



内点法

Interior-Point Methods · Chapter 11

梁经纬 Jingwei Liang

Institute of Natural Sciences
School of Mathematical Sciences

Lecture 06: 内点法

2026 年秋季学期

饮水思源 · 爱国荣校

本讲内容

问题与假设 §11.1
光滑凸约束、严格可行性与 KKT 条件



对数障碍与中心路径 §11.2
障碍函数、对偶点与间隙 m/t



障碍法 §11.3
中心化、参数更新与 Newton 方程



可行性与 Phase I §11.4
找严格可行点或给出不可行证书

自协调复杂度 §11.5
每次中心化与总 Newton 步数



广义不等式 §11.6
正交锥、二阶锥与半正定锥



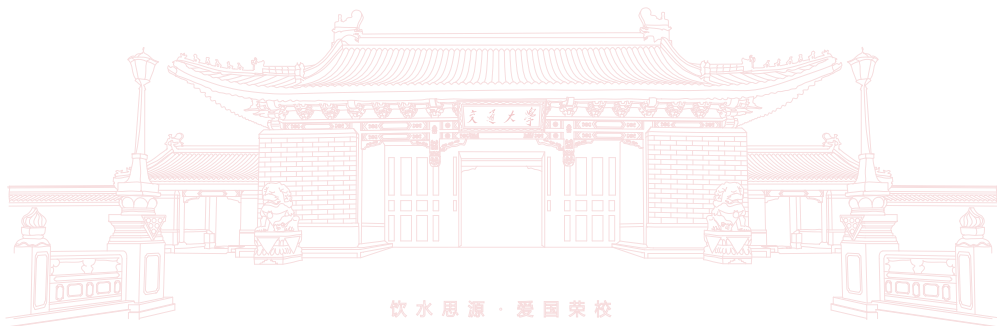
原始-对偶内点法 §11.7
残差、搜索方向与代理对偶间隙



实现 §11.8
稀疏 KKT 系统、LP、 ℓ_1 、SDP 与网络

01

问题、假设与前序知识



与前面课程的连接

Lecture 03
Newton 法、自协调函数、稀疏
线性代数

Lecture 04
Lagrangian、对偶函数、KKT 与
Slater 条件

Lecture 05
等式约束 Newton 法、KKT 线性
系统

Chapter 11
用一系列等式约束 Newton 问题
逼近含不等式约束的最优解

计算主线
保持在可行域内部；用对偶间隙判断精
度；每步求解结构化 KKT 系统。

课程 Lecture 03–05 回顾；Boyd & Vandenberghe, §11.1

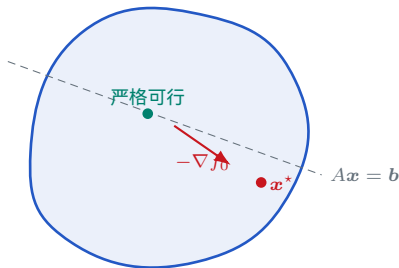
不等式约束凸优化

标准形式

$$\begin{aligned} & \text{minimize}_{x \in \mathbb{R}^n} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m, \\ & && Ax = b, \end{aligned} \quad (11.1)$$

其中 $f_0, \dots, f_m$ 凸且二阶连续可微, $A \in \mathbb{R}^{p \times n}$, $\text{rank } A = p < n$ 。

- 最优解 x^* 存在, $p^* = f_0(x^*)$ 。
- 存在严格可行点: $Ax = b$ 且 $f_i(x) < 0$ 。
- 严格可行性给出 Slater 条件与强对偶。



内点法让迭代点停留在不等式可行域的内部。

KKT 条件

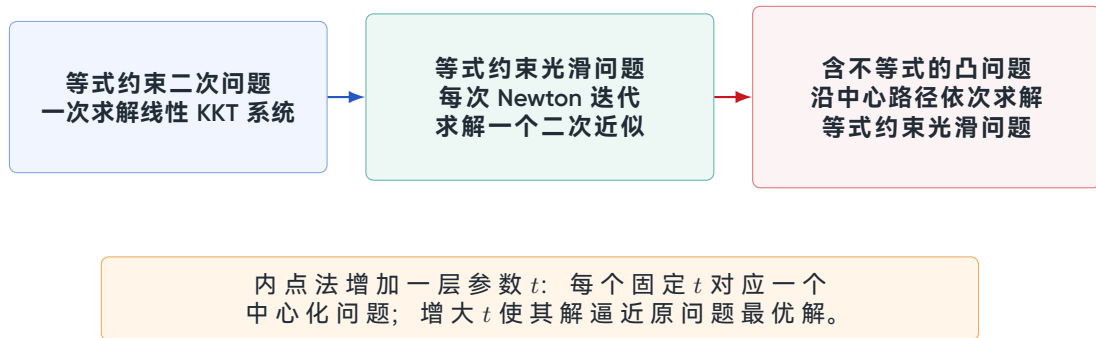
严格可行性保证存在 $\lambda^* \in \mathbb{R}^m$ 、 $\nu^* \in \mathbb{R}^p$ ，使

$$\begin{aligned} Ax^* &= b, & f_i(x^*) &\leq 0, \\ \lambda^* &\succeq 0, & \nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + A^\top \nu^* &= 0, \\ \lambda_i^* f_i(x^*) &= 0, & i &= 1, \dots, m. \end{aligned} \tag{11.2}$$

原始可行	对偶可行与驻点	互补松弛
约束满足, x^* 属于可行集。	乘子非负, Lagrangian 对 x 的梯度为零。	每个约束满足 $-\lambda_i^* f_i(x^*) = 0.$ 活跃约束可有正乘子; 松弛约束的乘子为零。

Boyd & Vandenberghe, *Convex Optimization*, §11.1, KKT conditions (11.2)

从 Newton 法到内点法



每一层都复用前一讲的核心计算: Hessian、Newton decrement、回溯线搜索与 KKT 消元。

可直接处理的问题

问题类型	模型	内点法所需结构
LP	$\min c^\top x, \quad Gx \preceq h, \quad Ax = b$	线性目标与仿射不等式
QP/QCQP	二次目标, 凸二次不等式	梯度与 Hessian 可显式计算
凸形式 GP	log-sum-exp 目标与约束	光滑凸复合函数
熵优化	$\min \sum_j x_j \log x_j, \quad Fx \preceq g$	定义域 $x \succ 0$ 与线性约束

非光滑目标需要先改写

$$\min_x \max_i (a_i^\top x + b_i) \quad \Longleftrightarrow \quad \min_{x, \tau} \tau \quad \text{subject to} \quad a_i^\top x + b_i \leq \tau.$$

