Université Grenoble Alpes

Franck Iutzeler

Jerome Malick

Asynchronous distributed optimization with unbounded delays

Konstantin Mishchenko
KAUST

$$\min_x F(x) := \frac{1}{M} \sum_{i=1}^{M} f_i(x) + r(x)$$

f_i

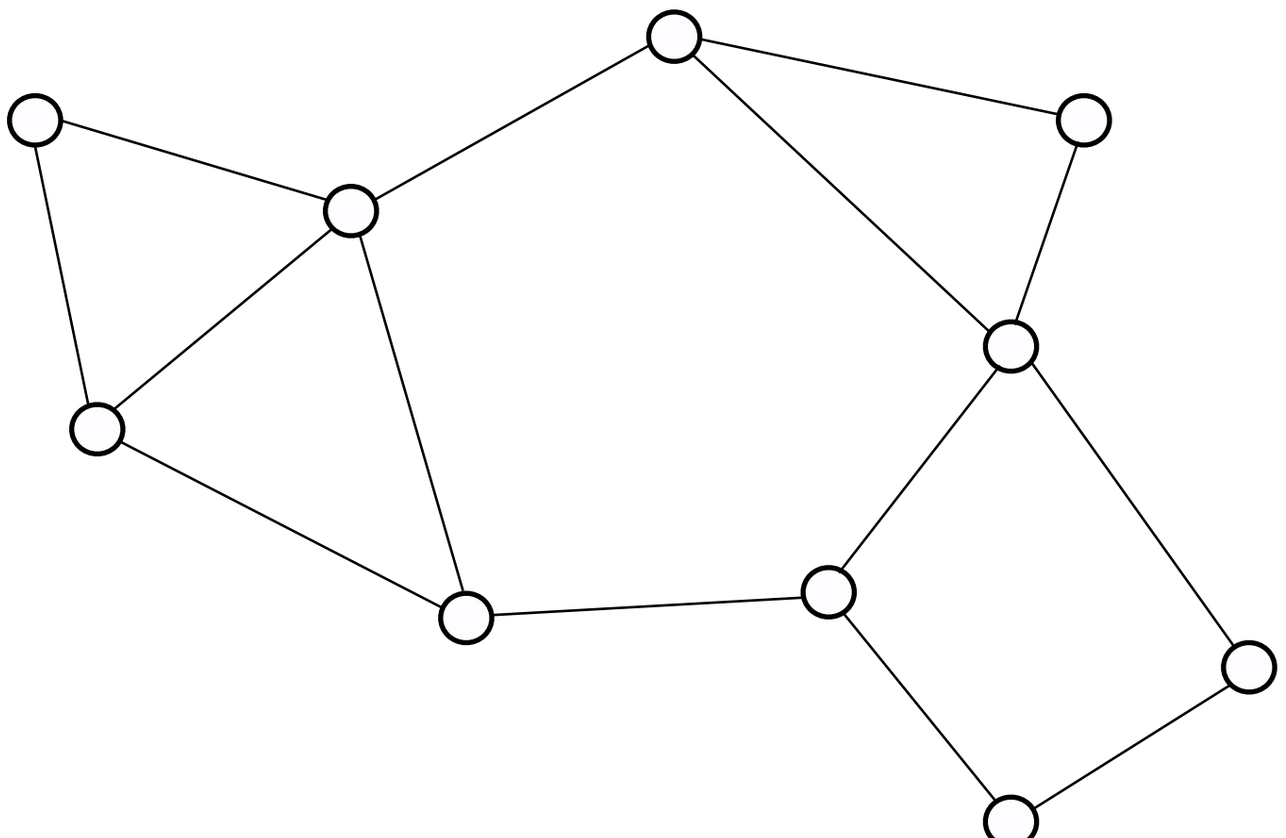
- Represents data on machine i
- Convex, smooth and differentiable

