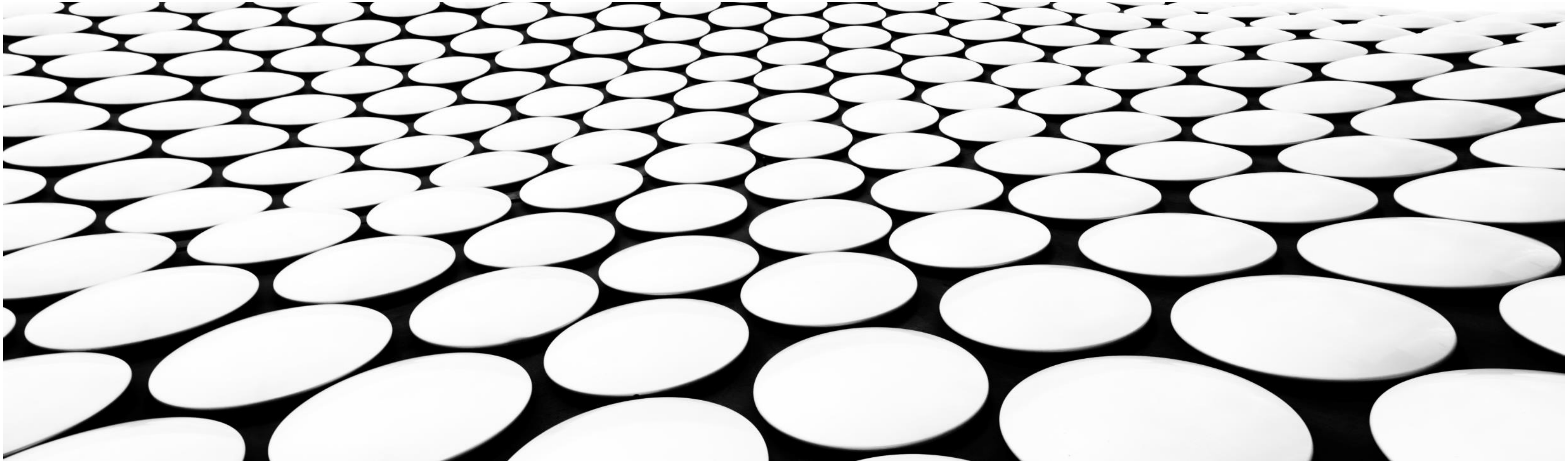
分布式计算

邱怡轩



今天的主题

- ADMM 算法（三）
- 一致性优化问题

通用框架

- “通用”的分布式计算框架
  - 一致性优化（Consensus）
  - 共享优化（Sharing）

一致性优化问题

优化问题

- 考虑一个可分的优化问题

$$\text{minimize} \;\; f(x) = \sum_{i=1}^{N} f_i(x)$$

- $x \in \mathbf{R}^n$，$f_i(x)$ 是凸函数

- 注意 x 指的是抽象的参数，不是数据
- 数据通常包括在 f_i 中

一致性问题

- 转换成 ADMM 形式

- Minimize $\sum_{i=1}^{N} f_i(x_i)$
- Subject to $x_i - z = 0,\ i = 1, \dots, N$

- 注意，此时需要被优化的参数包括 $z, x_1, \dots, x_N$， 共 $(N+1)n$ 个
- 全局一致性问题：所有局部变量相等

一致性问题

- Minimize $\sum_{i=1}^{N} f_i(x_i)$
- Subject to $x_i - z = 0,\ i = 1, \dots, N$

- 假设我们有 N 台机器
- 那么每台机器可以独立地计算 $\min_{x_i} f_i(x_i)$
- 但还有额外的约束 $x_1 = x_2 = \cdots = x_N$
- 因此机器之间需要交换信息

迭代算法

$$
\begin{aligned}
x_i^{k+1} &:= \underset{x_i}{\operatorname{argmin}} \left(f_i(x_i) + y_i^{kT}(x_i - z^k) + (\rho/2)\|x_i - z^k\|_2^2 \right) \\
z^{k+1} &:= \frac{1}{N} \sum_{i=1}^{N} \left(x_i^{k+1} + (1/\rho) y_i^k \right) \\
y_i^{k+1} &:= y_i^k + \rho(x_i^{k+1} - z^{k+1}).
\end{aligned}
$$

迭代算法

- 可以证明，$z^k = \bar{x}^k$
- $\bar{x}^k$ 是 $x_1^k, \dots, x_N^k$ 的平均
- 算法可以进一步化简

$$x_i^{k+1} := \underset{x_i}{\operatorname{argmin}} \left(f_i(x_i) + y_i^{kT}(x_i - \overline{x}^k) + (\rho/2)\|x_i - \overline{x}^k\|_2^2 \right)$$

$$y_i^{k+1} := y_i^k + \rho(x_i^{k+1} - \overline{x}^{k+1}).$$

意义

$$\text{minimize} \quad f(x) = \sum_{i=1}^N f_i(x),$$

$$x_i^{k+1} := \underset{x_i}{\operatorname{argmin}} \left(f_i(x_i) + y_i^{kT}(x_i - \overline{x}^k) + (\rho/2)\|x_i - \overline{x}^k\|_2^2 \right)$$

$$y_i^{k+1} := y_i^k + \rho(x_i^{k+1} - \overline{x}^{k+1}).$$

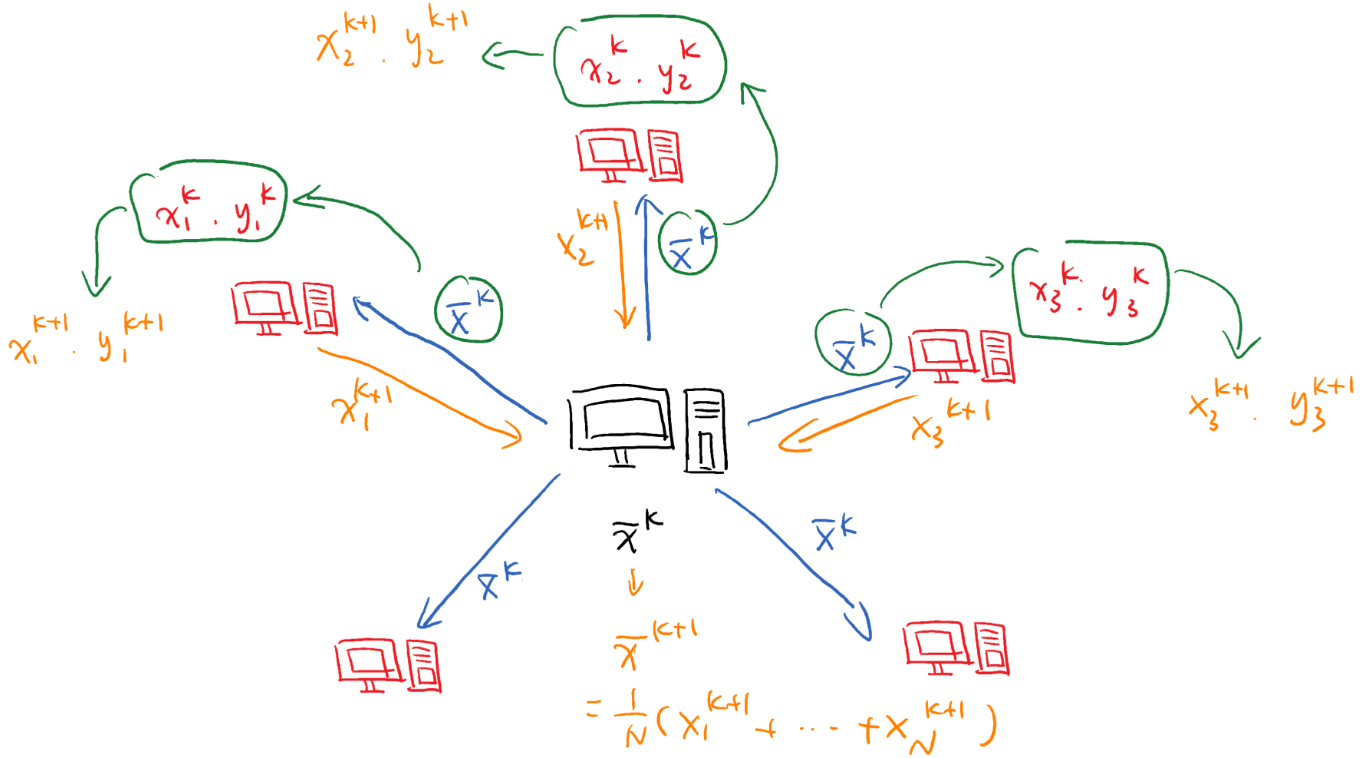
- 许多统计和机器学习模型都可以写成这种形式（似然函数平均）
- 每个 x_i^k 的更新是完全并行的（Map）
- $\bar{x}^k$ 负责收集每个分块的信息（Reduce）

意义

$$\text{minimize} \quad f(x) = \sum_{i=1}^{N} f_i(x),$$

$$x_i^{k+1} := \underset{x_i}{\text{argmin}} \left(f_i(x_i) + y_i^{kT}(x_i - \overline{x}^k) + (\rho/2)\|x_i - \overline{x}^k\|_2^2 \right)$$

$$y_i^{k+1} := y_i^k + \rho(x_i^{k+1} - \overline{x}^{k+1}).$$

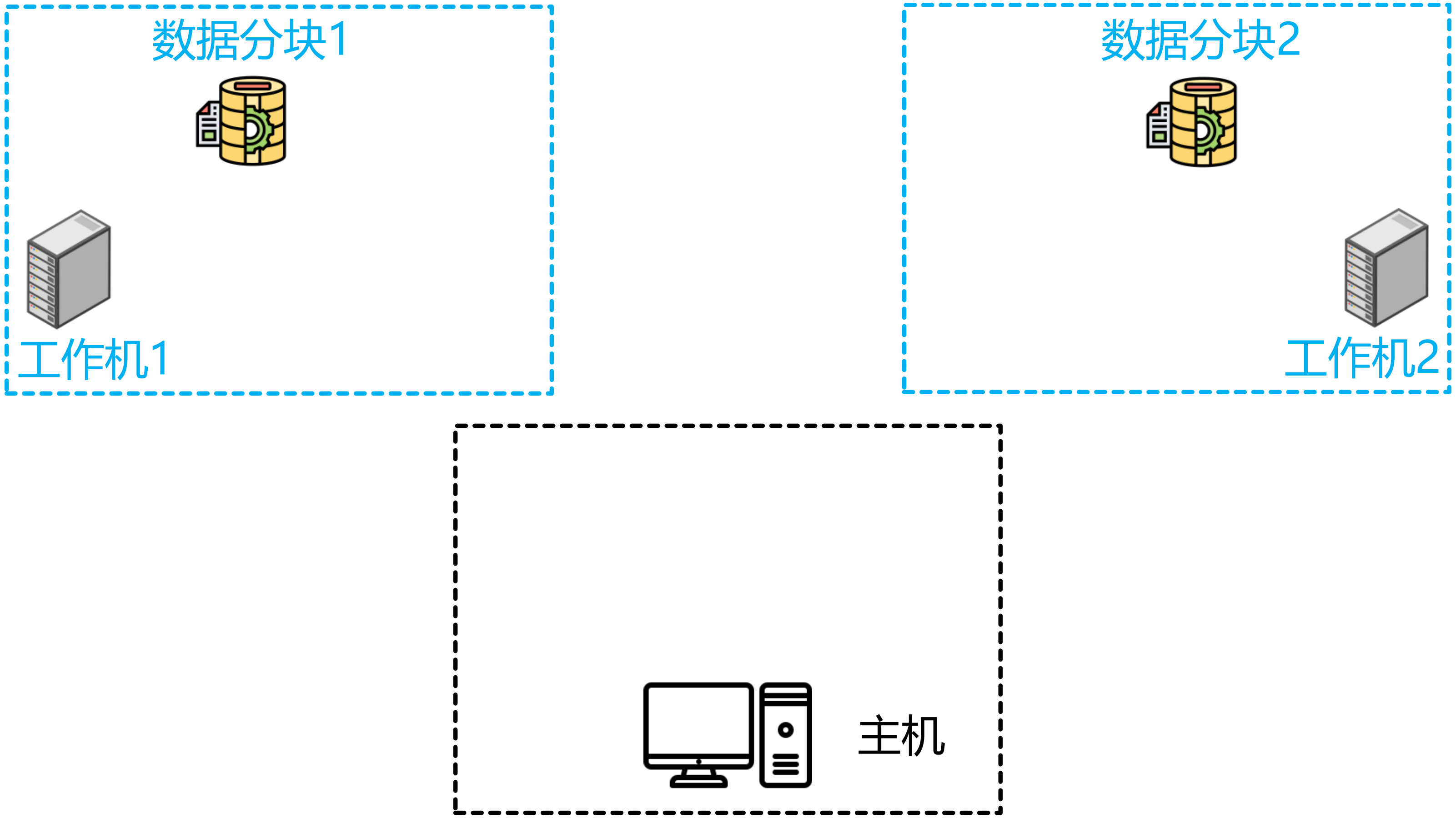


示意图

$$\text{minimize} \quad f(x) = \sum_{i=1}^{N} f_i(x),$$

$$x_i^{k+1} := \underset{x_i}{\text{argmin}} \left(f_i(x_i) + y_i^{kT}(x_i - \overline{x}^k) + (\rho/2)\|x_i - \overline{x}^k\|_2^2 \right)$$

