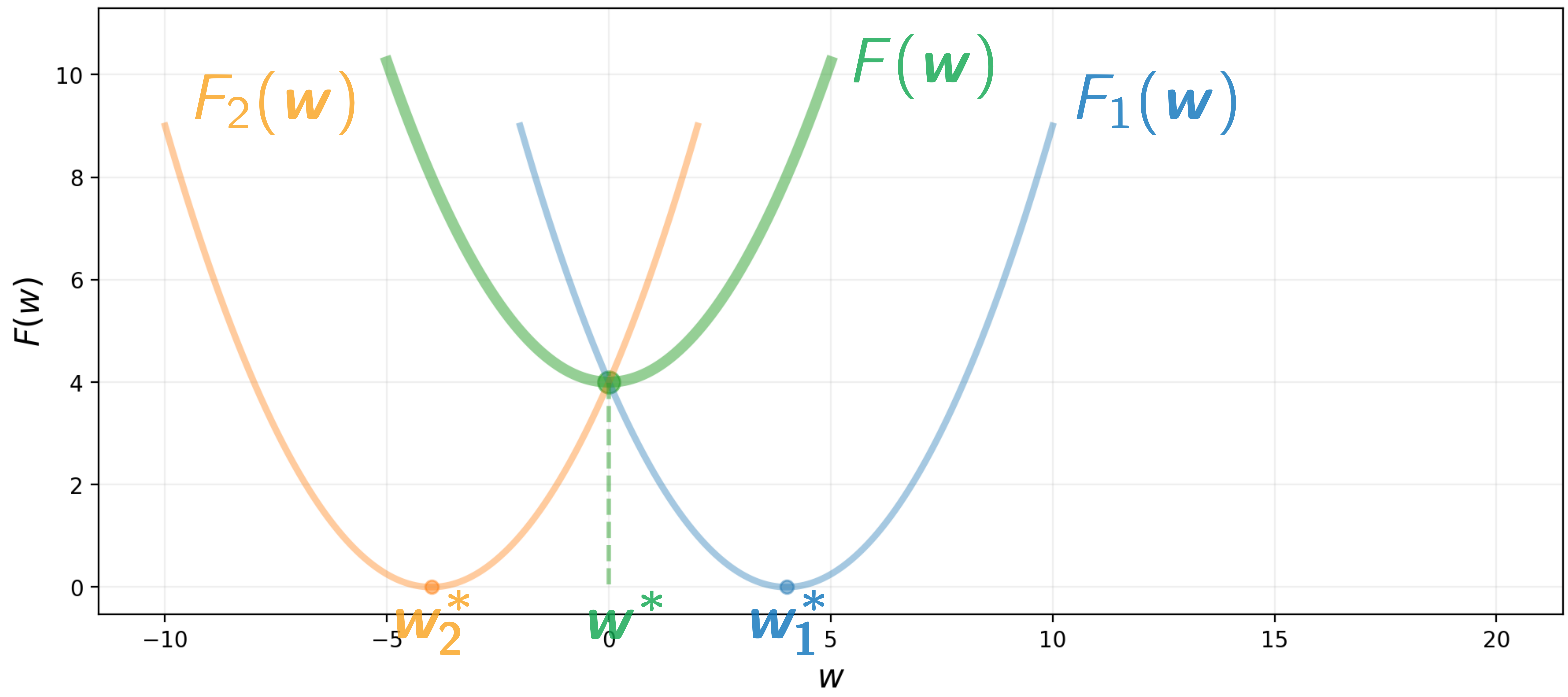
- Clients are always active or have uniform availability

We take into account the Heterogeneity and Correlation in Client Availability

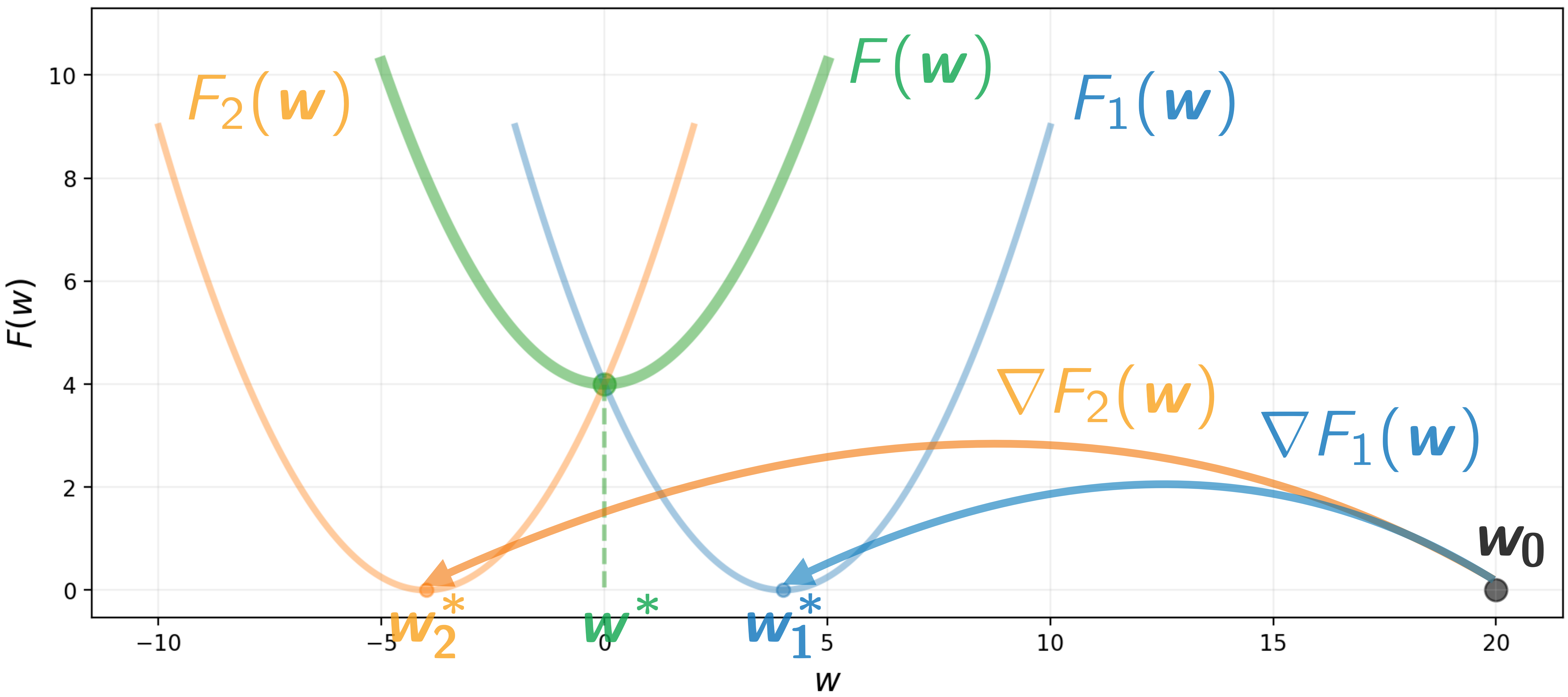


- Optimize training to make the best use of available client resources
- Minimize the impact of less available and correlated clients

