第28讲

交替极小化方法

教学提纲

- 交替极小化方法
- 凸优化交替极小化方法
- 二分快交替极小化方法
- 二分快线性化交替极小化方法

一、交替极小化方法

根据优化问题的结构将变量分块，然后依次对每组变量进行求解，通过“化整为零、各个击破”战术达到降低问题规模、简化问题难度、提高算法效率的目的.

$$\min_{\boldsymbol{x}_1\in\mathbb{R}^{n_1},\boldsymbol{x}_2\in\mathbb{R}^{n_2},\cdots,\boldsymbol{x}_s\in\mathbb{R}^{n_s}} \Psi(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_s)$$

以循环方式交替对 s 个模块变量求最小.

对迭代点 $\boldsymbol{x}^k=(\boldsymbol{x}_1^k,\boldsymbol{x}_2^k,\cdots,\boldsymbol{x}_s^k)$, 分别基于模块变量依次对 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_s$ 求极小，得到新的迭代点

$$\boldsymbol{x}^{k,1}=(\boldsymbol{x}_1^{k+1},\boldsymbol{x}_2^k,\cdots,\boldsymbol{x}_s^k),$$

$$\boldsymbol{x}^{k,2}=(\boldsymbol{x}_1^{k+1},\boldsymbol{x}_2^{k+1},\boldsymbol{x}_3^k,\cdots,\boldsymbol{x}_s^k),$$

$$\vdots$$

$$\boldsymbol{x}^{k,i}=(\boldsymbol{x}_1^{k+1},\boldsymbol{x}_2^{k+1},\cdots,\boldsymbol{x}_i^{k+1},\boldsymbol{x}_{i+1}^k,\cdots,\boldsymbol{x}_s^k),$$

$$\vdots$$

$$\boldsymbol{x}^{k,s}=\boldsymbol{x}^{k+1}=(\boldsymbol{x}_1^{k+1},\boldsymbol{x}_2^{k+1},\cdots,\boldsymbol{x}_s^{k+1}).$$

交替极小化算法

$$\min_{\boldsymbol{x}_1\in\mathbb{R}^{n_1},\boldsymbol{x}_2\in\mathbb{R}^{n_2},\cdots,\boldsymbol{x}_s\in\mathbb{R}^{n_s}} \Psi(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_s)$$

初始步: 取 $\boldsymbol{x}^0=(\boldsymbol{x}_1^0,\boldsymbol{x}_2^0,\cdots,\boldsymbol{x}_s^0)$, $k=0$;

迭代步: 对 $i=1,2,\cdots,s$, 依次求解子问题

$$\boldsymbol{x}_i^{k+1}=\arg\min_{\boldsymbol{x}_i\in\mathbb{R}^{n_i}} \Psi(\boldsymbol{x}_1^{k+1},\boldsymbol{x}_2^{k+1},\cdots,\boldsymbol{x}_{i-1}^{k+1},\boldsymbol{x}_i,\boldsymbol{x}_{i+1}^{k},\cdots,\boldsymbol{x}_s^{k})$$

定理 设目标函数 $\Psi:\mathbb{R}^n\to\mathbb{R}$ 连续可微, 目标函数水平集有界

且对任意的 $\boldsymbol{x}\in\mathbb{R}^n$ 和 $i\in\{1,2,\cdots,s\}$,

子问题 $\min_{\boldsymbol{y}\in\mathbb{R}^{n_i}}\Psi(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_{i-1},\boldsymbol{y},\boldsymbol{x}_{i+1},\cdots,\boldsymbol{x}_s)$ 有唯一最优解.

则算法产生迭代点列的任一聚点为优化问题的稳定点.

反例:连续不可微 $\min \Psi(x_1, x_2) = |3x_1 + 4x_2| + |x_1 - 2x_2|$

连续凸函数，水平集有界，且对任一分量有唯一最优解.

