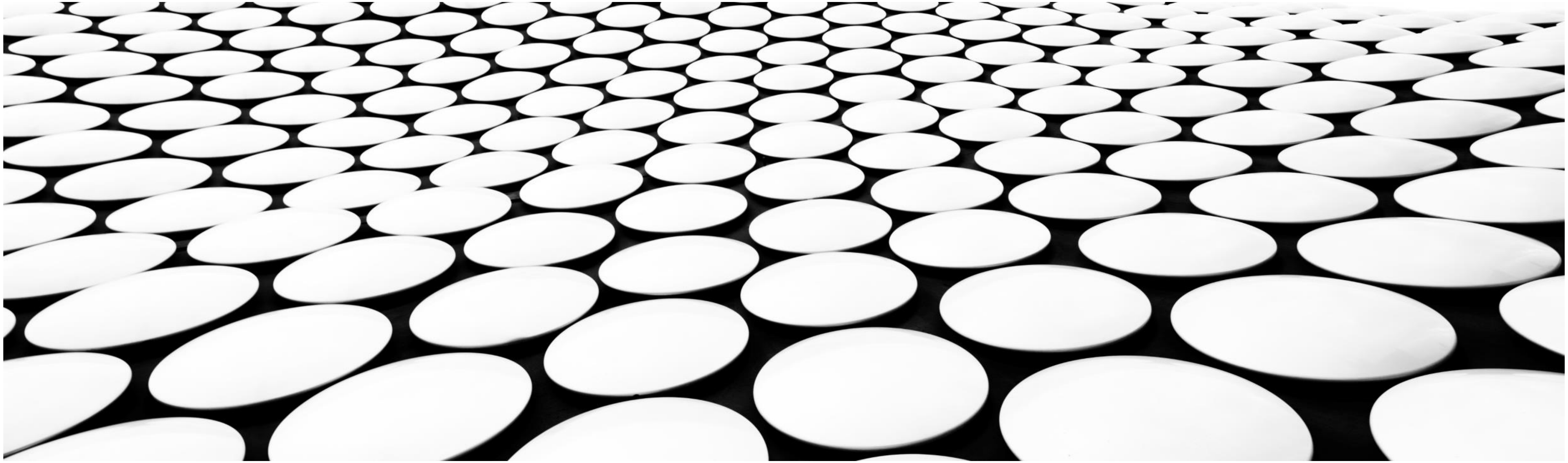
分布式计算

邱怡轩



今天的主题

- ADMM 算法（二）
- 一致性优化问题

回顾：ADMM

ADMM

- Minimize $f(x) + g(z)$
- Subject to $Ax + Bz = c$

- x: n 维向量
- z: m 维向量
- $A\ [p \times n]$，$B\ [p \times m]$，$c\ [p \times 1]$：约束条件
- $f(\cdot)$，$g(\cdot)$：凸函数

ADMM 算法

- $x^{k+1} = \text{argmin}_x \ f(x) + \frac{\rho}{2}\|Ax + Bz^k - c + u^k\|^2$
- $z^{k+1} = \text{argmin}_z \ \ g(z) + \frac{\rho}{2}\|Ax^{k+1} + Bz - c + u^k\|^2$
- $u^{k+1} = u^k + Ax^{k+1} + Bz^{k+1} - c$

例：LAD

- 原问题： $\min_{x} \|Ax - b\|_1$

- ADMM形式：
- Minimize $\|z\|_1$
- Subject to $Ax - z = b$

例：LAD

- $x^{k+1} = (A'A)^{-1}A'(b + z^k - u^k)$
- $z^{k+1} = S_{1/\rho}\,(Ax^{k+1} - b + u^k)$
- $u^{k+1} = u^k + Ax^{k+1} - z^{k+1} - b$

- $S_\kappa(a)$ 称为 Soft-thresholding 运算符

$$S_\kappa(a) = \begin{cases} a - \kappa & a > \kappa \\ 0 & |a| \le \kappa \\ a + \kappa & a < -\kappa, \end{cases}$$

例：Lasso

- 原问题： $\min\limits_{x} \frac{1}{2}\|Ax-b\|^2+\lambda\|x\|_1$

- ADMM形式：
- Minimize $\frac{1}{2}\|Ax-b\|^2+\lambda\|z\|_1$
- Subject to $x-z=0$

例：Lasso

- $x^{k+1} = (A'A + \rho I)^{-1}(A'b + \rho\left(z^k - u^k\right))$
- $z^{k+1} = S_{\lambda/\rho}\left(x^{k+1} + u^k\right)$
- $u^{k+1} = u^k + x^{k+1} - z^{k+1}$