r

- Regularization function
- Convex and proximal, possibly non-smooth

General structure

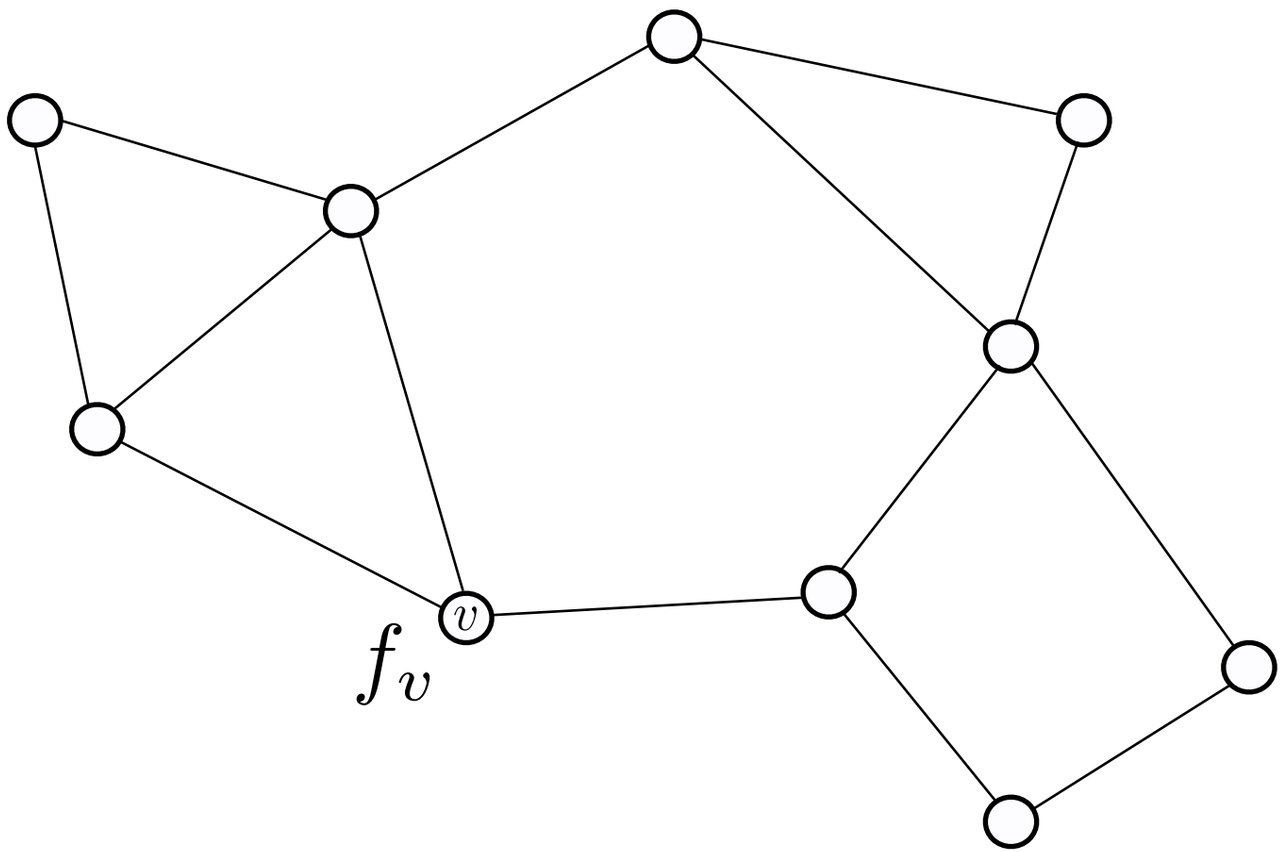
$$G = (V, E)$$



General structure

$$G = (V, E)$$

$$f_v, v \in V$$

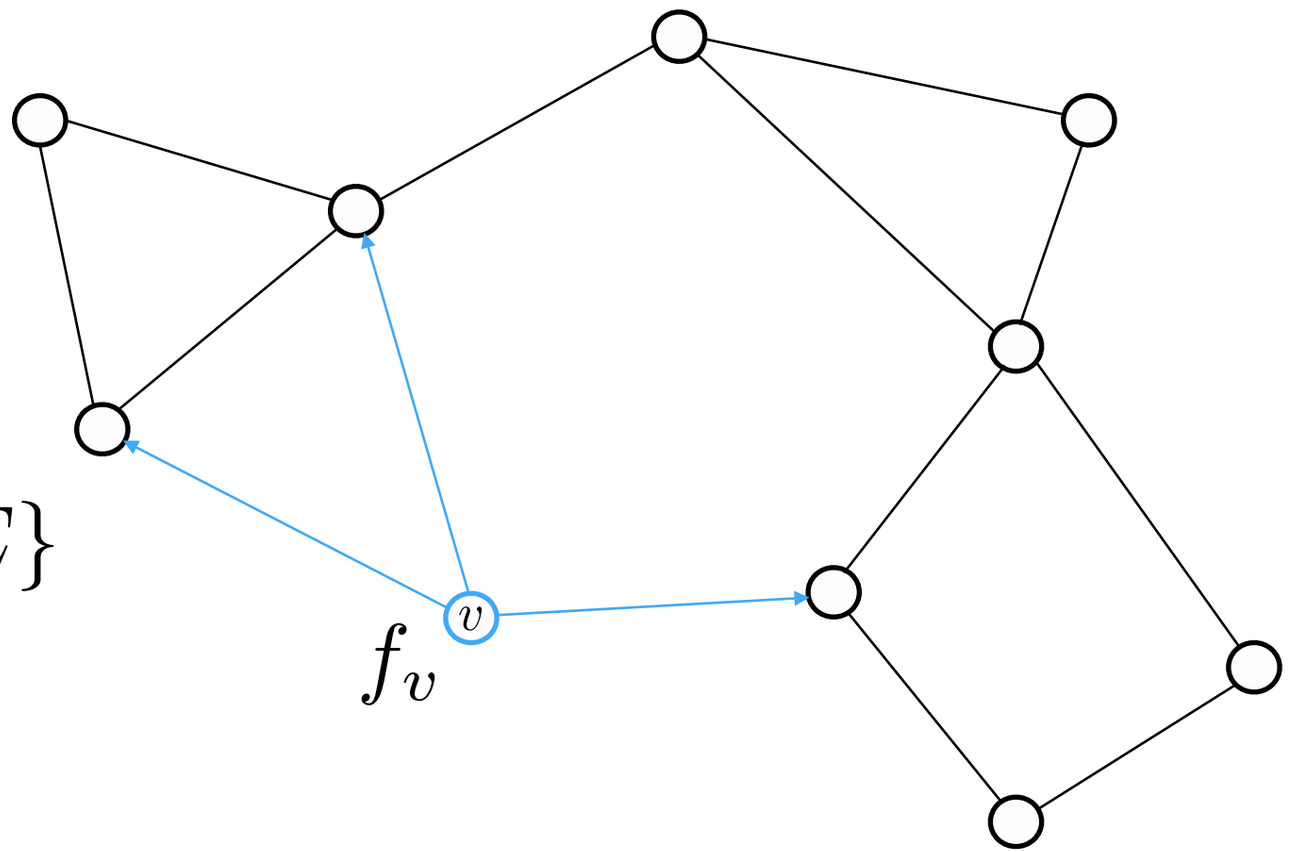


General structure

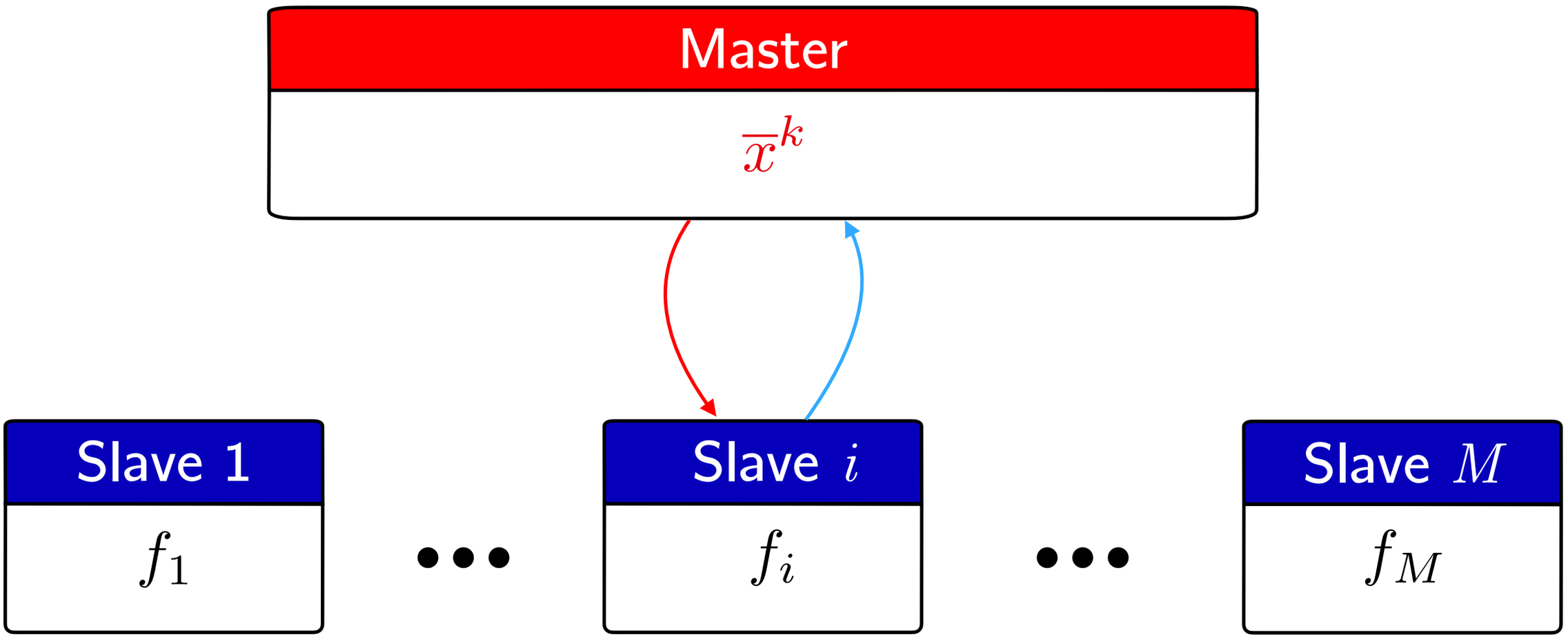
$$G = (V, E)$$