右侧具有光滑目标与仿射约束，可以直接使用本章方法。

应用例 1: 峰值负荷控制

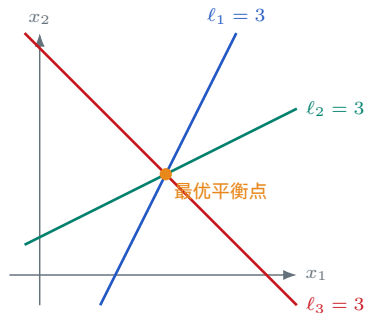
某工厂用 $x = (x_1, x_2)$ 表示两台设备的调节量。三个时段的超额负荷为

$$\ell_1 = 2x_1 - x_2 + 1, \quad \ell_2 = -x_1 + 2x_2 + 2, \quad \ell_3 = -x_1 - x_2 + 6.$$

原始目标 $\min \max\{\ell_1, \ell_2, \ell_3\}$ 在折点处不可微。引入峰值变量 τ :

$$\begin{array}{ll} \min_{x, \tau} & \tau \\ \text{subject to} & 2x_1 - x_2 + 1 \leq \tau, \\ & -x_1 + 2x_2 + 2 \leq \tau, \\ & -x_1 - x_2 + 6 \leq \tau. \end{array}$$

这是三条仿射不等式构成的 LP。

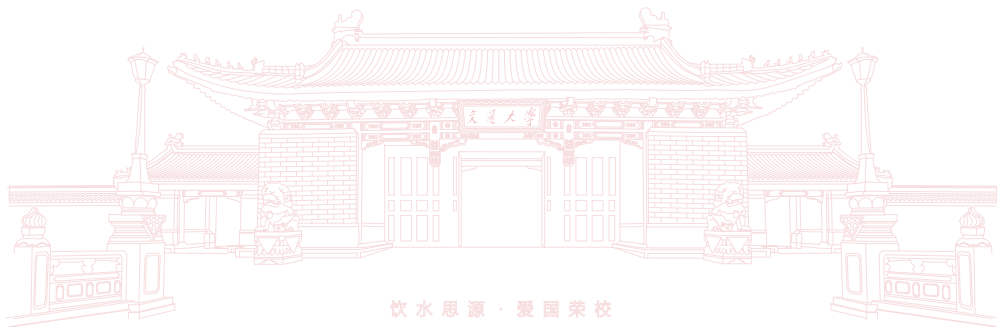


障碍项同时保持三条约束严格满足。

自编峰值负荷示例；模型变换参见 Boyd & Vandenberghe, §11.1

02

对数障碍与中心路径



从示性函数到对数障碍

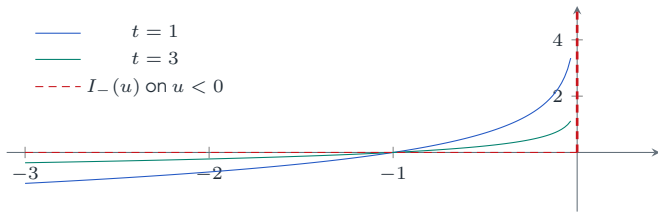
不等式 $f_i(\mathbf{x}) \leq 0$ 可写入目标:

$$\min f_0(\mathbf{x}) + \sum_{i=1}^m I_-(f_i(\mathbf{x})), \quad I_-(u) = \begin{cases} 0, & u \leq 0, \\ +\infty, & u > 0. \end{cases}$$

I_- 不可微。用

$$\hat{I}_-(u) = -\frac{1}{t} \log(-u), \quad u < 0, \quad t > 0$$

近似后, 可以调用 Newton 法。



Boyd & Vandenberghe, *Convex Optimization*, §11.2 and Figure 11.1

对数障碍函数

定义

$$\phi(\mathbf{x}) = -\sum_{i=1}^m \log(-f_i(\mathbf{x})), \quad \text{dom } \phi = \{\mathbf{x} \mid f_i(\mathbf{x}) < 0, \forall i\}. \quad (11.5)$$

中心化问题可写成两个等价形式:

$$\min_{A\mathbf{x}=\mathbf{b}} f_0(\mathbf{x}) + \frac{1}{t}\phi(\mathbf{x}) \quad \Longleftrightarrow \quad \min_{A\mathbf{x}=\mathbf{b}} t f_0(\mathbf{x}) + \phi(\mathbf{x}). \quad (11.6)$$

符号与导数

$$\begin{aligned} \nabla \phi(\mathbf{x}) &= \sum_{i=1}^m \frac{1}{-f_i(\mathbf{x})} \nabla f_i(\mathbf{x}), \\ \nabla^2 \phi(\mathbf{x}) &= \sum_{i=1}^m \frac{\nabla f_i(\mathbf{x}) \nabla f_i(\mathbf{x})^\top}{f_i(\mathbf{x})^2} + \sum_{i=1}^m \frac{1}{-f_i(\mathbf{x})} \nabla^2 f_i(\mathbf{x}). \end{aligned}$$

靠近任一边界 $f_i(\mathbf{x}) \uparrow 0$ 时, $\phi(\mathbf{x}) \rightarrow +\infty$ 。

Boyd & Vandenberghe, *Convex Optimization*, §11.2.1, equations (11.4)–(11.5)

一个可手算的中心路径

考虑

$$\min_{x \geq 0} x.$$

把约束写成 $f_1(x) = -x \leq 0$, 中心化目标为

$$q_t(x) = tx - \log x, \quad x > 0.$$

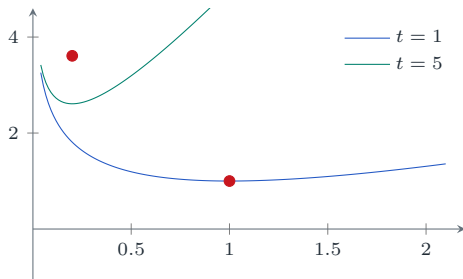
$$q'_t(x) = t - \frac{1}{x} = 0 \implies x^*(t) = \frac{1}{t}.$$

因此

$$f_0(x^*(t)) - p^* = \frac{1}{t},$$

恰好等于 m/t , 因为 $m = 1$ 。

t	$x^*(t)$	目标误差
1	1	1
5	0.2	0.2
20	0.05	0.05



自编数值例; 结论对应 Boyd & Vandenberghe, §11.2.2

中心路径

对每个 $t > 0$, 令 $x^*(t)$ 为中心化问题的唯一解。集合

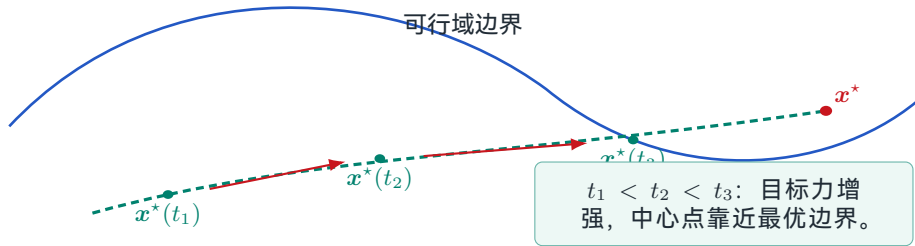
$$\{x^*(t) \mid t > 0\}$$

称为中心路径。中心点满足

$$Ax^*(t) = b, \quad f_i(x^*(t)) < 0,$$

且存在 $\hat{\nu}$ 使

$$t \nabla f_0(x^*(t)) + \sum_{i=1}^m \frac{1}{-f_i(x^*(t))} \nabla f_i(x^*(t)) + A^\top \hat{\nu} = 0. \quad (11.7)$$