$$y_i^{k+1} := y_i^k + \rho(x_i^{k+1} - \overline{x}^{k+1}).$$

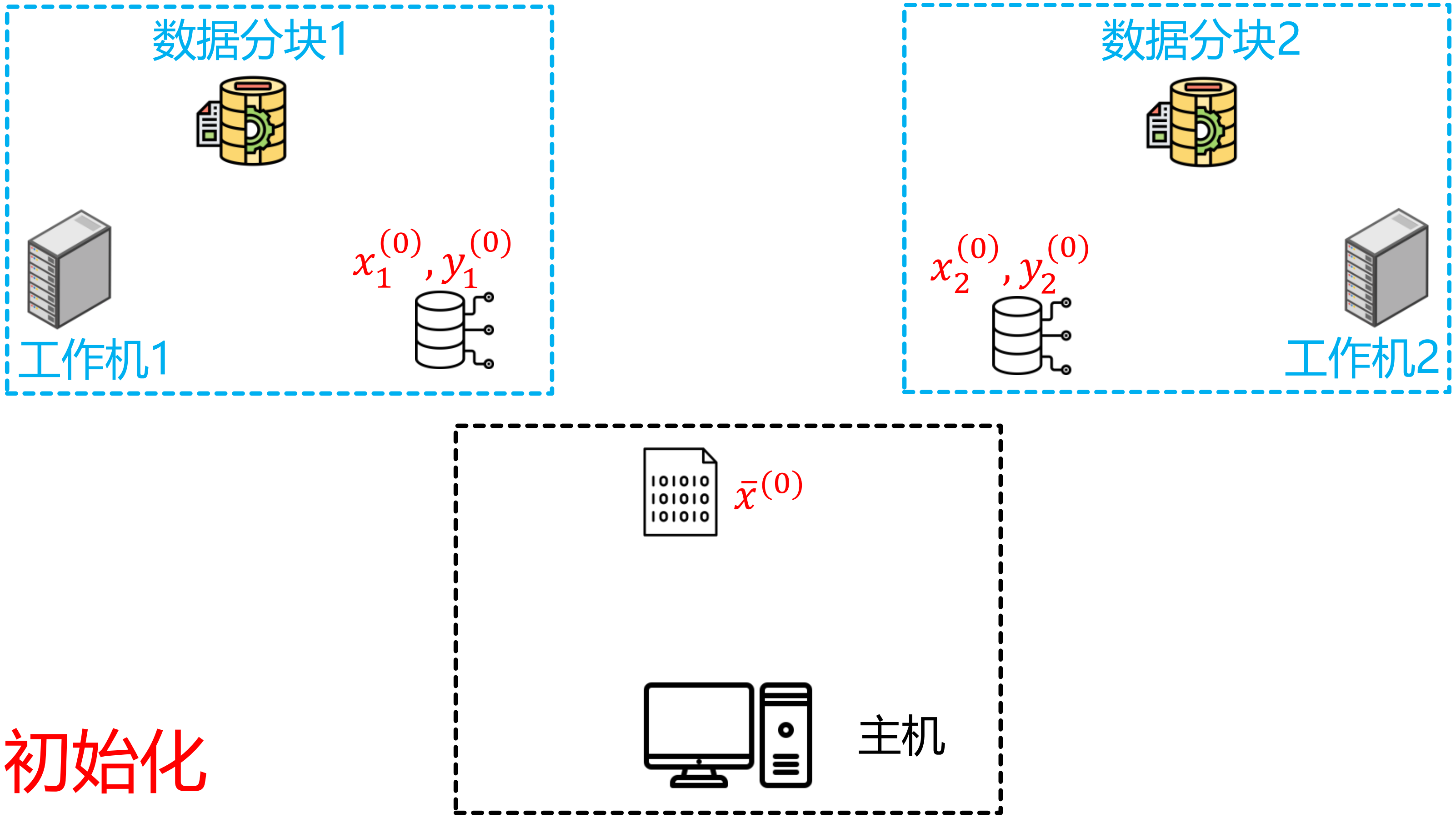


示意图

$$\text{minimize} \quad f(x) = \sum_{i=1}^{N} f_i(x),$$

$$x_i^{k+1} := \operatorname*{argmin}_{x_i} \left(f_i(x_i) + y_i^{kT}(x_i - \overline{x}^k) + (\rho/2)\|x_i - \overline{x}^k\|_2^2 \right)$$

$$y_i^{k+1} := y_i^k + \rho(x_i^{k+1} - \overline{x}^{k+1}).$$



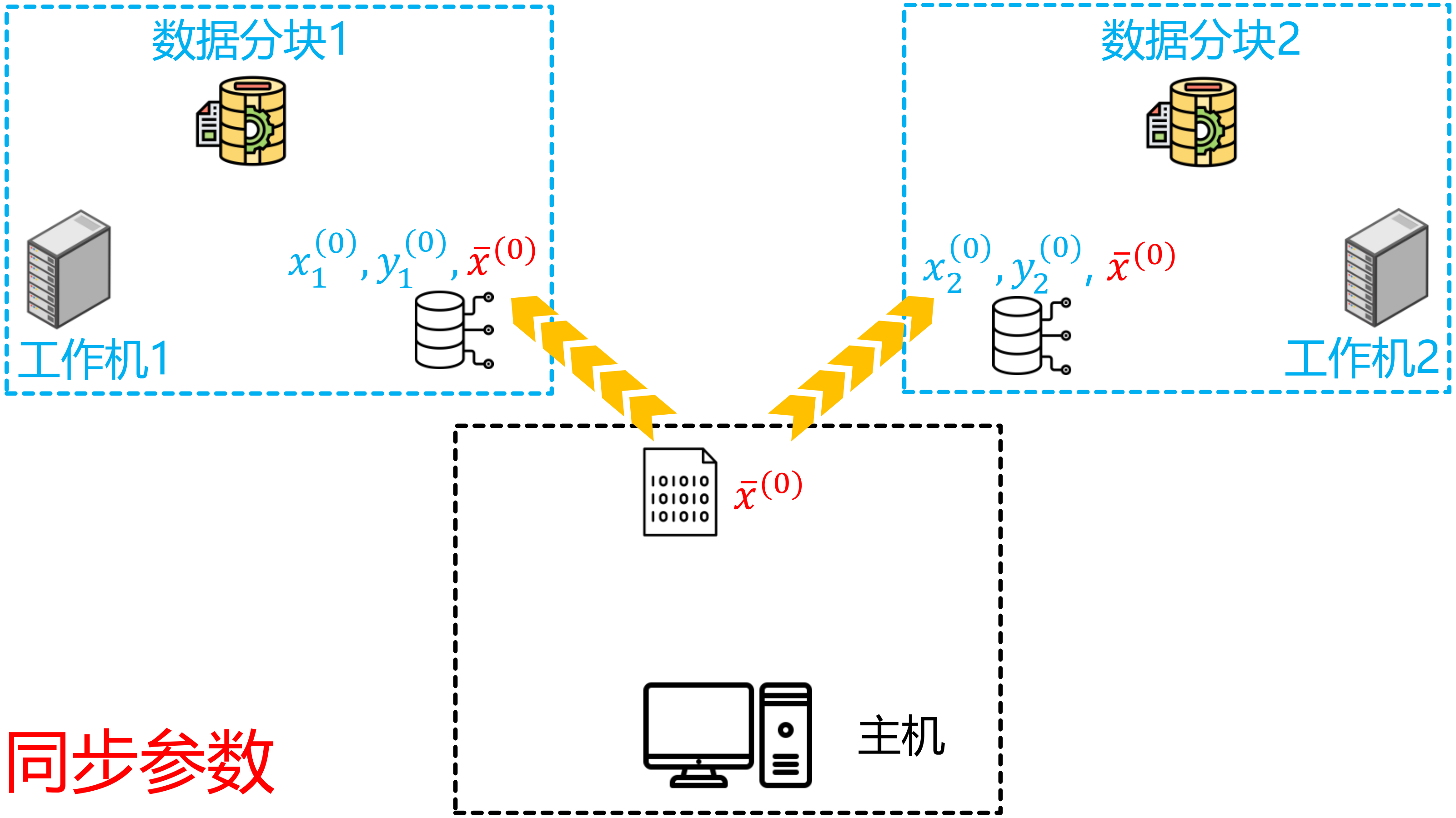
初始化

示意图

$$\text{minimize} \quad f(x) = \sum_{i=1}^{N} f_i(x),$$

$$x_i^{k+1} := \underset{x_i}{\text{argmin}} \left(f_i(x_i) + y_i^{kT}(x_i - \overline{x}^k) + (\rho/2)\|x_i - \overline{x}^k\|_2^2 \right)$$

$$y_i^{k+1} := y_i^k + \rho(x_i^{k+1} - \overline{x}^{k+1}).$$



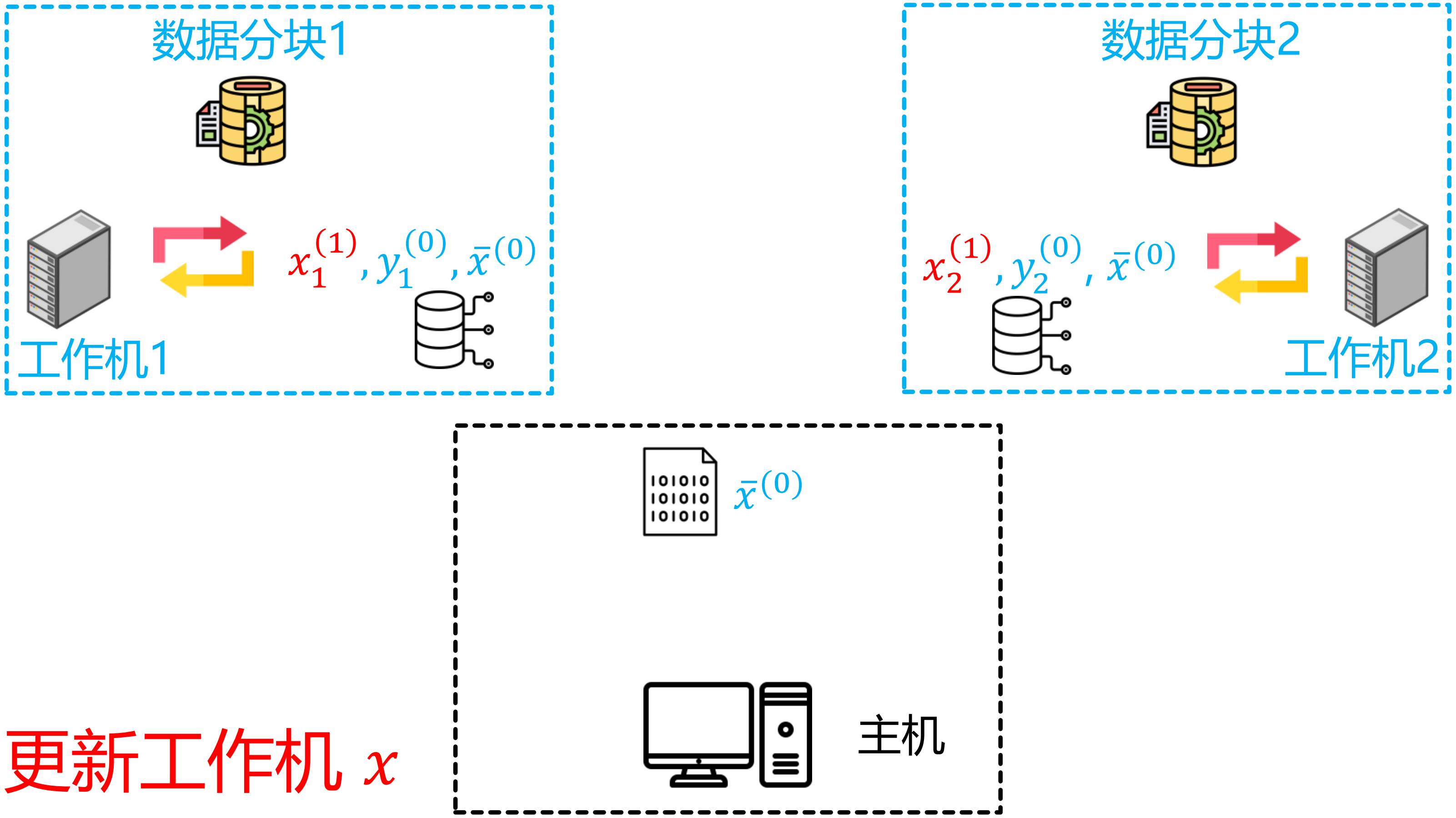
同步参数

示意图

$$\text{minimize} \quad f(x) = \sum_{i=1}^{N} f_i(x),$$

$$\boxed{x_i^{k+1} := \underset{x_i}{\text{argmin}} \left(f_i(x_i) + y_i^{kT}(x_i - \overline{x}^k) + (\rho/2)\|x_i - \overline{x}^k\|_2^2 \right)}$$

