

最优化简介

Hailiang ZHAO @ ZJU-CS

<http://hliangzhao.me>

2021 年 6 月 30 日

标准形式

$$\mathcal{P}_1 : \quad \min \quad f(x) \\ \text{s.t.} \quad x \in \mathcal{X}$$

1. $x \in \mathbb{R}^n$: 决策变量
2. $f: \mathbb{R}^n \rightarrow \mathbb{R}$: 目标函数
3. $\mathcal{X} \subseteq \mathbb{R}^n$: 可行域

$$\mathcal{X} \triangleq \{x \in \mathbb{R}^n \mid c_i(x) \leq 0, i = 1, \dots, m, \\ c_i(x) = 0, i = m + 1, \dots, m + l.\}$$

- ▶ $\min f(x)$ 不存在 $\Rightarrow \inf f(x)$ (最大下界)
- ▶ $\max f(x)$ 不存在 $\Rightarrow \sup f(x)$ (最小上界)

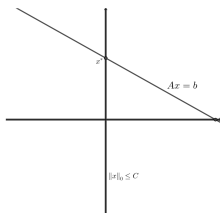
实例

► 稀疏优化 (信号还原)

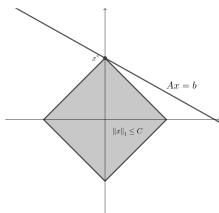
$$\begin{aligned} & \text{(NP-hard)} \quad \min_{x \in \mathbb{R}^n} \|x\|_0, \\ & \text{s.t.} \quad Ax = b \end{aligned}$$

变换为 l_1 -norm 优化问题求解 (l_0 -norm 和 l_1 -norm 优化一样, 均具备稀疏性):

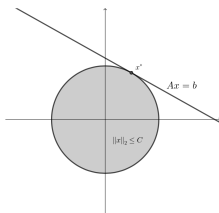
$$\begin{aligned} & \text{(solvable)} \quad \min_{x \in \mathbb{R}^n} \|x\|_1, \\ & \text{s.t.} \quad Ax = b \end{aligned}$$



(a) l_0 范数



(b) l_1 范数



(c) l_2 范数

实例

► 低秩矩阵恢复

$$\begin{aligned} \min_{X \in \mathbb{R}^{m \times n}} \text{rank}(X), \\ \text{s.t. } X_{ij} = M_{ij}, (i, j) \in \Omega \end{aligned}$$

根据稀疏优化思想，非零奇异值的个数 \Rightarrow 非零奇异值的和：

$$\begin{aligned} \min_{X \in \mathbb{R}^{m \times n}} \|X\|_*, \\ \text{s.t. } X_{ij} = M_{ij}, (i, j) \in \Omega \end{aligned}$$

二次罚函数形式（去掉约束项）：

$$\min_{X \in \mathbb{R}^{m \times n}} \mu \|X\|_* + \frac{1}{2} \sum_{(i,j) \in \Omega} (X_{ij} - M_{ij})^2$$

步骤与分类

▶ 步骤：构造模型 \Rightarrow 确定问题类型 \Rightarrow 设计算法 \Rightarrow 实现

▶ 分类：

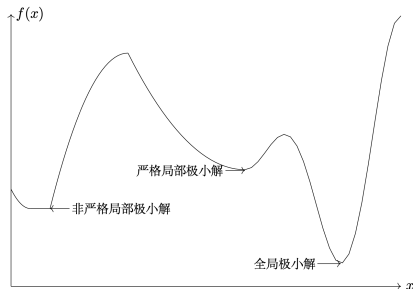
1. 连续：可以根据邻域信息来估计某点是否最优
 2. 离散：通常松弛成连续问题来求解
-
1. 无约束：欧几里得空间中寻找函数的最小点
 2. 有约束：将约束罚到目标函数上转换为无约束问题
-
1. 确定：不含未知参数
 2. 随机：目标函数是关于未知参数的期望
-
1. 线性：单纯形法、内点法
 2. 非线性：目标或约束至少有一个非线性
-
1. 凸：任何局部最优解都是全局最优解
 2. 非凸：转化为一系列凸子问题并逼近