$$f_v, v \in V$$

$$\mathcal{N}_v = \{u : (u, v) \in E\}$$



Parameter server model



Delays and stepsize

d_i^k **delay at Slave i at moment k**

γ **stepsize**

Delays and stepsize

d_i^k **delay at Slave** i **at moment** k

γ **stepsize**

Common assumption: $d_i^k \leq d = \text{const}$

Delays and stepsize

d_i^k **delay at Slave** i **at moment** k

γ **stepsize**

Common assumption: $d_i^k \leq d = \text{const}$

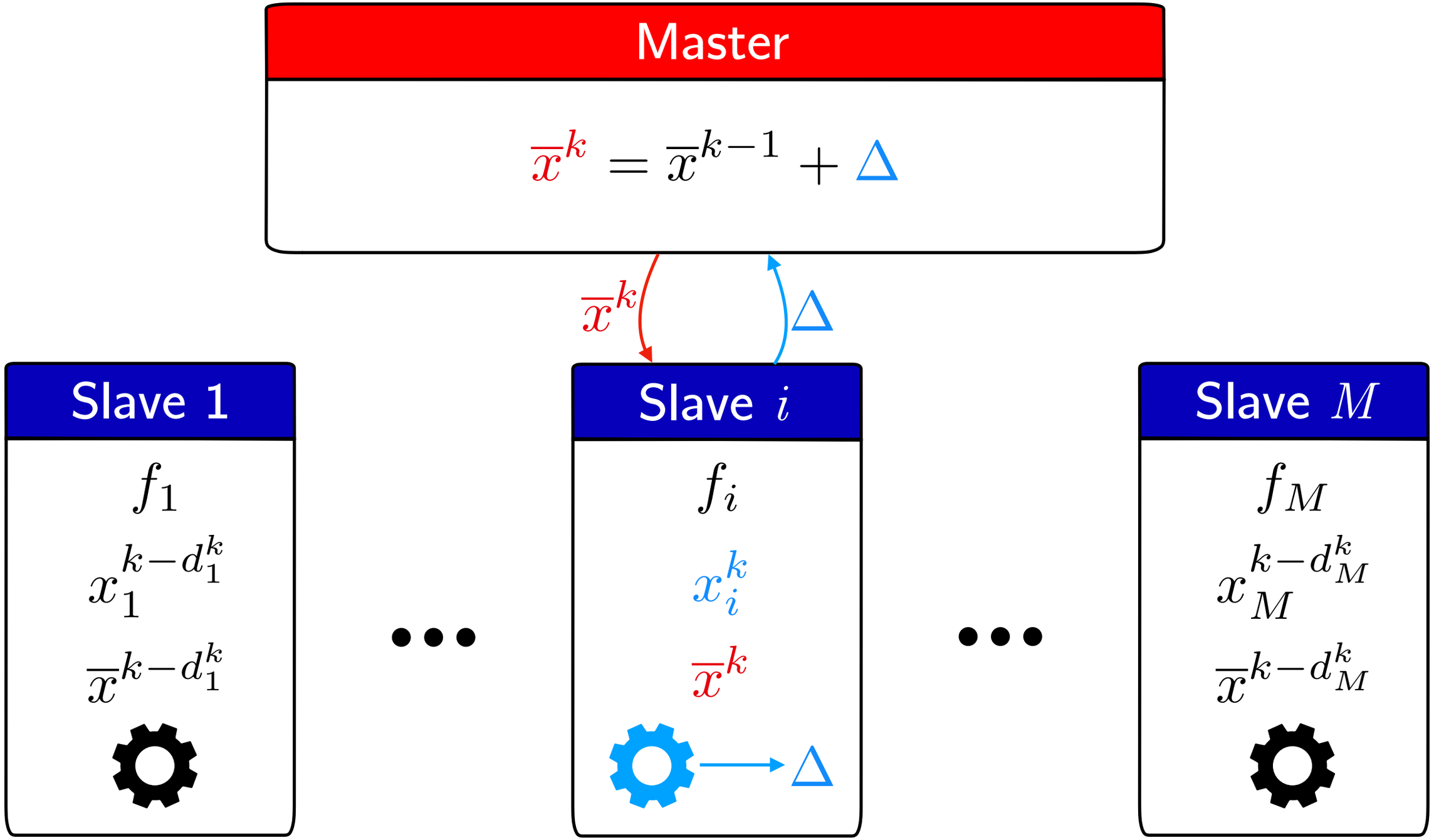
$$d \to +\infty \Longrightarrow \gamma \to 0$$

Convergence significantly depends on d

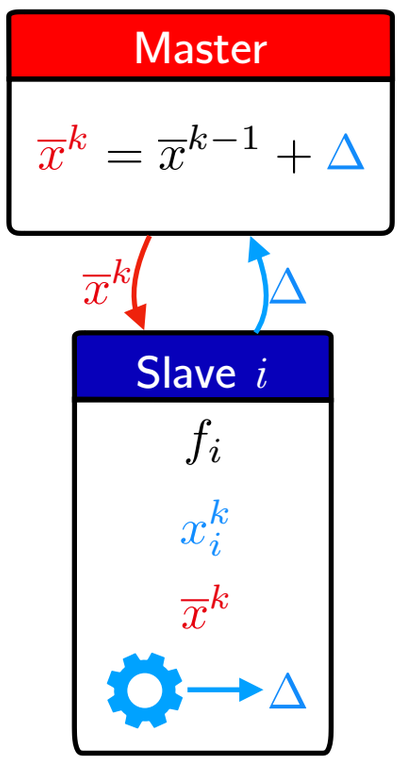
Main idea

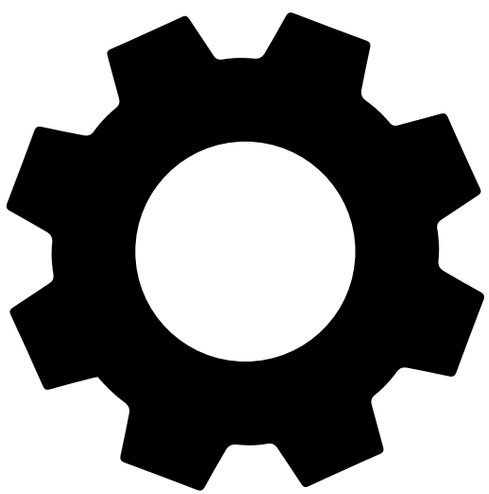
$$\overline{x}^k = \frac{1}{M} \sum_{i=1}^{M} x_i^k$$

$$x_i^k = \cdots = x_i^{k-d_i^k}$$


Master
$\overline{x}^k = \overline{x}^{k-1} + \Delta$
$\overline{x}^k$
Δ
Slave 1
f_1
$x_1^{k-d_1^k}$
$\overline{x}^{k-d_1^k}$
Slave i
f_i
x_i^k
$\overline{x}^k$
Δ
Slave M
f_M
$x_M^{k-d_M^k}$
$\overline{x}^{k-d_M^k}$