$$y_i^{k+1} := y_i^k + \rho(x_i^{k+1} - \overline{x}^{k+1}).$$

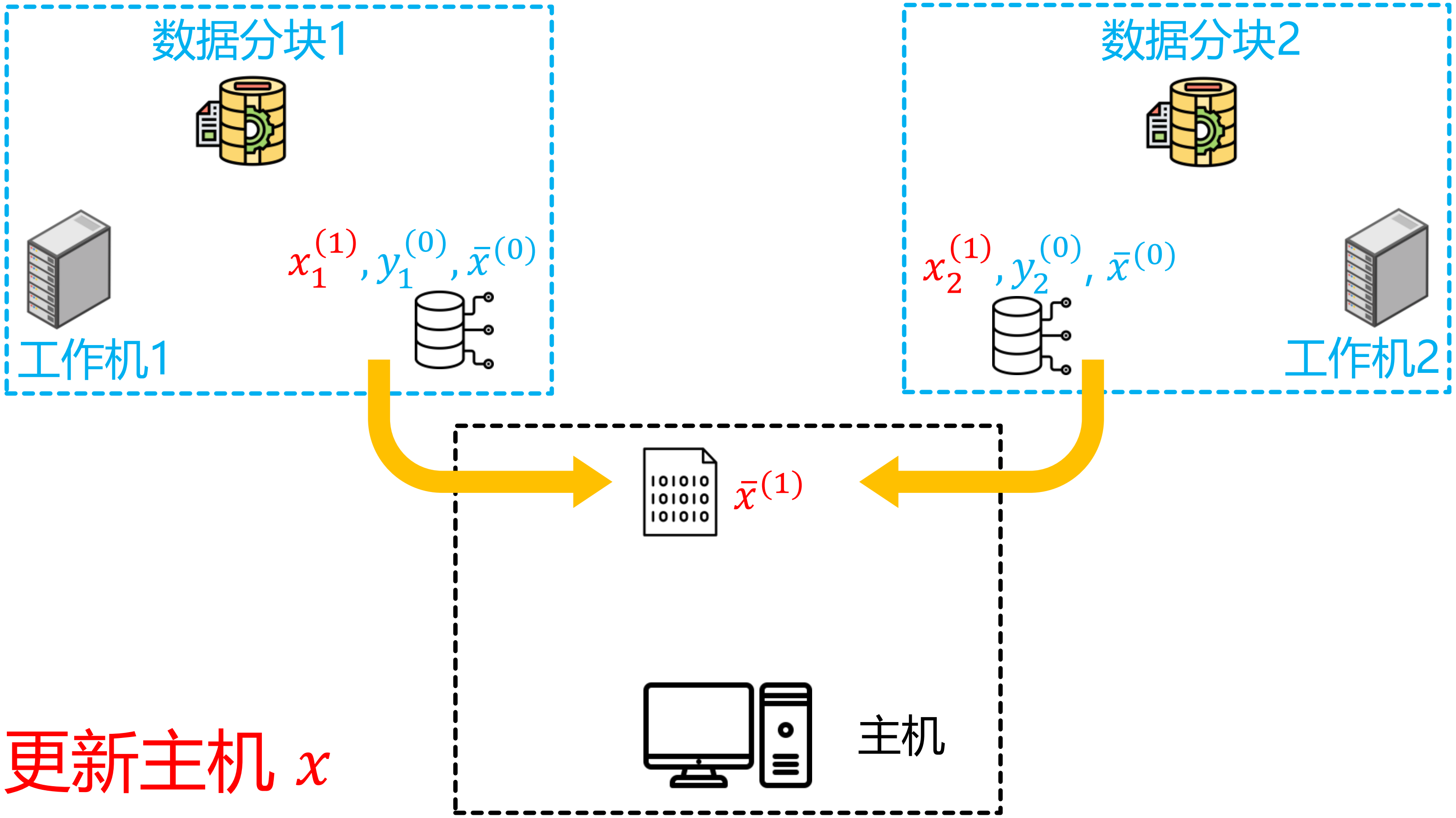


示意图

$$\text{minimize} \quad f(x) = \sum_{i=1}^N f_i(x),$$

$$x_i^{k+1} := \underset{x_i}{\text{argmin}} \left(f_i(x_i) + y_i^{kT}(x_i - \overline{x}^k) + (\rho/2)\|x_i - \overline{x}^k\|_2^2 \right)$$

$$y_i^{k+1} := y_i^k + \rho(x_i^{k+1} - \overline{x}^{k+1}).$$



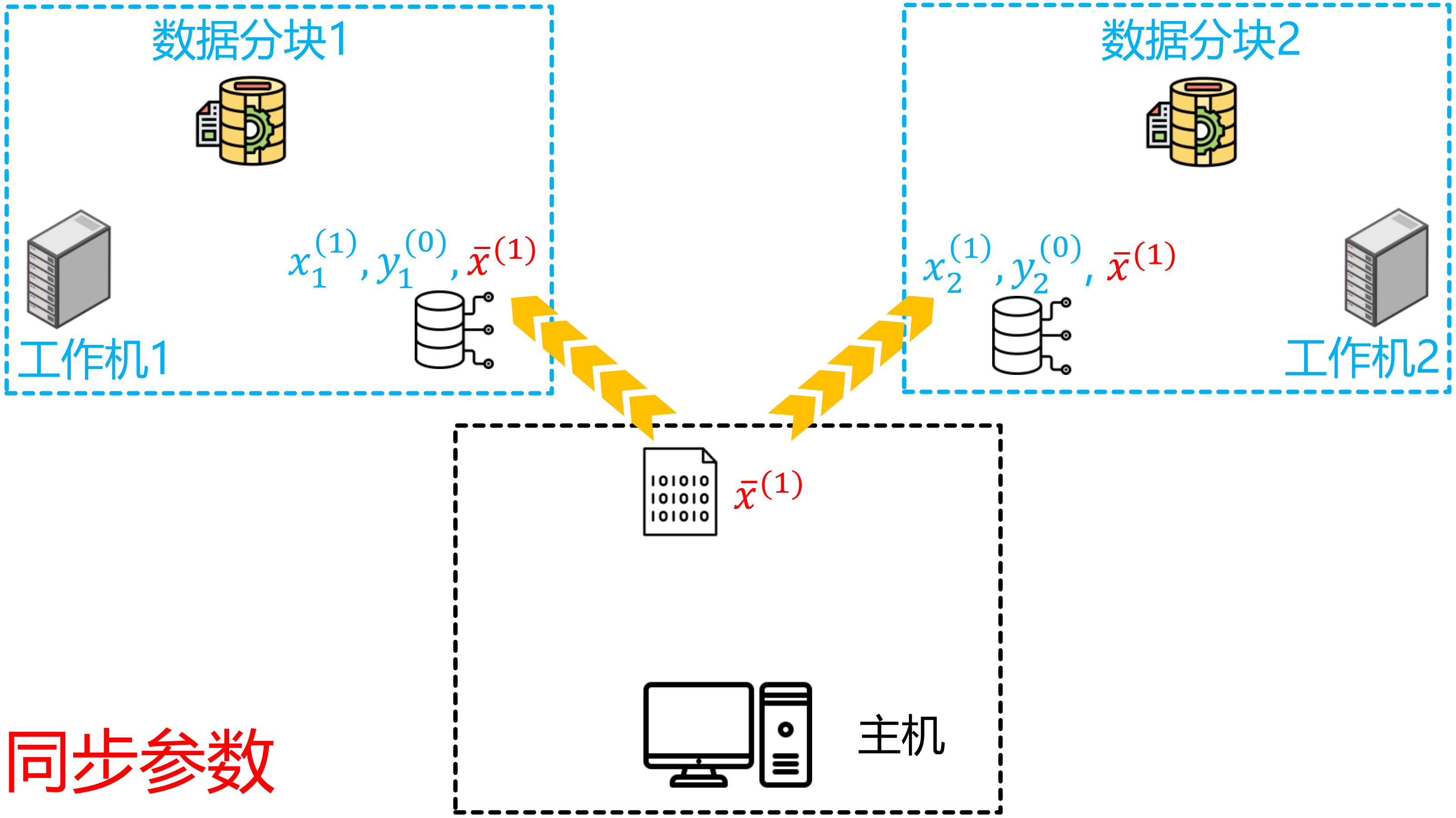
更新主机 x

示意图

$$\text{minimize} \quad f(x) = \sum_{i=1}^{N} f_i(x),$$

$$x_i^{k+1} := \underset{x_i}{\text{argmin}} \left(f_i(x_i) + y_i^{kT}(x_i - \overline{x}^k) + (\rho/2)\|x_i - \overline{x}^k\|_2^2 \right)$$

$$y_i^{k+1} := y_i^k + \rho(x_i^{k+1} - \overline{x}^{k+1}).$$



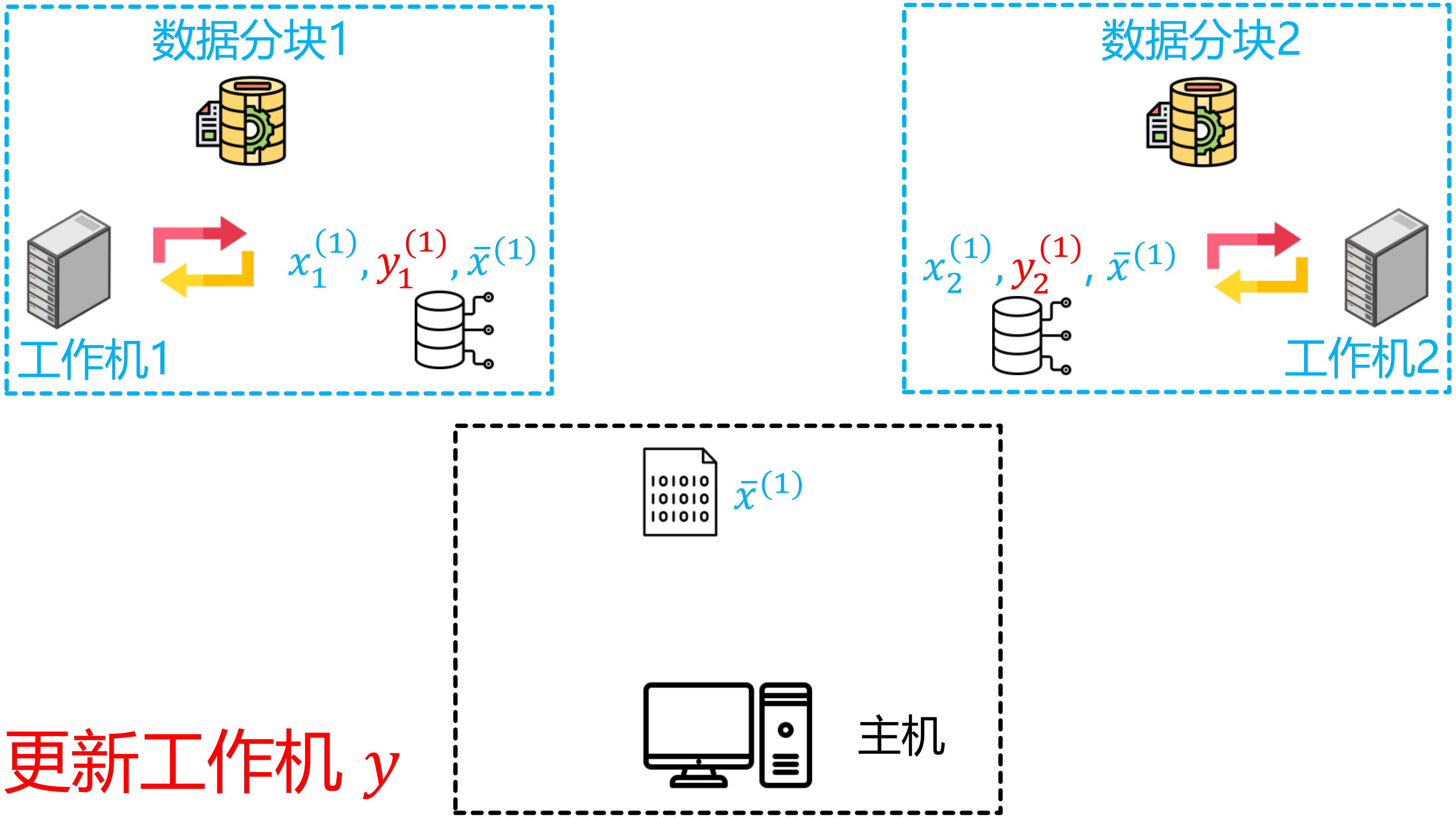
同步参数

示意图

$$\text{minimize} \quad f(x) = \sum_{i=1}^{N} f_i(x),$$

$$x_i^{k+1} := \underset{x_i}{\text{argmin}} \left(f_i(x_i) + y_i^{kT}(x_i - \overline{x}^k) + (\rho/2)\|x_i - \overline{x}^k\|_2^2 \right)$$