对任意 $\alpha > 0$,

$$\Psi(-4\alpha, t) = |4t - 12\alpha| + |2t + 4\alpha|$$
$$= \begin{cases} -6t + 8\alpha, & t < -2\alpha, \\ -2t + 16\alpha, & -2\alpha \leqslant t \leqslant 3\alpha, \\ 6t - 8\alpha, & t > 3\alpha, \end{cases}$$

最优解
$t = 3\alpha$

$$\Psi(t, 3\alpha) = |3t + 12\alpha| + |t - 6\alpha|$$
$$= \begin{cases} -4t - 6\alpha, & t < -4\alpha, \\ 2t + 18\alpha, & -4\alpha \leqslant t \leqslant 6\alpha, \\ 4t + 6\alpha, & t > 6\alpha, \end{cases}$$

最优解
$t = -4\alpha$

对任意的 $\alpha \leqslant 0$,

$$-4\alpha = \arg\min_{x_1 \in R} \Psi(x_1, 3\alpha),$$

$$3\alpha = \arg\min_{x_2 \in \mathbb{R}} \Psi(-4\alpha, x_2)$$

交替极小化方法 $\min \Psi(x_1, x_2) = |3x_1 + 4x_2| + |x_1 - 2x_2|$

聚点

若 x_1 非零，则在首次迭代后，算法滞留在 $(-4\alpha, 3\alpha)$ 点.

$\boldsymbol{x} = \mathbf{0}$ 为函数 $\Psi(x_1, x_2)$ 的唯一最小值点，$(-4\alpha,\ 3\alpha)$ 既不是该函数的最小值点，也不是其稳定点。

那它是什么性质的点呢？

定义：坐标轮换最小值点

若 $\boldsymbol{x}^* \in \mathbb{R}^n$ 满足

$$\Psi(\boldsymbol{x}^*) \leqslant \Psi(\boldsymbol{x}_1^*, \cdots, \boldsymbol{x}_{i-1}^*, \boldsymbol{x}_i, \boldsymbol{x}_{i+1}^*, \cdots, \boldsymbol{x}_s^*), \quad \forall i = 1, 2, \cdots, s, \ \boldsymbol{x}_i \in \mathbb{R}^{n_i},$$

则称 $\boldsymbol{x}^*$ 为函数 $\Psi(\boldsymbol{x})$ 的坐标轮换最小值点

连续不可微情形下的算法收敛性

定理：设分块优化问题 $\min\limits_{\boldsymbol{x}_1 \in \mathbb{R}^{n_1}, \boldsymbol{x}_2 \in \mathbb{R}^{n_2}, \cdots, \boldsymbol{x}_s \in \mathbb{R}^{n_s}} \Psi(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_s)$

目标函数下半连续，水平集有界，子问题

$$\min_{\boldsymbol{y} \in \mathbb{R}^{n_i}} \Psi(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{i-1}, \boldsymbol{y}, \boldsymbol{x}_{i+1}, \cdots, \boldsymbol{x}_s)$$

有唯一最优解。则迭代点列的任一聚点为坐标轮换最小值点.

由例 $\min \Psi(x_1, x_2) = |3x_1 + 4x_2| + |x_1 - 2x_2|$

连续不可微函数的坐标轮换极小值点未必是其最小值点或稳定点.

那么，在什么情况下是呢?

定理 对优化问题 $\min\limits_{\boldsymbol{x} \in \mathbb{R}^n} \Psi(\boldsymbol{x}) = f(\boldsymbol{x}) + g(\boldsymbol{x})$

$f : \mathbb{R}^n \to \mathbb{R}$ 连续可微, $g(\boldsymbol{x}) = \sum\limits_{i=1}^{s} g_i(\boldsymbol{x}_i)$, $g_i : \mathbb{R}^{n_i} \to \mathbb{R}$ 连续凸函数

坐标轮换极小值 $\Longleftrightarrow$ 稳定点

对分块和式优化问题，若子问题最优解不唯一,

算法收敛性不能保证。

反例：子问题最优解不唯一

$$f(x,y,z) = -xy - yz - zx + [x-1]_+^2 + [-x-1]_+^2 + [y-1]_+^2 + [-y-1]_+^2 + [z-1]_+^2 + [-z-1]_+^2.$$

连续可微

依次两两固定 y, z, x

$$\arg\min_x f(x,y,z) = \begin{cases} \operatorname{sgn}(y+z)(1+\frac{1}{2}|y+z|), & y+z \neq 0, \\ [-1,1], & y+z = 0. \end{cases}$$

$$\arg\min_y f(x,y,z) = \begin{cases} \operatorname{sgn}(x+z)(1+\frac{1}{2}|x+z|), & x+z \neq 0, \\ [-1,1], & x+z = 0, \end{cases}$$

$$\arg\min_z f(x,y,z) = \begin{cases} \operatorname{sgn}(x+y)(1+\frac{1}{2}|x+y|), & x+y \neq 0, \\ [-1,1], & x+y = 0. \end{cases}$$

