第一章：介绍

（1）课程介绍和数学基础

授课教师：曹语

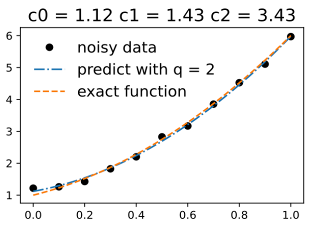
课程主页：`https://yucaoyc.github.io/math3806`

目录

优化问题无处不在

火车的时刻规划

实验数据处理



自动驾驶

等等问题都涉及复杂的优化过程

优化问题的基本框架

- 我们可选择的选项：**优化变量/备选解** $\boldsymbol{x} \in \mathbb{R}^n$（或者 $\boldsymbol{x} \in \{1, 2, 3, \cdots, n\}$）
- 更一般的，可行解的集合被标记为 Ω
- 我们希望优化的对象：**目标函数 (或费用函数)** $\boldsymbol{x} \mapsto f_0(\boldsymbol{x})$
- **优化问题**:

$$\min_{\boldsymbol{x} \in \Omega} f_0(\boldsymbol{x})$$

- 如果找到 $\boldsymbol{x}^\star$ 满足对于任何的 $\boldsymbol{x}$ 皆有 $f_0(\boldsymbol{x}) \geq f_0(\boldsymbol{x}^*)$，该 $\boldsymbol{x}^\star$ 被称为**最优解**

优化问题的标准形式

$$\begin{array}{ll} \text{minimize} & f_0(\boldsymbol{x}) \\ \text{subject to} & \begin{cases} f_i(\boldsymbol{x}) \leqslant 0, & i = 1, \cdots, m \\ h_i(\boldsymbol{x}) = 0, & i = 1, \cdots, p \end{cases} \end{array} \quad \text{（刻画可行解的集合 } \Omega\text{）}$$

惯例： 设不等式和等式约束的右端为零

问题：

- 若原始优化的形式是 $\max f_0(\boldsymbol{x})$，如何变成标准形式?
- 若原始优化的形式是 $\ell \leq f_1(\boldsymbol{x}) \leq r$，如何变成标准形式?

课程介绍

主要介绍凸优化的理论以及重要的优化算法

(I) 凸优化的数学基础

(II) 无约束优化问题

(III) 约束优化问题

基本流程:

具体问题 $\Longrightarrow$ 数学形式 $\Longrightarrow$ 数学理论（即重要定理和一些性质）
$\Longrightarrow$ 回归应用问题 $\Longrightarrow$ 写代码来实践

将使用 Canvas 作为平台:

考核形式: 见课内介绍，以及课程大纲（见 Canvas）

课程资料: 课内 PPT 和代码可见课程主页（Canvas 内有链接）;

作业需要在 Canvas 提交

课本:

- Boyd S, Vandenberghe L. Convex Optimization. Cambridge: Cambridge University Press. 【电子版课本见课程网站的链接】
- （中译本）Stephen Boyd, Lieven Vandenberghe 著王书宁，许鋆，黄晓霖译. 凸优化. 清华大学出版社

联系方式: 请直接通过 Canvas 发消息/或者发邮件至yucao@sjtu.edu.cn

目录

上确界和下确界

优化任务 $\min_{\boldsymbol{x}\in\Omega} f_0(\boldsymbol{x})$ 本质上是找到某个 $\boldsymbol{x}$ 使得 $f_0(\boldsymbol{x}) \leq f_0(\boldsymbol{y}), \forall \boldsymbol{y} \in \Omega$

问题：是否可能无法找到 x 呢?