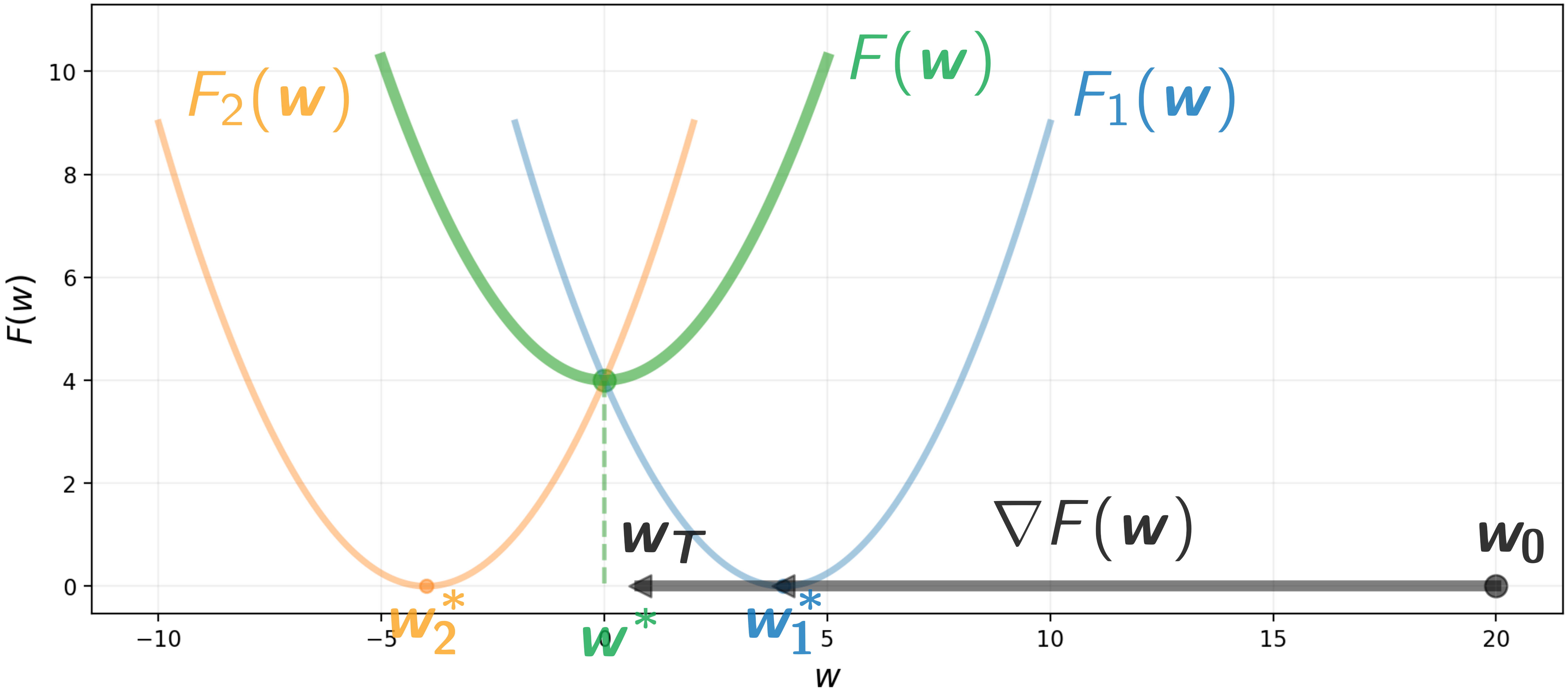
Quadratic Example



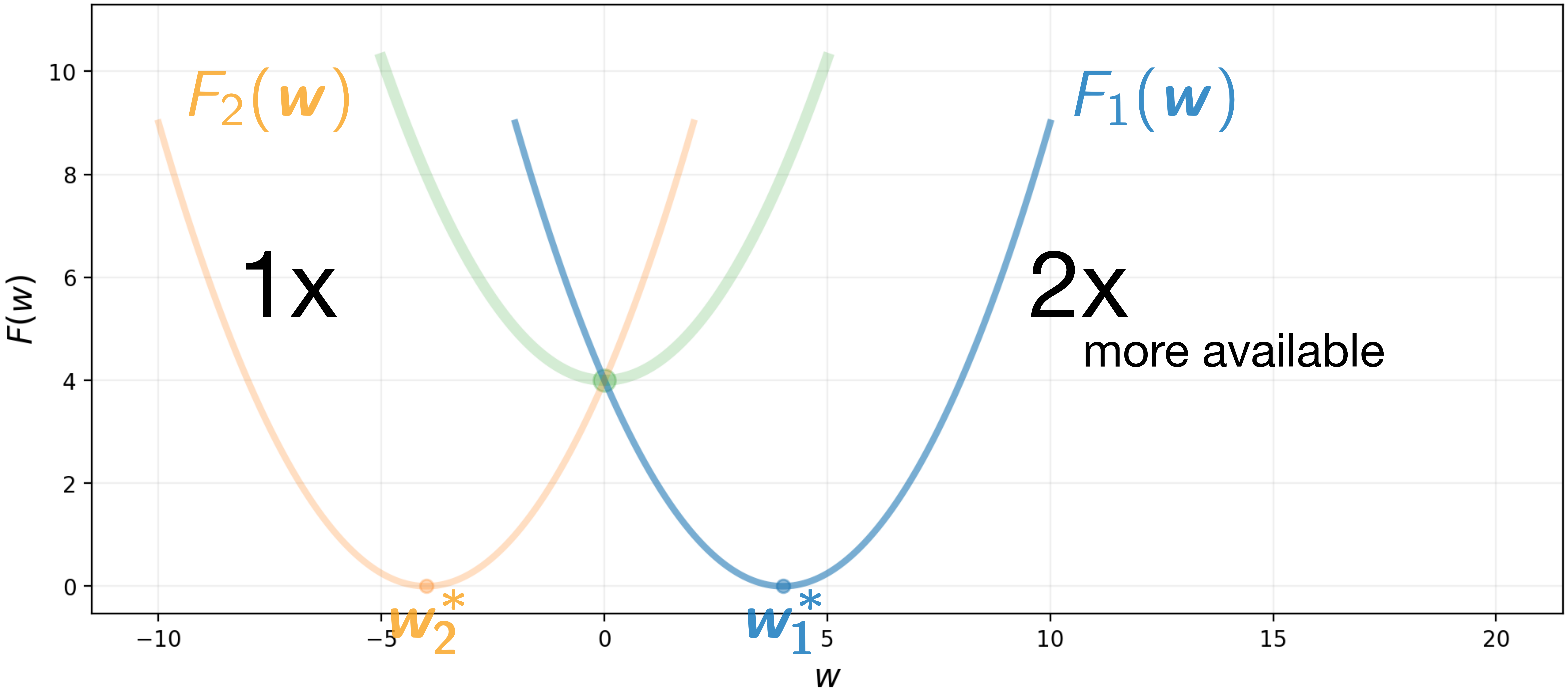
1) All Clients Available



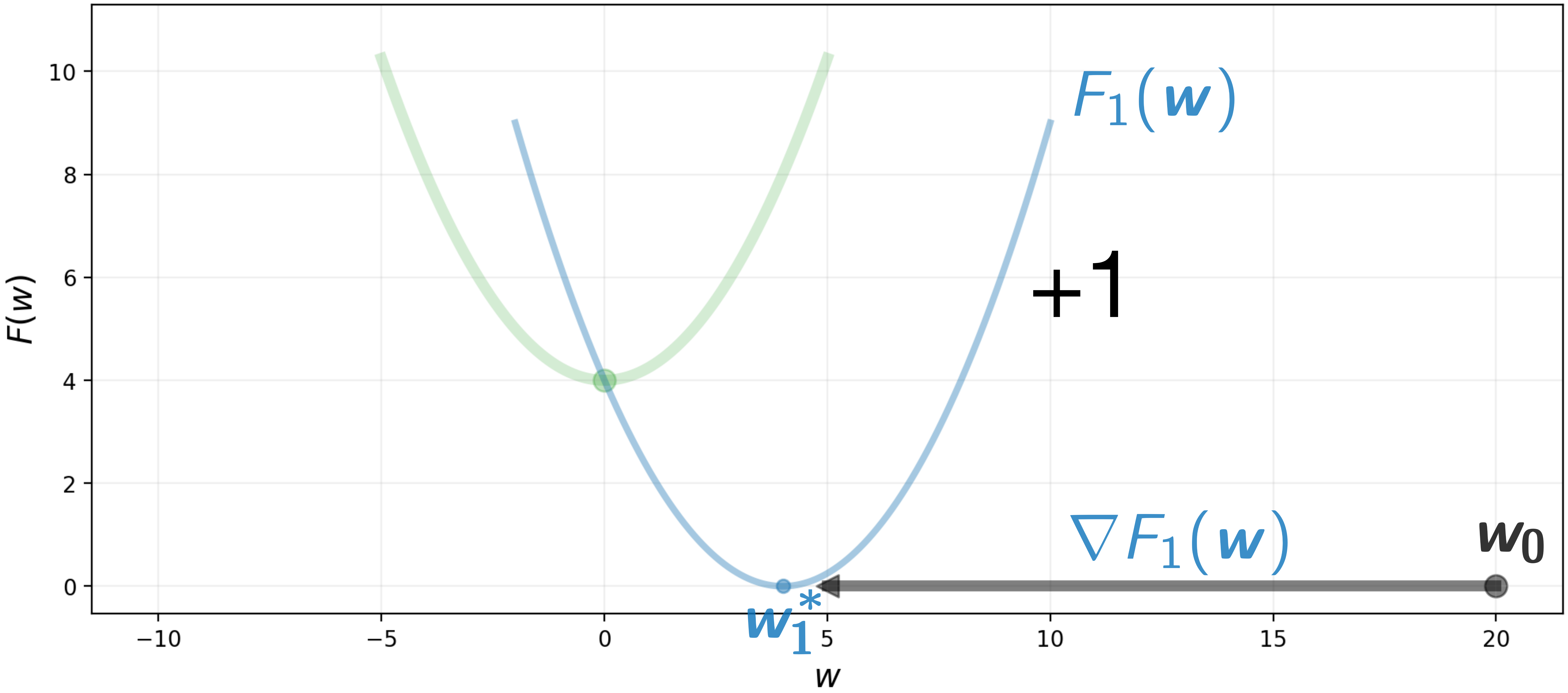
1) All Clients Available



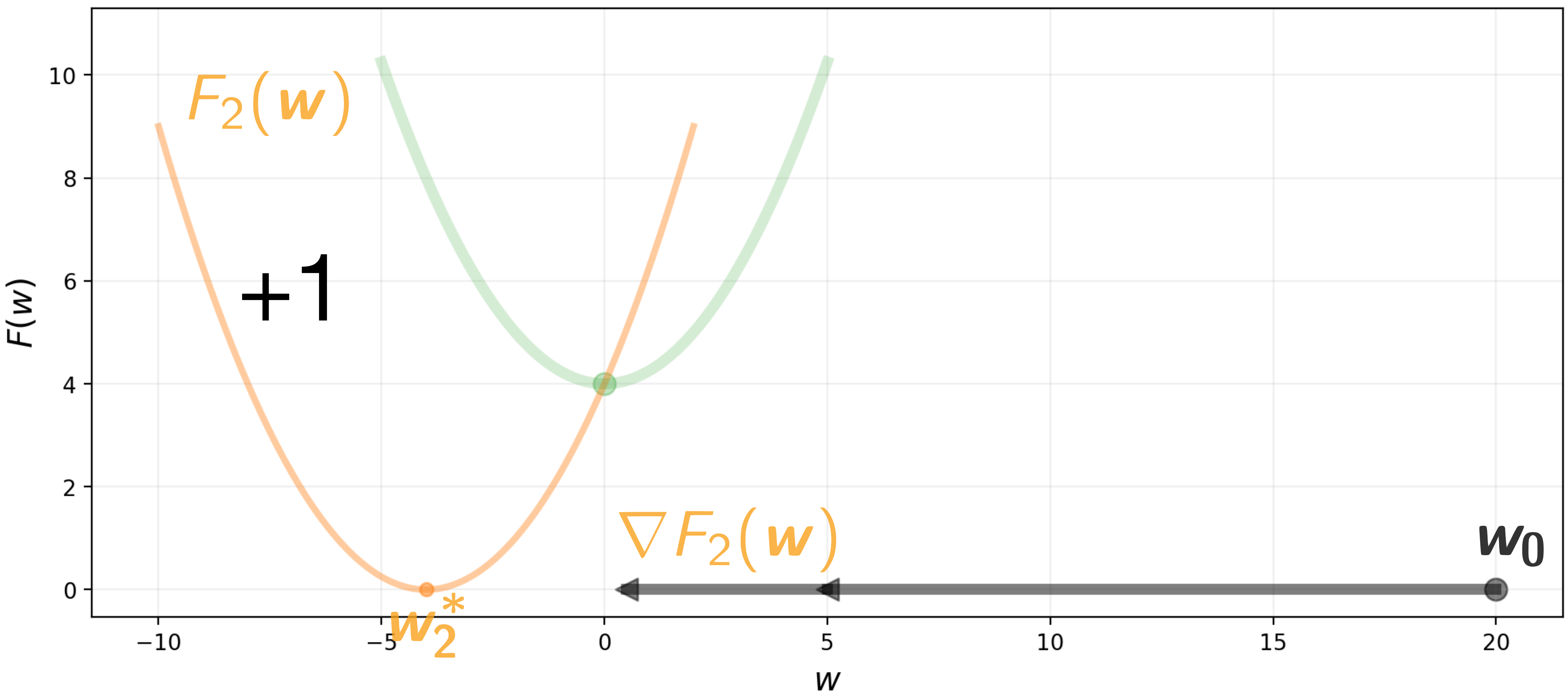
2) Heterogeneous Client Availability



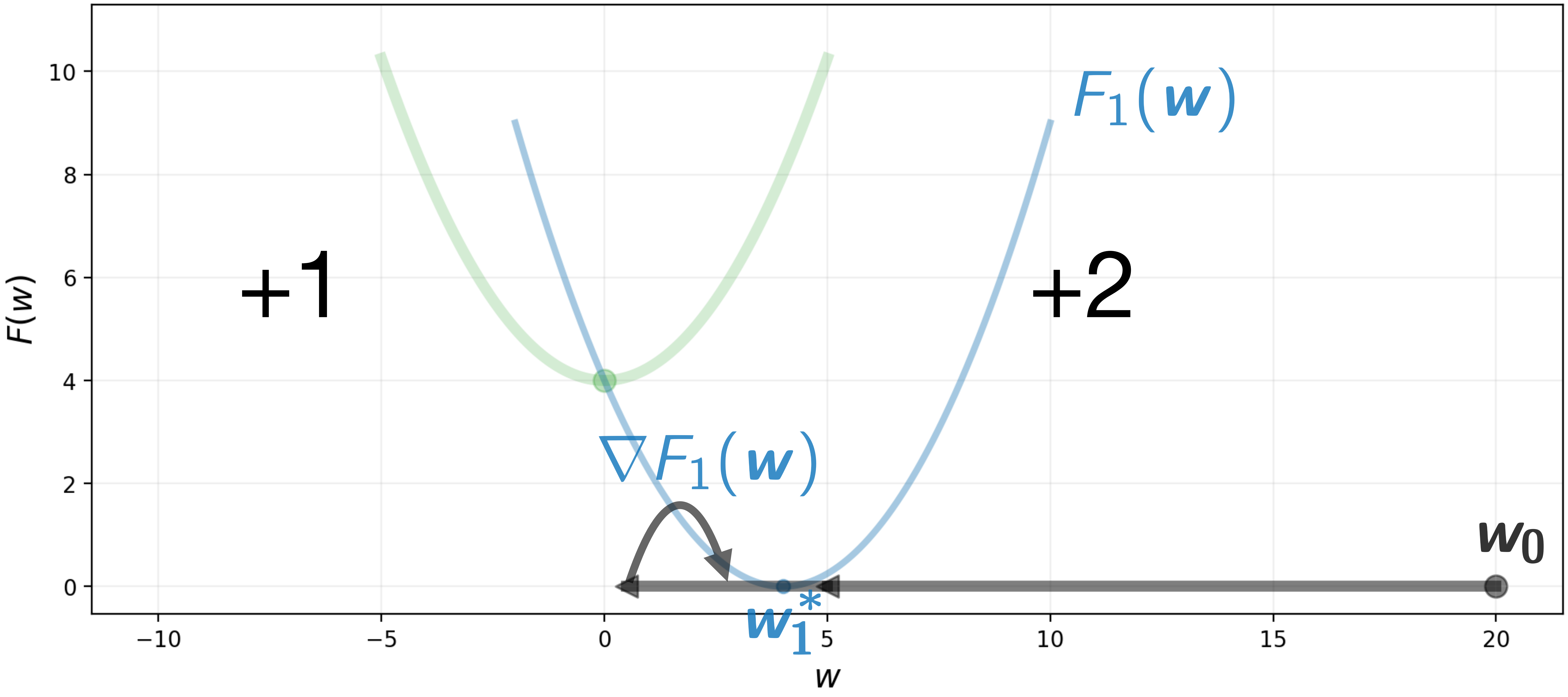
2) Heterogeneous Client Availability



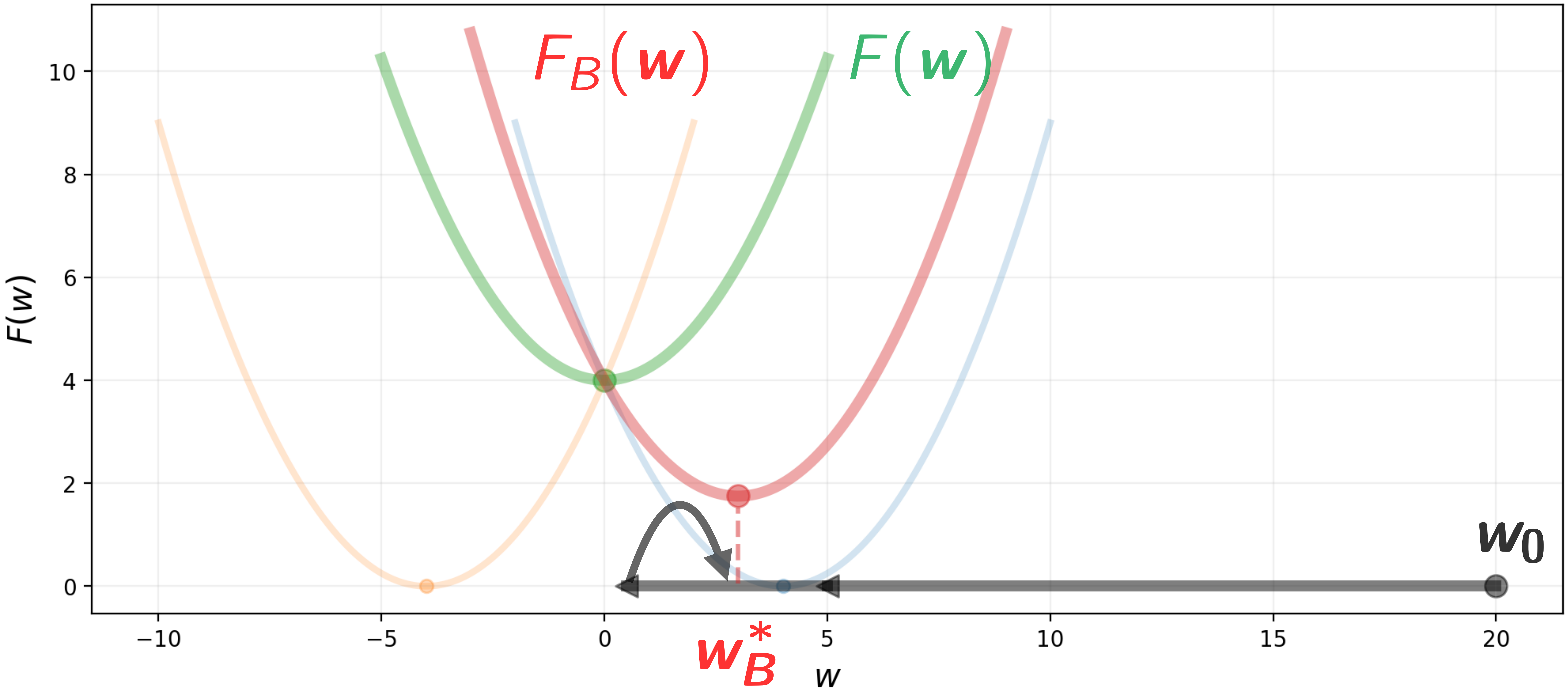
2) Heterogeneous Client Availability



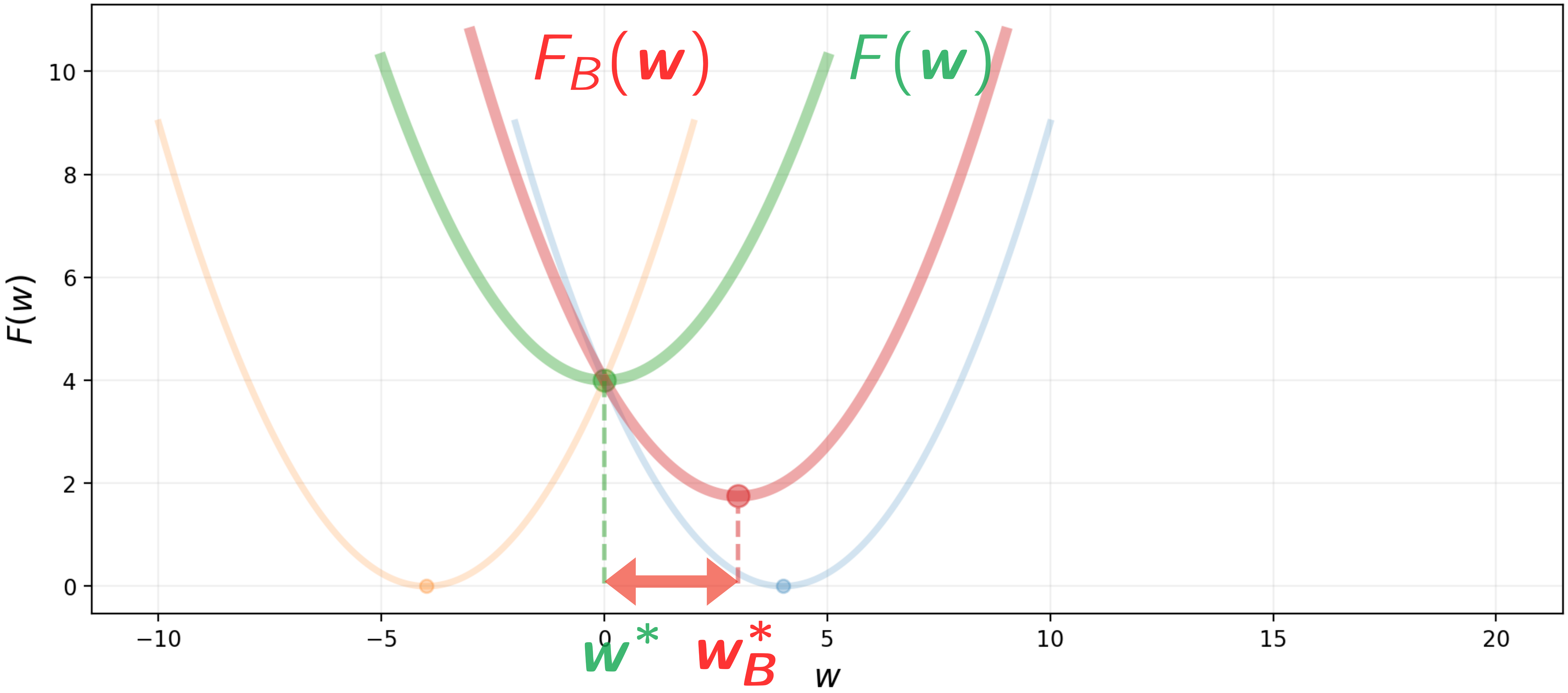
2) Heterogeneous Client Availability



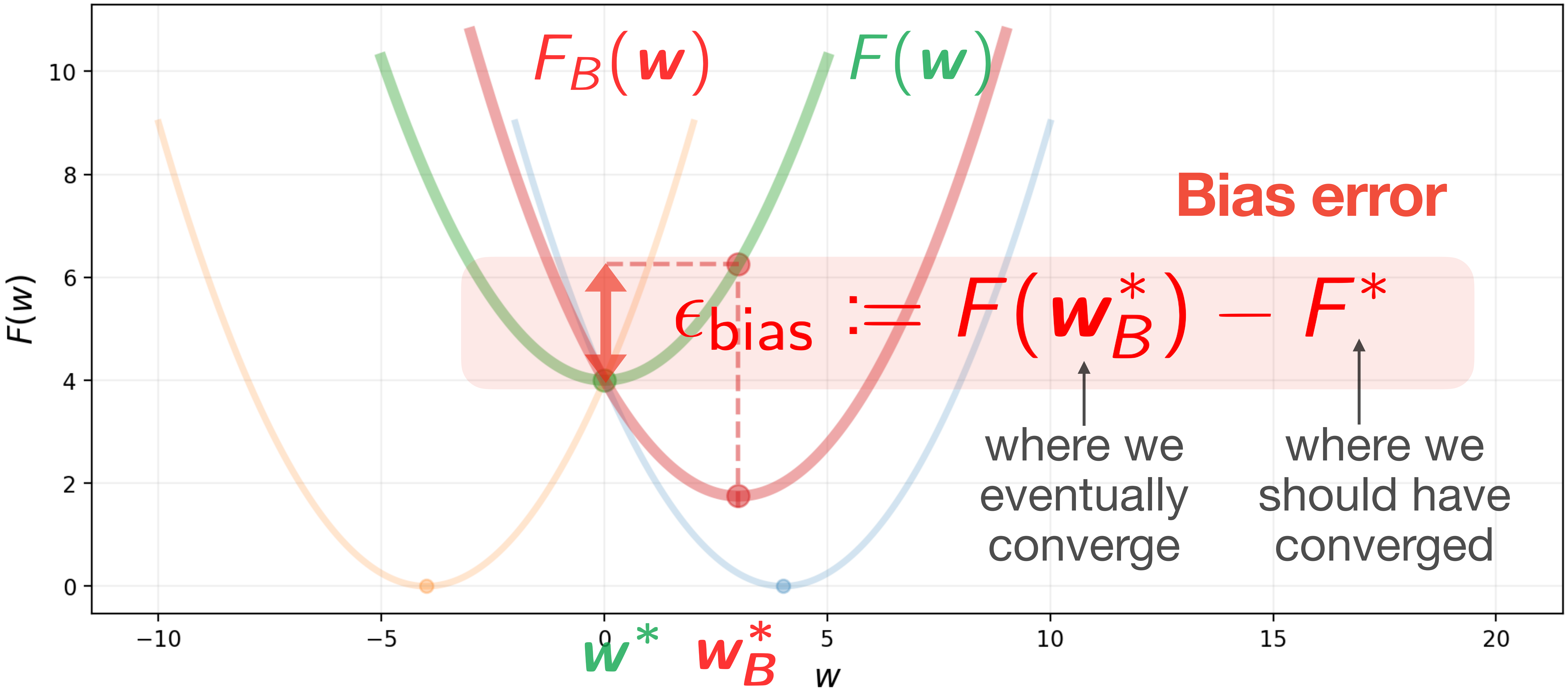
2) Heterogeneous Client Availability



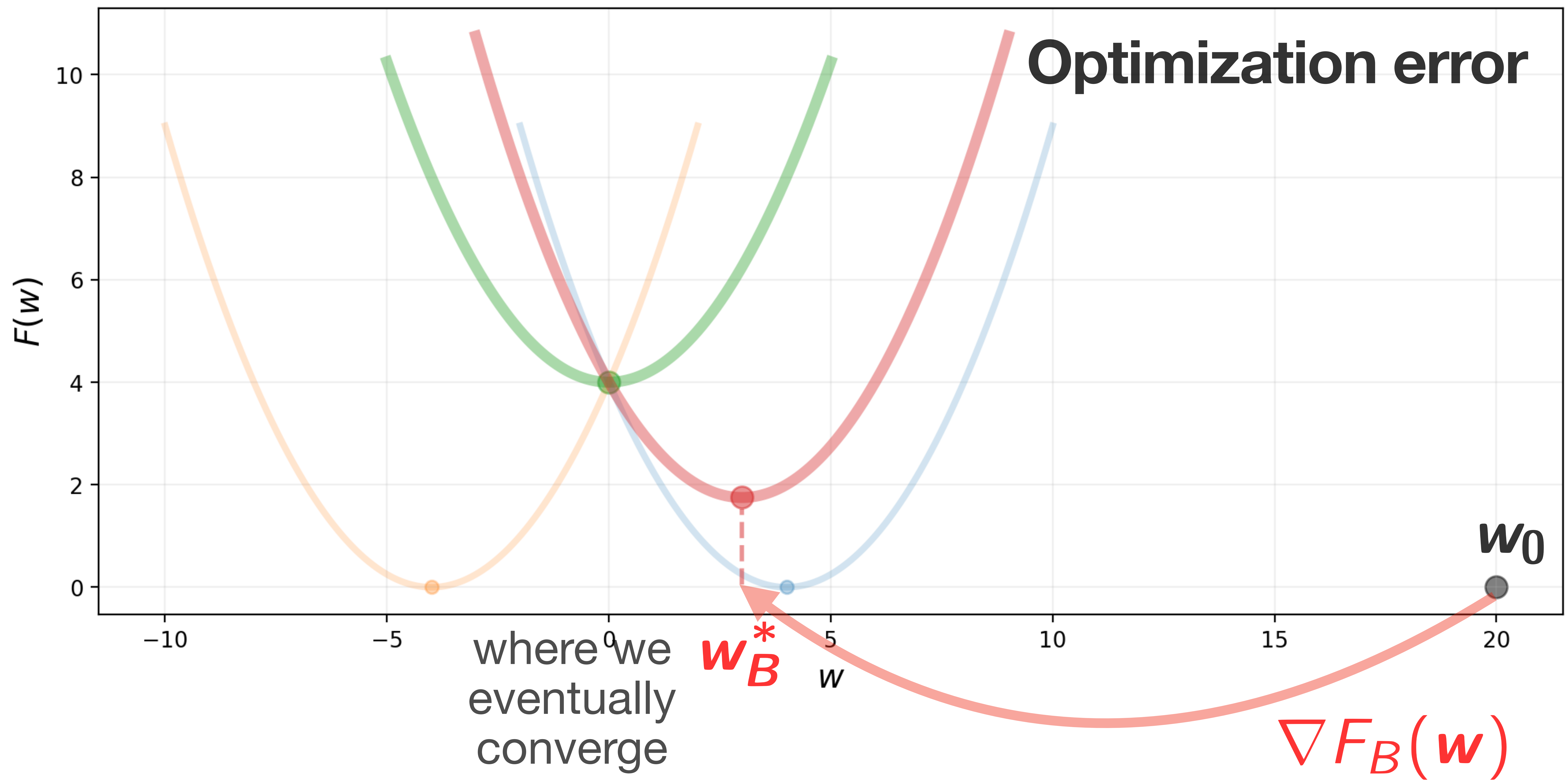
2) Heterogeneous Client Availability



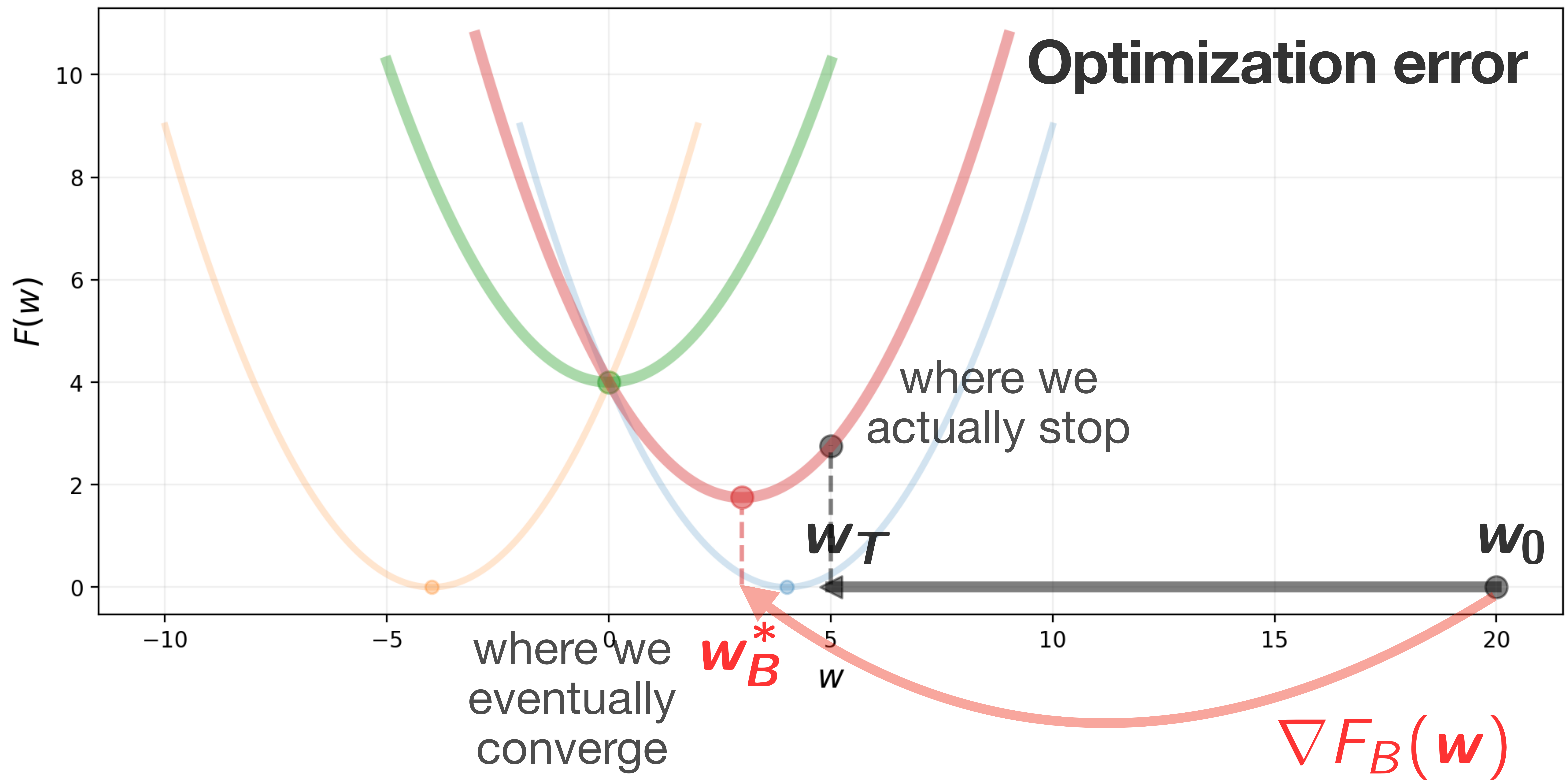