$$\boxed{y_i^{k+1} := y_i^k + \rho(x_i^{k+1} - \overline{x}^{k+1}).}$$



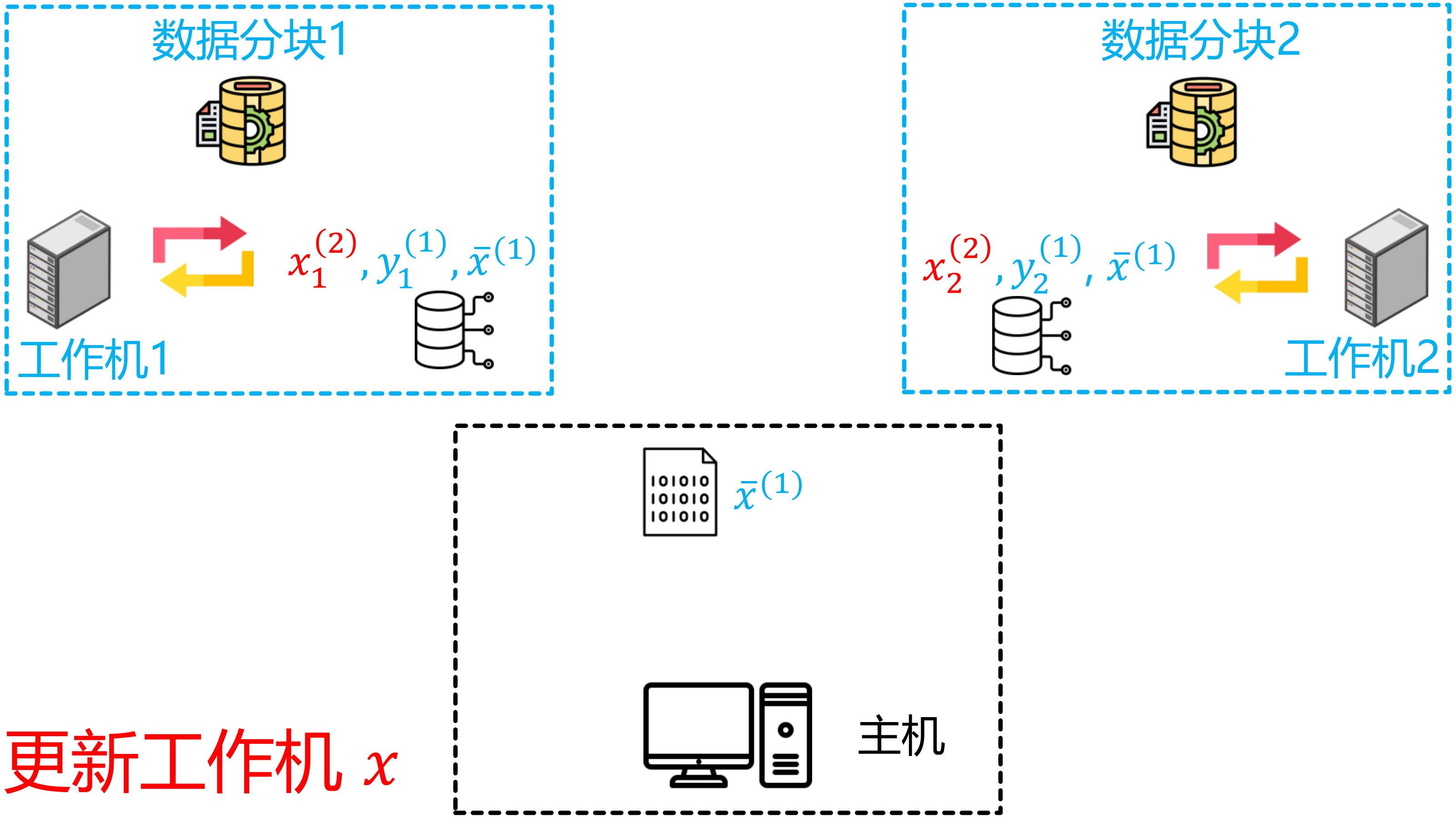
更新工作机 y

示意图

$$\text{minimize} \quad f(x) = \sum_{i=1}^{N} f_i(x),$$

$$\boxed{x_i^{k+1} := \underset{x_i}{\text{argmin}}\left(f_i(x_i) + y_i^{kT}(x_i - \overline{x}^k) + (\rho/2)\|x_i - \overline{x}^k\|_2^2\right)}$$

$$y_i^{k+1} := y_i^k + \rho(x_i^{k+1} - \overline{x}^{k+1}).$$



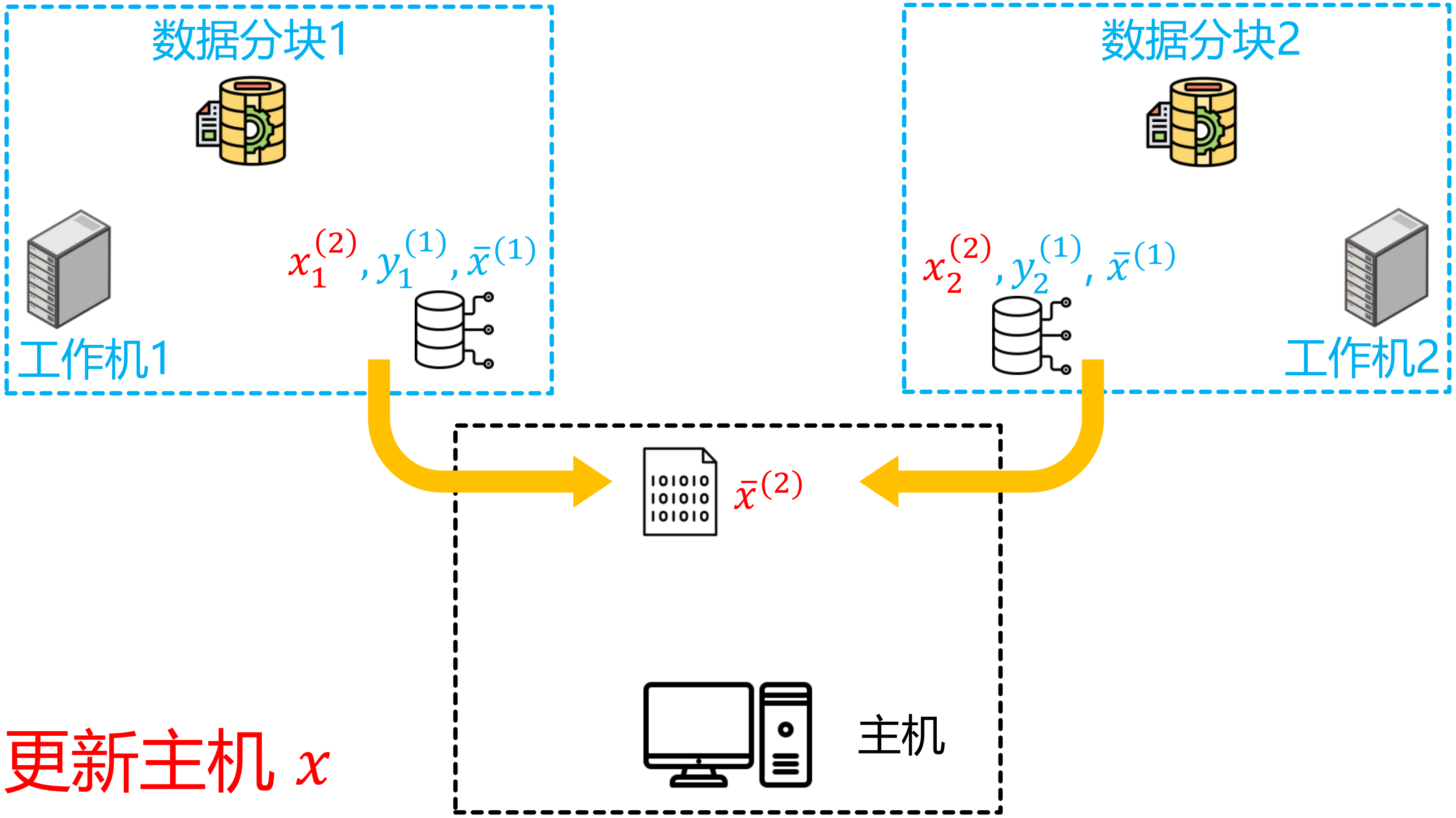
更新工作机 x

示意图

$$\text{minimize} \quad f(x) = \textstyle\sum_{i=1}^{N} f_i(x),$$

$$x_i^{k+1} := \underset{x_i}{\text{argmin}} \left(f_i(x_i) + y_i^{kT}(x_i - \overline{x}^k) + (\rho/2)\|x_i - \overline{x}^k\|_2^2 \right)$$

$$y_i^{k+1} := y_i^k + \rho(x_i^{k+1} - \overline{x}^{k+1}).$$



例：线性回归

$$\text{minimize} \quad f(x) = \sum_{i=1}^{N} f_i(x),$$

$$x_i^{k+1} := \underset{x_i}{\operatorname{argmin}} \left(f_i(x_i) + y_i^{kT}(x_i - \overline{x}^k) + (\rho/2)\|x_i - \overline{x}^k\|_2^2 \right)$$

$$y_i^{k+1} := y_i^k + \rho(x_i^{k+1} - \overline{x}^{k+1}).$$

- 如果原问题是最小二乘回归
- 将数据按观测切为 N 块
- 那么每个 f_i 就是每个分块上的损失函数
- 每个分块上各自求解一个线性方程组

正则项

- 有时我们需要对参数加入全局的正则项
- 优化问题

$$\text{minimize}\ \ f(x) = g(x) + \sum_{i=1}^{N} f_i(x)$$

- $f_i(x)$，$g(x)$ 是凸函数

- 例：Lasso

$$\min_x \frac{1}{2}\|Ax - b\|^2 + \lambda\|x\|_1$$

正则项

- 优化问题

$$\text{minimize } f(x) = g(x) + \sum_{i=1}^{N} f_i(x)$$

- 转换成 ADMM 形式
- Minimize $\quad g(z) + \sum_{i=1}^{N} f_i(x_i)$
- Subject to $\ x_i - z = 0,\ i = 1, \ldots, N$

迭代算法

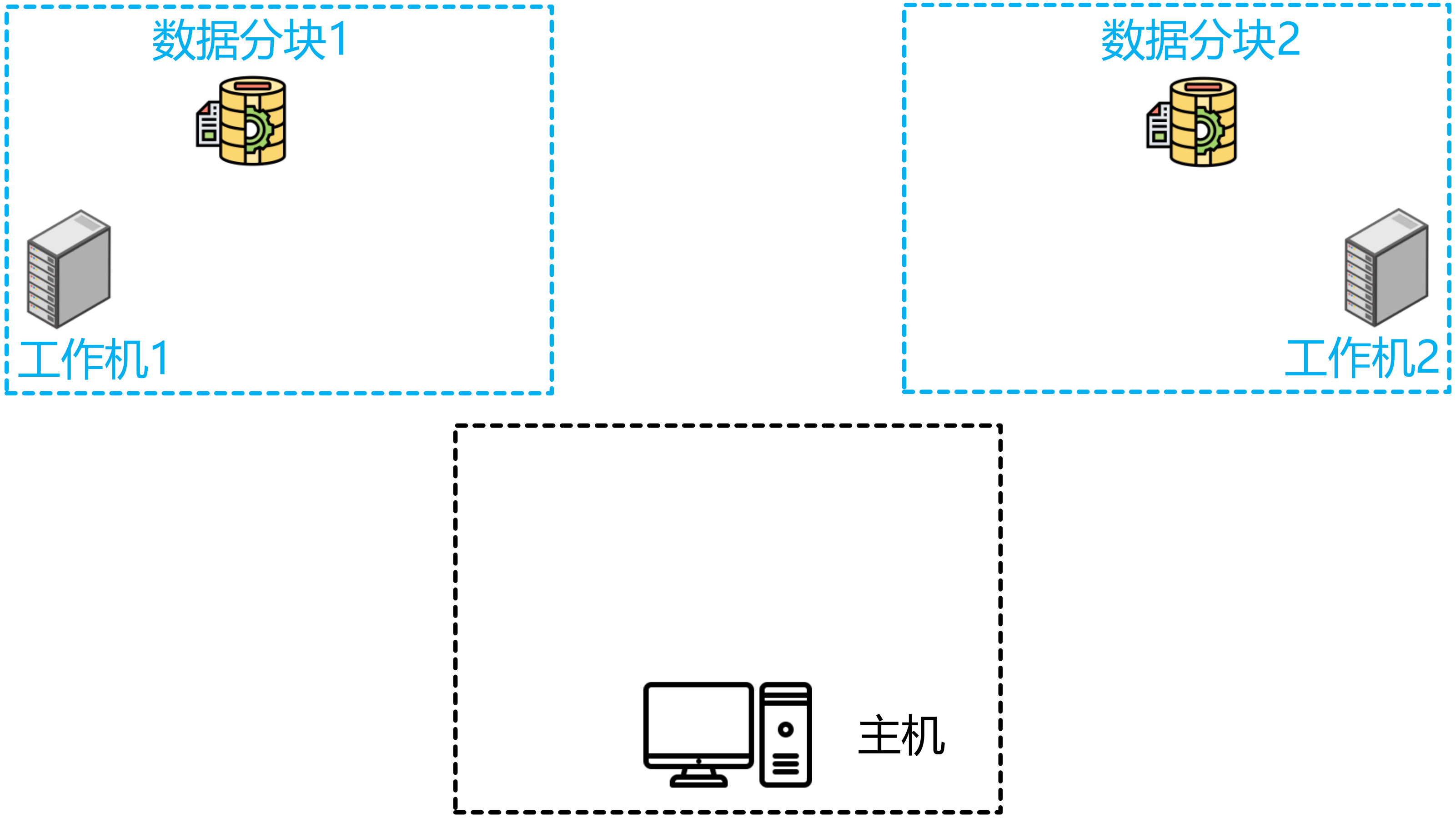
$$
\begin{aligned}
x_i^{k+1} &:= \operatorname*{argmin}_{x_i}\left(f_i(x_i) + y_i^{kT}(x_i - z^k) + (\rho/2)\|x_i - z^k\|_2^2\right)\\
z^{k+1} &:= \operatorname*{argmin}_{z}\left(g(z) + \sum_{i=1}^{N}(-y_i^{kT} z + (\rho/2)\|x_i^{k+1} - z\|_2^2)\right)\\
y_i^{k+1} &:= y_i^k + \rho(x_i^{k+1} - z^{k+1}).
\end{aligned}
$$