取 $\varepsilon>0$, 以 $(-1-\varepsilon;1+\frac{1}{2}\varepsilon;-1-\frac{1}{4}\varepsilon)$ 为初始点，

用交替极小化方法求解，迭代6次

$$(1+\frac{1}{8}\varepsilon;1+\frac{1}{2}\varepsilon;-1-\frac{1}{4}\varepsilon),\quad (1+\frac{1}{8}\varepsilon;-1-\frac{1}{16}\varepsilon;-1-\frac{1}{4}\varepsilon),$$

$$(1+\frac{1}{8}\varepsilon;-1-\frac{1}{16}\varepsilon;1+\frac{1}{32}\varepsilon),\quad (-1-\frac{1}{64}\varepsilon;-1-\frac{1}{16}\varepsilon;1+\frac{1}{32}\varepsilon),$$

$$(-1-\frac{1}{64}\varepsilon;1+\frac{1}{128}\varepsilon;1+\frac{1}{32}\varepsilon),\quad (-1-\frac{1}{64}\varepsilon;1+\frac{1}{128}\varepsilon;-1-\frac{1}{256}\varepsilon).$$

将初始点中的 ε 换成 $\frac{1}{64}\varepsilon$，则算法产生的迭代点列围绕如下6点循环

$$\boxed{\begin{array}{ccc}(1;1;-1), & (1;-1;-1), & (1;-1;1),\\ (-1;-1;1), & (-1;1;1), & (-1;1;-1).\end{array}}$$

聚点但非稳定点

非坐标轮换最小值点

二、和式凸优化交替极小化方法

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \Psi(\boldsymbol{x}) = f(\boldsymbol{x}) + g(\boldsymbol{x})$$

$f:\mathbb{R}^n \to \mathbb{R}$ 连续可微的凸函数

$g(\boldsymbol{x}) = \sum_{i=1}^{s} g_i(\boldsymbol{x}_i),\ g_i:\mathbb{R}^{n_i} \to \mathbb{R}$ 连续凸函数

对该优化问题，无需子问题最优解唯一，就能建立算法的全局收敛性.

定理： 对上述优化问题，若目标函数水平集有界，

则交替极小化方法产生迭代点列的任一聚点为问题的最优解.

定理 设 $\min_{\boldsymbol{x}\in\mathbb{R}^n} \Psi(\boldsymbol{x}) = f(\boldsymbol{x}) + g(\boldsymbol{x})$

$f:\mathbb{R}^n \to \mathbb{R}$ 连续可微的凸函数，梯度函数Lipschitz连续，常数为 L_f

$g(\boldsymbol{x}) = \sum_{i=1}^{s} g_i(\boldsymbol{x}_i),\ g_i:\mathbb{R}^{n_i} \to \mathbb{R}$ 下半连续的凸函数

目标函数水平集有界。

则算法产生的迭代点列满足

目标函数线性下降

$$\Psi(\boldsymbol{x}^k) - \Psi^* \leqslant \max\left\{\left(\frac{1}{2}\right)^{(k-1)/2}(\Psi(\boldsymbol{x}^0) - \Psi^*), \frac{8s^2 L_f R^2_{\Psi(\boldsymbol{x}^0)}}{k-1}\right\},$$

其中 $R_{\Psi(\boldsymbol{x}^0)} = \max\{\|\boldsymbol{x} - \boldsymbol{x}^*\| \mid \Psi(\boldsymbol{x}) \leqslant \Psi(\boldsymbol{x}^0)\}.$

三、二分块交替极小化方法

$$\min_{\boldsymbol{x}_1\in\mathbb{R}^{n_1},\boldsymbol{x}_2\in\mathbb{R}^{n_2}}\{\Psi(\boldsymbol{x}_1,\boldsymbol{x}_2)=f(\boldsymbol{x}_1,\boldsymbol{x}_2)+g_1(\boldsymbol{x}_1)+g_2(\boldsymbol{x}_2)\}$$