Boyd & Vandenberghe, *Convex Optimization*, §11.2.2, central path and equation (11.7)

例 11.1: 不等式形式 LP

对于

$$\min \mathbf{c}^\top \mathbf{x} \quad \text{subject to} \quad G\mathbf{x} \preceq \mathbf{h},$$

记 $g_i^\top$ 为 G 的第 i 行, 松弛量 $s_i = h_i - g_i^\top \mathbf{x} > 0$ 。则

$$\phi(\mathbf{x}) = -\sum_{i=1}^m \log s_i.$$

梯度与 Hessian

令 $d_i = 1/s_i$:

$$\nabla \phi(\mathbf{x}) = G^\top \mathbf{d}, \quad \nabla^2 \phi(\mathbf{x}) = G^\top \text{diag}(\mathbf{d})^2 G.$$

若 G 满列秩, 则 Hessian 正定。

每个约束贡献一个秩一项 $d_i^2 g_i g_i^\top$; 靠近边界时, 对应项迅速增大。

中心性条件

$$t\mathbf{c} + G^\top \mathbf{d} = 0.$$

几何上, $\mathbf{c}^\top \mathbf{x} = \mathbf{c}^\top \mathbf{x}^*(t)$ 与经过 $\mathbf{x}^*(t)$ 的障碍函数等值线相切。

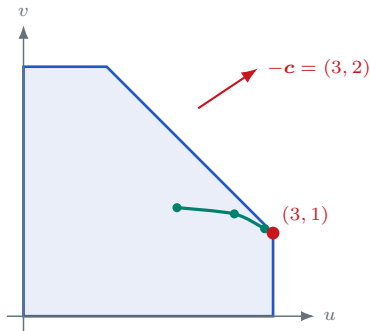
二维 LP 的中心路径

考虑

$$\begin{aligned} \min_{u,v} \quad & -3u - 2v \\ \text{subject to} \quad & u + v \leq 4, \quad u \leq 3, \quad v \leq 3, \\ & u \geq 0, \quad v \geq 0. \end{aligned}$$

最优点为 $(3, 1)$ ，最优值 -11 。障碍法从区域内部逼近这个顶点。

$$\phi = -\log(4-u-v) - \log(3-u) - \log(3-v) - \log u - \log v.$$



坐标采用相同单位长度。

自编二维 LP；理论与几何解释参见 Boyd & Vandenberghe, Example 11.1 and Figure 11.2

中心点给出对偶可行点

由中心性条件定义

$$\lambda_i^*(t) = -\frac{1}{tf_i(\mathbf{x}^*(t))} > 0, \quad \boldsymbol{\nu}^*(t) = \frac{\hat{\boldsymbol{\nu}}}{t}. \quad (11.10)$$

于是

$$\nabla f_0(\mathbf{x}^*(t)) + \sum_{i=1}^m \lambda_i^*(t) \nabla f_i(\mathbf{x}^*(t)) + A^\top \boldsymbol{\nu}^*(t) = 0.$$

Lagrangian 在 $\mathbf{x}^*(t)$ 取极小

$$g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) = f_0(\mathbf{x}^*) + \sum_i \lambda_i^* f_i(\mathbf{x}^*).$$

每个互补乘积相同

$$-\lambda_i^*(t) f_i(\mathbf{x}^*(t)) = \frac{1}{t}.$$

故

$$g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) = f_0(\mathbf{x}^*(t)) - \frac{m}{t}.$$

Boyd & Vandenberghe, *Convex Optimization*, §11.2.2, equations (11.10) and dual points from the central path

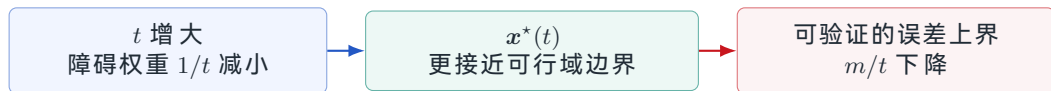
对偶间隙为 m/t

中心点与相应对偶点的对偶间隙满足

$$f_0(\mathbf{x}^*(t)) - g(\boldsymbol{\lambda}^*(t), \boldsymbol{\nu}^*(t)) = - \sum_{i=1}^m \lambda_i^*(t) f_i(\mathbf{x}^*(t)) = \frac{m}{t}.$$

由弱对偶,

$$0 \leq f_0(\mathbf{x}^*(t)) - p^* \leq \frac{m}{t}. \quad (11.12)$$

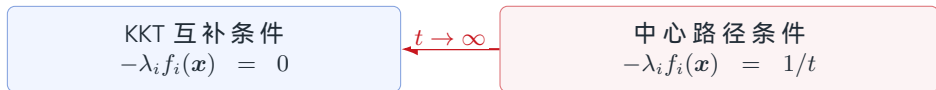


这个界无需知道 p^* 。它同时给出停止准则和对偶证书。

中心路径是扰动后的 KKT 系统

$x = x^*(t)$ 当且仅当存在 λ, ν 满足

$$\begin{aligned} Ax &= b, & f_i(x) < 0, \quad \lambda_i > 0, \\ \nabla f_0(x) + \sum_i \lambda_i \nabla f_i(x) + A^\top \nu &= 0, \\ -\lambda_i f_i(x) &= \frac{1}{t}, \quad i = 1, \dots, m. \end{aligned} \tag{11.11}$$



- ▣ 小 t : 每个约束保持明显松弛。
- ▣ 大 t : 活跃约束的松弛趋近零，乘子承担相应边界压力。

Boyd & Vandenberghe, *Convex Optimization*, §11.2.2, interpretation via KKT conditions

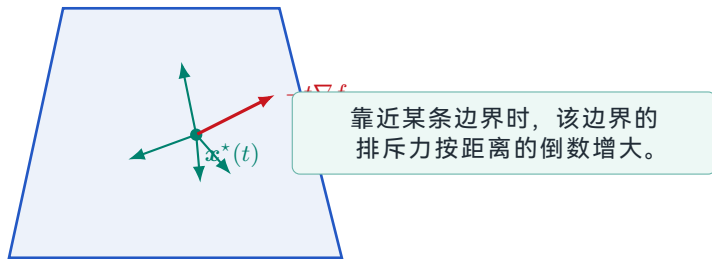
力场解释

暂不考虑等式约束。第 i 个约束产生排斥力

$$F_i(\mathbf{x}) = -\nabla[-\log(-f_i(\mathbf{x}))] = \frac{1}{f_i(\mathbf{x})} \nabla f_i(\mathbf{x}),$$

目标产生力 $F_0(\mathbf{x}) = -t \nabla f_0(\mathbf{x})$ 。中心点满足

$$F_0(\mathbf{x}^*(t)) + \sum_i F_i(\mathbf{x}^*(t)) = 0.$$



Boyd & Vandenberghe, *Convex Optimization*, §11.2.2, force field interpretation and Example 11.3