2) Heterogeneous Client Availability



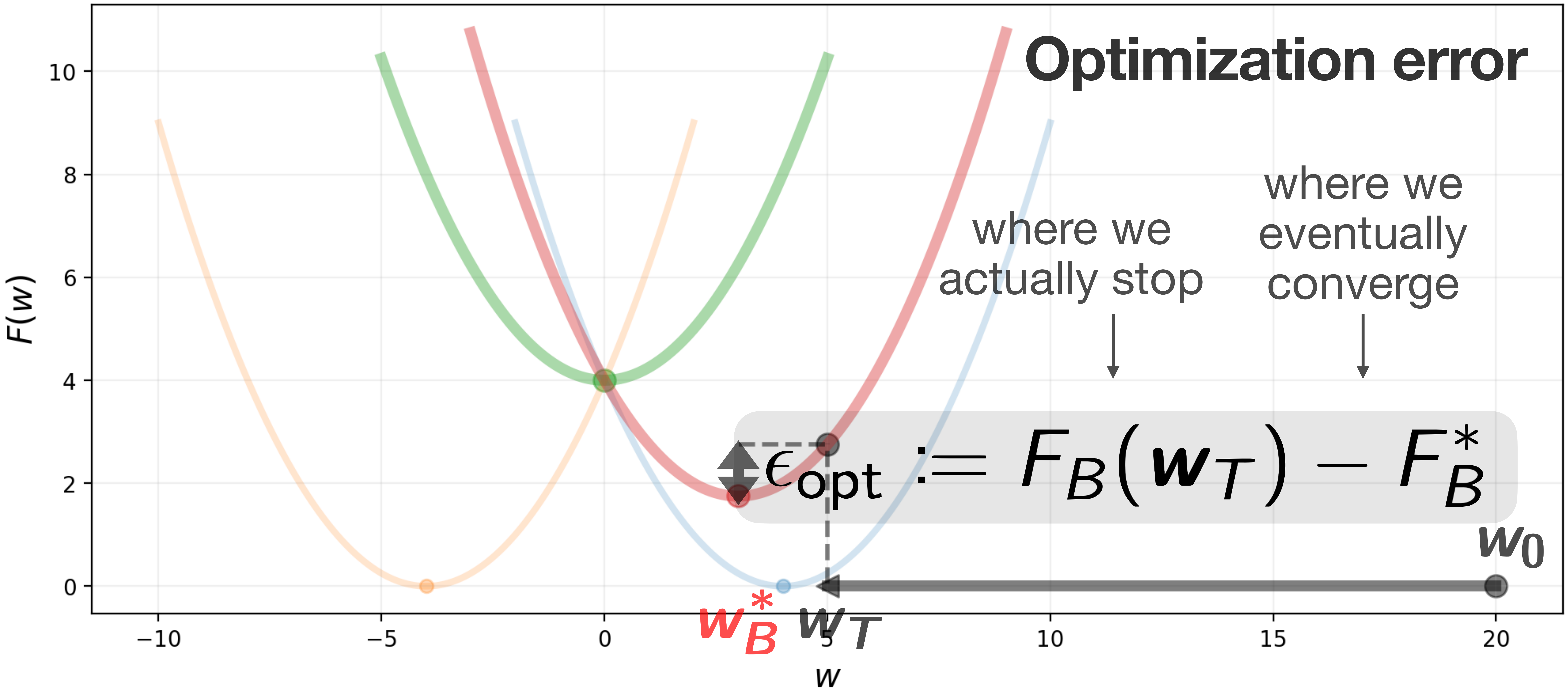
2) Heterogeneous Client Availability



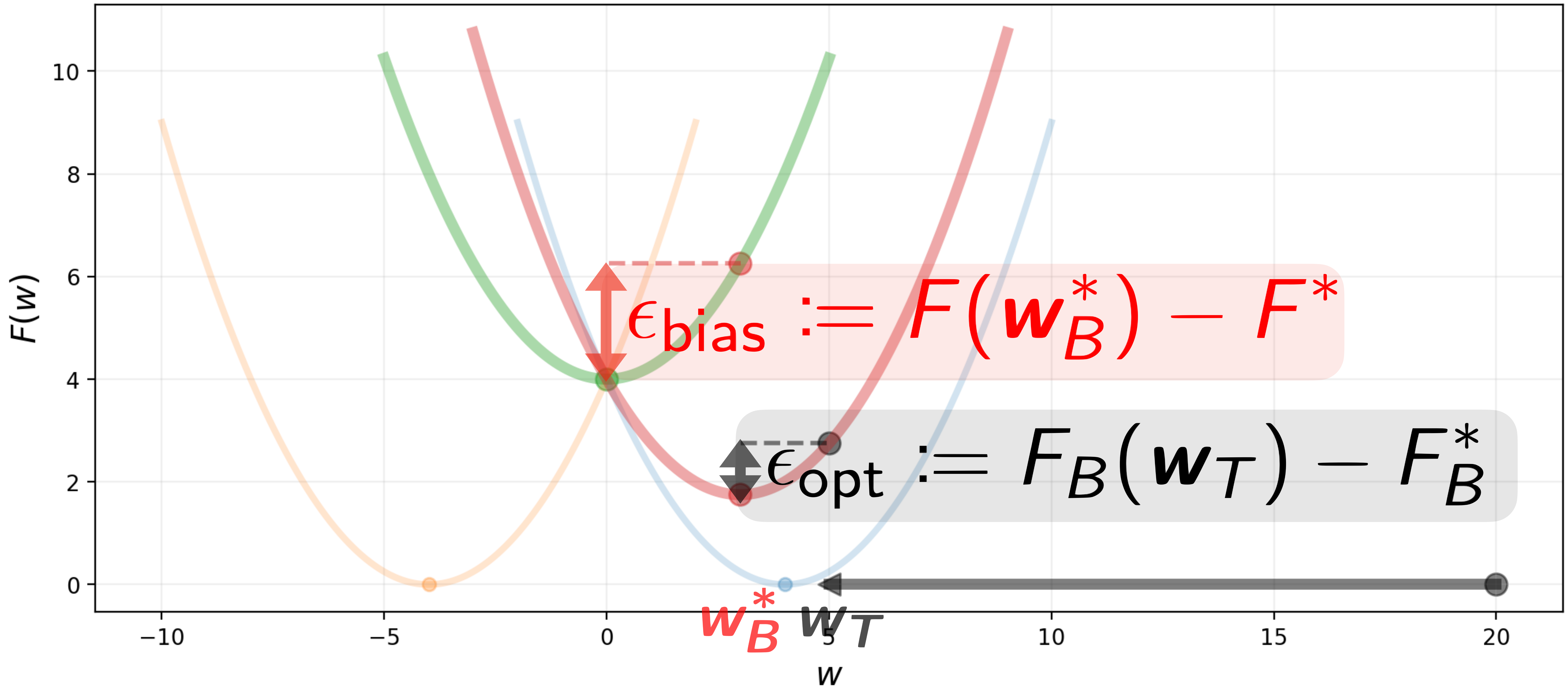
2) Heterogeneous Client Availability



2) Heterogeneous Client Availability

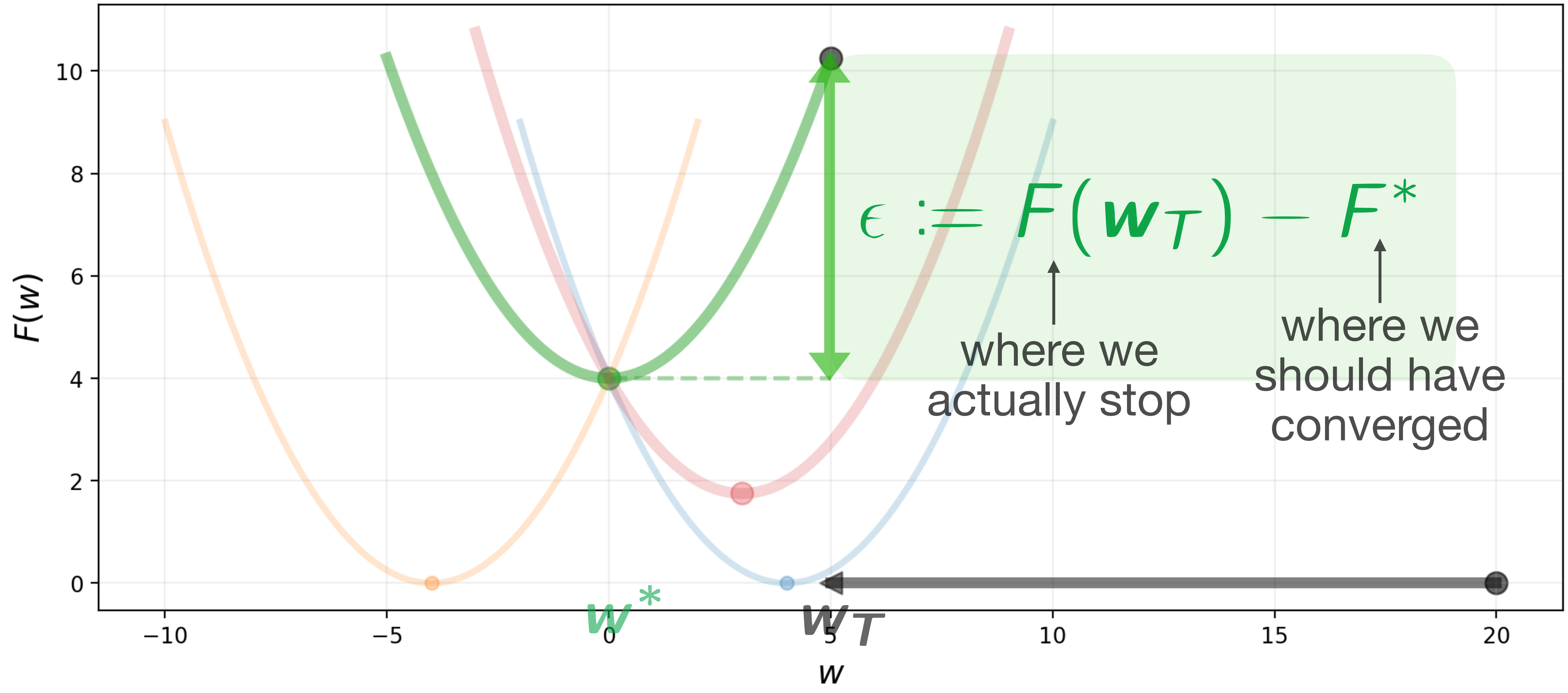


2) Heterogeneous Client Availability



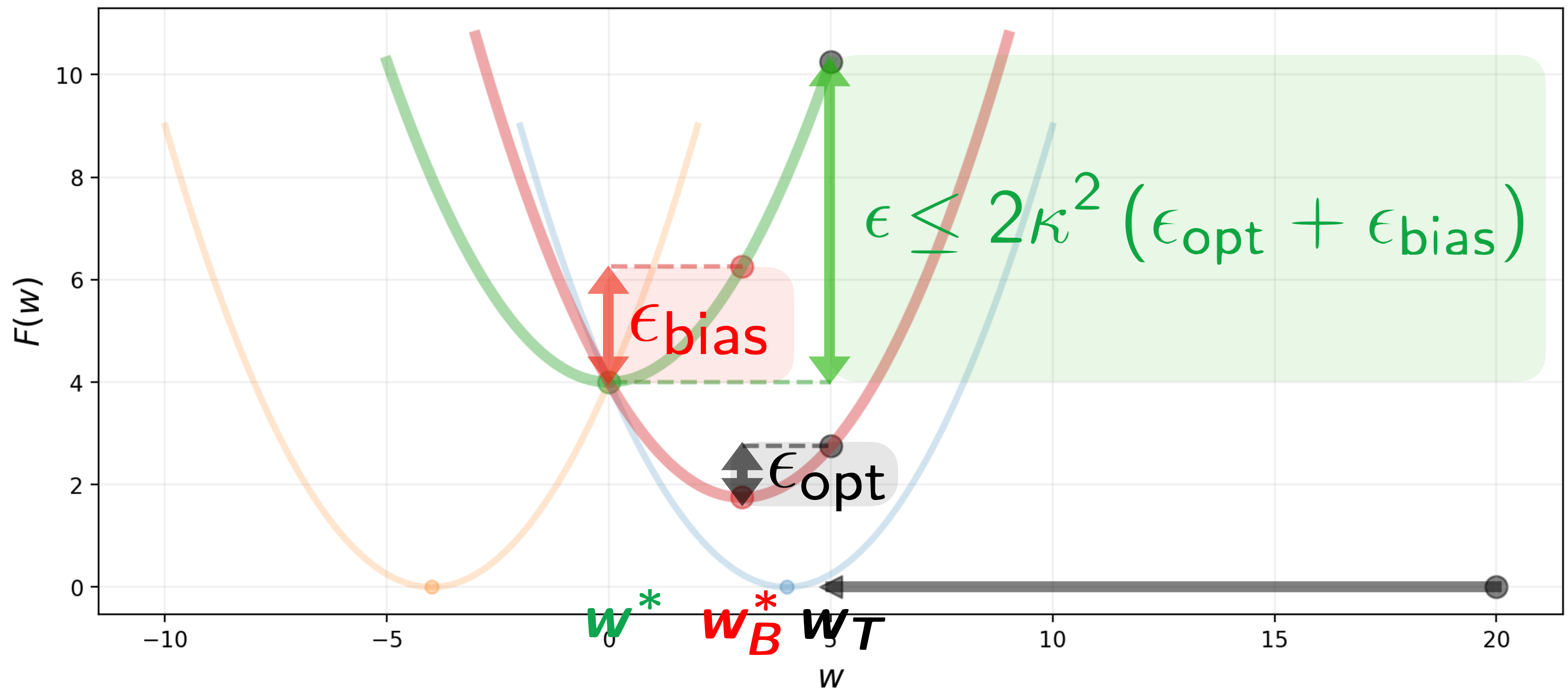
2) Heterogeneous Client Availability

Total error

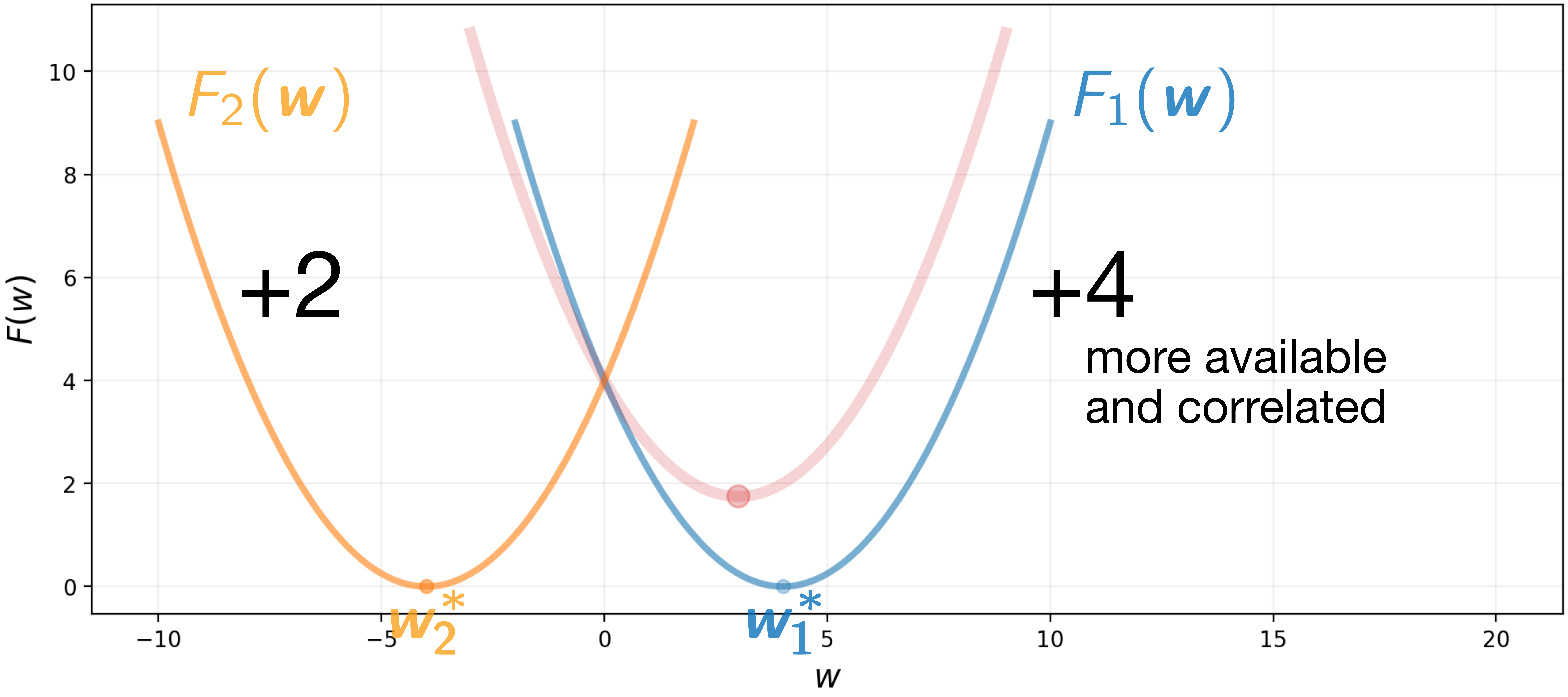


2) Heterogeneous Client Availability

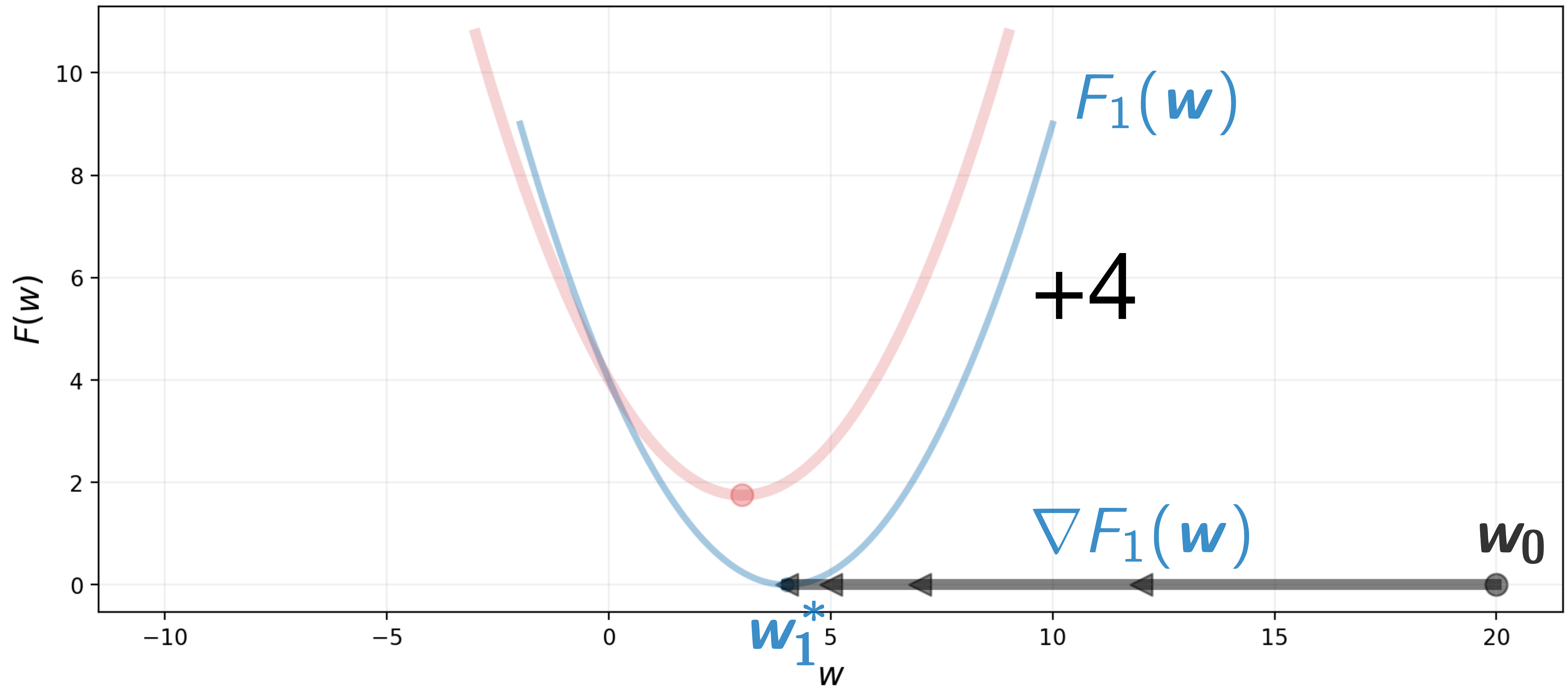
Total error



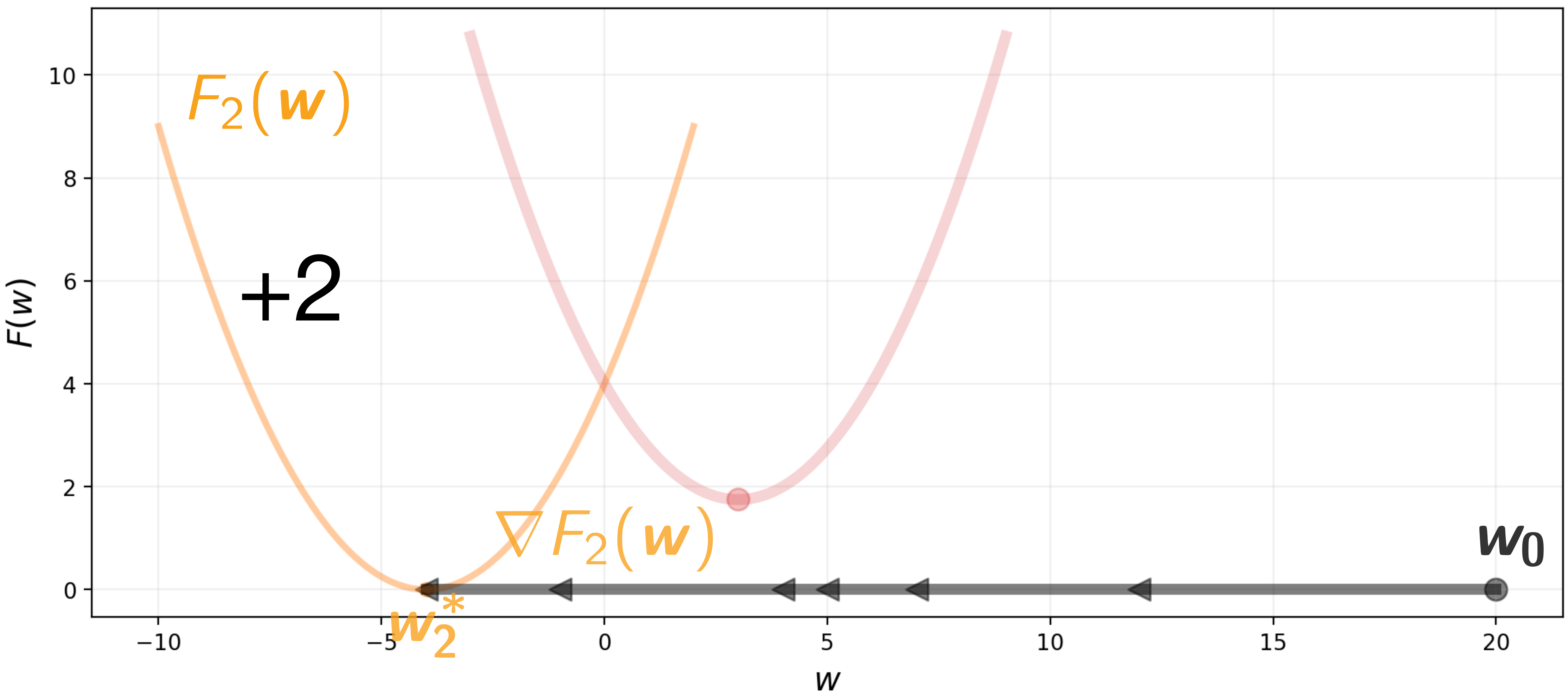
3) Heterogeneous and Correlated Client Availability



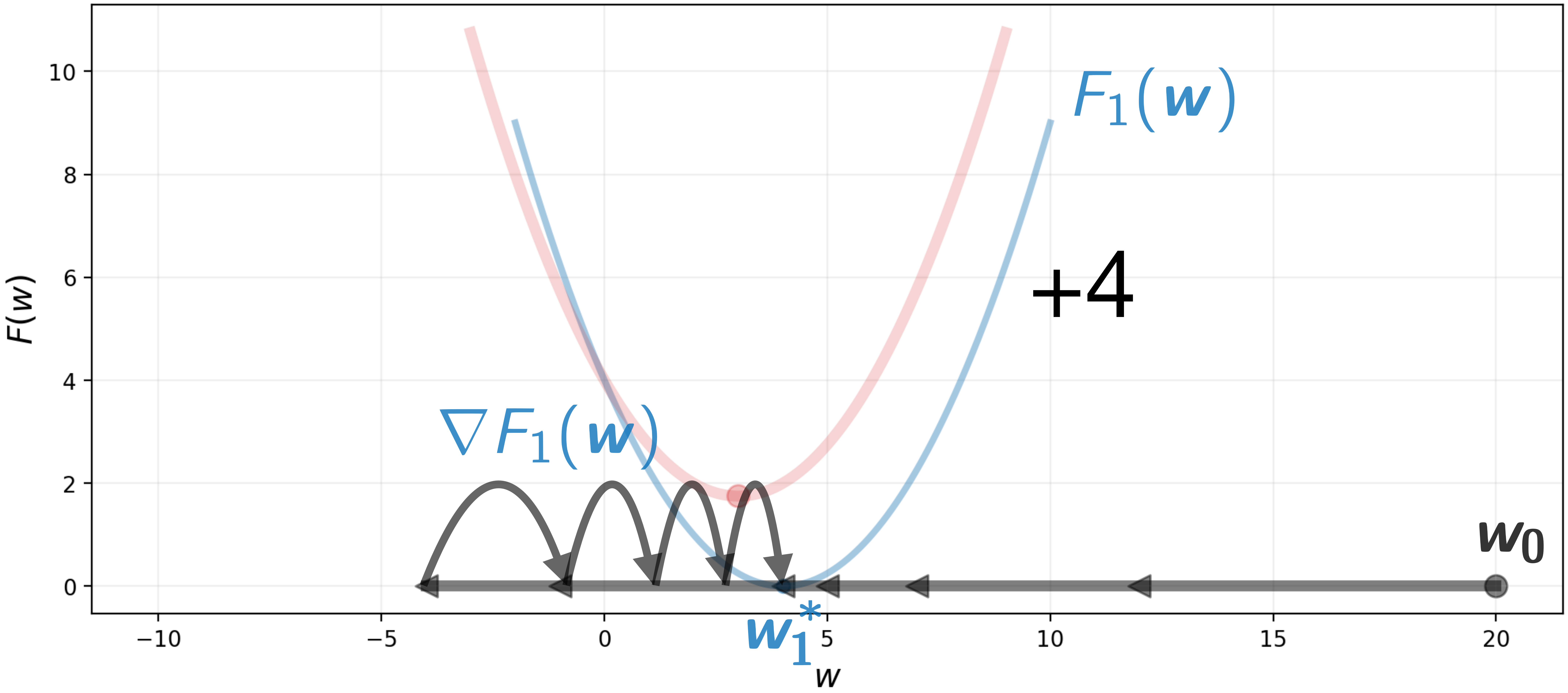
3) Heterogeneous and Correlated Client Availability