实现

- 参见 `lec12-admm-lad.ipynb`
- 练习 `lec12-admm-lasso.ipynb`

适用范围

- “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

迭代算法

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

$$z^{k+1} := \frac{1}{N}\sum_{i=1}^{N}\left(x_i^{k+1} + (1/\rho)y_i^k \right)$$

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

迭代算法

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

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

意义

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

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

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

- 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）

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

迭代算法

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

简化形式

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

例：Lasso

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

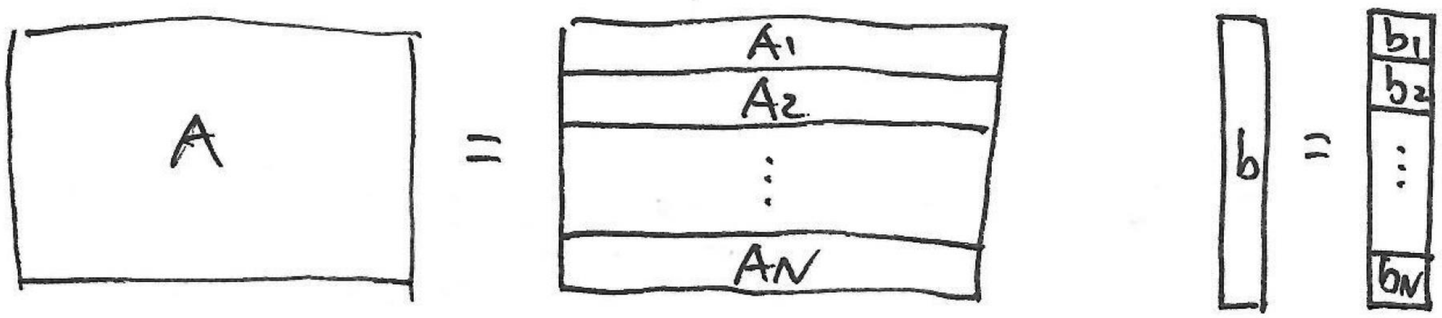
$$A=\begin{bmatrix} A_1 \\ \vdots \\ A_N \end{bmatrix}, \qquad b=\begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix},$$

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

典型问题

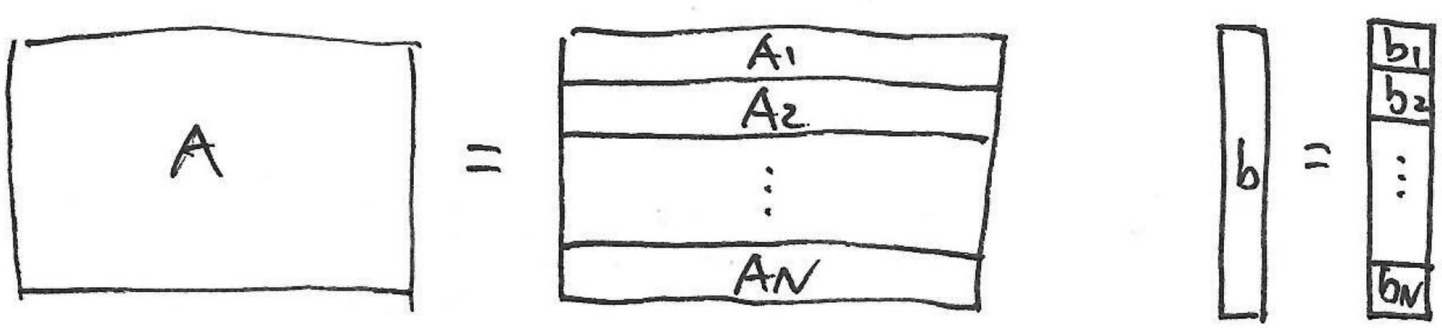
- $\min\limits_{x} \underbrace{\boxed{l(Ax-b)}}_{\text{损失函数}} + \underbrace{\boxed{r(x)}}_{\text{正则项}}$
- x：参数向量
- A, b：数据矩阵/向量

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 $x_i - z = 0,\ i=1,\dots,N$

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/