例子： $\Omega = (0, 1]$，$f_0(\boldsymbol{x}) = \boldsymbol{x}$。此时我们可以无限让 $\boldsymbol{x}$ 趋紧 0，但是由于 0 不是可行解，因此该优化问题其实无解。

直觉：对于开区间或无界区域，最小值可能“取不到”。为了在数学上严谨地描述这种“无限逼近的底线”，我们需要引入下确界的概念。

上确界和下确界的定义

定义 (上确界和下确界)

- 对于任意的一个集合 $C \subset \mathbb{R}$，我们定义其**上确界（标记为 $\sup C$）**为集合 C 的所有上界中最小的值。若 C 有上界，根据上确界性质（least-upper-bound property），$\sup C$ 存在。
- 对于任意的一个集合 $C \subset \mathbb{R}$，我们定义其**下确界（标记为 $\inf C$）**为集合 C 的所有下界中最大的值。若 C 有下界，则 $\inf C$ 存在。

例子： 对于集合 $C = (0, 1]$，$\sup C = 1$, $\inf C = 0$

练习： 请写出集合 $C = \{1\} \cup [2, 3] \cup (4, 5)$ 的上/下确界

对于上述问题，更一般性的符号为

$$\sup_{\boldsymbol{x}\in\Omega} f_0(\boldsymbol{x}) \qquad \text{或} \qquad \inf_{\boldsymbol{x}\in\Omega} f_0(\boldsymbol{x})$$

好处：我们永远可以计算这个数值（有时可能是 $\pm\infty$）

- 若 $\inf_{\boldsymbol{x}\in\Omega} f_0(\boldsymbol{x}) = -\infty$，或不存在 $\boldsymbol{x} \in \Omega$ 满足 $f_0(\boldsymbol{x}) = \inf_{\boldsymbol{x}\in\Omega} f_0(\boldsymbol{x})$，则该优化问题无可行解。
- 若存在 $\boldsymbol{x} \in \Omega$ 满足 $f_0(\boldsymbol{x}) = \inf_{\boldsymbol{x}\in\Omega} f_0(\boldsymbol{x})$，该问题等价于 $\min_{\boldsymbol{x}\in\Omega} f_0(\boldsymbol{x})$。即严格意义下，我们使用 min 的时候，我们（一般而言）已经假设/验证该问题可以有解。
- 对于 sup 的情况类似。

目录

偏导数

设 $f:\mathbb{R}^n\to\mathbb{R}$，$\boldsymbol{x}=\begin{bmatrix}x_1 & \cdots & x_n\end{bmatrix}^\top$。

偏导数定义为

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x})=\lim_{t\to 0}\frac{f(x_1,\cdots,x_i+t,\cdots,x_n)-f(\boldsymbol{x})}{t}.$$

例子： 若 $f(x_1,x_2)=x_1^2+3x_1x_2+x_2^2$，则

$$\frac{\partial f}{\partial x_1}=2x_1+3x_2,\qquad \frac{\partial f}{\partial x_2}=3x_1+2x_2.$$

偏导数的链式法则

设

$$z = f(u_1, \cdots, u_m), \qquad u_j = g_j(x_1, \cdots, x_n),\ j = 1, \cdots, m.$$

则对任意 $i = 1, \cdots, n$，有

$$\frac{\partial z}{\partial x_i} = \sum_{j=1}^{m} \frac{\partial f}{\partial u_j} \frac{\partial u_j}{\partial x_i}.$$

练习： 设

$$f(u, v) = u^2 + uv, \qquad u = x^2 + y, \qquad v = e^{xy}.$$

请用链式法则计算 $\frac{\partial f}{\partial x}$ 与 $\frac{\partial f}{\partial y}$，并求在 $(x, y) = (1, 0)$ 处的数值。

梯度、Hessian

把所有偏导数组合成向量，得到梯度：

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\boldsymbol{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\boldsymbol{x}) \end{bmatrix}.$$

Hessian 矩阵是二阶偏导数组成的矩阵：

$$\nabla^2 f(\boldsymbol{x}) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x})\right]_{i,j}.$$

例子： 令 $f(x_1, x_2) = x_1^2 + 3x_1x_2 + x_2^2$，则

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 + 3x_2 \\ 3x_1 + 2x_2 \end{bmatrix},$$

$$\nabla^2 f(x_1, x_2) = \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix}.$$