$f:\mathbb{R}^n\to\mathbb{R}$ 连续可微的凸函数

$\nabla_i f(\boldsymbol{x})$ Lipschitz连续，常数为 L_i, $i=1,2$

$g_i:\mathbb{R}^{n_i}\to\mathbb{R}$ 下半连续的凸函数

二分块交替极小化算法

初始步: 取 $\boldsymbol{x}_1^0 \in \mathbb{R}^{n_1}$，计算 $\boldsymbol{x}_2^0 = \arg\min_{\boldsymbol{x}_2} f(\boldsymbol{x}_1^0, \boldsymbol{x}_2) + g_2(\boldsymbol{x}_2)$

迭代步: 依次计算

$$
\begin{aligned}
\boldsymbol{x}_1^{k+1} &= \arg\min_{\boldsymbol{x}_1} f(\boldsymbol{x}_1, \boldsymbol{x}_2^k) + g_1(\boldsymbol{x}_1), \\
\boldsymbol{x}_2^{k+1} &= \arg\min_{\boldsymbol{x}_2} f(\boldsymbol{x}_1^{k+1}, \boldsymbol{x}_2) + g_2(\boldsymbol{x}_2).
\end{aligned}
$$

定理 $$\min_{\boldsymbol{x}_1\in\mathbb{R}^{n_1},\boldsymbol{x}_2\in\mathbb{R}^{n_2}}\{\Psi(\boldsymbol{x}_1,\boldsymbol{x}_2)=f(\boldsymbol{x}_1,\boldsymbol{x}_2)+g_1(\boldsymbol{x}_1)+g_2(\boldsymbol{x}_2)\}$$

$f:\mathbb{R}^n\to\mathbb{R}$ 连续可微、凸函数

$\nabla_i f(\boldsymbol{x})$ Lipschitz连续，常数为 $L_i,\ i=1,2$

$g_i:\mathbb{R}^{n_i}\to\mathbb{R}$下半连续、凸函数，

目标函数水平集有界。

则算法产生的迭代点列满足

$$\Psi(\boldsymbol{x}^k)-\Psi^*\leqslant\max\left\{\left(\frac{1}{2}\right)^{(k-1)/2}(\Psi(\boldsymbol{x}^0)-\Psi^*),\frac{8\min\{L_1,L_2\}R^2_{\Psi(\boldsymbol{x}^0)}}{k-1}\right\}.$$

其中 $R_{\Psi(\boldsymbol{x}^0)}=\max\{\|\boldsymbol{x}-\boldsymbol{x}^*\|\mid\Psi(\boldsymbol{x})\leqslant\Psi(\boldsymbol{x}^0)\}.$

与多分块收敛性比较

定理 $\min_{\boldsymbol{x}\in\mathbb{R}^n} \Psi(\boldsymbol{x}) = f(\boldsymbol{x}) + g(\boldsymbol{x})$

$f:\mathbb{R}^n \to \mathbb{R}$ 梯度函数**Lipschitz**连续，常数为 L_f

$g(\boldsymbol{x}) = \sum_{i=1}^{s} g_i(\boldsymbol{x}_i),\ g_i:\mathbb{R}^{n_i}\to\mathbb{R}$ 下半连续、凸函数，……

则 $\boxed{\Psi(\boldsymbol{x}^k) - \Psi^* \leqslant \max\left\{\left(\frac{1}{2}\right)^{(k-1)/2}(\Psi(\boldsymbol{x}^0) - \Psi^*), \frac{8s^2 \boxed{L_f} R^2_{\Psi(\boldsymbol{x}^0)}}{k-1}\right\}}$

定理 $\min_{\boldsymbol{x}_1\in\mathbb{R}^{n_1},\boldsymbol{x}_2\in\mathbb{R}^{n_2}} \{\Psi(\boldsymbol{x}_1,\boldsymbol{x}_2) = f(\boldsymbol{x}_1,\boldsymbol{x}_2) + g_1(\boldsymbol{x}_1) + g_2(\boldsymbol{x}_2)\}$

$\nabla_i f(\boldsymbol{x})$ Lipschitz连续，常数为 $L_i,\ i=1,2$

$g_i:\mathbb{R}^{n_i}\to\mathbb{R}$ 下半连续、凸函数，……

则 $\boxed{\Psi(\boldsymbol{x}^k) - \Psi^* \leqslant \max\left\{\left(\frac{1}{2}\right)^{(k-1)/2}(\Psi(\boldsymbol{x}^0) - \Psi^*), \frac{8\boxed{\min\{L_1,L_2\}} R^2_{\Psi(\boldsymbol{x}^0)}}{k-1}\right\}}$