数值例 2: 二维 LP 的中心点

对前述 LP 数值求解中心化问题, 得到

t	$u^*(t)$	$v^*(t)$	$-3u^* - 2v^*$
0.5	1.8449	1.3050	-8.1447
2	2.5342	1.2304	-10.0634
10	2.8991	1.0520	-10.8012
50	2.9799	1.0101	-10.9600

此问题有 $m = 5$ 个标量不等式, 因此保证

$$f_0(\mathbf{x}^*(t)) - p^* \leq \frac{5}{t}.$$

例如 $t = 50$ 时, 实际误差约 0.0400, 理论上界为 0.1。

数值检查

$t = 50$ 时

$$4 - u - v = 0.00995,$$

$$3 - u = 0.02006,$$

$$u, v, 3 - v > 0.$$

点仍严格可行, 但两个即将活跃的约束松弛已经很小。

自编数值计算; 误差界参见 Boyd & Vandenberghe, §11.2.2

应用例 3: 概率分配与熵

考虑

$$\min_{\mathbf{x}} \sum_{j=1}^n x_j \log x_j \quad \text{subject to} \quad \mathbf{1}^\top \mathbf{x} = 1, \quad \mathbf{x} \succeq 0.$$

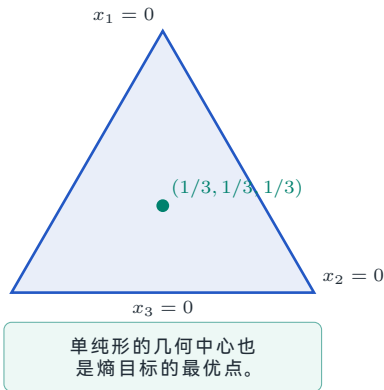
加入正性障碍后,

$$q_t(\mathbf{x}) = t \sum_j x_j \log x_j - \sum_j \log x_j.$$

对称性与约束 $\sum_j x_j = 1$ 给出

$$x_1^*(t) = \cdots = x_n^*(t) = \frac{1}{n}.$$

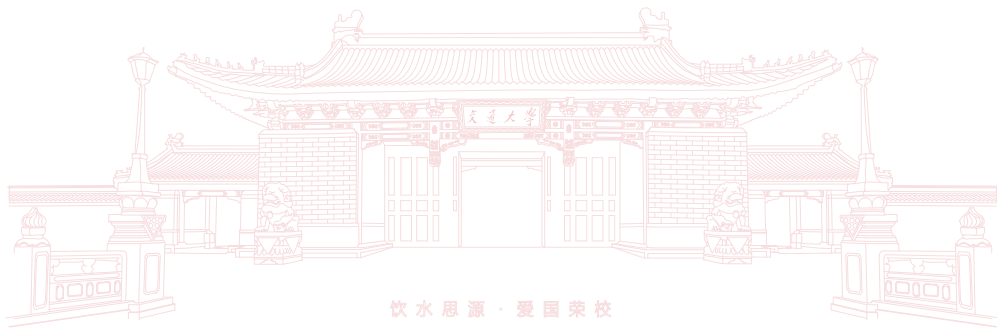
该例中中心路径退化为一个点: 最大熵分布始终是均匀分布。



Boyd & Vandenberghe, *Convex Optimization*, §11.1 entropy example; §11.2 logarithmic barrier

03

障碍法

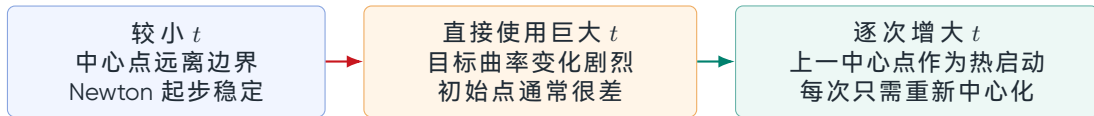


为何不直接取很大的 t

若希望误差不超过 ε ，可以一次取 $t = m/\varepsilon$ ，然后求解

$$\min_{Ax=b} \frac{m}{\varepsilon} f_0(\mathbf{x}) + \phi(\mathbf{x}).$$

但当 ε 很小时， $\mathbf{x}^*(t)$ 靠近边界，Hessian 在不同方向上的尺度差异很大。



障碍法的关键不是单次求解，而是沿中心路径进行一系列热启动 Newton 求解。

障碍法

Algorithm 11.1

给定严格可行 x 、 $t = t^{(0)} > 0$ 、 $\mu > 1$ 与容差 $\varepsilon > 0$:

1. **中心化**: 从当前 x 出发, 求解

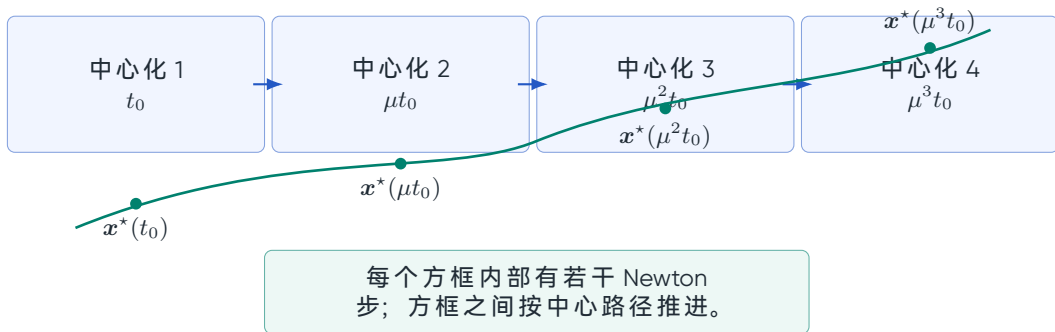
$$\min_{Ax=b} tf_0(x) + \phi(x),$$

得到 $x^*(t)$ 。

2. **更新**: $x \leftarrow x^*(t)$ 。
3. **停止**: 若 $m/t < \varepsilon$, 返回当前原始解和对偶证书。
4. **增大参数**: $t \leftarrow \mu t$, 重复。

- ▣ 一次中心化称为一次外层迭代。
- ▣ 中心化内部执行的 Newton 步称为内层迭代。
- ▣ 除首次中心化外, 每次都由上一中心点热启动。

外层与内层迭代



小 μ 带来更多外层迭代和较少的每次内层步数；较大 μ 的效果相反。

Boyd & Vandenberghe, *Convex Optimization*, §11.3.1, centering steps and choice of μ

中心化的 Newton 方程

在当前严格可行点 x , 令

$$H_t = t \nabla^2 f_0(x) + \nabla^2 \phi(x), \quad g_t = t \nabla f_0(x) + \nabla \phi(x).$$

等式约束 Newton 步由 KKT 系统给出:

$$\begin{bmatrix} H_t & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\text{nt}} \\ \nu_{\text{nt}} \end{bmatrix} = - \begin{bmatrix} g_t \\ 0 \end{bmatrix}. \quad (11.14)$$