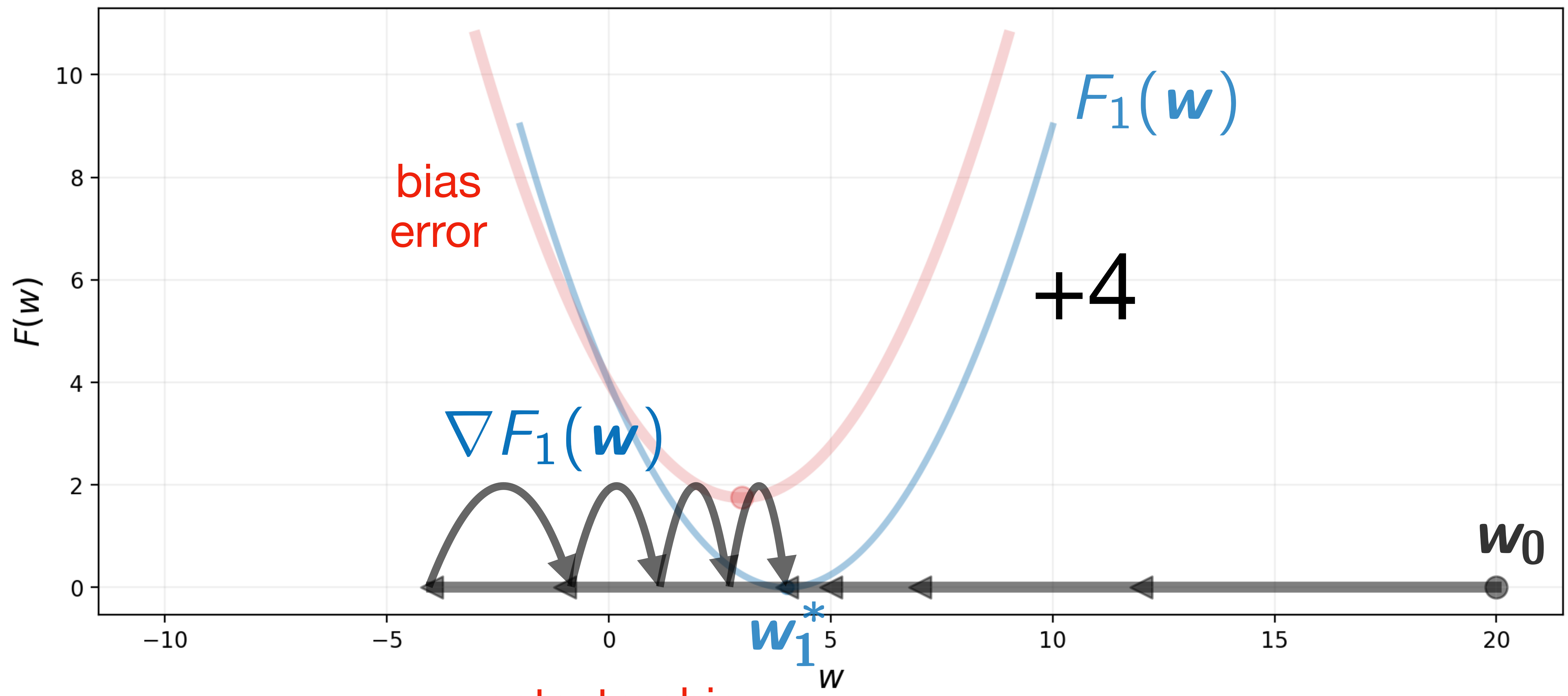
3) Heterogeneous and Correlated Client Availability



3) Heterogeneous and Correlated Client Availability



3) Heterogeneous and Correlated Client Availability



catastrophic forgetting : forget previously learned models

Assumption to model the heterogeneous and correlated client availability

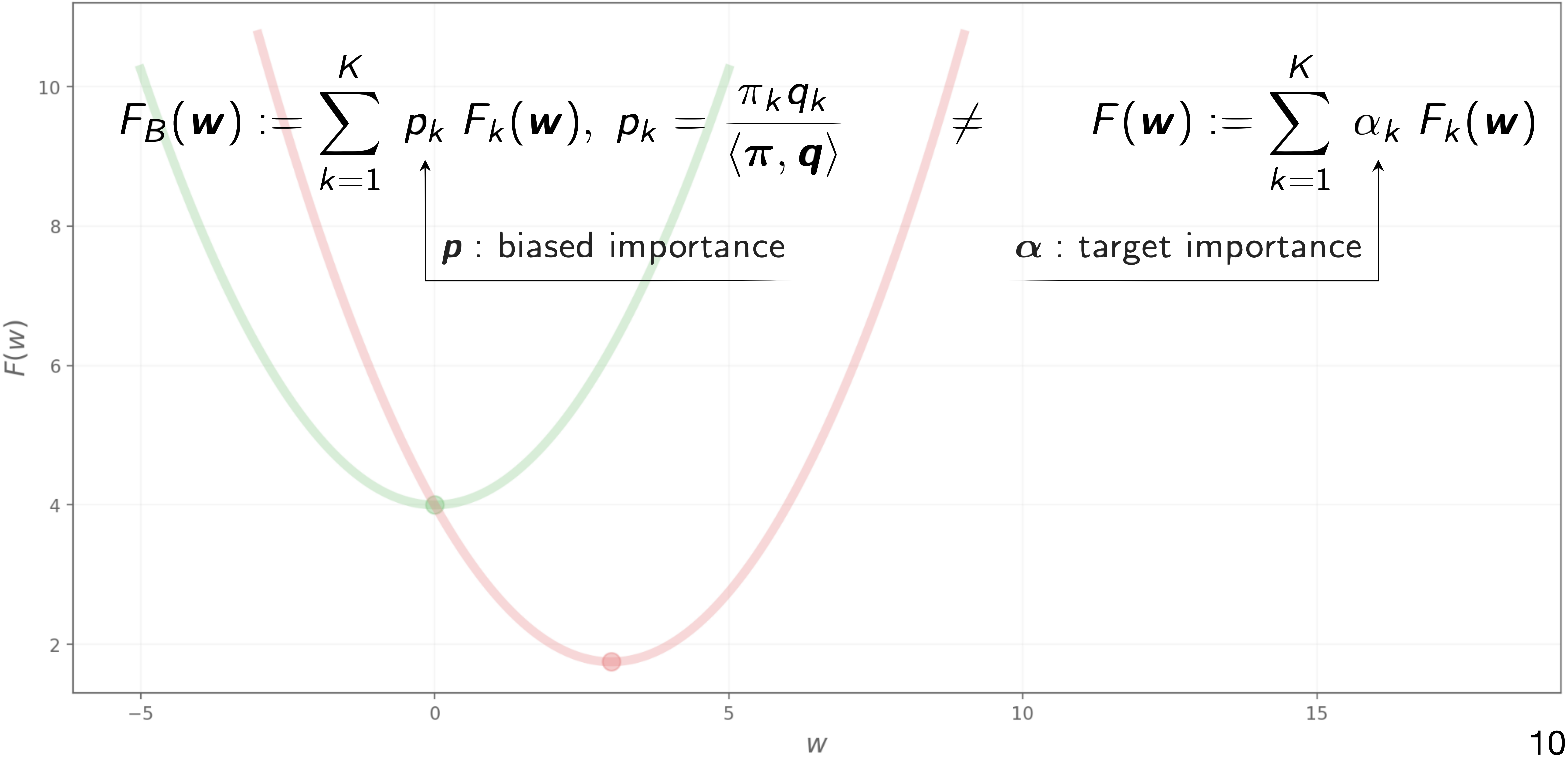
- A_t: set of active clients at time t
- $(A_t)_{t \geq 0}$ is a discrete-time Markov chain
    - transition matrix $\boldsymbol{P}$
    - stationary distribution $\boldsymbol{\pi}$ $\leftrightarrow$ avg. availability
    - largest 2nd eigenvalue $\lambda(\boldsymbol{P})$ $\leftrightarrow$ correlation

Assumption to model the heterogeneous and correlated client availability

- A_t: set of active clients at time t
- $(A_t)_{t \geq 0}$ is a discrete-time Markov chain
  - transition matrix $\boldsymbol{P}$
  - stationary distribution $\boldsymbol{\pi}$ $\leftrightarrow$ avg. availability
  - largest 2nd eigenvalue $\lambda(\boldsymbol{P})$ $\leftrightarrow$ correlation

If clients' availabilities are independent: $\lambda(\boldsymbol{P}) = \max_{k \in [K]} \lambda(\boldsymbol{P_k})$

Bias error

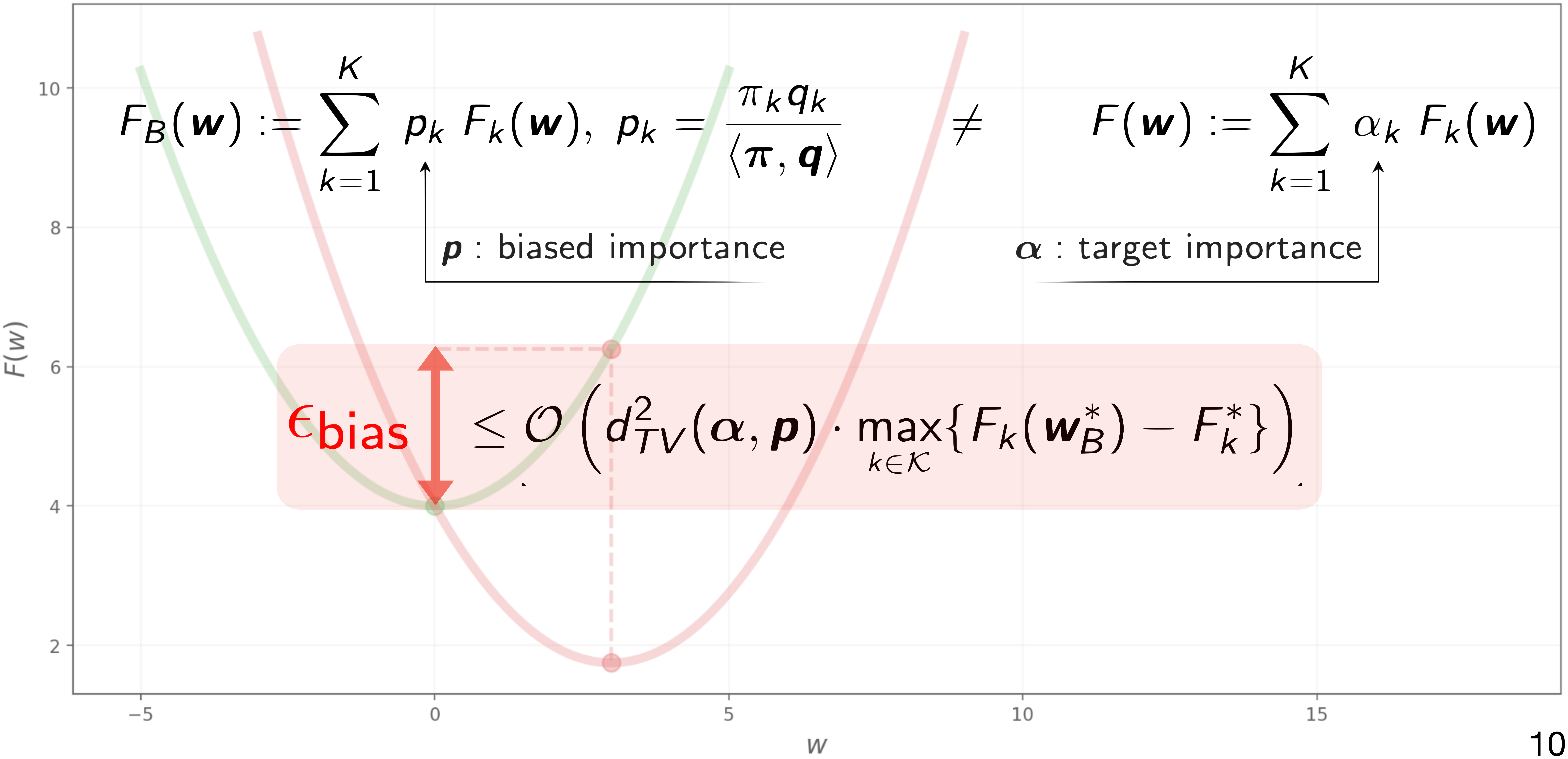


$$F_B(\boldsymbol{w}) := \sum_{k=1}^{K} p_k \, F_k(\boldsymbol{w}), \; p_k = \frac{\pi_k q_k}{\langle \boldsymbol{\pi}, \boldsymbol{q} \rangle} \neq F(\boldsymbol{w}) := \sum_{k=1}^{K} \alpha_k \, F_k(\boldsymbol{w})$$

$\boldsymbol{p}$: biased importance

$\boldsymbol{\alpha}$: target importance

Bias error



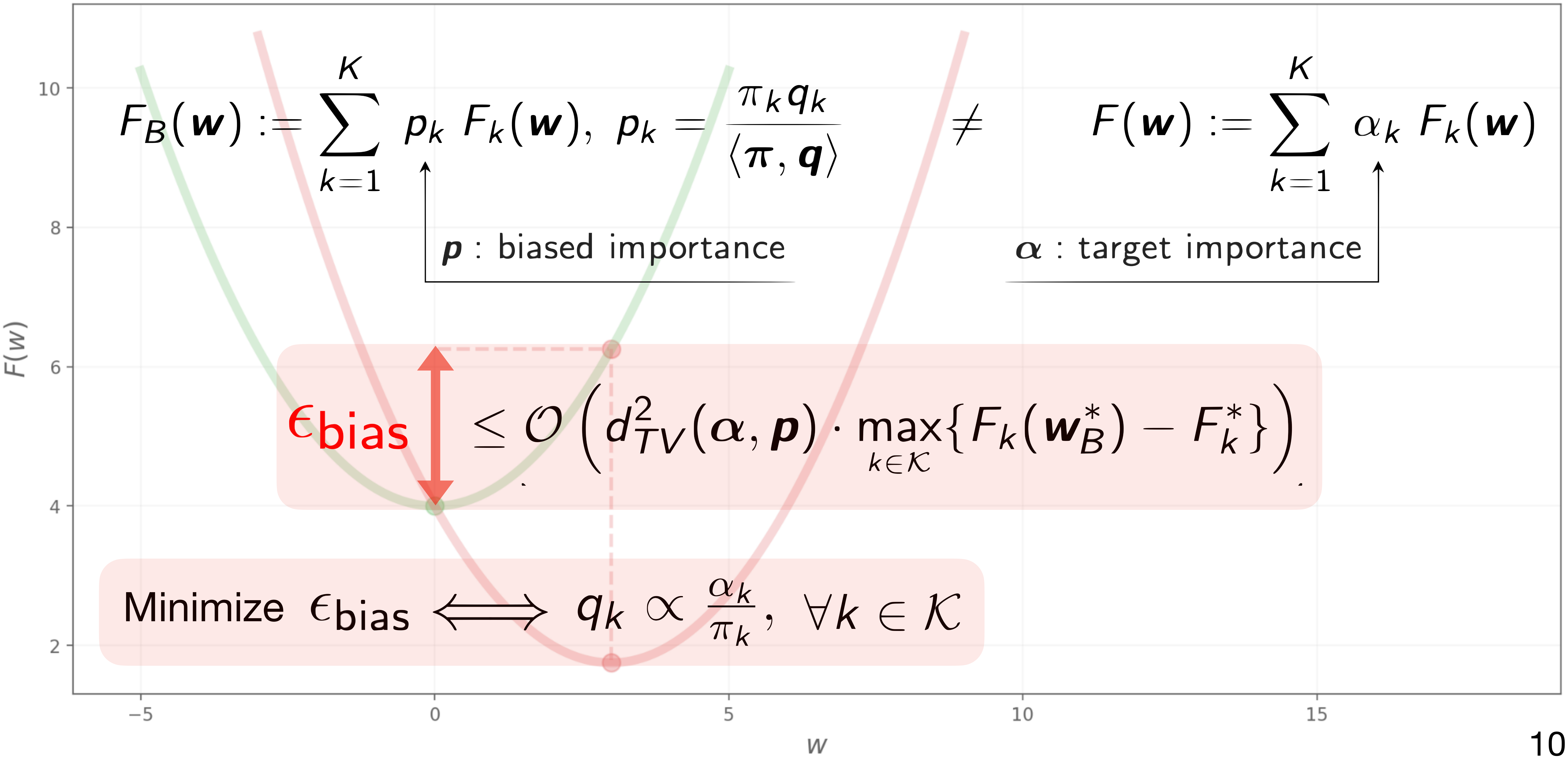
$$F_B(\mathbf{w}) := \sum_{k=1}^{K} p_k \, F_k(\mathbf{w}), \;\; p_k = \frac{\pi_k q_k}{\langle \boldsymbol{\pi}, \mathbf{q} \rangle} \neq F(\mathbf{w}) := \sum_{k=1}^{K} \alpha_k \, F_k(\mathbf{w})$$

$\mathbf{p}$: biased importance

$\boldsymbol{\alpha}$: target importance

$$\epsilon_{\text{bias}} \leq \mathcal{O}\left(d_{TV}^2(\boldsymbol{\alpha}, \mathbf{p}) \cdot \max_{k \in \mathcal{K}} \{ F_k(\mathbf{w}_B^*) - F_k^* \} \right)$$

Bias error



$$F_B(\boldsymbol{w}) := \sum_{k=1}^{K} p_k \, F_k(\boldsymbol{w}), \; p_k = \frac{\pi_k q_k}{\langle \boldsymbol{\pi}, \boldsymbol{q} \rangle} \neq F(\boldsymbol{w}) := \sum_{k=1}^{K} \alpha_k \, F_k(\boldsymbol{w})$$

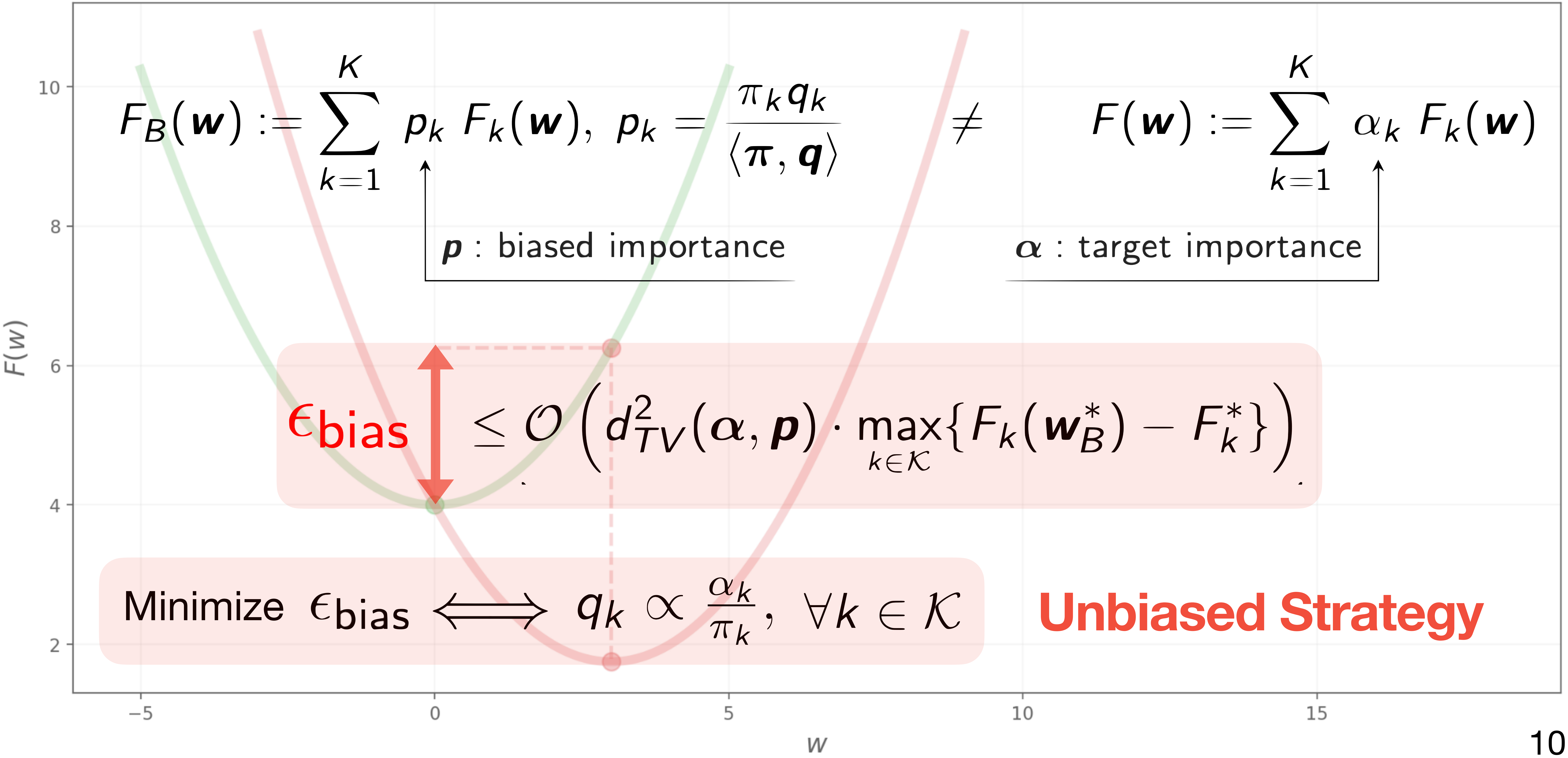
$\boldsymbol{p}$: biased importance

$\boldsymbol{\alpha}$: target importance

$$\epsilon_{\text{bias}} \leq \mathcal{O}\left(d_{TV}^2(\boldsymbol{\alpha}, \boldsymbol{p}) \cdot \max_{k \in \mathcal{K}} \{ F_k(\boldsymbol{w}_B^*) - F_k^* \} \right)$$

$$\text{Minimize } \epsilon_{\text{bias}} \iff q_k \propto \frac{\alpha_k}{\pi_k}, \; \forall k \in \mathcal{K}$$