简化形式

$$
\begin{aligned}
x_i^{k+1} &:= \underset{x_i}{\operatorname{argmin}} \left(f_i(x_i) + (\rho/2)\|x_i - z^k + u_i^k\|_2^2 \right) \\
z^{k+1} &:= \underset{z}{\operatorname{argmin}} \left(g(z) + (N\rho/2)\|z - \overline{x}^{k+1} - \overline{u}^k\|_2^2 \right) \\
u_i^{k+1} &:= u_i^k + x_i^{k+1} - z^{k+1}.
\end{aligned}
$$

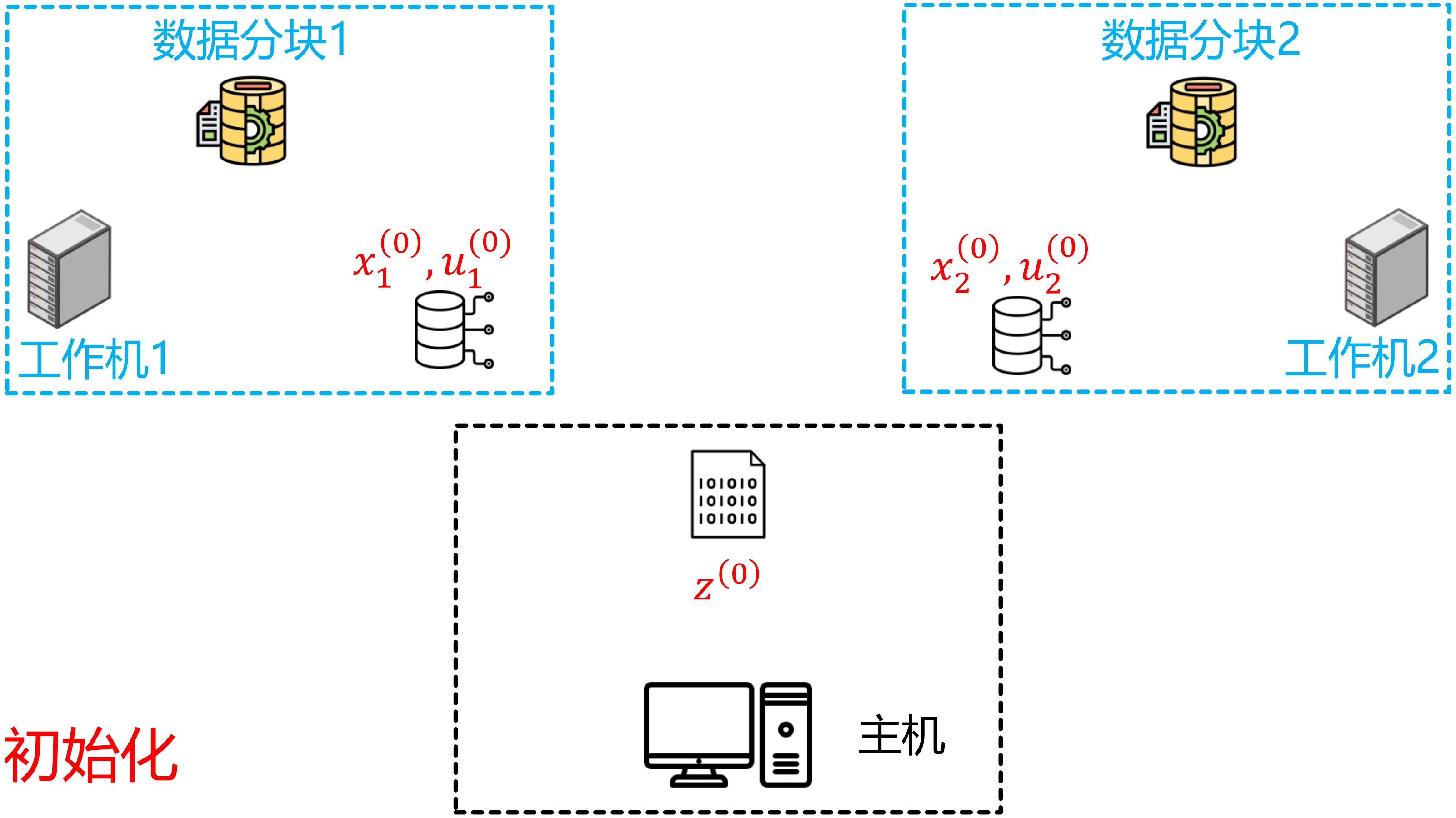
示意图

$$
\begin{aligned}
x_i^{k+1} &:= \operatorname*{argmin}_{x_i}\left(f_i(x_i) + (\rho/2)\|x_i - z^k + u_i^k\|_2^2\right)\\
z^{k+1} &:= \operatorname*{argmin}_{z}\left(g(z) + (N\rho/2)\|z - \overline{x}^{k+1} - \overline{u}^k\|_2^2\right)\\
u_i^{k+1} &:= u_i^k + x_i^{k+1} - z^{k+1}.
\end{aligned}
$$



示意图

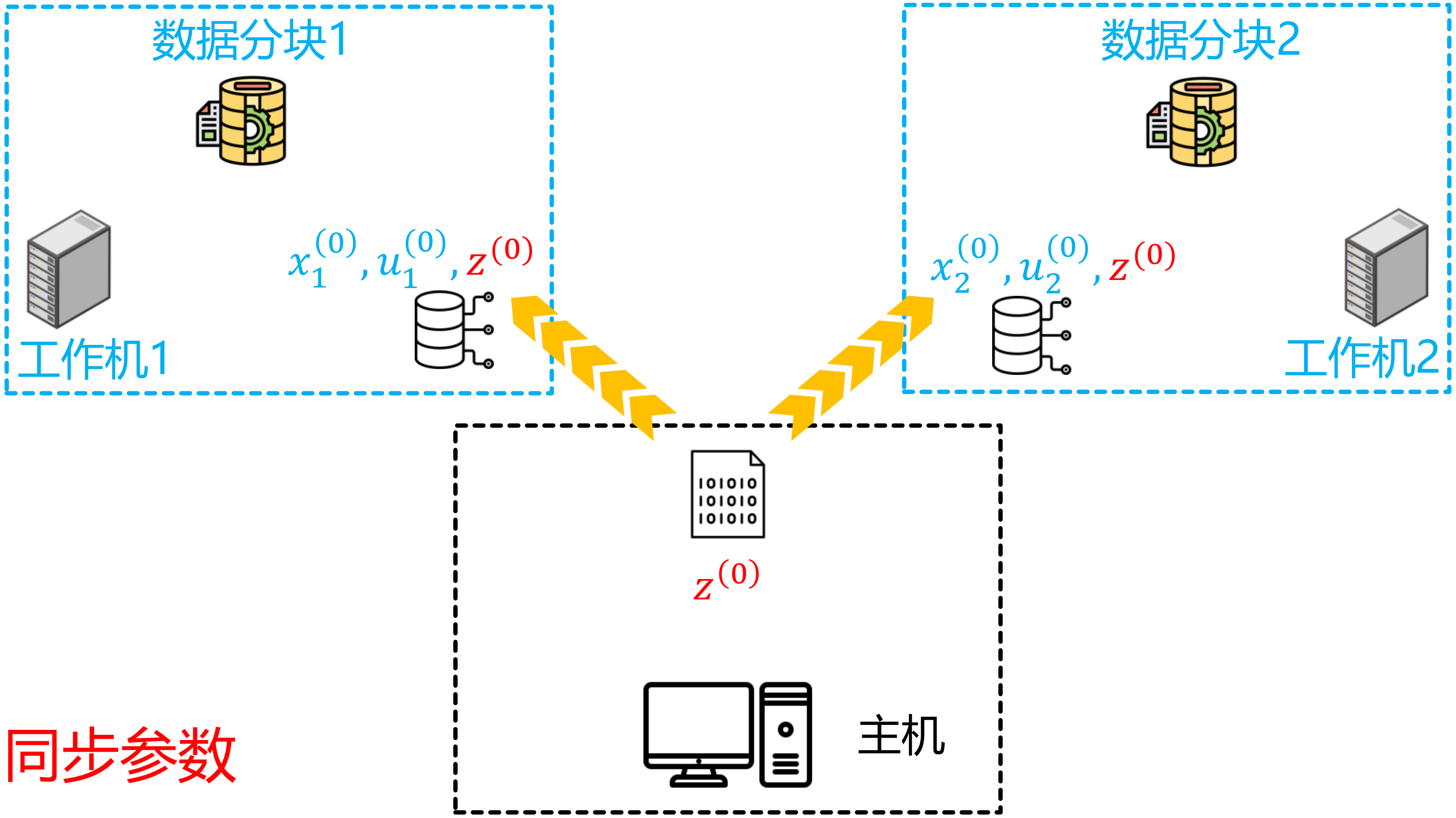
$$
\begin{aligned}
x_i^{k+1} &:= \underset{x_i}{\operatorname{argmin}} \left(f_i(x_i) + (\rho/2)\|x_i - z^k + u_i^k\|_2^2 \right) \\
z^{k+1} &:= \underset{z}{\operatorname{argmin}} \left(g(z) + (N\rho/2)\|z - \overline{x}^{k+1} - \overline{u}^k\|_2^2 \right) \\
u_i^{k+1} &:= u_i^k + x_i^{k+1} - z^{k+1}.
\end{aligned}
$$



初始化

示意图

$$
\begin{aligned}
x_i^{k+1} &:= \underset{x_i}{\operatorname{argmin}}\left(f_i(x_i) + (\rho/2)\|x_i - z^k + u_i^k\|_2^2\right)\\
z^{k+1} &:= \underset{z}{\operatorname{argmin}}\left(g(z) + (N\rho/2)\|z - \overline{x}^{k+1} - \overline{u}^k\|_2^2\right)\\
u_i^{k+1} &:= u_i^k + x_i^{k+1} - z^{k+1}.
\end{aligned}
$$