保持等式可行

$$A \Delta x_{\text{nt}} = 0 \implies A(x + s \Delta x_{\text{nt}}) = b.$$

保持严格不等式

回溯线搜索首先缩短 s , 直到

$$f_i(x + s \Delta x_{\text{nt}}) < 0, \quad \forall i,$$

再检查充分下降。

中心化精度与对偶证书

理论叙述常把中心化看作精确求得 $x^*(t)$ 。实际实现用 Newton decrement 停止：

$$\frac{\lambda_t(x)^2}{2} \leq \varepsilon_{\text{nt}}.$$

适度精度已足够

中心路径只是通往原问题解的工具。近似中心化仍会产生收敛的外层序列。

若中心化不精确，直接使用 $\lambda_i = -1/(t f_i(x))$ 得到的点一般不完全对偶可行；可增加修正项，或以残差共同判断精度。

高精度代价通常不大

进入二次收敛区后，再增加若干 Newton 步即可显著降低中心化误差。

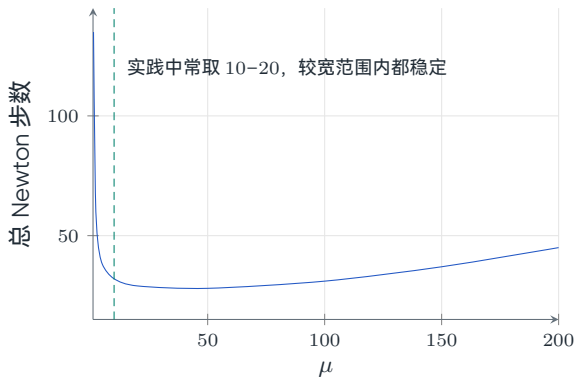
参数 μ 的取舍

μ 接近 1

- ▣ 相邻中心点距离小;
- ▣ 每次中心化 Newton 步少;
- ▣ 外层迭代次数多。

μ 较大

- ▣ 单次对偶间隙下降快;
- ▣ 热启动离新中心点较远;
- ▣ 每次中心化 Newton 步多。



Boyd & Vandenberghe, *Convex Optimization*, §11.3.1, Choice of μ ; qualitative version of Figure 11.5

初始参数 $t^{(0)}$

$t^{(0)}$ 太大，首次中心化困难；太小，则浪费外层迭代。

已有原始一对偶可行点

若初始对偶间隙为

$$\eta = f_0(\mathbf{x}^{(0)}) - g(\boldsymbol{\lambda}, \boldsymbol{\nu}),$$

可取

$$t^{(0)} = \frac{m}{\eta}.$$

首次中心化产生与现有证书同量级的间隙。

只有原始点

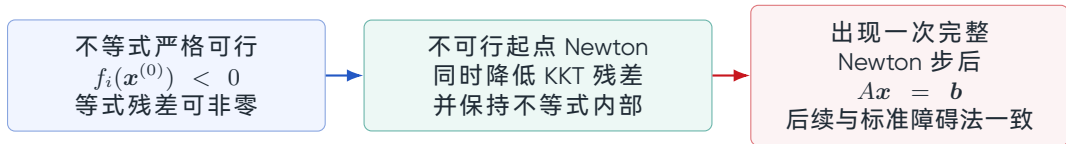
可选择 $t, \boldsymbol{\nu}$ 使中心性残差较小：

$$\min_{t, \boldsymbol{\nu}} \left\| t \nabla f_0(\mathbf{x}^{(0)}) + \nabla \phi(\mathbf{x}^{(0)}) + A^\top \boldsymbol{\nu} \right\|_2.$$

这是关于 $t, \boldsymbol{\nu}$ 的最小二乘问题；也可用 Hessian 度量构造仿射不变版本。

以不可行起点求中心点

若 $x^{(0)}$ 满足 $f_i(x^{(0)}) < 0$, 但 $Ax^{(0)} \neq b$, 可以对中心化问题采用 Lecture 05 的不可行起点 Newton 法。



该变体放宽了等式可行起点要求, 但仍需知道一个处于所有不等式定义域内部的点。

外层迭代次数

初始中心化后再完成 k 次中心化, 对偶间隙为

$$\frac{m}{\mu^k t^{(0)}}.$$

达到容差 ε 所需的额外中心化次数为

$$k = \left\lceil \frac{\log(m/(\varepsilon t^{(0)}))}{\log \mu} \right\rceil. \quad (11.13)$$

例

$m = 100$, $t^{(0)} = 1$, $\varepsilon = 10^{-6}$, $\mu = 10$:

$$k = \left\lceil \frac{\log 10^8}{\log 10} \right\rceil = 8.$$

含义

外层次数仅按精度倒数的对数增长。总成本还取决于每次中心化需要多少 Newton 步。

Boyd & Vandenberghe, *Convex Optimization*, §11.3.3, equation (11.13)

Newton 步与扰动 KKT

从扰动 KKT 中消去

$$\lambda_i = -\frac{1}{tf_i(x)}$$

后, 得到关于 x, ν 的非线性方程。其 Newton 线性化为

$$H\Delta x + A^\top \Delta \nu = -g, \quad A\Delta x = 0,$$

其中

$$H = \nabla^2 f_0 + \sum_i \frac{\nabla^2 f_i}{-tf_i} + \sum_i \frac{\nabla f_i \nabla f_i^\top}{tf_i^2},$$

$$g = \nabla f_0 + \sum_i \frac{\nabla f_i}{-tf_i}.$$

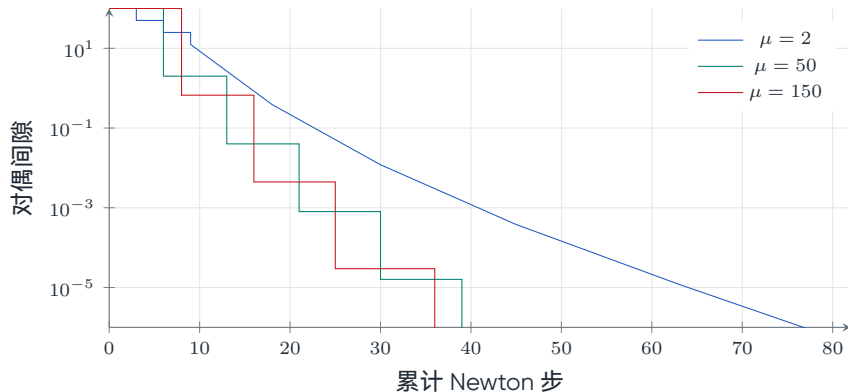
乘以 t 后恰好得到中心化问题的 Newton 方程。

结论

障碍法可以解释为: 先消去乘子 λ , 再对扰动 KKT 方程使用 Newton 法。§11.7 将不再消去 λ 。

教材例：小规模 LP

教材考虑 $G \in \mathbb{R}^{100 \times 50}$ 的随机严格可行 LP，初始对偶间隙 100，终止间隙 10^{-6} 。



台阶横段表示中心化 Newton 步；竖段表示 t 更新后证书间隙按 μ 缩小。曲线为教材现象的示意重绘。

Boyd & Vandenberghe, *Convex Optimization*, §11.3.2, Linear programming in inequality form; Figure 11.4