Bias error



$$F_B(\boldsymbol{w}) := \sum_{k=1}^{K} p_k \, F_k(\boldsymbol{w}), \;\; p_k = \frac{\pi_k q_k}{\langle \boldsymbol{\pi}, \boldsymbol{q} \rangle} \qquad \neq \qquad F(\boldsymbol{w}) := \sum_{k=1}^{K} \alpha_k \, F_k(\boldsymbol{w})$$

$\boldsymbol{p}$: biased importance

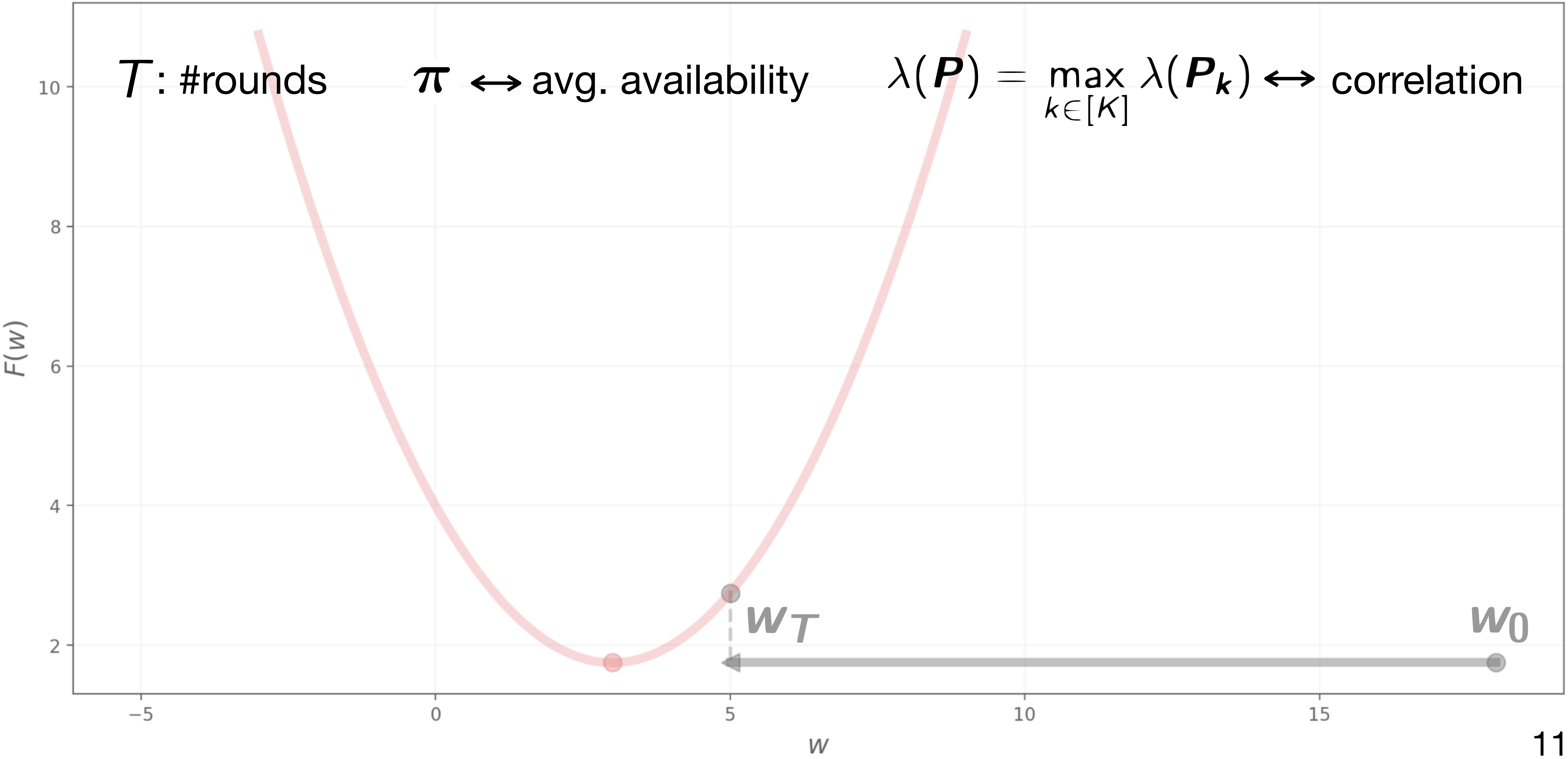
$\boldsymbol{\alpha}$: target importance

$$\epsilon_{\text{bias}} \leq \mathcal{O}\left(d_{TV}^2(\boldsymbol{\alpha}, \boldsymbol{p}) \cdot \max_{k \in \mathcal{K}} \{F_k(\boldsymbol{w}_B^*) - F_k^*\}\right)$$

$$\text{Minimize } \epsilon_{\text{bias}} \iff q_k \propto \frac{\alpha_k}{\pi_k}, \;\; \forall k \in \mathcal{K}$$

Unbiased Strategy

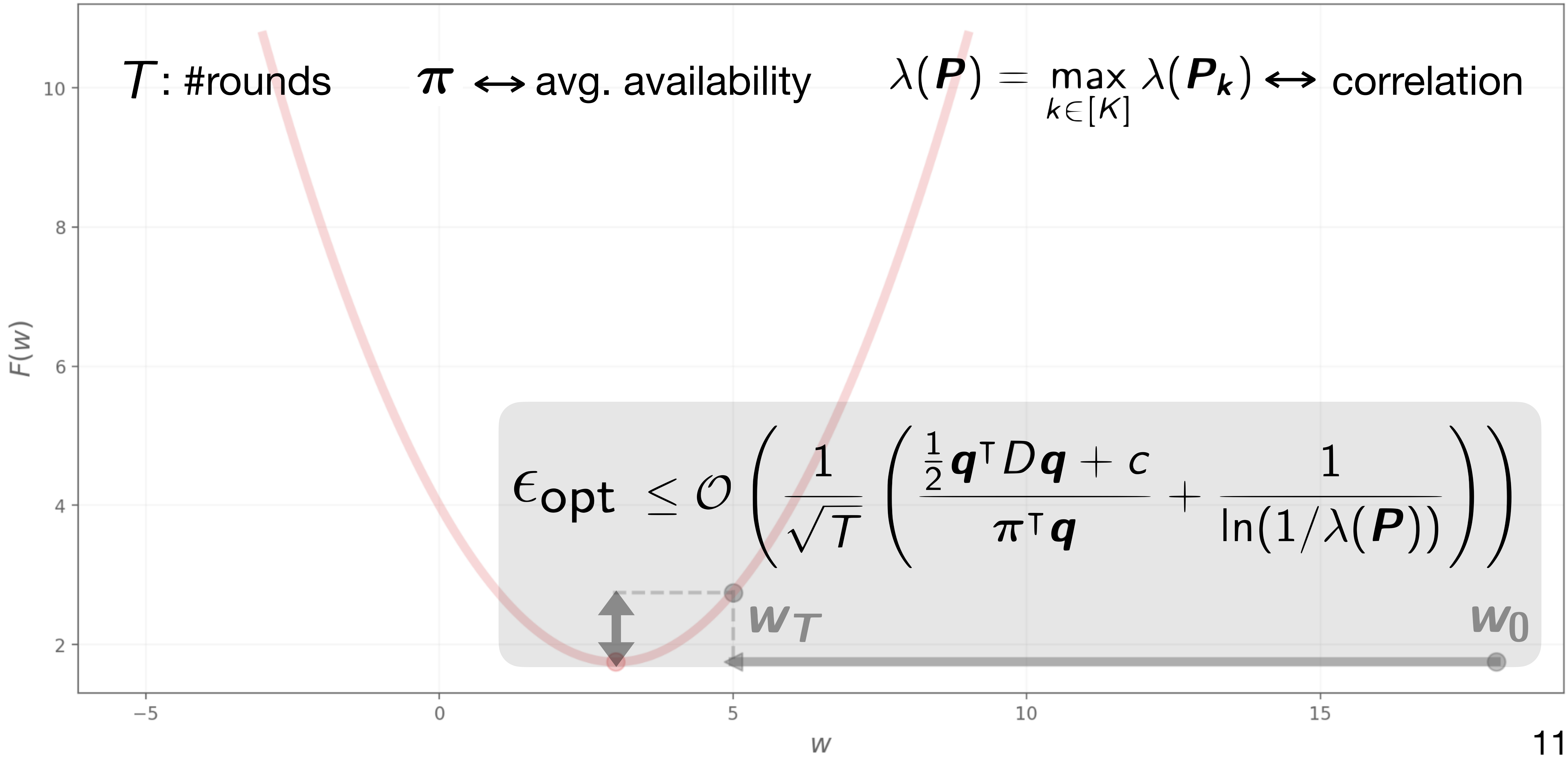
Optimization error



Optimization error

T: #rounds $\quad \boldsymbol{\pi} \leftrightarrow$ avg. availability $\quad \lambda(\boldsymbol{P}) = \max_{k\in[K]} \lambda(\boldsymbol{P_k}) \leftrightarrow$ correlation

$$\epsilon_{\text{opt}} \le \mathcal{O}\left(\frac{1}{\sqrt{T}}\left(\frac{\frac{1}{2}\boldsymbol{q}^\intercal D\boldsymbol{q} + c}{\boldsymbol{\pi}^\intercal \boldsymbol{q}} + \frac{1}{\ln(1/\lambda(\boldsymbol{P}))}\right)\right)$$



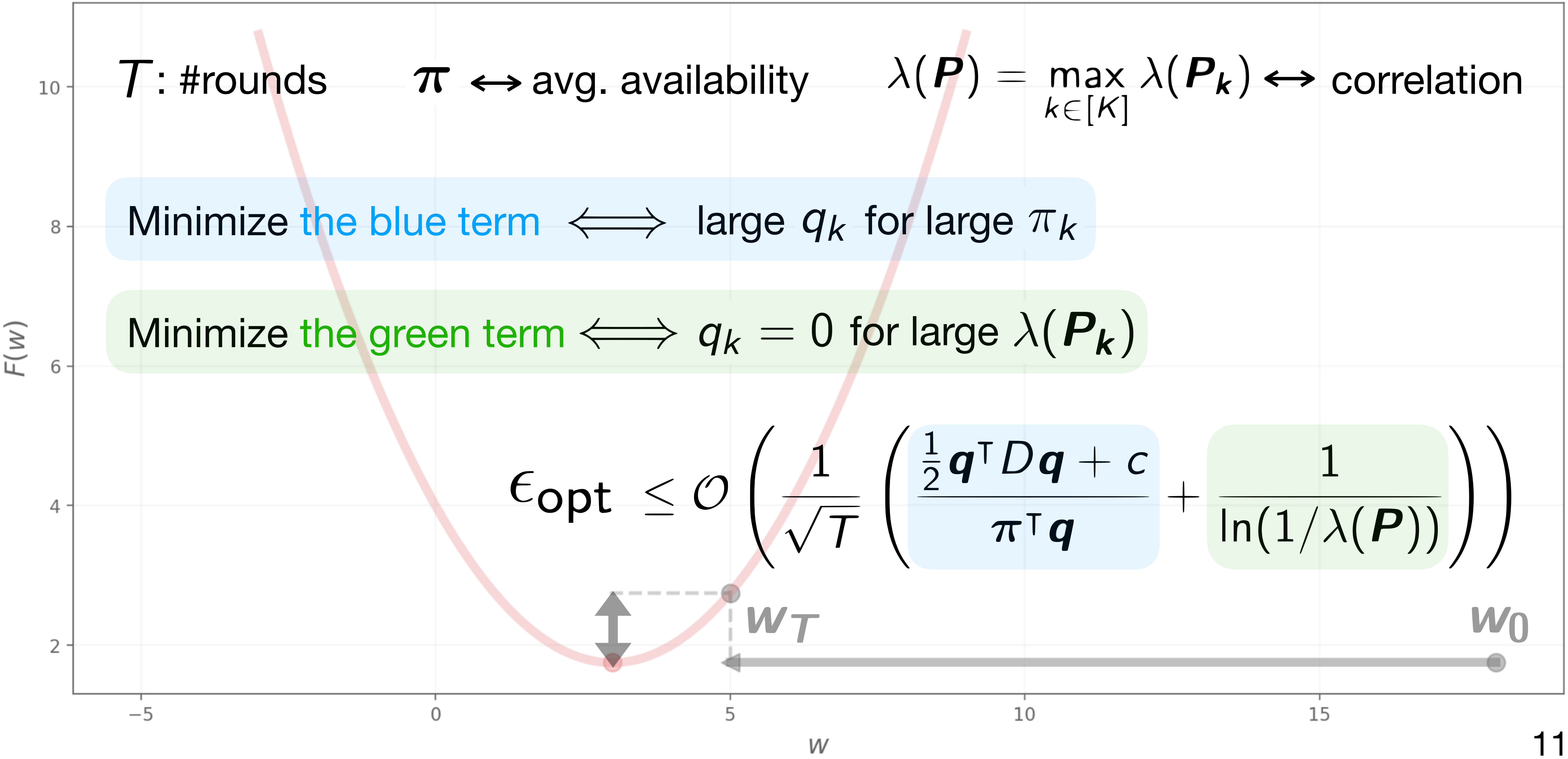
Optimization error

T: #rounds $\boldsymbol{\pi} \longleftrightarrow$ avg. availability $\lambda(\boldsymbol{P}) = \max_{k \in [K]} \lambda(\boldsymbol{P_k}) \longleftrightarrow$ correlation

Minimize the blue term $\Longleftrightarrow$ large q_k for large π_k

Minimize the green term $\Longleftrightarrow$ $q_k = 0$ for large $\lambda(\boldsymbol{P_k})$

$$\epsilon_{\text{opt}} \leq \mathcal{O}\left(\frac{1}{\sqrt{T}}\left(\frac{\frac{1}{2}\boldsymbol{q}^\top D \boldsymbol{q} + c}{\boldsymbol{\pi}^\top \boldsymbol{q}} + \frac{1}{\ln(1/\lambda(\boldsymbol{P}))}\right)\right)$$



Optimization error

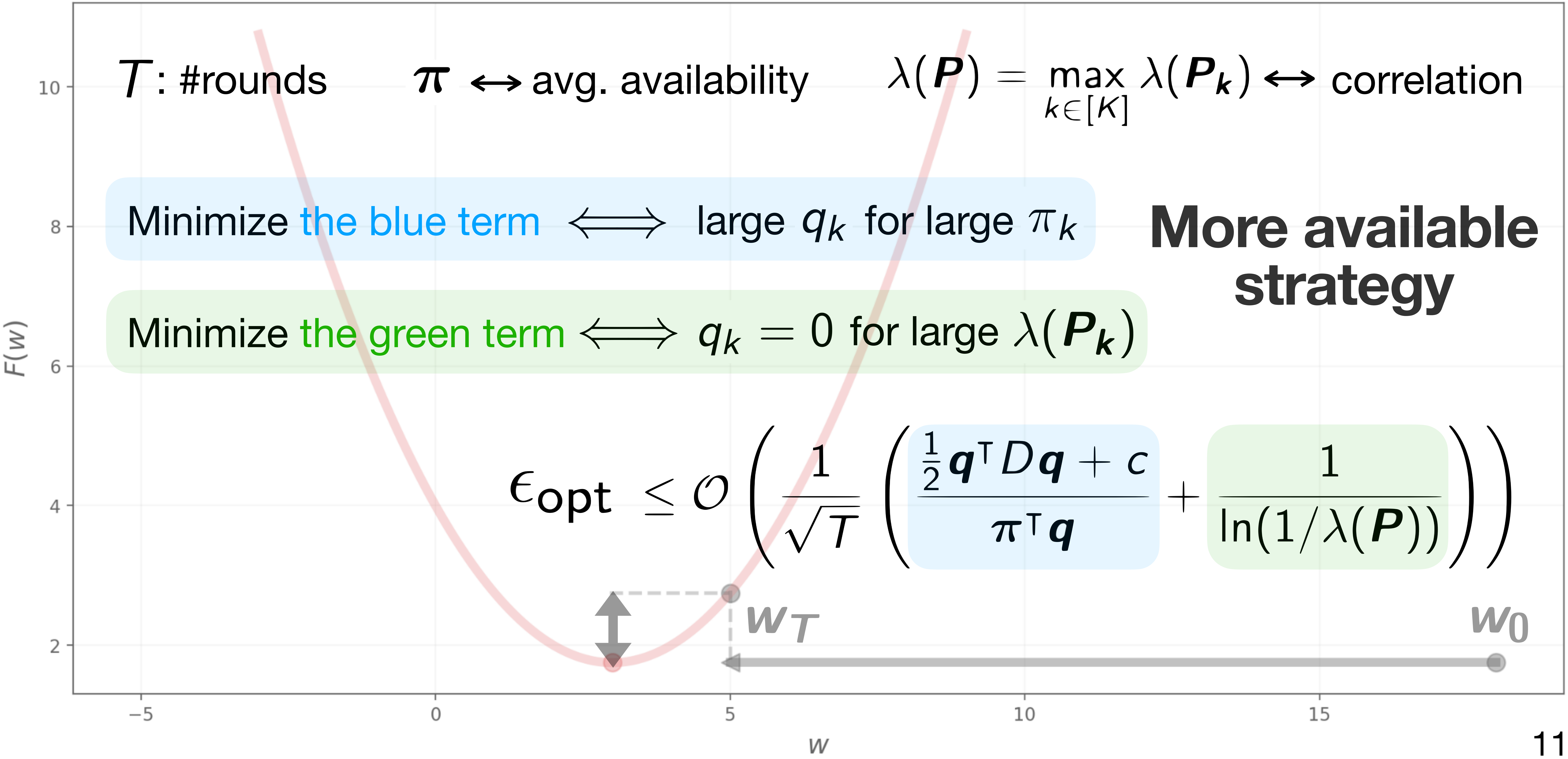
T: #rounds $\quad$ $\boldsymbol{\pi} \leftrightarrow$ avg. availability $\quad$ $\lambda(\boldsymbol{P}) = \max_{k \in [K]} \lambda(\boldsymbol{P_k}) \leftrightarrow$ correlation

Minimize the blue term $\Longleftrightarrow$ large q_k for large π_k

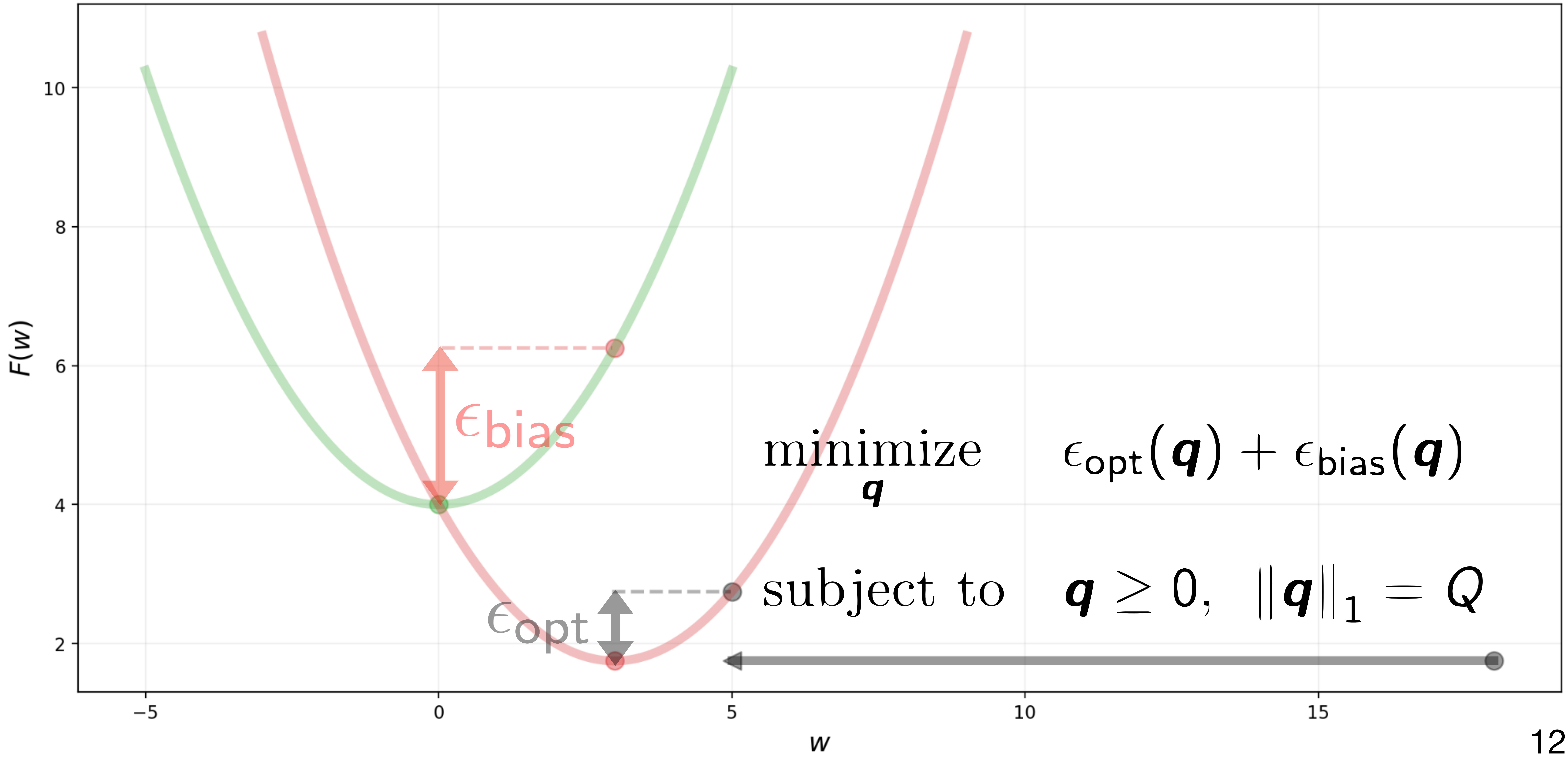
Minimize the green term $\Longleftrightarrow$ $q_k = 0$ for large $\lambda(\boldsymbol{P_k})$