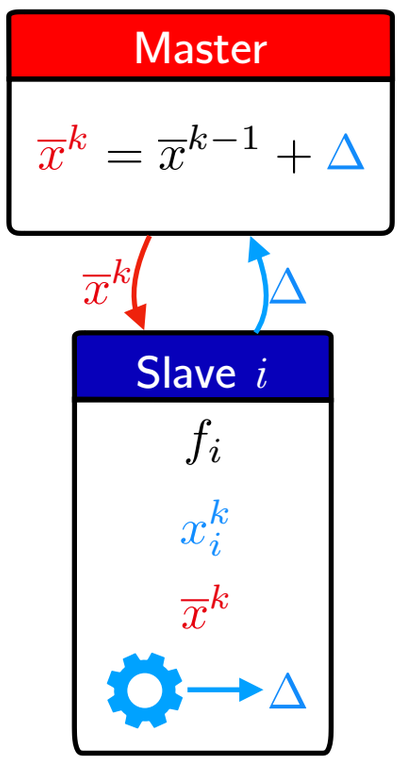
Distributed Gradient Descent

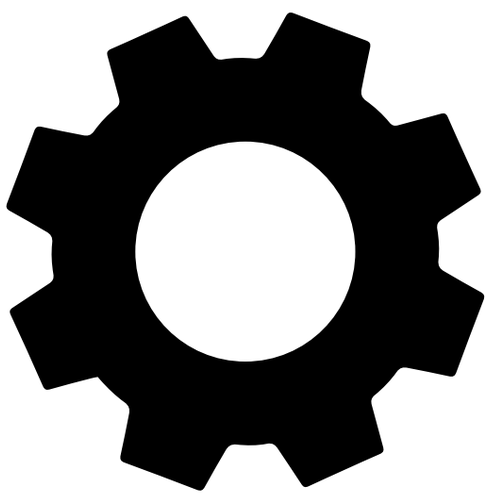




$$
\begin{aligned}
x^+ &\leftarrow \overline{x} - \gamma \nabla f_i(\overline{x}) \\
\Delta &\leftarrow \frac{1}{M}\left(x^+ - x\right) \\
x &\leftarrow x^+
\end{aligned}
$$