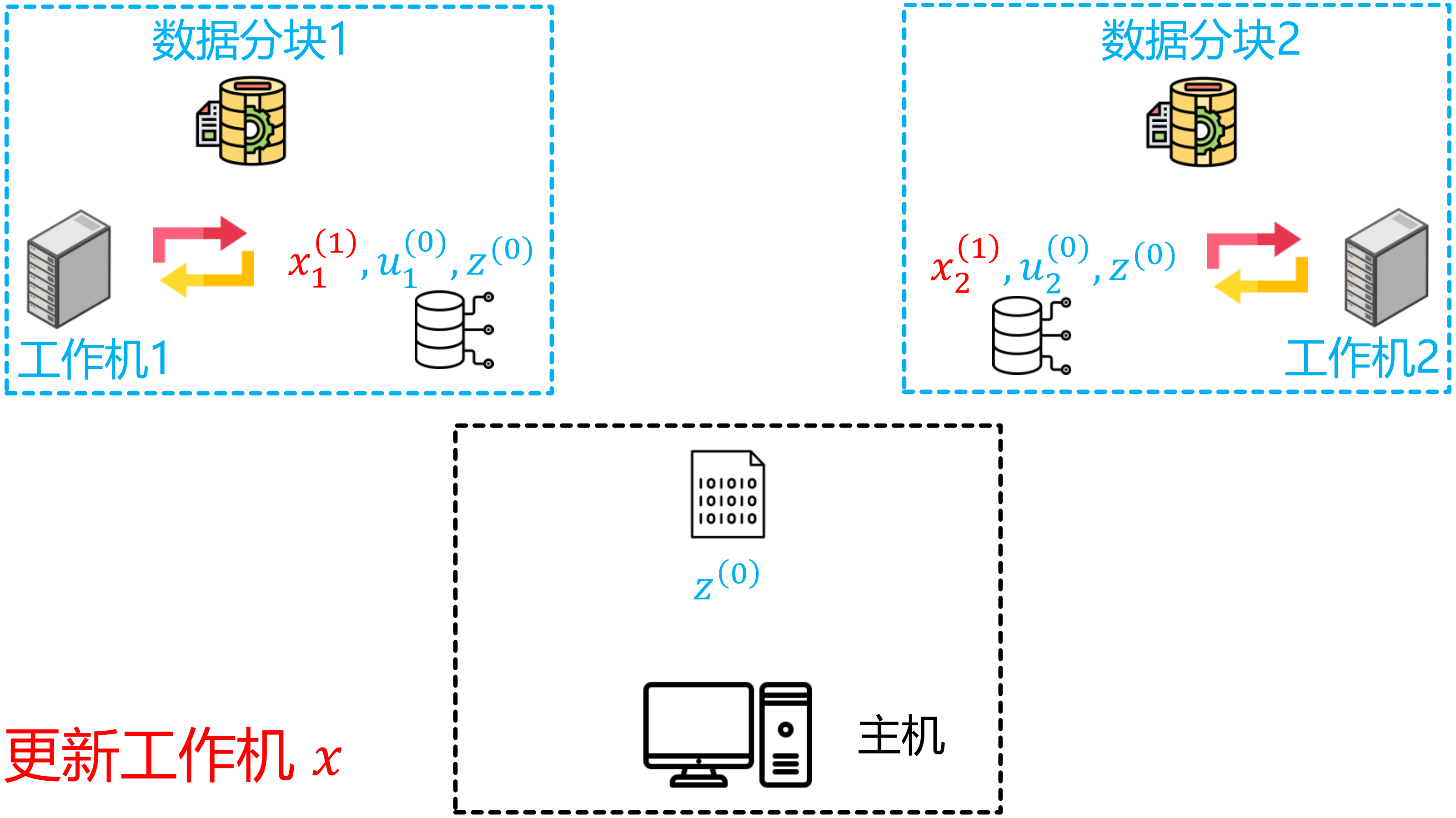


示意图

$$\boxed{x_i^{k+1} := \underset{x_i}{\operatorname{argmin}} \left(f_i(x_i) + (\rho/2)\|x_i - z^k + u_i^k\|_2^2 \right)}$$

$$z^{k+1} := \underset{z}{\operatorname{argmin}} \left(g(z) + (N\rho/2)\|z - \overline{x}^{k+1} - \overline{u}^k\|_2^2 \right)$$

$$u_i^{k+1} := u_i^k + x_i^{k+1} - z^{k+1}.$$



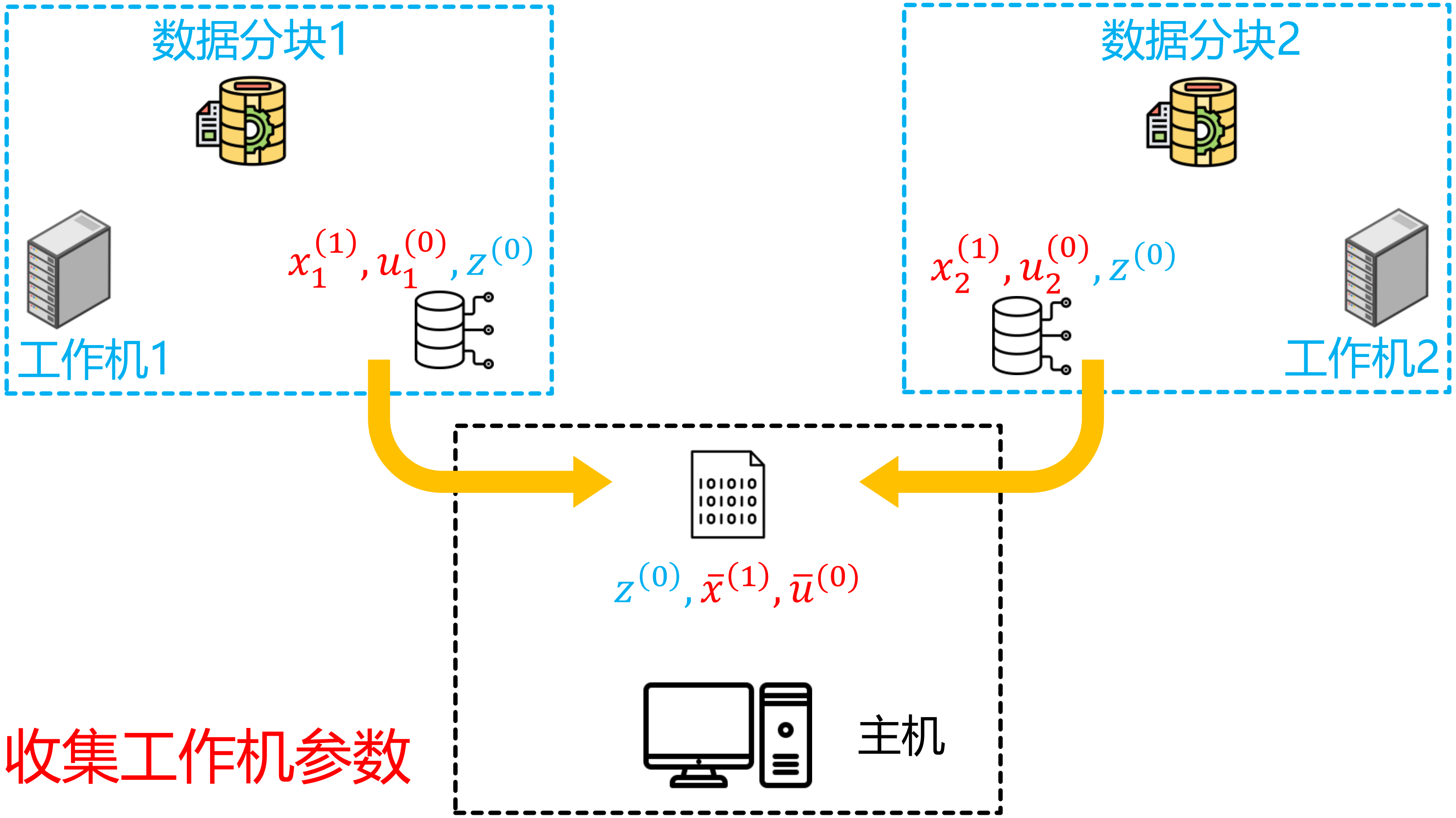
更新工作机 x

示意图

$$x_i^{k+1} := \underset{x_i}{\operatorname{argmin}}\left(f_i(x_i) + (\rho/2)\|x_i - z^k + u_i^k\|_2^2\right)$$

$$z^{k+1} := \underset{z}{\operatorname{argmin}}\left(g(z) + (N\rho/2)\|z - \overline{x}^{k+1} - \overline{u}^k\|_2^2\right)$$

$$u_i^{k+1} := u_i^k + x_i^{k+1} - z^{k+1}.$$



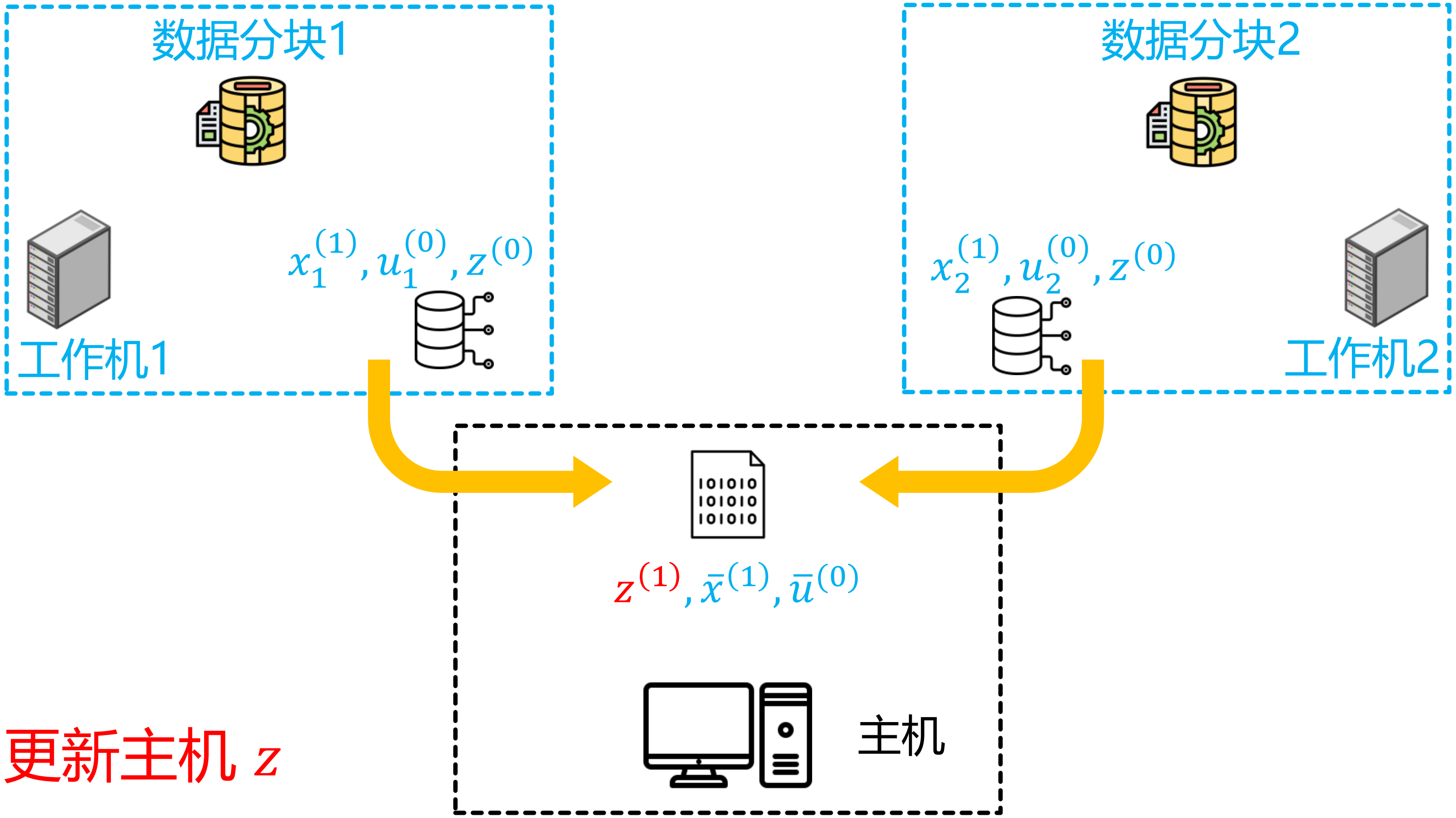
收集工作机参数

示意图

$$x_i^{k+1} := \underset{x_i}{\operatorname{argmin}} \left(f_i(x_i) + (\rho/2)\|x_i - z^k + u_i^k\|_2^2 \right)$$

$$\boxed{z^{k+1} := \underset{z}{\operatorname{argmin}} \left(g(z) + (N\rho/2)\|z - \overline{x}^{k+1} - \overline{u}^k\|_2^2 \right)}$$