教材例：几何规划

凸形式 GP:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \log \left(\sum_{k=1}^{K_0} e^{\mathbf{a}_{0k}^\top \mathbf{x} + b_{0k}} \right) \\ \text{subject to} \quad & \log \left(\sum_{k=1}^{K_i} e^{\mathbf{a}_{ik}^\top \mathbf{x} + b_{ik}} \right) \leq 0, \quad i = 1, \dots, m. \end{aligned}$$

对应障碍

$$\phi(\mathbf{x}) = - \sum_{i=1}^m \log \left[- \log \left(\sum_{k=1}^{K_i} e^{\mathbf{a}_{ik}^\top \mathbf{x} + b_{ik}} \right) \right].$$

教材用 $n = 50, m = 100, K_i = 5$ 的随机实例比较不同 μ 。

观察

- ▣ 对偶间隙随累计 Newton 步近似线性收敛;
- ▣ μ 在较宽范围内对总步数影响不大;
- ▣ LP 与 GP 呈现相似的外层行为。

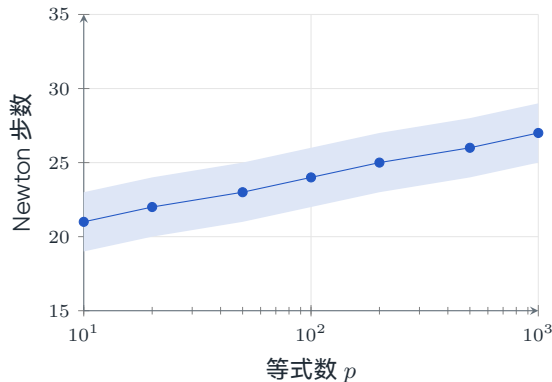
教材例：问题规模与 Newton 步数

教材生成标准形式 LP

$$\min c^T x \quad \text{subject to}$$

$$Ax = b, \quad x \succeq 0,$$

并令变量数 $n = 2p$ ，把等式数 p 从 10 增加到 1000。



变量规模扩大 100 倍时，观察到的 Newton 步数只从约 21 增至 27。

计算成本的真正来源

迭代次数增长缓慢，但每次 Newton 步需要解更大的线性系统。因此必须同时分析矩阵结构与分解成本。

Boyd & Vandenberghe, *Convex Optimization*, §11.3.2, A family of standard form LPs; Figures 11.7–11.8

数值例 4：中心路径的热启动

仍考虑二维 LP，依次取 $t = 0.5, 2, 10, 50$ ，并以上一个中心点作为下一次 Newton 法起点。

t	起点	Newton 步	最终目标
0.5	(1.5, 1.5)	5	-8.1447
2	上一中心点	7	-10.0634
10	上一中心点	10	-10.8012
50	上一中心点	8	-10.9600

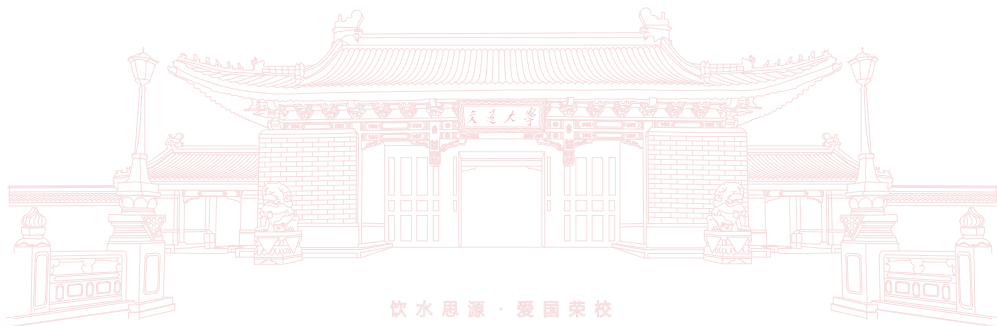
观察

- ▣ 迭代点始终严格可行；
- ▣ 对偶间隙上界按 m/t 缩小；
- ▣ 继续增大 t 主要改善靠近边界方向的精度。

数值步数取决于停止容差与回溯参数；表中的重点是热启动链，而非某个固定步数。

04

可行性与 Phase I

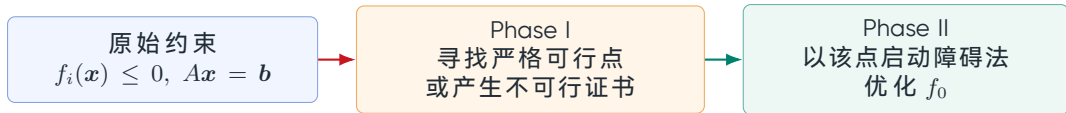


为何需要 Phase I

障碍法要求一个严格可行起点:

$$Ax^{(0)} = b, \quad f_i(x^{(0)}) < 0.$$

很多模型只给出约束，无法直接写出这样的点。Phase I 先解决一个辅助问题:



Phase I 的计算结果本身也有解释：它量化约束系统距离严格可行或不可行边界有多远。

基本 Phase I 问题

给定 $Ax^{(0)} = b$ 且 $x^{(0)}$ 位于所有 f_i 的定义域, 引入标量 s :

$$\begin{aligned} \min_{x, s} \quad & s \\ \text{subject to} \quad & f_i(x) \leq s, \quad i = 1, \dots, m, \\ & Ax = b. \end{aligned} \tag{11.19}$$

容易构造严格可行起点

保持 $x = x^{(0)}$, 只需选

$$s > \max_i f_i(x^{(0)}).$$

于是所有约束 $f_i(x) < s$ 。

s 的含义

对固定 x , 最小可行 s 为

$$\max_i f_i(x).$$

因此 Phase I 最小化最大约束违反量。

Phase I 最优值的三种情况

记辅助问题最优值为 $\bar{p}^*$ 。

情况	结论	可用的停止方式
$\bar{p}^* < 0$	原约束严格可行	一旦找到可行 (x, s) 且 $s < 0$, 即可停止并把 x 交给 Phase II。
$\bar{p}^* > 0$	原约束不可行	一旦得到正的对偶下界, 即可用对偶点给出不可行证书。
$\bar{p}^* = 0$	边界情形	若最小值达到, 约束可行但不严格可行; 若不达到, 约束不可行。数值上只能给出容差结论。

算法只能确定 $|\bar{p}^*| < \varepsilon$, 可断言 $f_i(x) \leq -\varepsilon$ 不可行, 而 $f_i(x) \leq \varepsilon$ 可行。

Boyd & Vandenberghe, *Convex Optimization*, §11.4.1, three cases for the phase I optimal value