最优解

定义

\mathcal{P}_1 的最优解 对于 $\bar{x} \in \mathcal{X}$:

1. $\forall x \in \mathcal{X}, f(\bar{x}) \leq f(x) \Rightarrow \bar{x}$ 为全局极小解/最优解/最小值点;
2. 存在 \bar{x} 的 ε -邻域 $\mathcal{N}_\varepsilon(\bar{x})$ 使得 $\forall x \in \mathcal{N}_\varepsilon(\bar{x}) \cap \mathcal{X}, f(\bar{x}) \leq f(x) \Rightarrow \bar{x}$ 为局部极小解/局部最优解;
3. $\forall x \in \mathcal{N}_\varepsilon(\bar{x}) \cap \mathcal{X}$ 且 $x \neq \bar{x}, f(\bar{x}) < f(x) \Rightarrow \bar{x}$ 为严格局部极小解。



解的形式

- ▶ 解析解（分析函数性质得到）
- ▶ 数值解（迭代算法）

(1) (无约束优化) 收敛准则（需要某种方法对 x^* 进行估计）:

$$\frac{f(x^k) - f^*}{\max\{|f^*|, 1\}} \leq \varepsilon_1, \quad \|\nabla f(x^k)\| \leq \varepsilon_2$$

判定约束优化问题是否收敛还需考虑约束条件的违反度。

(2) 停机准则:

$$\frac{\|x^{k+1} - x^k\|}{\max\{\|x^k\|, 1\}} \leq \varepsilon_3, \quad \frac{|f(x^{k+1}) - f(x^k)|}{\max\{|f(x^k)|, 1\}} \leq \varepsilon_4$$

依点列收敛到局部（全局）最优解:

$$\lim_{k \rightarrow \infty} \|x^k - x^*\| = 0$$

渐进收敛速度

Q-收敛速度: 对于充分大的 k , 若有

1. $\frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} \leq a, a \in (0, 1) \implies Q\text{-线性收敛}$

2. $\frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} = 0 \implies Q\text{-超线性收敛 (快)}$

3. $\frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} \leq 1 \implies Q\text{-次线性收敛}$

4. $\frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|^2} \leq a, a \in (0, 1) \implies Q\text{-二次收敛 (快)}$

差距越小, 速度越快, 注意 $\|x^{k+1} - x^*\| \ll 1$ 和 $a \ll 1$ 。

R-线性收敛 (同理可定义 R -超线性/次线性/二次收敛):

$$\|x^k - x^*\| \leq t_k \longmapsto \text{记为 } \mathcal{O}(t_k)$$

其中 $\{t_k\}$ 是非负的 Q -线性收敛序列。

复杂度

复杂度通常用迭代次数来描述，可结合渐进收敛速度得到。

示例：若已知

$$f(x^k) - f(x^*) \leq \frac{c}{\sqrt{k}}, \forall k > 0,$$

则满足精度 $f(x^k) - f(x^*) \leq \varepsilon$ 的迭代次数 k 满足

$$k \geq \frac{c^2}{\varepsilon^2}.$$

因此，算法对应的复杂度为 $\mathcal{O}(\frac{1}{\varepsilon^2})$ 。

优化算法设计技巧 *

1. 泰勒展开 (多项式可拟合任意光滑曲线)

在局部用多项式函数逼近目标函数 \Rightarrow 计算极小值点 \Rightarrow 迭代更新。

2. 对偶

对偶问题通常比原始问题容易求解, Primal-Dual Update, ADMM, etc.

3. 拆分

$$\min_x f(x) + g(x) \quad \Longrightarrow \quad \min_{x,y} f(x) + g(y), \quad s.t. \quad x = y$$

4. 块坐标下降

通过逐步求解分量转化为多个低维空间的优化问题。