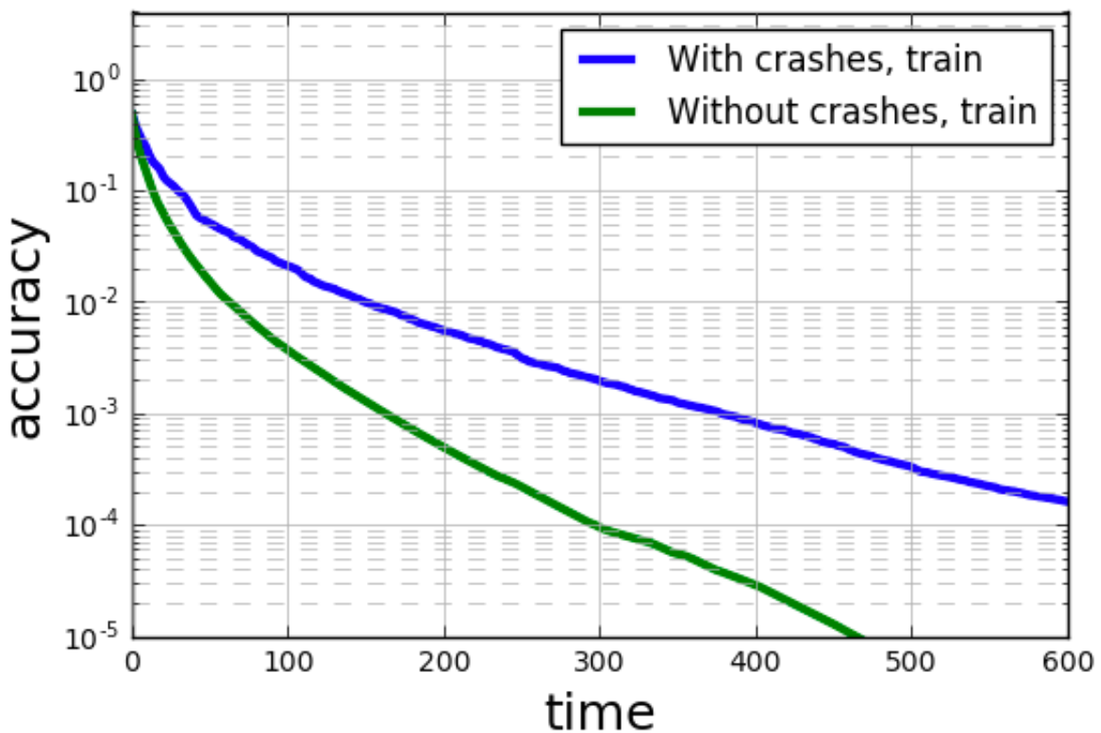
Distributed Gradient Descent



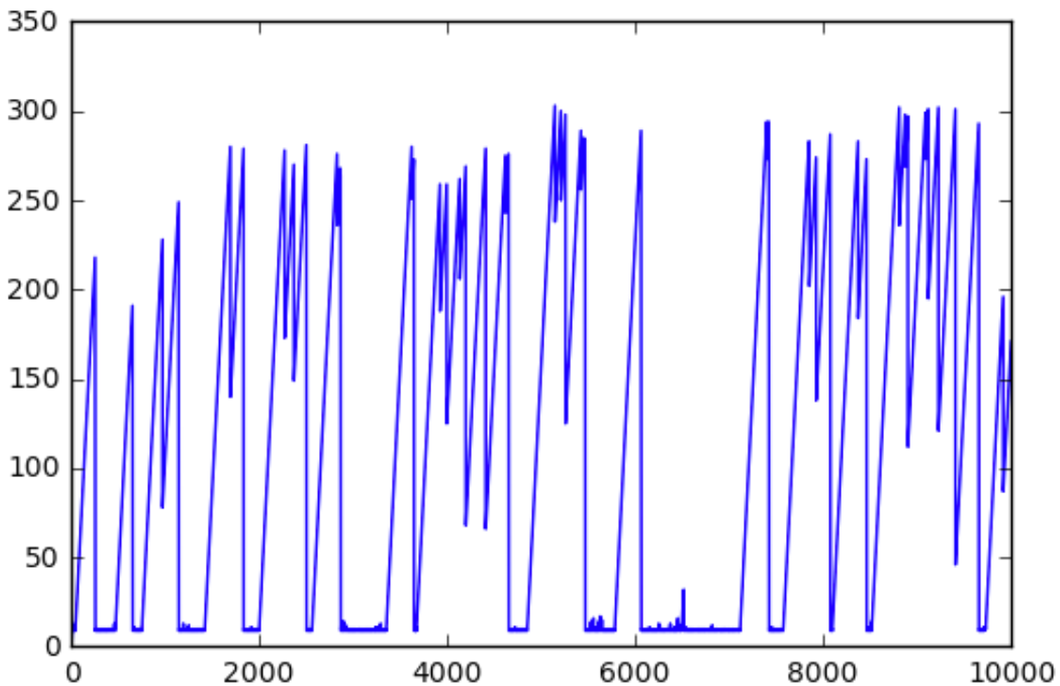


$$
\begin{aligned}
z &\leftarrow \operatorname{prox}_{\gamma r}(\overline{x}) \\
\hline
x^+ &\leftarrow z - \gamma \nabla f_i(z) \\
\Delta &\leftarrow \frac{1}{M}(x^+ - x) \\
x &\leftarrow x^+
\end{aligned}
$$

Convergence



Delays



RCV1 dataset, 10 slaves, 10% of the data on each machine

Open questions

- **General structure**
- **Other types of local computations**
- **Acceleration**
- **Non-convex case**