$$u_i^{k+1} := u_i^k + x_i^{k+1} - z^{k+1}.$$



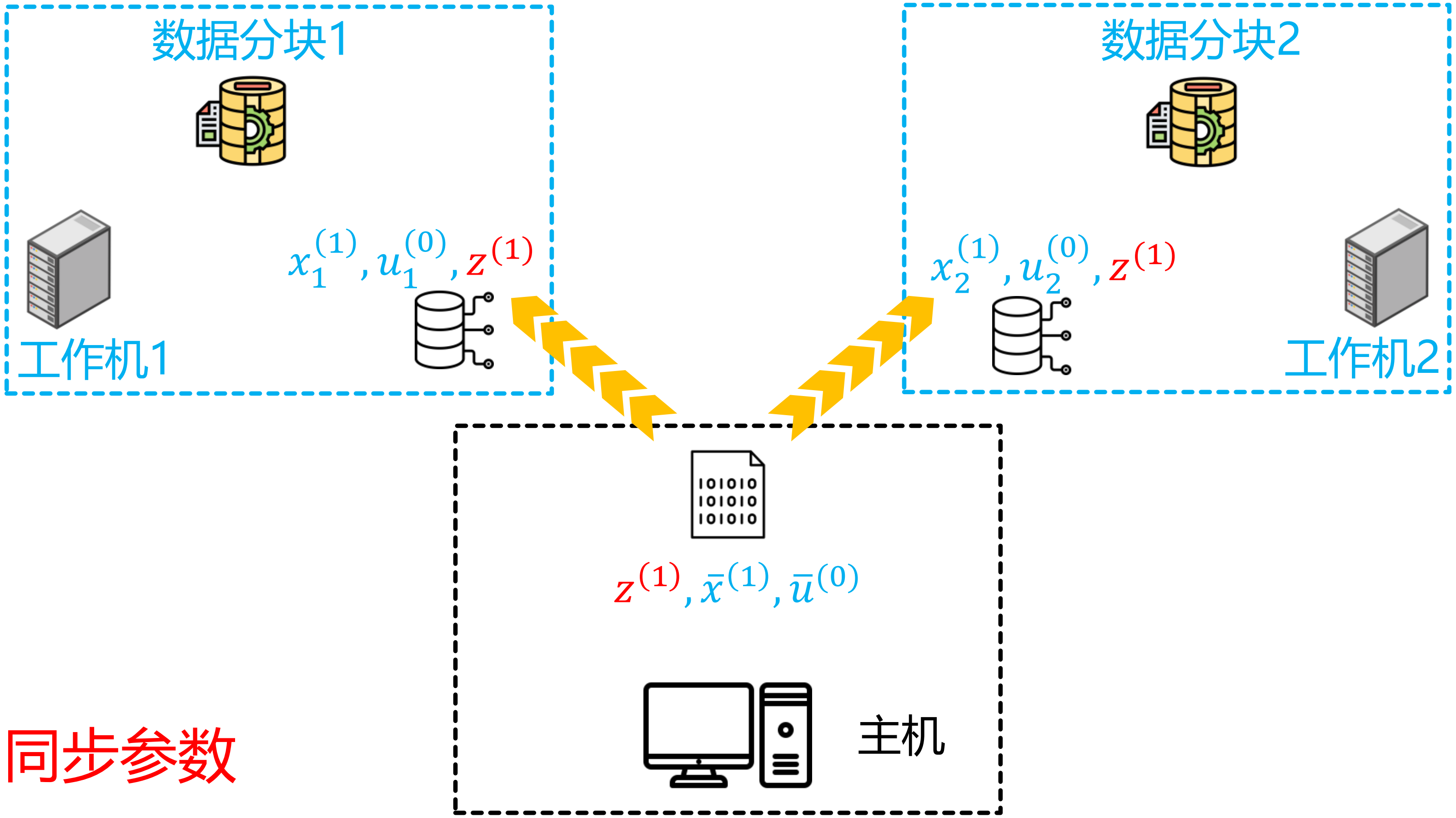
更新主机 z

示意图

$$x_i^{k+1} := \underset{x_i}{\mathrm{argmin}}\left(f_i(x_i) + (\rho/2)\|x_i - z^k + u_i^k\|_2^2\right)$$

$$z^{k+1} := \underset{z}{\mathrm{argmin}}\left(g(z) + (N\rho/2)\|z - \overline{x}^{k+1} - \overline{u}^k\|_2^2\right)$$

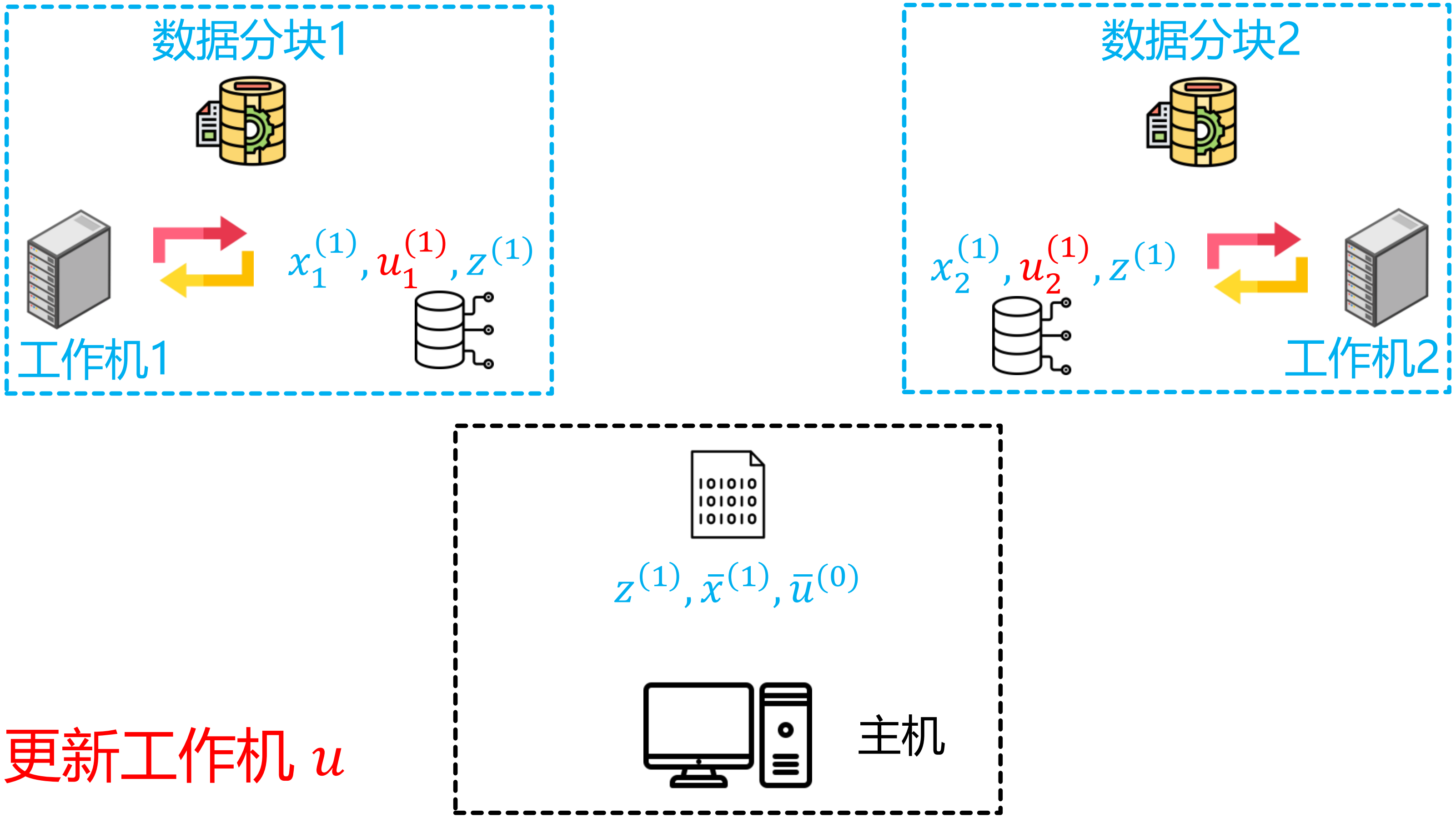
$$u_i^{k+1} := u_i^k + x_i^{k+1} - z^{k+1}.$$



同步参数

示意图

$$
\begin{aligned}
x_i^{k+1} &:= \underset{x_i}{\operatorname{argmin}} \left(f_i(x_i) + (\rho/2)\|x_i - z^k + u_i^k\|_2^2 \right) \\
z^{k+1} &:= \underset{z}{\operatorname{argmin}} \left(g(z) + (N\rho/2)\|z - \overline{x}^{k+1} - \overline{u}^k\|_2^2 \right) \\
&\boxed{u_i^{k+1} := u_i^k + x_i^{k+1} - z^{k+1}.}
\end{aligned}
$$



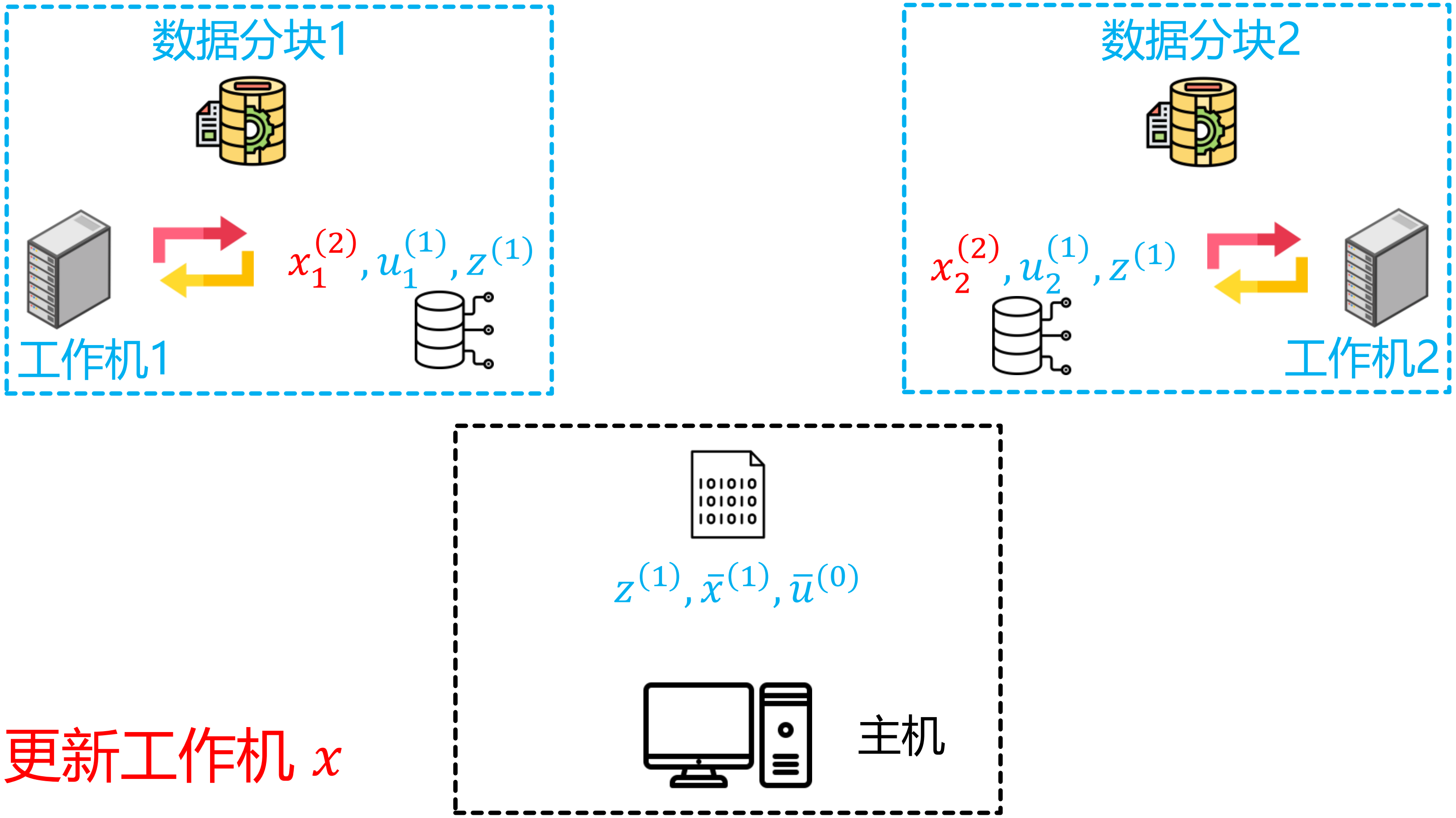
更新工作机 u

示意图

$$\boxed{x_i^{k+1} := \underset{x_i}{\operatorname{argmin}}\left(f_i(x_i) + (\rho/2)\|x_i - z^k + u_i^k\|_2^2\right)}$$

$$z^{k+1} := \underset{z}{\operatorname{argmin}}\left(g(z) + (N\rho/2)\|z - \overline{x}^{k+1} - \overline{u}^k\|_2^2\right)$$