数值例 5: 可行性由 $\bar{p}^*$ 判定

考虑两组只有常数不同的约束:

$$x_1 + x_2 \leq c, \quad x_1 \geq 1, \quad x_2 \geq 1.$$

Phase I 为

$$\min_{x_1, x_2, s} s \quad \text{subject to} \quad x_1 + x_2 - c \leq s, \quad 1 - x_1 \leq s, \quad 1 - x_2 \leq s.$$

由对称性令 $x_1 = x_2 = a$, 最优点满足

$$2a - c = 1 - a \implies a = \frac{c+1}{3}, \quad \bar{p}^* = \frac{2-c}{3}.$$

$c = 3$

$$a = \frac{4}{3}, \quad \bar{p}^* = -\frac{1}{3} < 0.$$

点 $(4/3, 4/3)$ 严格可行。

$c = 1$

$$a = \frac{2}{3}, \quad \bar{p}^* = \frac{1}{3} > 0.$$

三条约束不能同时满足。

自编数值例; 判别依据参见 Boyd & Vandenberghe, §11.4.1

最小化违反量之和

另一种 Phase I 形式是

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{s}} \quad & \mathbf{1}^\top \mathbf{s} \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq s_i, \quad i = 1, \dots, m, \\ & A\mathbf{x} = \mathbf{b}, \quad \mathbf{s} \succeq 0. \end{aligned} \tag{11.20}$$

对固定 $\mathbf{x}$, $s_i = \max\{f_i(\mathbf{x}), 0\}$, 所以目标是总违反量。

最大违反量

$\min \max_i f_i(\mathbf{x})$ 倾向于平衡最严重的几个约束。

违反量之和

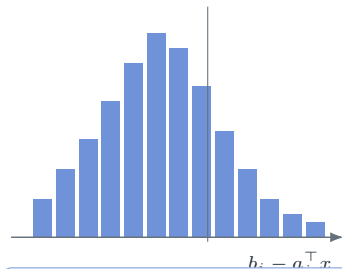
$\min \sum_i \max\{f_i(\mathbf{x}), 0\}$ 常产生稀疏违反：很多约束完全满足，仅少量约束承担冲突。

当系统不可行时，后者有助于识别一个尽可能大的可行约束子集。

例 11.4: 两种 Phase I 的结果

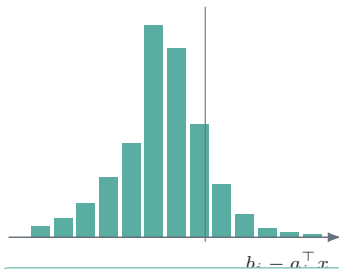
教材对一个不可行系统 $Gx \preceq h$ 取 $m = 100, n = 50$:

最小最大违反量



满足 39 条约束

最小违反量之和



满足 79 条约束

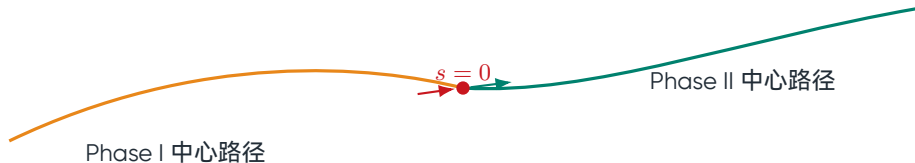
图为教材结论的示意重绘。两种目标都确认不可行，但第二种更清楚地定位了冲突约束。

Phase I 与 Phase II 中心路径的连接

加入一个目标上界 $f_0(\mathbf{x}) \leq M$:

$$\begin{aligned} \min_{\mathbf{x}, s} \quad & s \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq s, \quad i = 1, \dots, m, \\ & f_0(\mathbf{x}) \leq M, \quad A\mathbf{x} = \mathbf{b}. \end{aligned} \tag{11.21}$$

其中 $M > \max\{f_0(\mathbf{x}^{(0)}), p^*\}$ 。



在 $s = 0$ 处切换到 Phase II; 此时 $t = 1/(M - f_0(\mathbf{x}))$, 初始间隙不超过 $m(M - p^*)$ 。

Boyd & Vandenberghe, *Convex Optimization*, §11.4.1, Termination near the phase II central path and (11.22)

不可行起点 Newton 的 Phase I 变体

把原问题等价写成

$$f_i(\mathbf{x}) \leq s, \quad A\mathbf{x} = \mathbf{b}, \quad s = 0.$$

对

$$\min_{\mathbf{x}, s} t^{(0)} f_0(\mathbf{x}) - \sum_i \log(s - f_i(\mathbf{x})) \quad \text{subject to} \quad A\mathbf{x} = \mathbf{b}, s = 0$$

使用不可行起点 Newton 法。

- ▣ 可用任意 $\mathbf{x}$ 位于共同定义域内，并选 $s > \max_i f_i(\mathbf{x})$ 。
- ▣ 起点不满足 $s = 0$ ，但残差下降后最终会出现完整 Newton 步。
- ▣ 此后 $s = 0$ 且 $\mathbf{x}$ 严格可行，算法与标准障碍法相同。