More available strategy

$$\epsilon_{\text{opt}} \leq \mathcal{O}\left(\frac{1}{\sqrt{T}}\left(\frac{\frac{1}{2}\boldsymbol{q}^\top D \boldsymbol{q} + c}{\boldsymbol{\pi}^\top \boldsymbol{q}} + \frac{1}{\ln(1/\lambda(\boldsymbol{P}))}\right)\right)$$



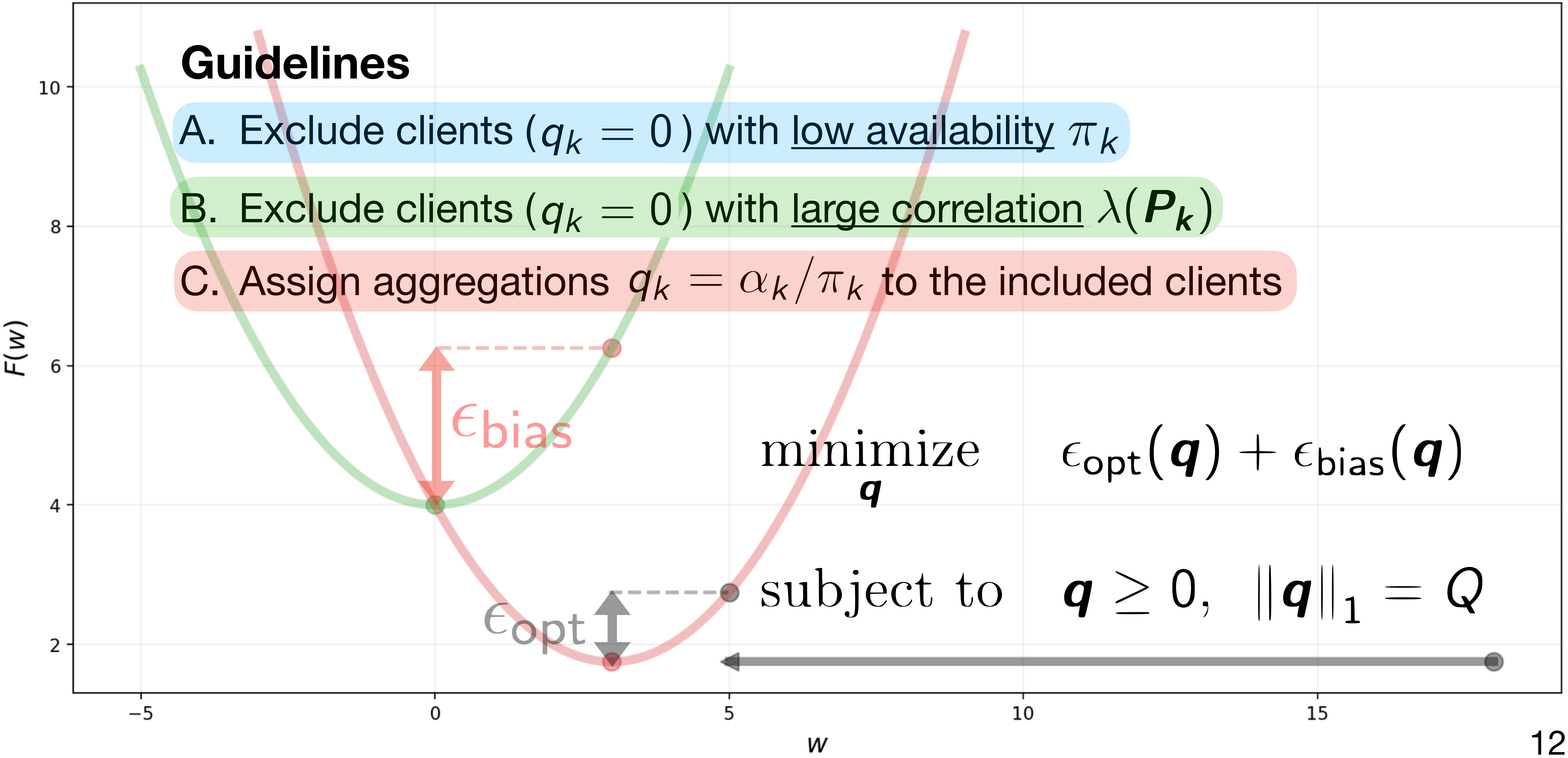
Total error : optimization error + bias error



$$\underset{\boldsymbol{q}}{\text{minimize}} \quad \epsilon_{\mathsf{opt}}(\boldsymbol{q}) + \epsilon_{\mathsf{bias}}(\boldsymbol{q})$$

$$\text{subject to} \quad \boldsymbol{q} \geq 0, \ \ \|\boldsymbol{q}\|_1 = Q$$

Total error : optimization error + bias error



Guidelines

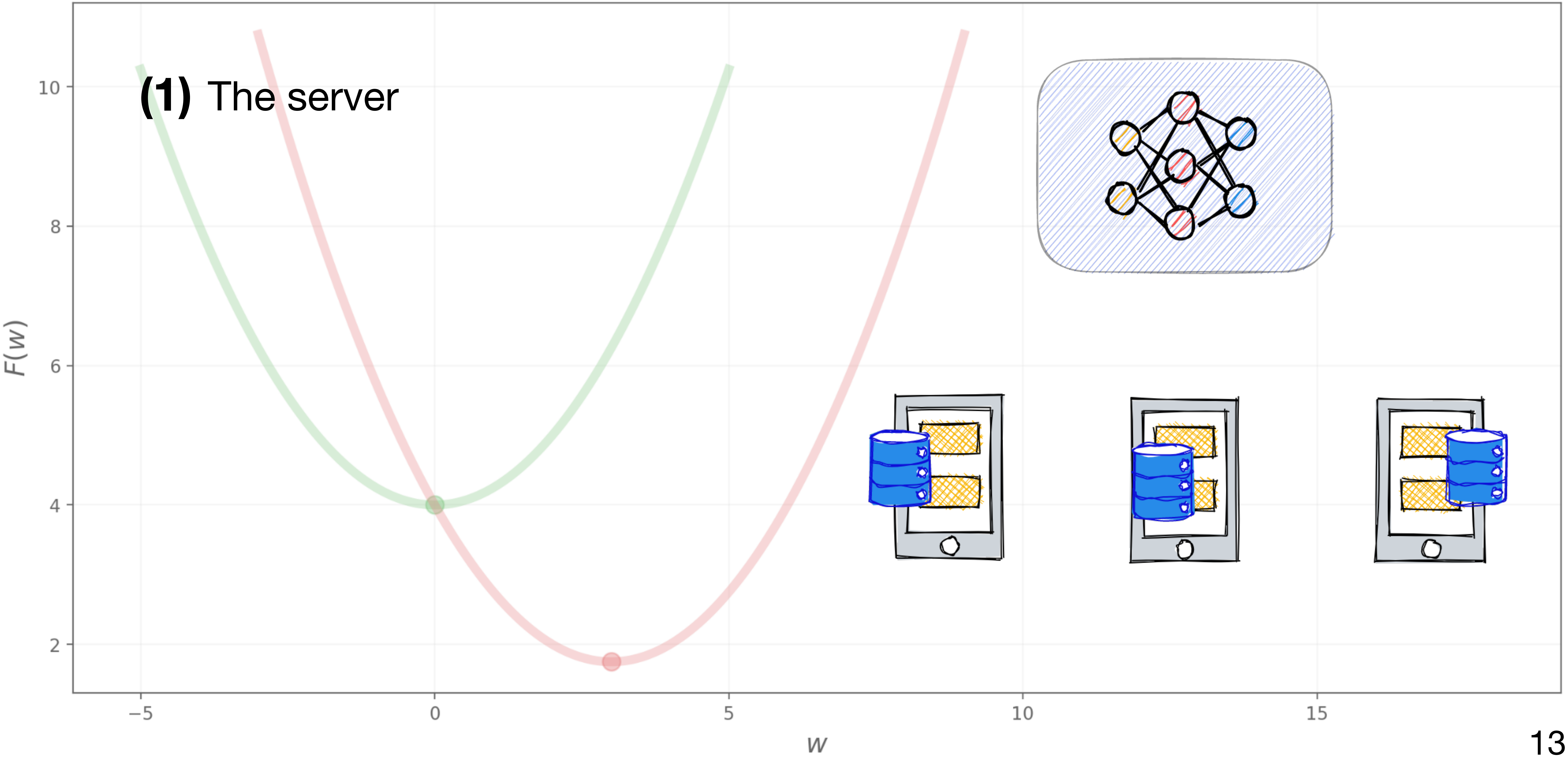
A. Exclude clients ($q_k = 0$) with low availability π_k

B. Exclude clients ($q_k = 0$) with large correlation $\lambda(\boldsymbol{P_k})$

C. Assign aggregations $q_k = \alpha_k / \pi_k$ to the included clients

$$\begin{aligned} \underset{\boldsymbol{q}}{\text{minimize}} \quad & \epsilon_{\text{opt}}(\boldsymbol{q}) + \epsilon_{\text{bias}}(\boldsymbol{q}) \\ \text{subject to} \quad & \boldsymbol{q} \geq 0, \ \|\boldsymbol{q}\|_1 = Q \end{aligned}$$

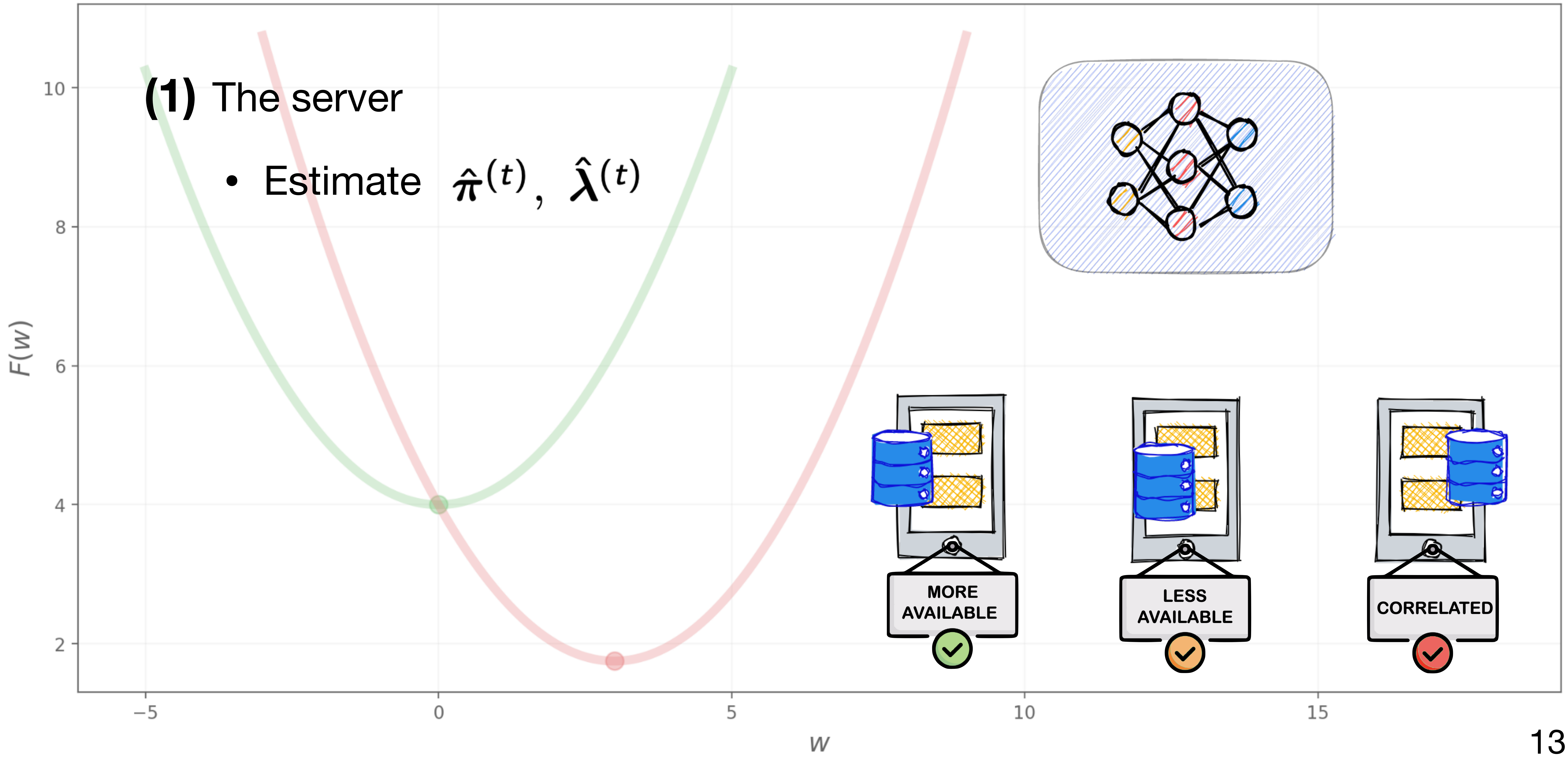
Correlation-Aware FL (CA-Fed)



Correlation-Aware FL (CA-Fed)

(1) The server

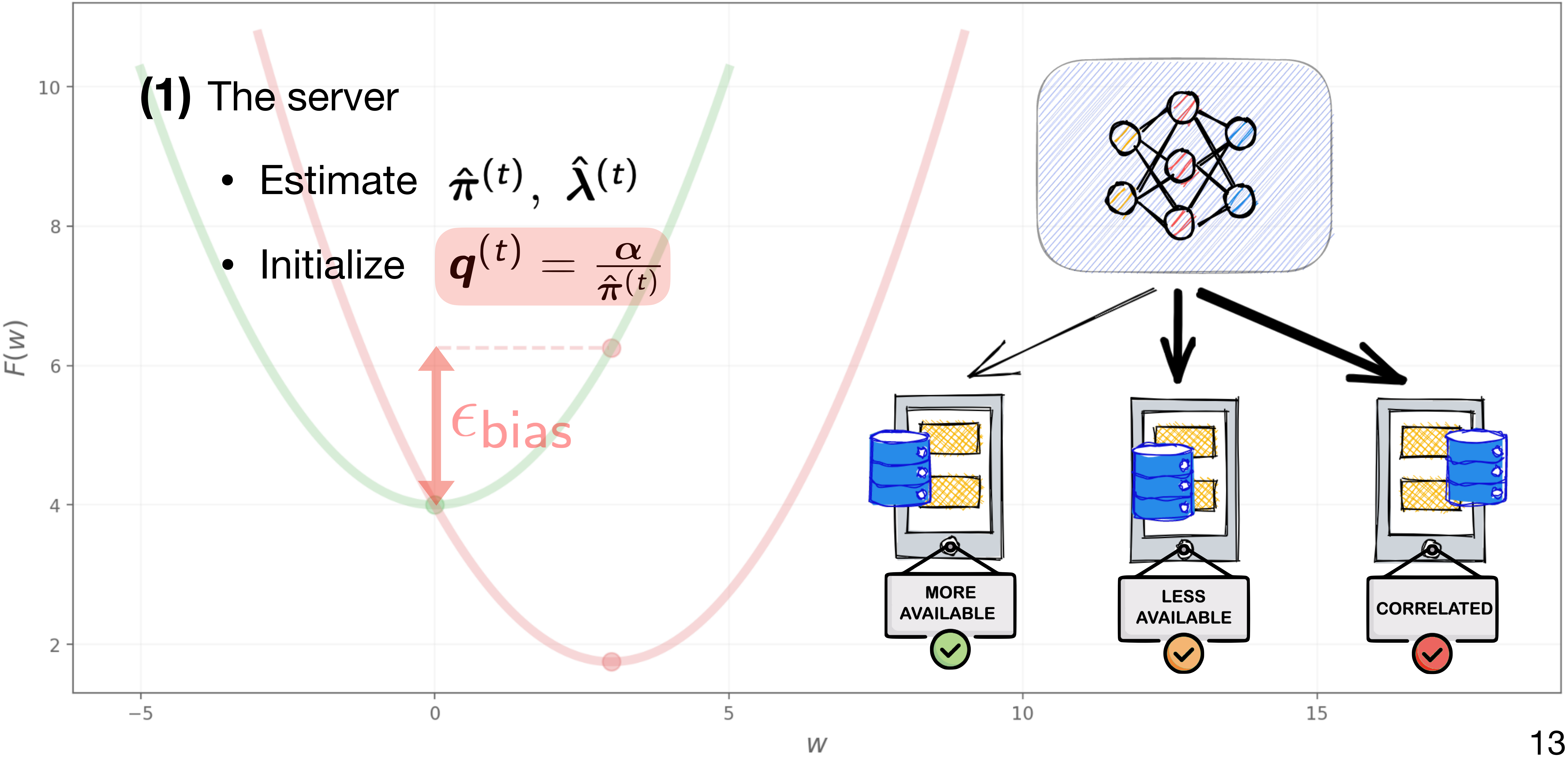
- Estimate $\hat{\pi}^{(t)}, \hat{\lambda}^{(t)}$



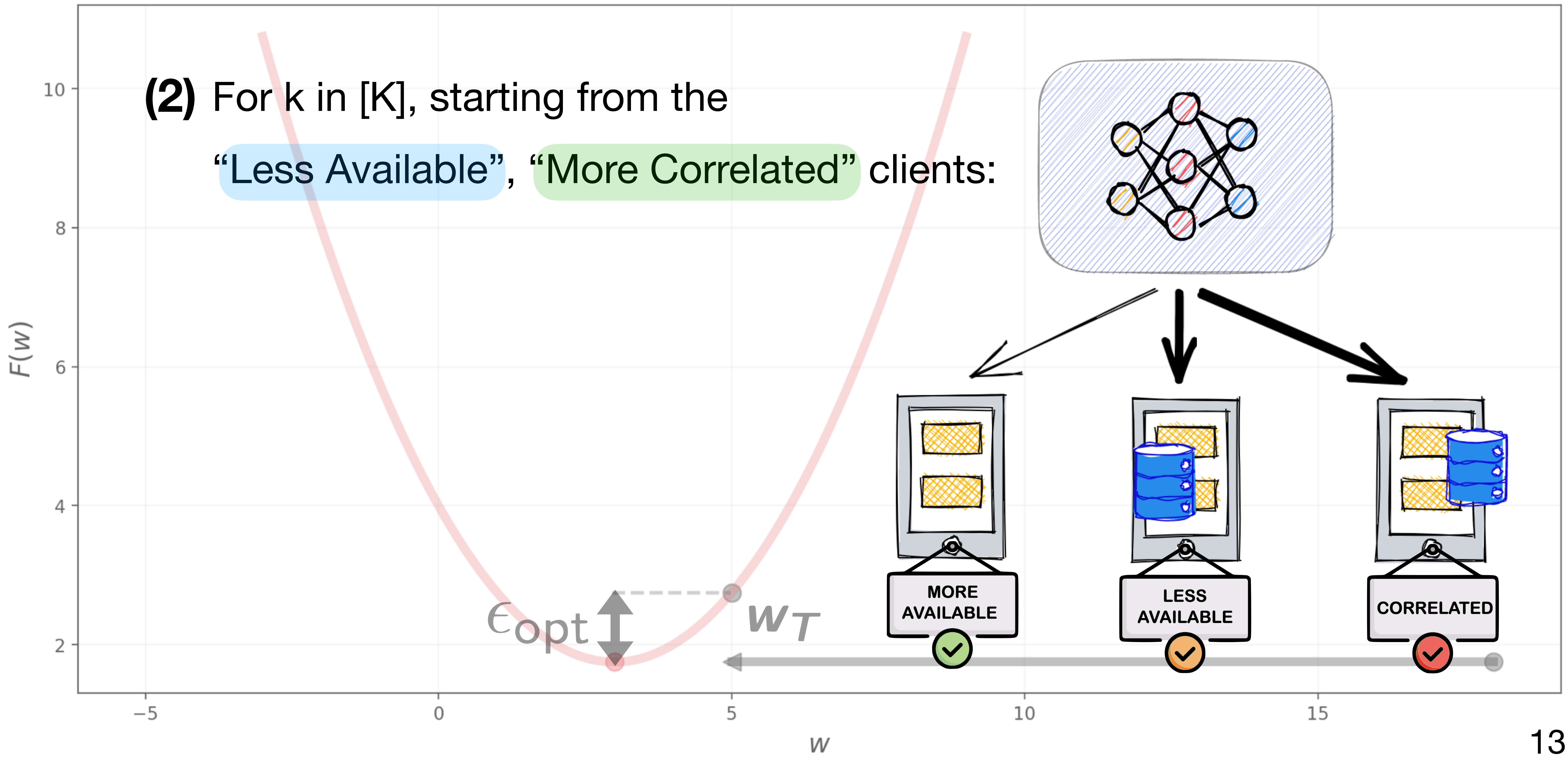
Correlation-Aware FL (CA-Fed)

(1) The server

- Estimate $\hat{\boldsymbol{\pi}}^{(t)}, \ \hat{\boldsymbol{\lambda}}^{(t)}$
- Initialize $\boldsymbol{q}^{(t)} = \frac{\boldsymbol{\alpha}}{\hat{\boldsymbol{\pi}}^{(t)}}$