$$u_i^{k+1} := u_i^k + x_i^{k+1} - z^{k+1}.$$



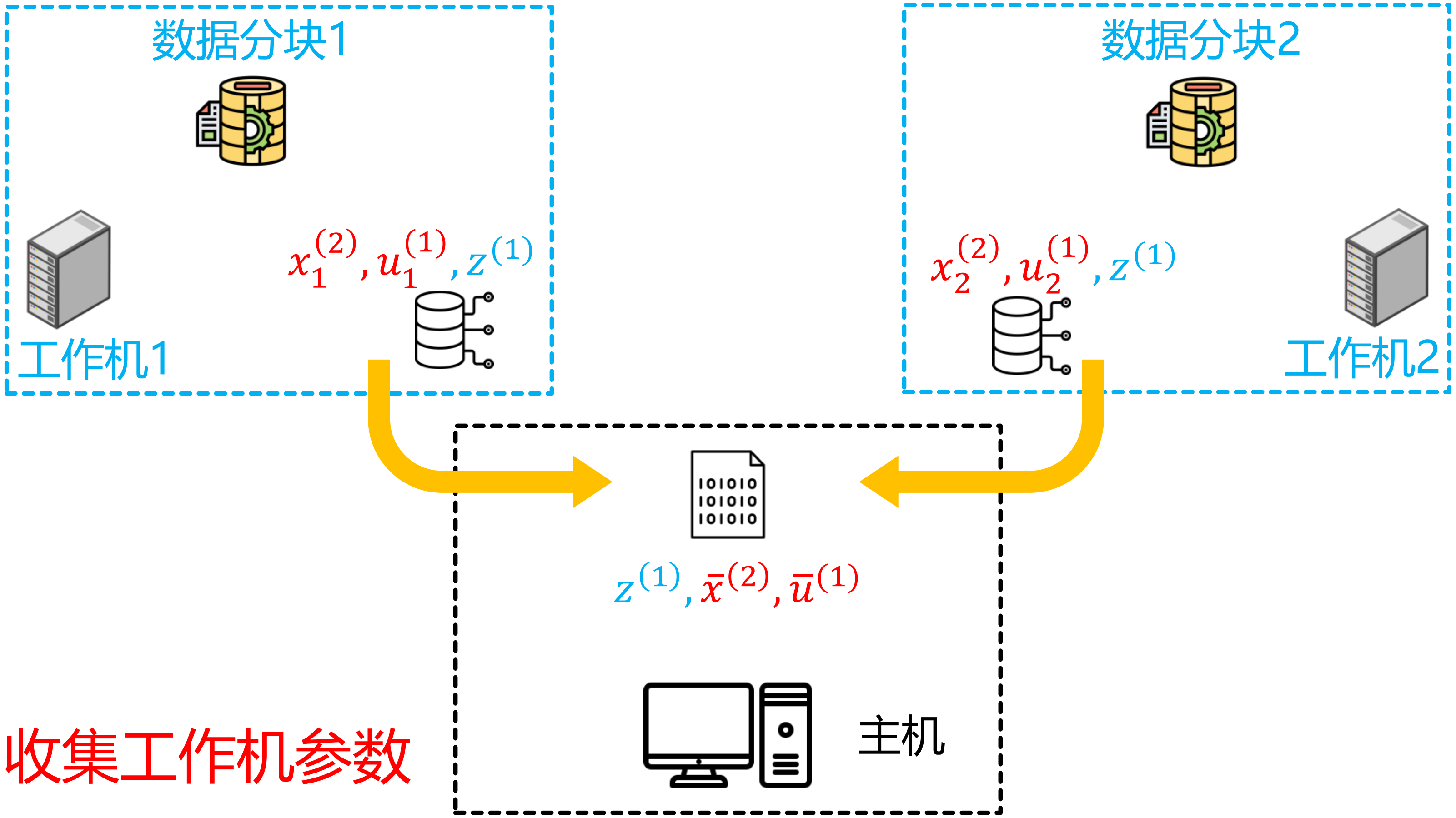
更新工作机 x

示意图

$$x_i^{k+1} := \underset{x_i}{\operatorname{argmin}}\left(f_i(x_i) + (\rho/2)\|x_i - z^k + u_i^k\|_2^2\right)$$

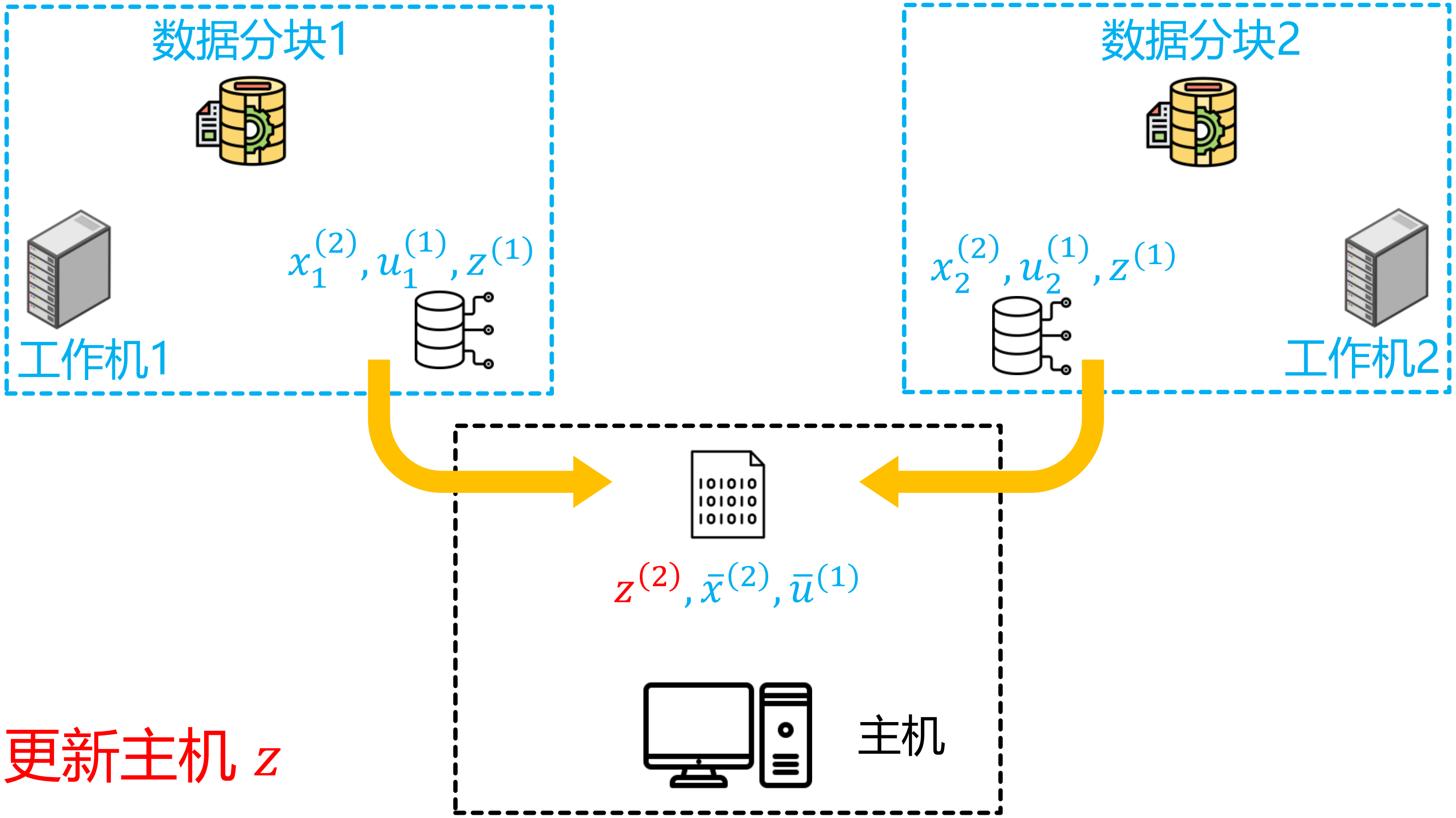
$$z^{k+1} := \underset{z}{\operatorname{argmin}}\left(g(z) + (N\rho/2)\|z - \overline{x}^{k+1} - \overline{u}^k\|_2^2\right)$$

$$u_i^{k+1} := u_i^k + x_i^{k+1} - z^{k+1}.$$



示意图

$$
\begin{aligned}
x_i^{k+1} &:= \underset{x_i}{\operatorname{argmin}} \left(f_i(x_i) + (\rho/2)\|x_i - z^k + u_i^k\|_2^2 \right) \\
&\boxed{z^{k+1} := \underset{z}{\operatorname{argmin}} \left(g(z) + (N\rho/2)\|z - \overline{x}^{k+1} - \overline{u}^k\|_2^2 \right)} \\
u_i^{k+1} &:= u_i^k + x_i^{k+1} - z^{k+1}.
\end{aligned}
$$



更新主机 z

例：Lasso

- $\min_{x} \frac{1}{2}\|Ax-b\|^2 + \lambda\|x\|_1$

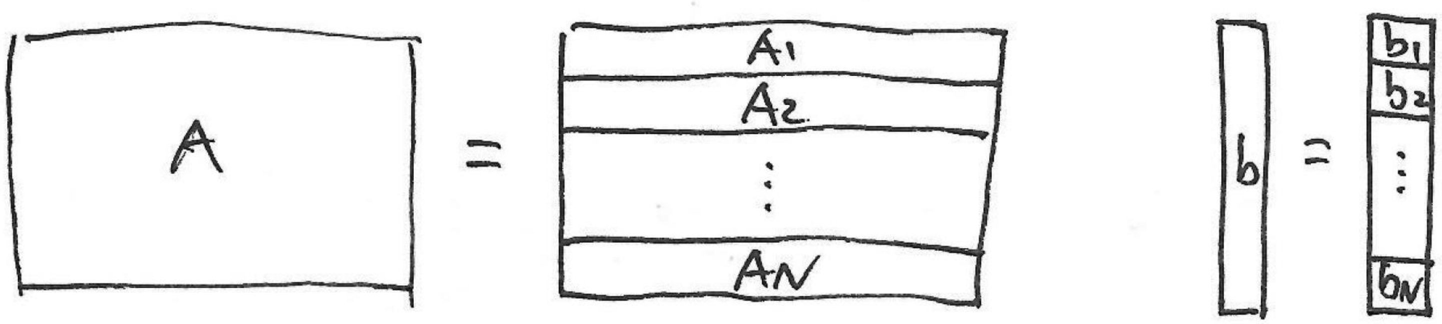
$$A = \begin{bmatrix} A_1 \\ \vdots \\ A_N \end{bmatrix}, \qquad b = \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix},$$

$$\begin{aligned} x_i^{k+1} &:= (A_i^T A_i + \rho I)^{-1}(A_i^T b_i + \rho(z^k - u_i^k)) \\ z^{k+1} &:= S_{\lambda/\rho N}(\overline{x}^{k+1} + \overline{u}^k) \\ u_i^{k+1} &:= u_i^k + x_i^{k+1} - z^{k+1} \end{aligned}$$

典型问题

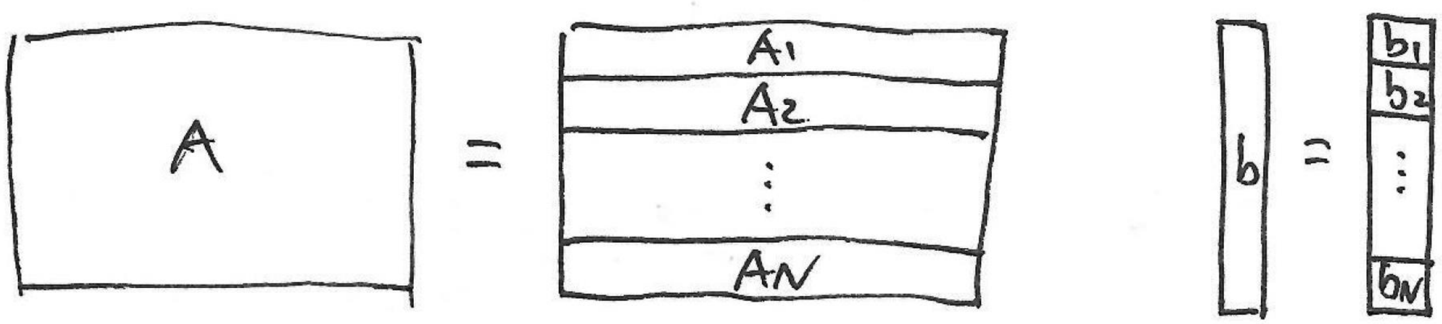
- $\min\limits_{x} \overset{\text{损失函数}}{\boxed{l(Ax-b)}} + \overset{\text{正则项}}{\boxed{r(x)}}$
- x：参数向量
- A, b：数据矩阵/向量

数据切分



- 按行切分
- 每个分块包含一部分观测
- 每个分块包含所有的变量

数据切分



- $l(Ax - b) = \sum_{i=1}^{N} l_i(Ax_i - b_i)$
- Minimize $\sum_{i=1}^{N} l_i(Ax_i - b_i) + r(z)$
- Subject to $x_i - z = 0, \ i = 1, \ldots, N$

数据切分

$$A = \begin{bmatrix} A_1 \\ \vdots \\ A_N \end{bmatrix}, \qquad b = \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix},$$

- $l(Ax - b) = \sum_{i=1}^{N} l_i(Ax_i - b_i)$
- Minimize $\sum_{i=1}^{N} l_i(Ax_i - b_i) + r(z)$
- Subject to $x_i - z = 0,\ i = 1, \ldots, N$

迭代算法

$$
\begin{aligned}
x_i^{k+1} &:= \underset{x_i}{\operatorname{argmin}} \left(l_i(A_i x_i - b_i) + (\rho/2)\|x_i - z^k + u_i^k\|_2^2 \right) \\
z^{k+1} &:= \underset{z}{\operatorname{argmin}} \left(r(z) + (N\rho/2)\|z - \overline{x}^{k+1} - \overline{u}^k\|_2^2 \right) \\
u_i^{k+1} &:= u_i^k + x_i^{k+1} - z^{k+1}.
\end{aligned}
$$

扩展阅读

- https://joegaotao.github.io/2014/02/11/admm-stat-compute/