方向导数

沿方向 $\boldsymbol{d}$ 的方向导数为

$$D_{\boldsymbol{d}} f(\boldsymbol{x}) = \lim_{t \to 0} \frac{f(\boldsymbol{x} + t\boldsymbol{d}) - f(\boldsymbol{x})}{t}.$$

练习：：请通过链式法则验证

$$D_{\boldsymbol{d}} f(\boldsymbol{x}) = \nabla f(\boldsymbol{x})^\top \boldsymbol{d}.$$

一维函数的泰勒展开

设 $f:\mathbb{R}\to\mathbb{R}$ 在点 x 附近足够光滑，则在 x 处对 $x+t$ 做展开：

$$f(x+t)=f(x)+f'(x)t+\frac{1}{2}f''(x)t^2+o(t^2),\qquad t\to 0.$$

常用近似：

- 一阶近似（线性化）：$f(x+t)\approx f(x)+f'(x)t$.
- 二阶近似（局部二次模型）：$f(x+t)\approx f(x)+f'(x)t+\frac{1}{2}f''(x)t^2$.

例子： 对 $f(x)=e^x$，在 $x=0$ 处有

$$e^t=1+t+\frac{1}{2}t^2+o(t^2).$$

此处 $o(t^2)$ 表示当 $t\to 0$ 时，$\frac{o(t^2)}{t^2}\to 0$。

目录

矩阵乘法

若 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，$\boldsymbol{x} \in \mathbb{R}^{n}$，则

$$\boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^{m}, \qquad (\boldsymbol{A}\boldsymbol{x})_i = \sum_{j=1}^{n} A_{ij} x_j.$$

一般地，若 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，$\boldsymbol{B} \in \mathbb{R}^{n\times p}$，则

$$\boldsymbol{A}\boldsymbol{B} \in \mathbb{R}^{m\times p}, \qquad (\boldsymbol{A}\boldsymbol{B})_{ik} = \sum_{j=1}^{n} A_{ij} B_{jk}.$$

练习：：若矩阵 $\boldsymbol{A} = \left[\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right]$，向量 $\boldsymbol{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]$，请写出 $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ 的具体表达式。

分块矩阵乘法

设

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}, \qquad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{bmatrix},$$

且各分块维度可乘，则

$$\boldsymbol{AB} = \begin{bmatrix} \boldsymbol{A}_{11}\boldsymbol{B}_{11} + \boldsymbol{A}_{12}\boldsymbol{B}_{21} & \boldsymbol{A}_{11}\boldsymbol{B}_{12} + \boldsymbol{A}_{12}\boldsymbol{B}_{22} \\ \boldsymbol{A}_{21}\boldsymbol{B}_{11} + \boldsymbol{A}_{22}\boldsymbol{B}_{21} & \boldsymbol{A}_{21}\boldsymbol{B}_{12} + \boldsymbol{A}_{22}\boldsymbol{B}_{22} \end{bmatrix}.$$

练习： 设

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 3 & 4 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix},$$

按分块矩阵乘法计算 $\boldsymbol{AB}$。

矩阵转置和矩阵的逆

- **转置**：对于 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 矩阵，$\boldsymbol{A}^\top$ 是指某个 $n \times m$ 的矩阵，其中元素

$$(\boldsymbol{A}^\top)_{ij} = A_{ji}, \qquad i = 1, \cdots, n,\ j = 1, \cdots, m.$$

- **矩阵的逆**： n 阶矩阵 $\boldsymbol{A}$ 可逆是指存在矩阵 $\boldsymbol{A}^{-1}$，使得

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I},$$

  其中 $\boldsymbol{I}$ 为 n 阶单位矩阵。
- **对称矩阵**：若矩阵 $A^\top = A$，则称 A 为**对称矩阵**。
- **规则**：

$$(\boldsymbol{A}\boldsymbol{B})^\top = \boldsymbol{B}^\top\boldsymbol{A}^\top, \qquad (\boldsymbol{A}^{-1})^\top = (\boldsymbol{A}^\top)^{-1}, \qquad (\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}.$$