四、二分块线性化交替极小化方法

$$\min_{\boldsymbol{x}_1\in\mathbb{R}^{n_1},\boldsymbol{x}_2\in\mathbb{R}^{n_2}}\{\Psi(\boldsymbol{x}_1,\boldsymbol{x}_2)=f(\boldsymbol{x}_1,\boldsymbol{x}_2)+g_1(\boldsymbol{x}_1)+g_2(\boldsymbol{x}_2)\}$$

$f:\mathbb{R}^n\to\mathbb{R}$ 连续可微，

$\nabla_i f(\boldsymbol{x})$ 一致Lipschitz连续，常数为 $L_i,\ i=1,2$

$$\|\nabla_{\boldsymbol{x}_1}f(\boldsymbol{y},\boldsymbol{x}_2)-\nabla_{\boldsymbol{x}_1}f(\boldsymbol{z},\boldsymbol{x}_2)\|\leqslant L_1\|\boldsymbol{y}-\boldsymbol{z}\|,\quad \forall\boldsymbol{y},\boldsymbol{z}\in\mathbb{R}^{n_1},\boldsymbol{x}_2\in\mathbb{R}^{n_2}$$

$$\|\nabla_{\boldsymbol{x}_2}f(\boldsymbol{x}_1,\boldsymbol{y})-\nabla_{\boldsymbol{x}_2}f(\boldsymbol{x}_1,\boldsymbol{z})\|\leqslant L_2\|\boldsymbol{y}-\boldsymbol{z}\|,\quad \forall\boldsymbol{y},\boldsymbol{z}\in\mathbb{R}^{n_2}.\boldsymbol{x}_1\in\mathbb{R}^{n_1}$$

$g_i:\mathbb{R}^{n_i}\to\mathbb{R}$ 下半连续

问题求解 $\min\limits_{\boldsymbol{x}_1\in\mathbb{R}^{n_1},\boldsymbol{x}_2\in\mathbb{R}^{n_2}} \Psi(\boldsymbol{x}_1,\boldsymbol{x}_2)=f(\boldsymbol{x}_1,\boldsymbol{x}_2)+g_1(\boldsymbol{x}_1)+g_2(\boldsymbol{x}_2)$

将函数 $f(\boldsymbol{x}_1,\boldsymbol{x}_2)$ 在 $(\boldsymbol{x}_1^k,\boldsymbol{x}_2^k)$ 点关于 $\boldsymbol{x}_1$ 线性化，再添加临近点正则项, 关于 $\boldsymbol{x}_1$ 求最小得 $\boldsymbol{x}_1^{k+1}$；

再将该函数在 $(\boldsymbol{x}_1^{k+1},\boldsymbol{x}_2^k)$ 点关于 $\boldsymbol{x}_2$ 线性化，添加临近点正则项, 关于 $\boldsymbol{x}_2$ 求最小得 $\boldsymbol{x}_2^{k+1}$.

由此得迭代过程

$$\boldsymbol{x}_1^{k+1}=\arg\min_{\boldsymbol{x}_1\in\mathbb{R}^{n_1}}\left\{\nabla_{\boldsymbol{x}_1}f(\boldsymbol{x}_1^k,\boldsymbol{x}_2^k),\boldsymbol{x}_1-\boldsymbol{x}_1^k\rangle+\frac{L_k^1}{2}\|\boldsymbol{x}_1-\boldsymbol{x}_1^k\|^2+g_1(\boldsymbol{x}_1)\right\}$$

$$\boldsymbol{x}_2^{k+1}=\arg\min_{\boldsymbol{x}_2\in\mathbb{R}^{n_2}}\left\{\langle\nabla_{\boldsymbol{x}_2}f(\boldsymbol{x}_1^{k+1},\boldsymbol{x}_2^k),\boldsymbol{x}_2-\boldsymbol{x}_2^k\rangle+\frac{L_k^2}{2}\|\boldsymbol{x}_2-\boldsymbol{x}_2^k\|^2+g_2(\boldsymbol{x}_2)\right\}$$

其中 $L_k^1=\gamma_1L_1,\ L_k^2=\gamma_2L_2,\ \gamma_1,\gamma_2>1$

定理 线性化临近点交替极小化方法产生的迭代点列的任一聚点为优化问题的稳定点。