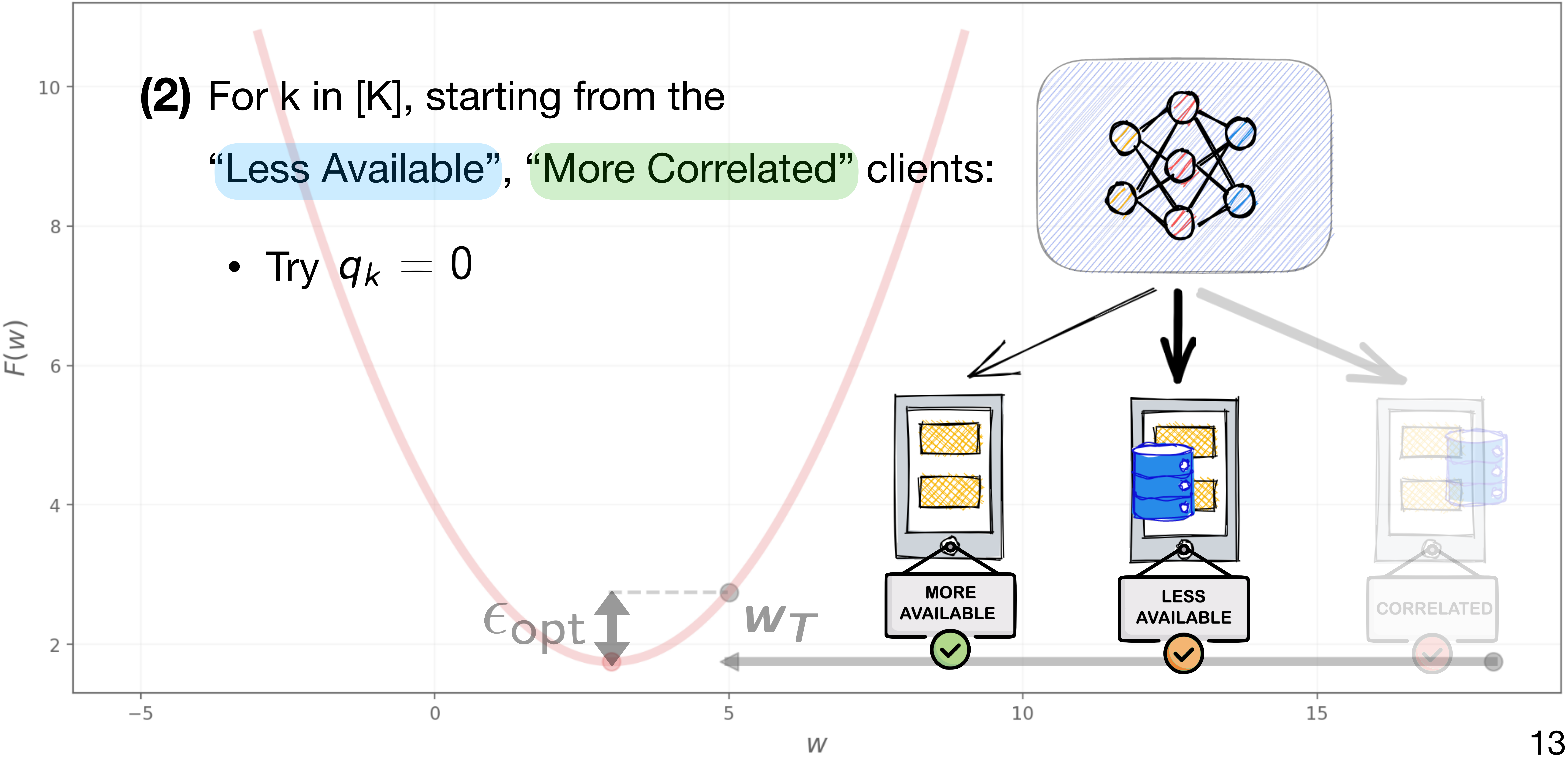
Correlation-Aware FL (CA-Fed)



Correlation-Aware FL (CA-Fed)

(2) For k in [K], starting from the "Less Available", "More Correlated" clients:

- Try $q_k = 0$

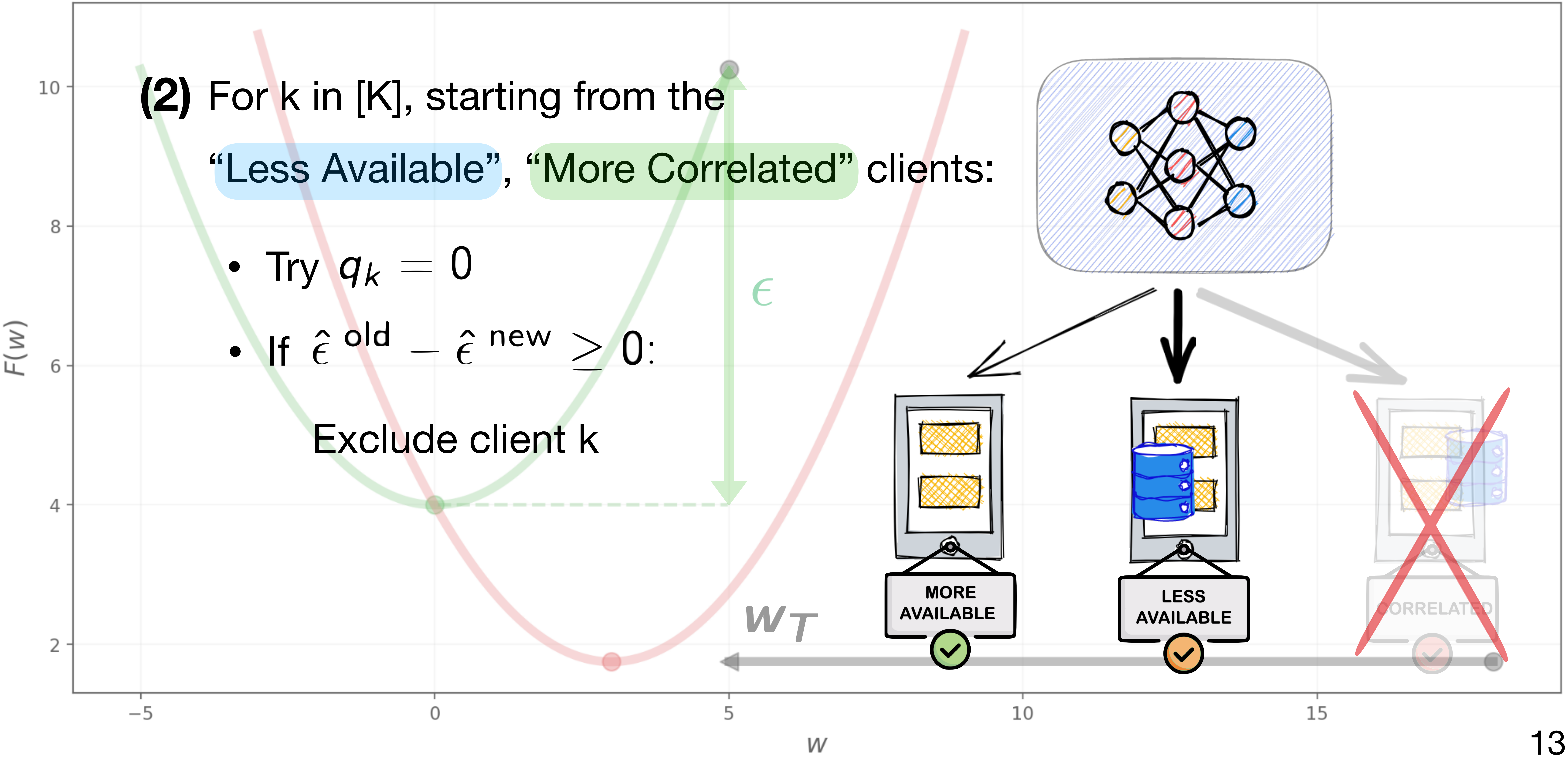


Correlation-Aware FL (CA-Fed)

(2) For k in [K], starting from the "Less Available", "More Correlated" clients:

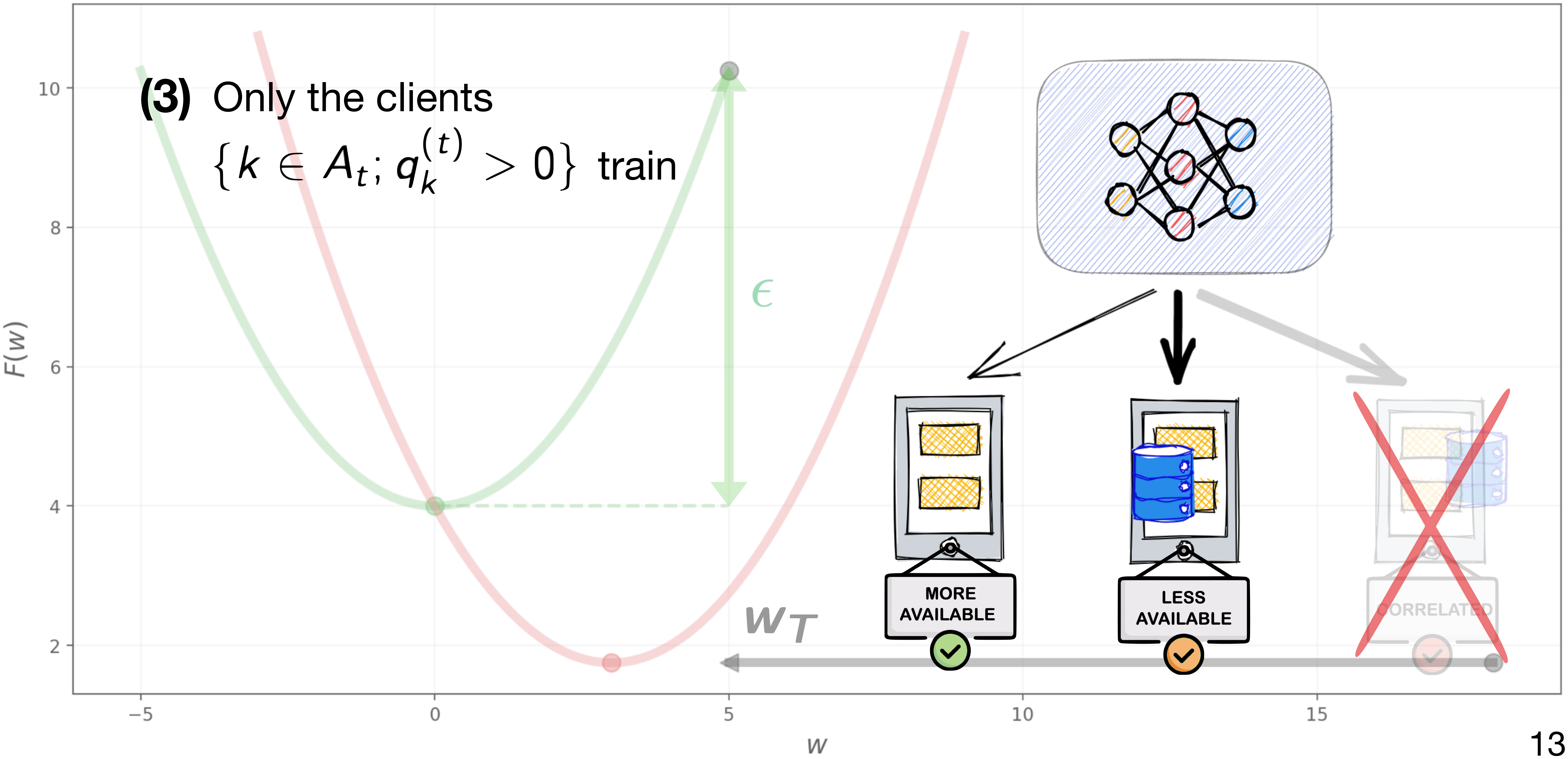
- Try $q_k = 0$
- If $\hat{\epsilon}^{\text{old}} - \hat{\epsilon}^{\text{new}} \geq 0$:

  Exclude client k



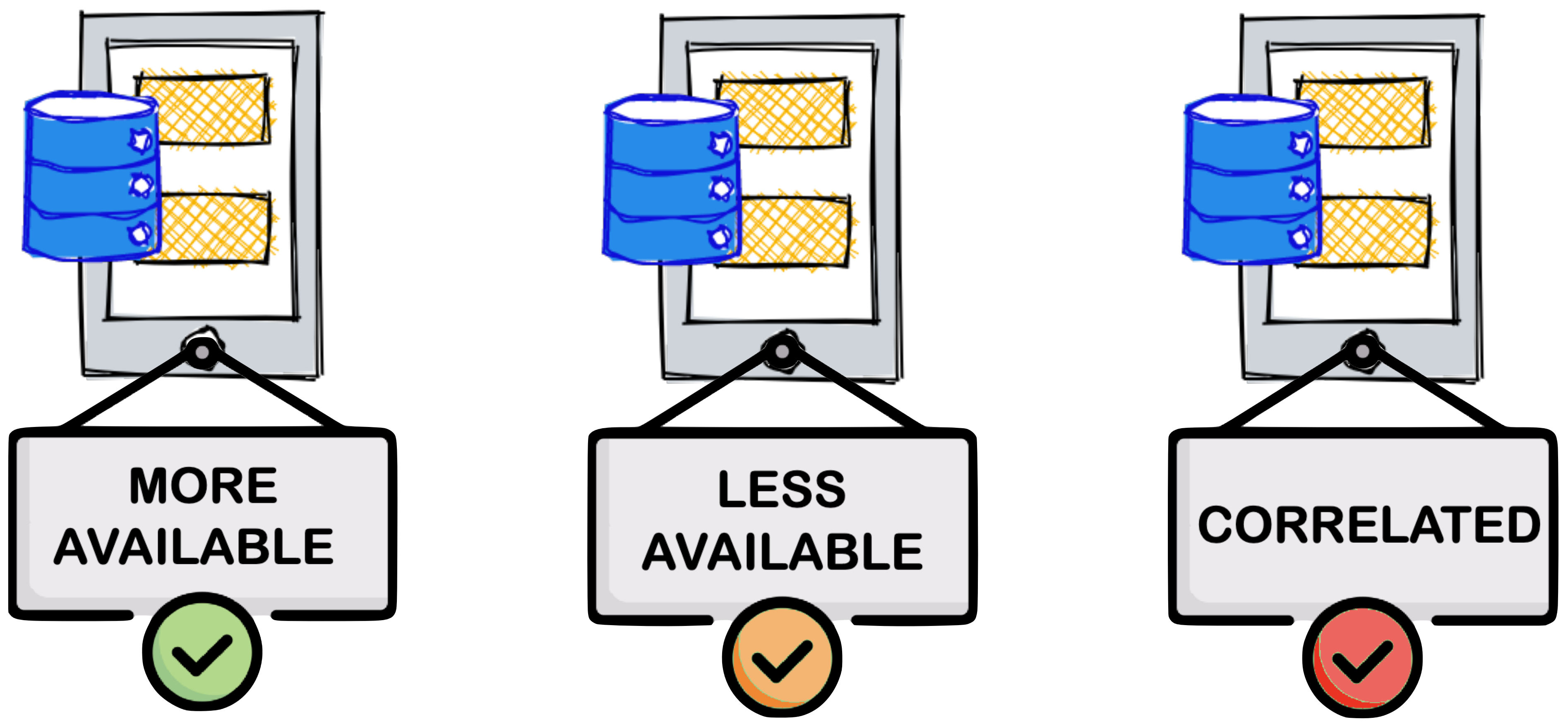
Correlation-Aware FL (CA-Fed)

(3) Only the clients $\{k \in A_t; q_k^{(t)} > 0\}$ train



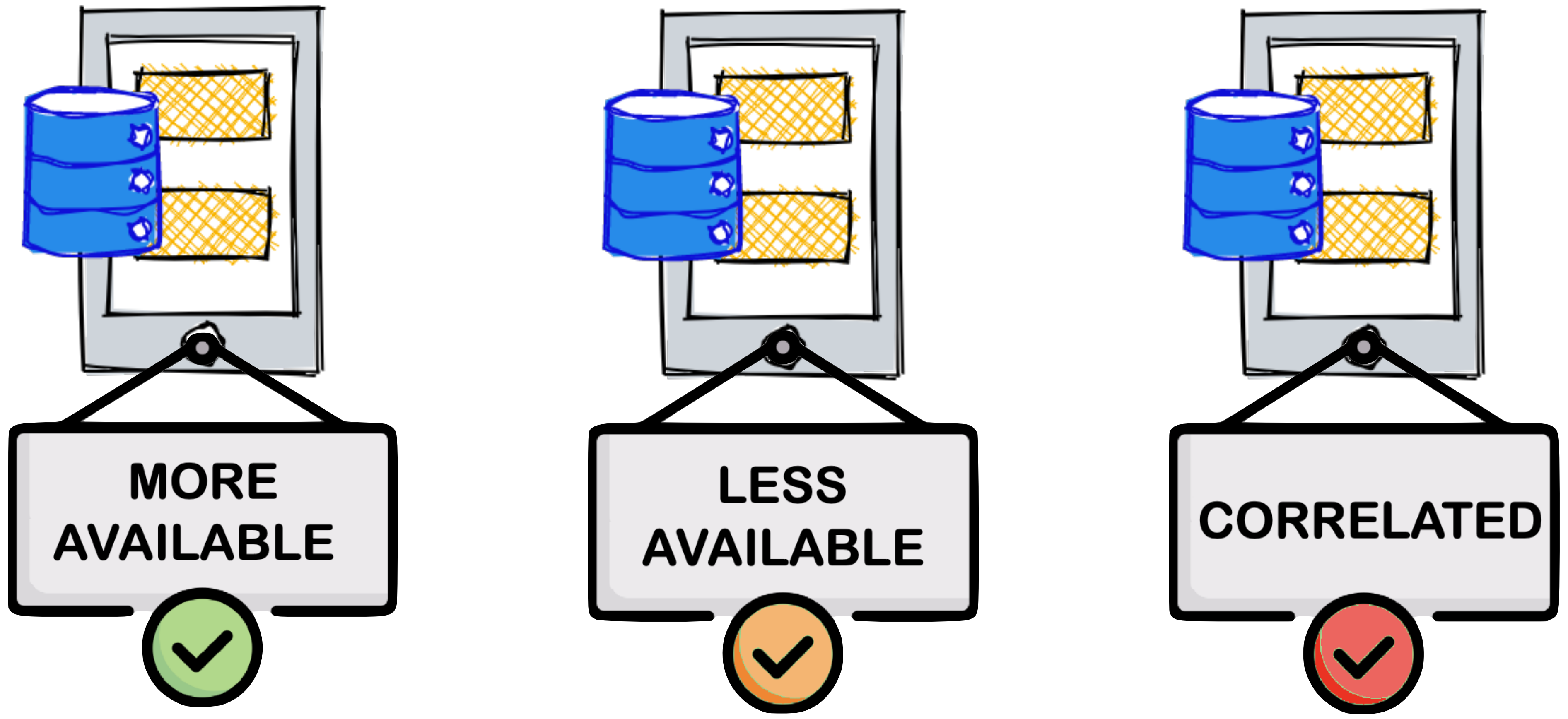
Experimental setting

- Population of K=100 clients, divided in:

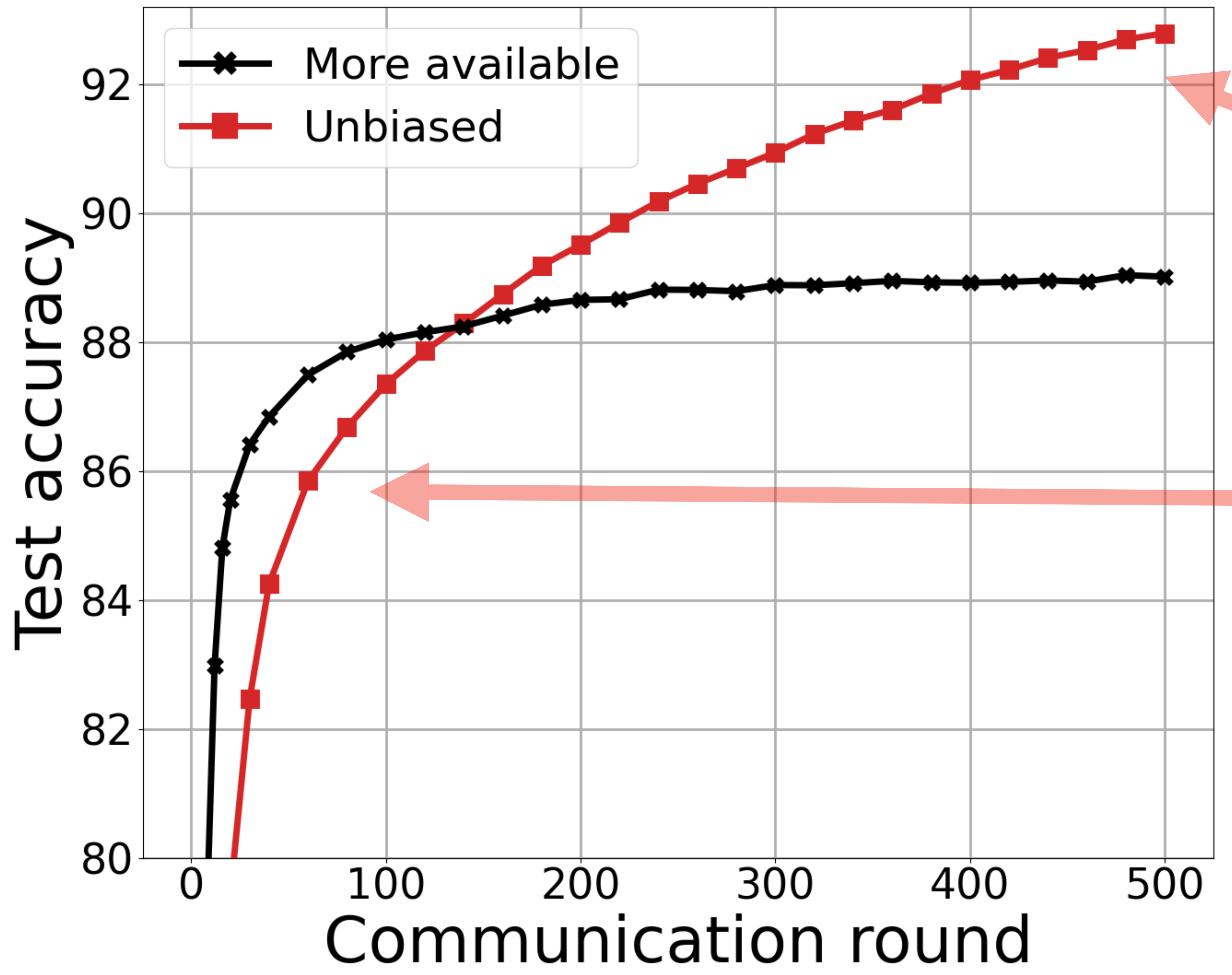


Experimental setting

- Population of K=100 clients, divided in:



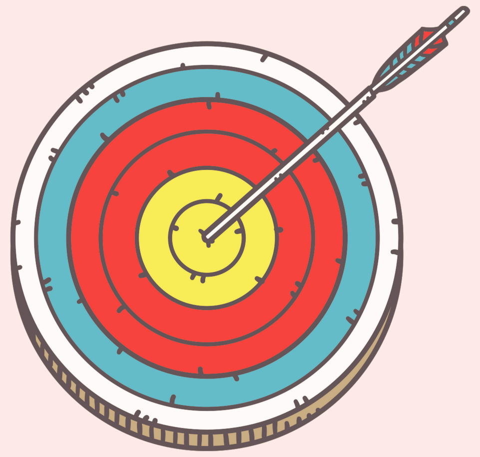
- Trade-off:



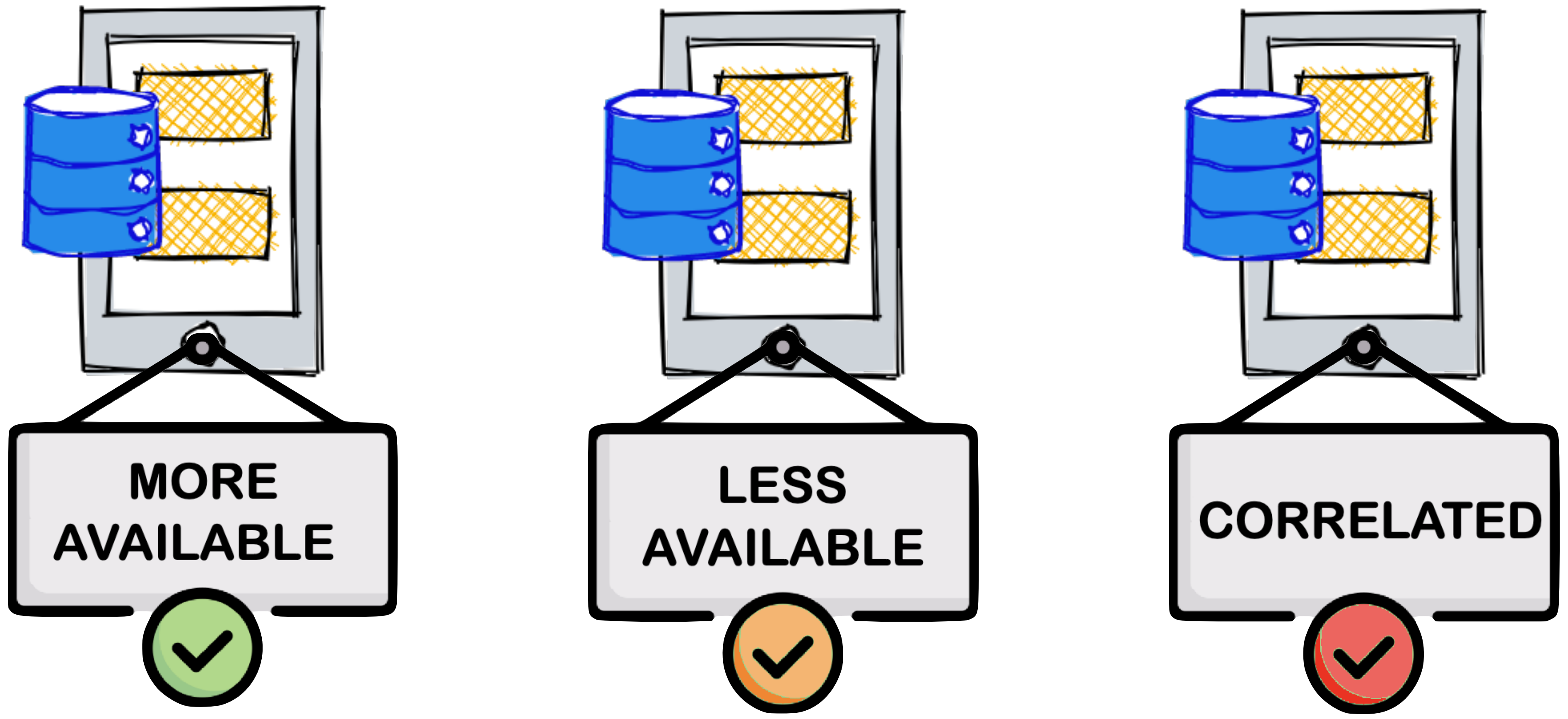
Unbiased

Minimize the bias error

Slower convergence to the target objective

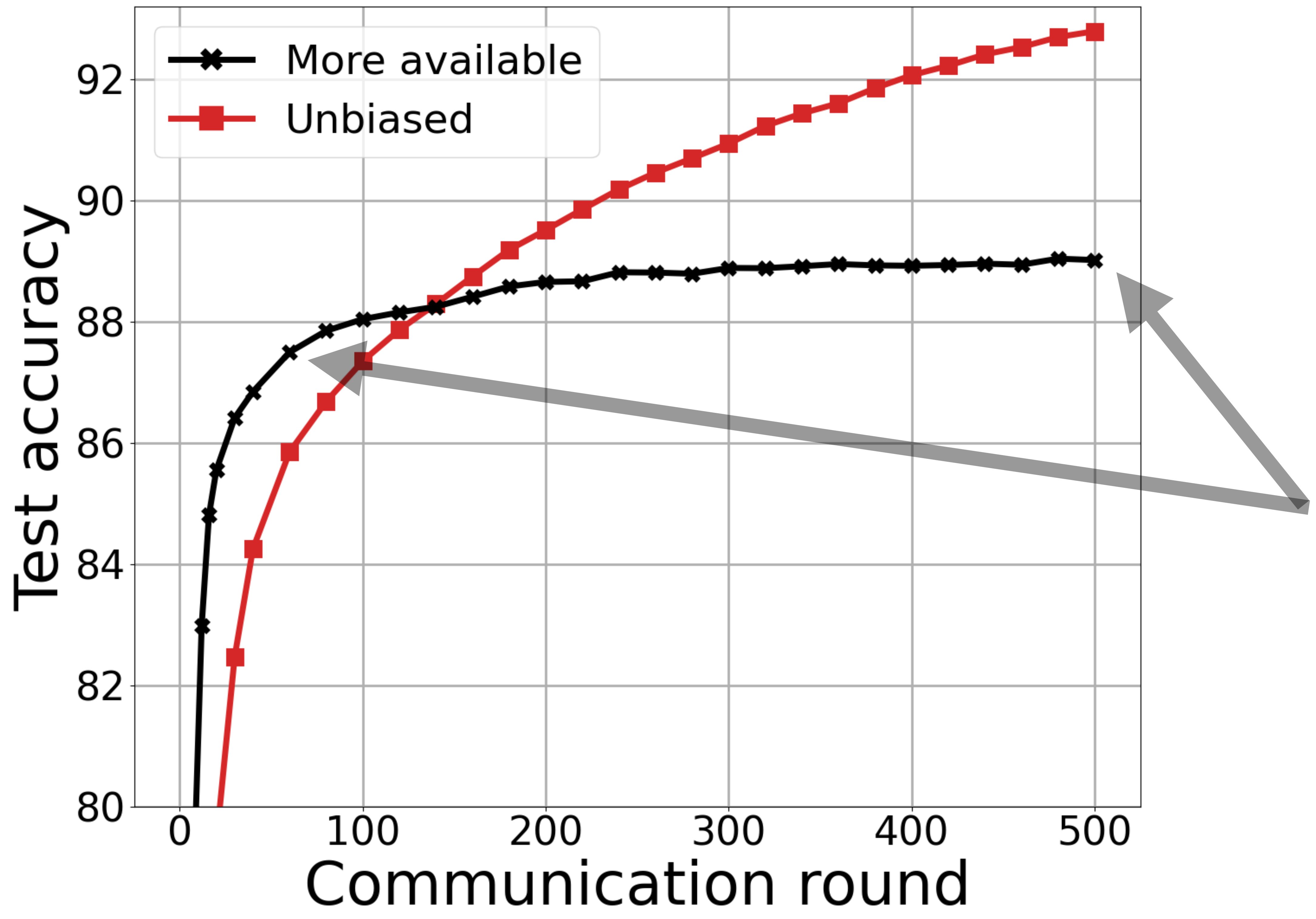


Experimental setting



- Population of K=100 clients, divided in:

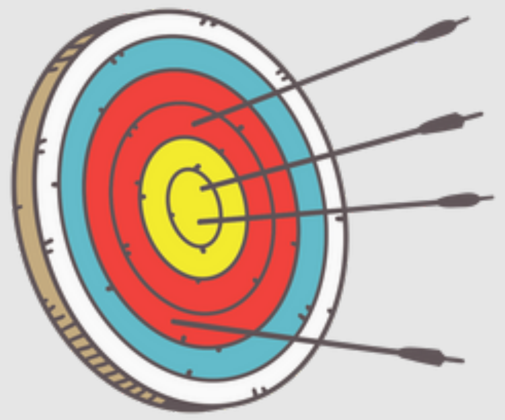
- Trade-off:



More Available

Minimize the optimization error

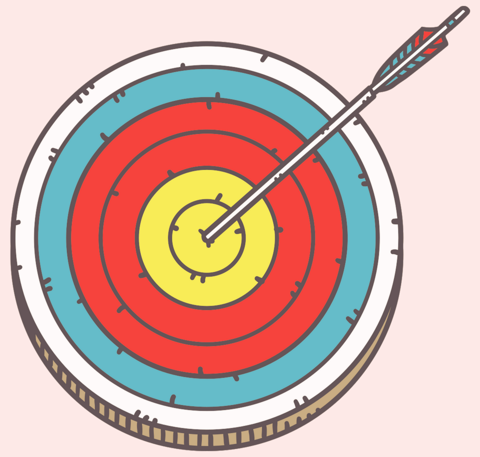
Faster convergence to a biased objective



Unbiased

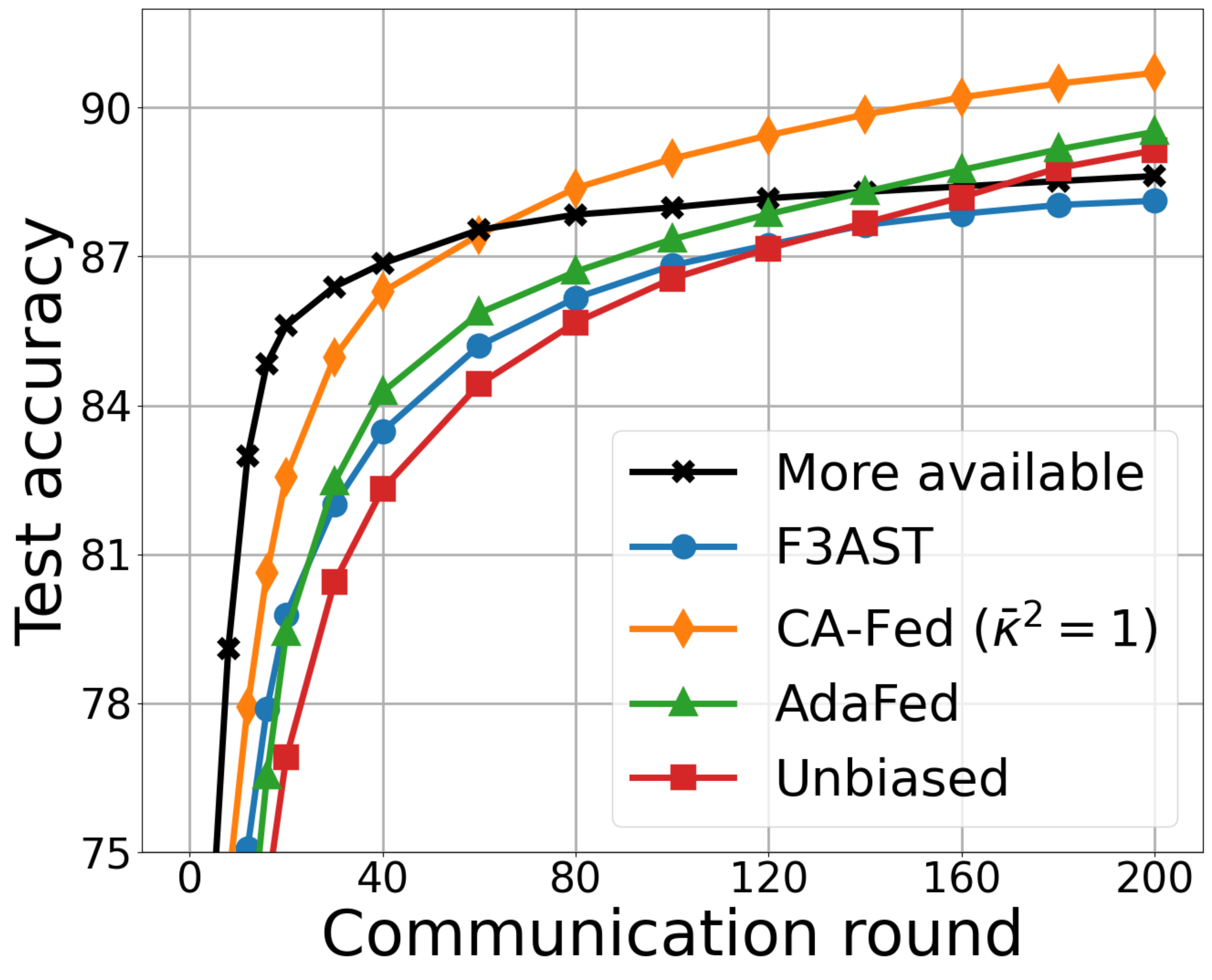
Minimize the bias error

Slower convergence to the target objective

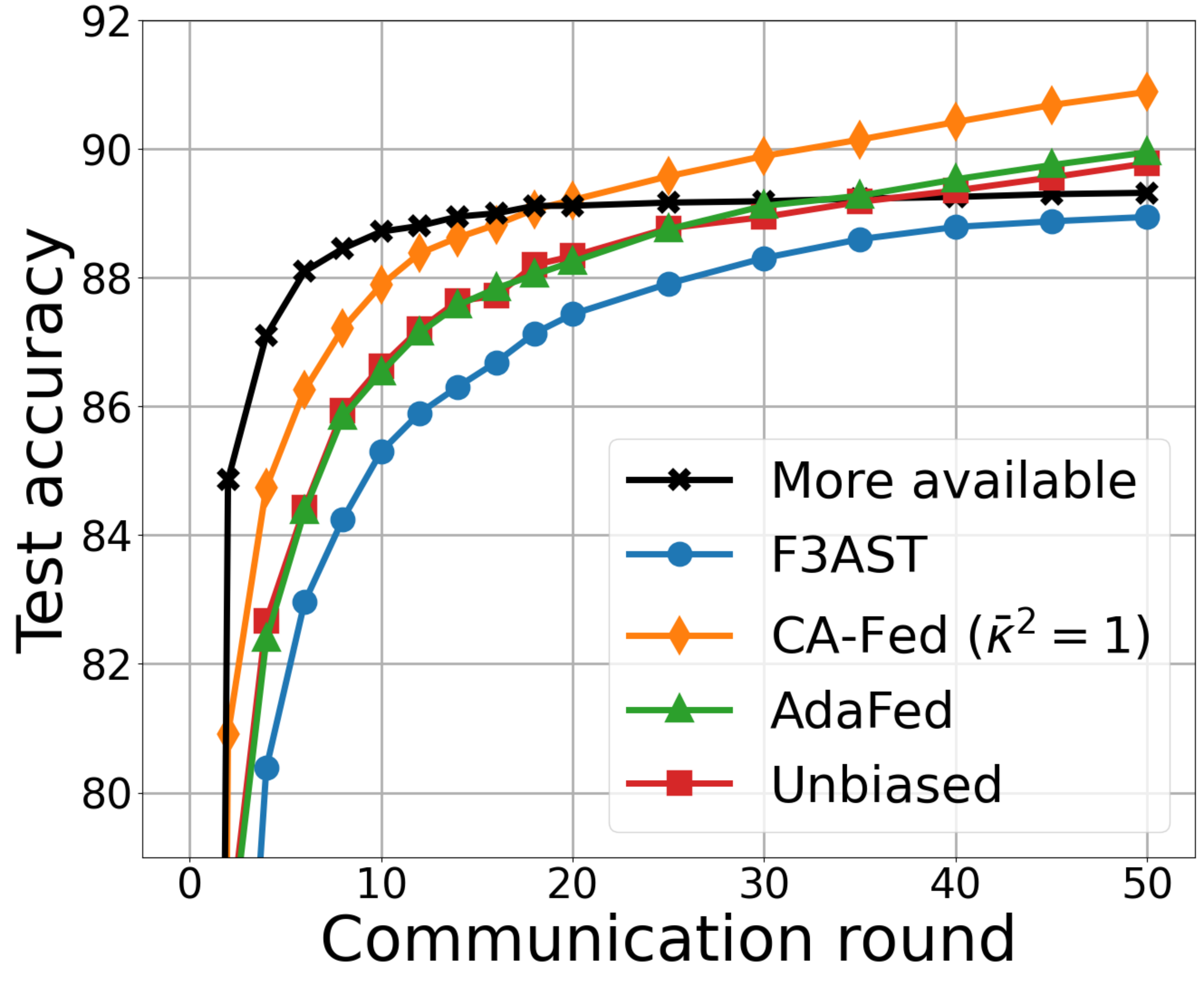


Experimental results

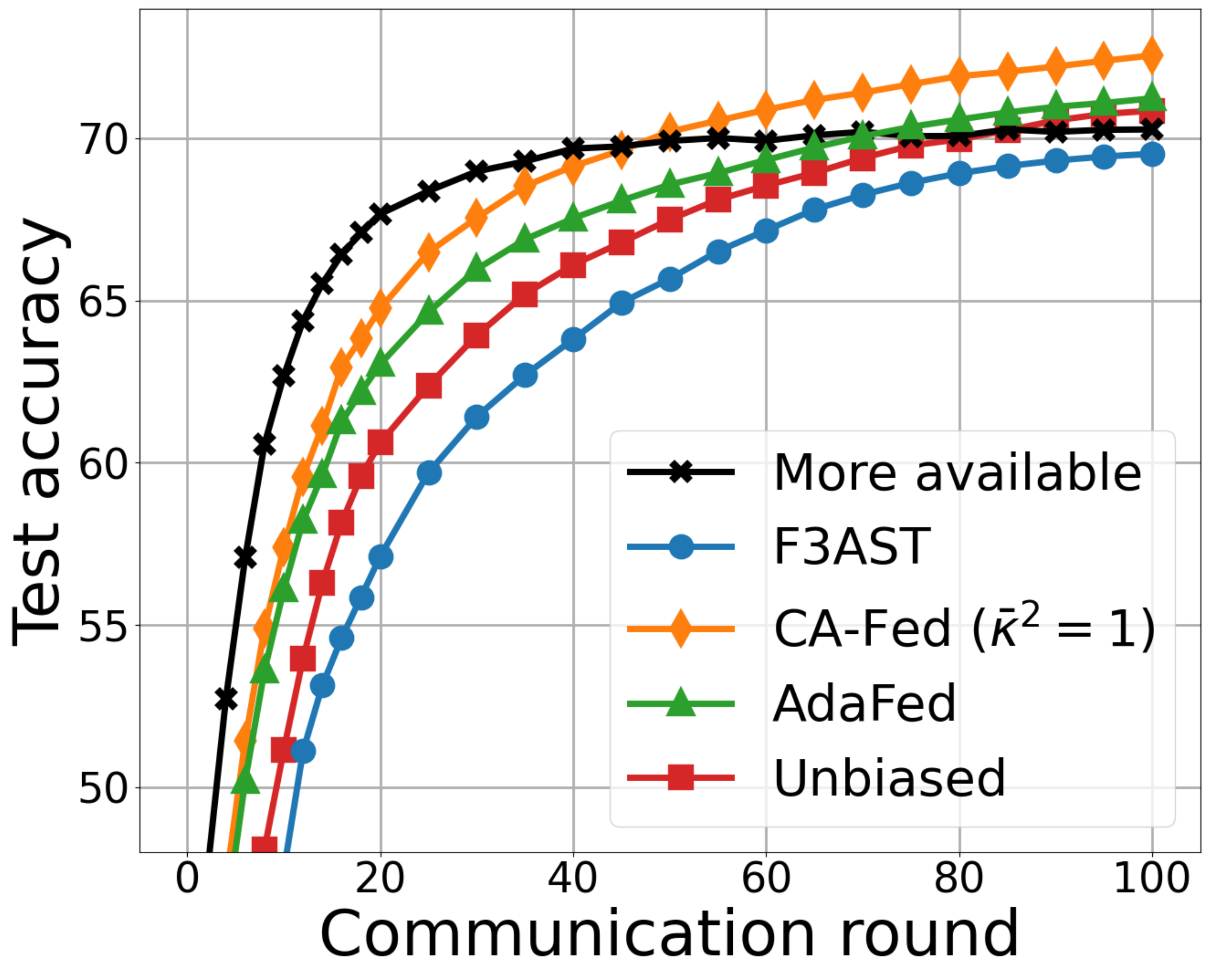
- We compare CA-Fed with More Available, Unbiased, AdaFed, and F3AST



(a) Synthetic LEAF



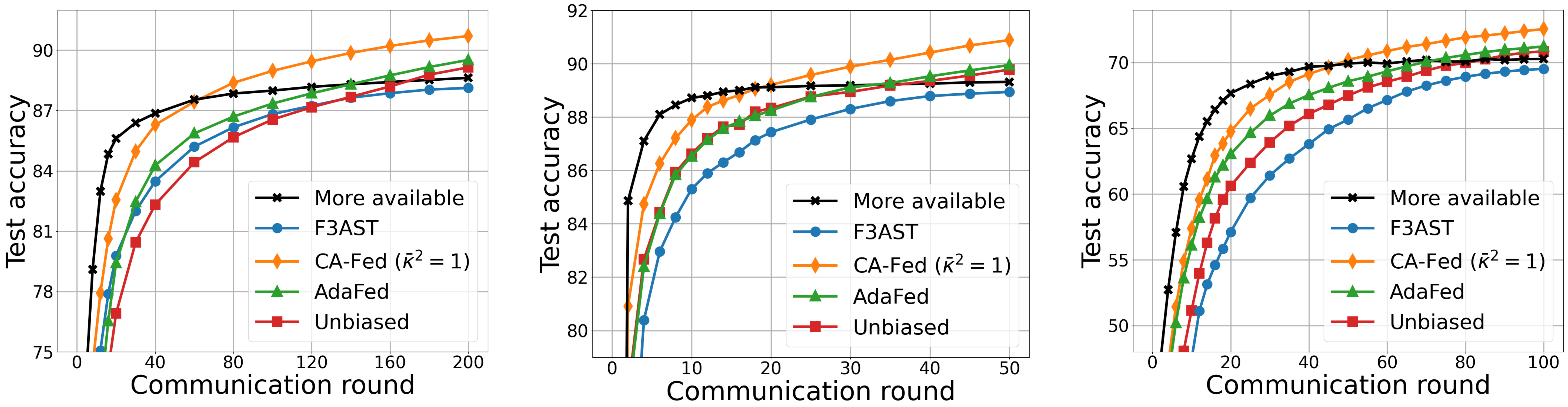
(b) MNIST



(c) CIFAR10

Experimental results

- We compare CA-Fed with More Available, Unbiased, AdaFed, and F3AST



(a) Synthetic LEAF (b) MNIST (c) CIFAR10

CA-Fed achieves higher accuracy in a shorter time

Conclusions

- First convergence analysis under heterogeneous and correlated client availability
- Adaptively excluding less available and highly correlated clients can be effective
- Further discussions and experiments in our paper!

Thank you for your attention!

Paper Code