练习：

- 若 $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = 3x_1^2 + 4x_1x_2 + 5x_2^2$，且矩阵 $\boldsymbol{A}$ 为对称矩阵，请写出 $\boldsymbol{A}$。

零空间和值域/Null and Range

- 对于一个矩阵 $\boldsymbol{A} \in \mathbb{R}^{n\times p}$，我们定义

$$
\begin{array}{lll}
\text{零空间（核）} & \mathrm{Null}(\boldsymbol{A}) & := \left\{\boldsymbol{x} \in \mathbb{R}^p : \ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}_n\right\} \\
\text{值域} & \mathrm{Range}(\boldsymbol{A}) & := \left\{\boldsymbol{y} \in \mathbb{R}^n : \ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y} \text{ 对于某个向量} \boldsymbol{x} \in \mathbb{R}^p\right\}
\end{array}
$$

  上述的 $\mathrm{Null}(\boldsymbol{A})$ 和 $\mathrm{Range}(\boldsymbol{A})$ 都是子空间。

- $\dim \mathrm{Range}(\boldsymbol{A})$ 被定义为 $\mathrm{Rank}(\boldsymbol{A})$。

练习： 设 $\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$. 请计算 $\mathrm{Null}(\boldsymbol{A})$, $\mathrm{Range}(\boldsymbol{A})$ 和 $\mathrm{Rank}(\boldsymbol{A})$。

正交子空间

若 $S, T \subset \mathbb{R}^n$ 是某个子空间：

- 若对于任意的 $\boldsymbol{x} \in S$, $\boldsymbol{y} \in T$，我们都有 $\boldsymbol{x} \cdot \boldsymbol{y} = 0$，则我们说这两个子空间**正交/垂直**。
- S 的**正交补 (orthogonal complement)**被定义为
$$S^{\perp} = \{\boldsymbol{x} \in \mathbb{R}^n : \ \boldsymbol{x} \cdot \boldsymbol{y} = 0, \ \ \forall \boldsymbol{y} \in S\}$$
  例子： $S = \text{span}\{\boldsymbol{e}_1\} \subset \mathbb{R}^3$ 的正交补是 $\text{span}\{\boldsymbol{e}_2, \boldsymbol{e}_3\}$
- $\dim(S) + \dim(S^{\perp}) = n$
- $(S^{\perp})^{\perp} = S$

零空间与值域的关系

定理

对于矩阵 $\boldsymbol{A} \in \mathbb{R}^{n\times p}$，$\mathit{Null}(\boldsymbol{A}) = \mathit{Range}(\boldsymbol{A}^\top)^\perp \subseteq \mathbb{R}^p$

练习： 请验证该结果对于上述的矩阵 $\boldsymbol{A} = \left[\begin{smallmatrix} 1 & 2 \\ 2 & 4 \end{smallmatrix}\right]$ 成立。

证明.

$$
\begin{aligned}
\boldsymbol{x} \in \mathrm{Null}(\boldsymbol{A}) \quad &\leftrightarrow \quad \boldsymbol{A}\boldsymbol{x} = \mathbf{0}_n \quad \leftrightarrow \quad \boldsymbol{x}^\top \boldsymbol{A}^\top = \mathbf{0}_n^\top \\
&\leftrightarrow \quad \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{y} = 0,\ \forall \boldsymbol{y} \in \mathbb{R}^n \quad \leftrightarrow \quad \boldsymbol{x} \in \mathrm{Range}(\boldsymbol{A}^\top)^\perp
\end{aligned}
$$

□

目录

总结

- 掌握优化问题的框架，以及一些基本术语
- 回顾复习偏导数、方向导数、矩阵乘法等数学基础

阅读作业 & 参考资料：

- 课本第 1 章
- Stephen Boyd and Sanjay Lall. Range and Null Space. Available at `https://ee263.stanford.edu/lectures/rangenull.pdf`
- 课本附录 A.5.1

参考答案

- 第5页：
  - $\max f_0(\boldsymbol{x})$ 可以等价写成 $\min -f_0(\boldsymbol{x})$；两者的最优解相同。
  - $\ell \leq f_1(\boldsymbol{x}) \leq r$ 等价于

$$f_1(\boldsymbol{x}) - r \leq 0, \qquad \text{且} \qquad -f_1(\boldsymbol{x}) + \ell \leq 0.$$

- 第10页：$\sup C = 5$, $\inf C = 1$

参考答案（续）

- 第14页：令 $F(x,y)=f(u(x,y),v(x,y))$，其中

$$f(u,v)=u^2+uv, \qquad u=x^2+y, \qquad v=e^{xy},$$

则

$$\begin{aligned}
\frac{\partial F}{\partial x} &= \frac{\partial F}{\partial u}\frac{\partial u}{\partial x}+\frac{\partial F}{\partial v}\frac{\partial v}{\partial x}=(2u+v)\frac{\partial u}{\partial x}+(u)\frac{\partial v}{\partial x} \\
&= 2x\Big(2(x^2+y)+e^{xy}\Big)+(x^2+y)ye^{xy}, \\
\frac{\partial F}{\partial y} &= \frac{\partial F}{\partial u}\frac{\partial u}{\partial y}+\frac{\partial F}{\partial v}\frac{\partial v}{\partial y}=(2u+v)\frac{\partial u}{\partial y}+(u)\frac{\partial v}{\partial y} \\
&= \Big(2(x^2+y)+e^{xy}\Big)+(x^2+y)xe^{xy}.
\end{aligned}$$

在 $(x,y)=(1,0)$ 处有 $\frac{\partial F}{\partial x}=6$, $\frac{\partial F}{\partial y}=4$.

参考答案（续）

- 第19页：

$$\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$
$$= a_{11}x_1^2 + (a_{12} + a_{21})x_1x_2 + a_{22}x_2^2.$$

- 第22页（对称矩阵题）：若 $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = 3x_1^2 + 4x_1x_2 + 5x_2^2$，且 $\boldsymbol{A} = \boldsymbol{A}^\top$，则

$$\boldsymbol{A} = \left[\begin{smallmatrix} 3 & 2 \\ 2 & 5 \end{smallmatrix}\right].$$

- （Null/Range 题）设 $\boldsymbol{A} = \left[\begin{smallmatrix} 1 & 2 \\ 2 & 4 \end{smallmatrix}\right]$. 则 Null($\boldsymbol{A}$) $= \mathrm{span}\left\{\left[\begin{smallmatrix} -2 \\ 1 \end{smallmatrix}\right]\right\}$, Range($\boldsymbol{A}$) $= \mathrm{span}\left\{\left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]\right\}$, Rank($\boldsymbol{A}$) $= 1$.

参考答案（续）

- （零空间与值域关系验证）对 $\boldsymbol{A} = \left[\begin{smallmatrix} 1 & 2 \\ 2 & 4 \end{smallmatrix}\right]$, 有

$$\mathsf{Range}(\boldsymbol{A}^\top) = \mathsf{Range}(\boldsymbol{A}) = \operatorname{span}\{\left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]\}.$$

因此

$$\mathsf{Range}(\boldsymbol{A}^\top)^\perp = \left\{\boldsymbol{x} \in \mathbb{R}^2 : \ \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]^\top \boldsymbol{x} = 0\right\}$$
$$= \left\{\boldsymbol{x} \in \mathbb{R}^2 : \ x_1 + 2x_2 = 0\right\} = \operatorname{span}\{\left[\begin{smallmatrix} -2 \\ 1 \end{smallmatrix}\right]\}.$$

而这正是 $\mathsf{Null}(\boldsymbol{A})$，故 $\mathsf{Null}(\boldsymbol{A}) = \mathsf{Range}(\boldsymbol{A}^\top)^\perp$.