局限

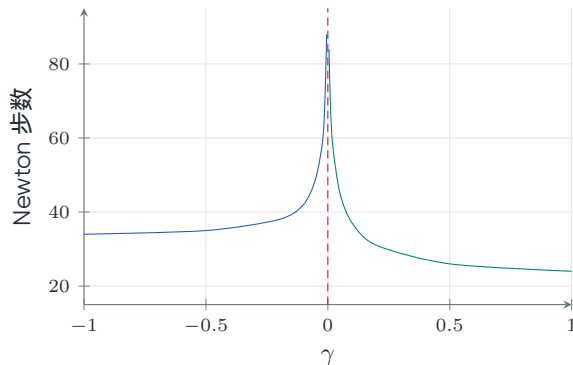
若原问题不可行，残差只会停在正值附近，难以给出可靠不可行停止准则。

接近可行性边界时更困难

教材考虑参数化系统

$$Gx \preceq h + \gamma \Delta h.$$

数据构造为： $\gamma > 0$ 时严格可行， $\gamma < 0$ 时不可行， $\gamma = 0$ 时可行但不严格。



基本 Phase I

远离 $\gamma = 0$ 时，用约几十个 Newton 步即可找到严格可行点或不可行证书。

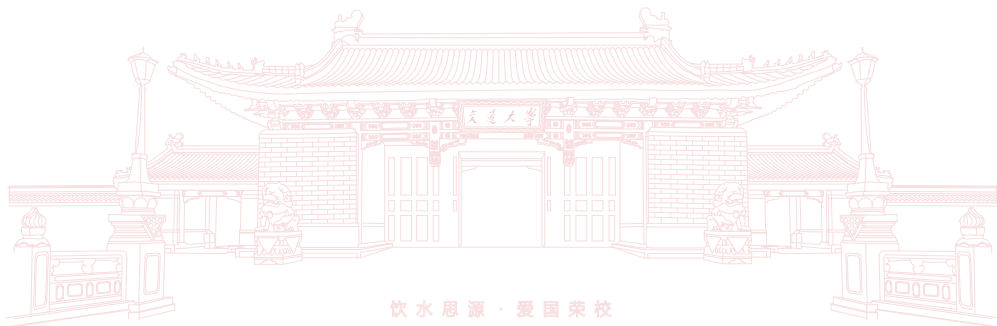
不可行起点法

对明显严格可行的问题可能更快；当 $\gamma \downarrow 0$ 时，步数可能急剧上升。

Boyd & Vandenberghe, *Convex Optimization*, §11.4.3, Figures 11.10–11.12

05

自协调复杂度



复杂度分析的假设

§11.5 用 Lecture 03 的自协调 Newton 理论回答：增大 t 后，中心化会不会越来越难？

两个假设

1. 对所有 $t \geq t^{(0)}$ ，函数

$$tf_0 + \phi$$

是闭的自协调函数。

2. 原问题的次水平集有界。

第二个假设保证中心化问题可解，并使相应 Hessian 正定。

- ▣ LP、QP、QCQP 满足该框架。
- ▣ 障碍法在更一般问题上仍可能表现很好，但以下最坏复杂度界未必适用。

Boyd & Vandenberghe, *Convex Optimization*, §11.5.1, self-concordance assumption

熵问题的自协调改写

原问题

$$\min \sum_{j=1}^n x_j \log x_j \quad \text{subject to} \quad F\mathbf{x} \preceq \mathbf{g}, \quad A\mathbf{x} = \mathbf{b}$$

的自然目标加障碍不一定闭且自协调。显式加入定义域约束 $\mathbf{x} \succeq 0$:

$$\min \sum_j x_j \log x_j \quad \text{subject to} \quad F\mathbf{x} \preceq \mathbf{g}, \quad A\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \succeq 0. \quad (11.23)$$

于是

$$tf_0 + \phi = t \sum_j x_j \log x_j - \sum_j \log x_j - \sum_i \log(g_i - \mathbf{f}_i^\top \mathbf{x}).$$

作用

函数 $ty \log y - \log y$ 在 $\mathbb{R}_{++}$ 上自协调。冗余约束不改变原问题，却使复杂度理论适用。

Boyd & Vandenberghe, *Convex Optimization*, §11.5.1, entropy maximization reformulation (11.23)

GP 的自协调改写

对每个单项式 $e^{\mathbf{a}_{ik}^\top \mathbf{x} + b_{ik}}$ 引入上界 y_{ik} :

$$e^{\mathbf{a}_{ik}^\top \mathbf{x} + b_{ik}} \leq y_{ik} \iff \mathbf{a}_{ik}^\top \mathbf{x} + b_{ik} - \log y_{ik} \leq 0.$$

凸形式 GP 可改写为

$$\begin{aligned} \min \quad & \sum_{k=1}^{K_0} y_{0k} \\ \text{subject to} \quad & \sum_{k=1}^{K_i} y_{ik} \leq 1, \quad i = 1, \dots, m, \\ & \mathbf{a}_{ik}^\top \mathbf{x} + b_{ik} - \log y_{ik} \leq 0, \quad y_{ik} \geq 0. \end{aligned}$$

原始 log-sum-exp 形式
障碍法可用
自协调性不明显

引入 y_{ik}

提升变量后的形式
目标线性
标准障碍闭且自协调

Boyd & Vandenberghe, *Convex Optimization*, §11.5.1, geometric programming reformulation

每次中心化的 Newton 步数

闭且严格凸的自协调函数 q , 从起点 x 求至规定 Newton 精度所需步数满足

$$\frac{q(x) - q^*}{\gamma} + c, \quad \frac{1}{\gamma} = \frac{20 - 8\alpha}{\alpha\beta(1 - 2\alpha)^2}, \quad c \approx 6. \quad (11.24)$$

从 $x^*(t)$ 热启动求 $x^*(\mu t)$ 时, 中心化目标差可估计为

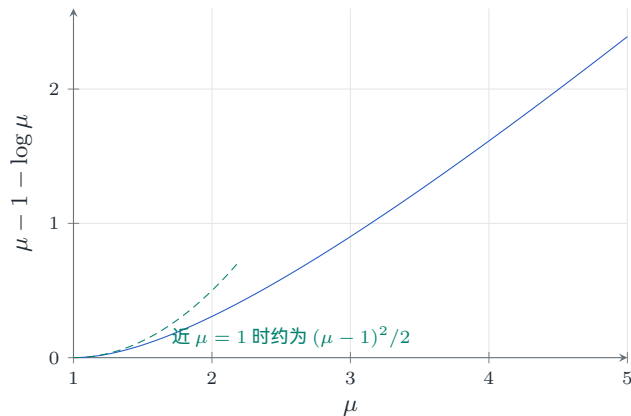
$$\mu t f_0(x^*(t)) + \phi(x^*(t)) - \mu t f_0(x^*(\mu t)) - \phi(x^*(\mu t)) \leq m(\mu - 1 - \log \mu).$$

因此每次外层迭代至多需要

$$\boxed{\frac{m(\mu - 1 - \log \mu)}{\gamma} + c} \quad (11.26)$$

个 Newton 步。这个界与 t 、变量维数 n 和具体数据无关。

$$\mu - 1 - \log \mu$$



- ▣ μ 接近 1: 单次中心化容易。
- ▣ μ 增大: 单次中心化的理论上界近似线性增长。
- ▣ 界虽然保守, 却证明中心化难度不会随 $t \rightarrow \infty$ 发散。

Boyd & Vandenberghe, *Convex Optimization*, §11.5.2 and Figure 11.13

总 Newton 步数

把外层次数与每次中心化上界相乘:

$$N = \left\lceil \frac{\log(m/(t^{(0)}\varepsilon))}{\log \mu} \right\rceil \left(\frac{m(\mu - 1 - \log \mu)}{\gamma} + c \right). \quad (11.27)$$

关于精度

固定 m, μ 时,

$$N = O\left(\log \frac{1}{\varepsilon}\right).$$

这对应至少线性收敛的最坏界。

理论界用于说明可解性与维数依赖，不用于预测实际的几十次 Newton 步。

关于 μ

$\mu \downarrow 1$ 时外层次数大; μ 很大时单次中心化界大。理论最优值通常非常接近 1。

理论上的 $\sqrt{m}$ 复杂度

取

$$\mu = 1 + \frac{1}{\sqrt{m}}, \quad (11.28)$$

利用

$$\mu - 1 - \log \mu \leq \frac{1}{2m}, \quad \log \mu \geq \frac{\log 2}{\sqrt{m}},$$

可得

$$N \leq \left\lceil \sqrt{m} \log_2 \frac{m}{t^{(0)} \varepsilon} \right\rceil \left(\frac{1}{2\gamma} + c \right) = O\left(\sqrt{m} \log \frac{1}{\varepsilon} \right). \quad (11.29)$$

理论参数与实践参数

$\mu = 1 + 1/\sqrt{m}$ 只用于改善最坏界；实际通常选明显更大的 μ ，例如 10–20，并观察到更少的总 Newton 步。

可行性问题的复杂度

令

$$F = \max_i f_i(0) > 0, \quad G = \max_i \|\nabla f_i(0)\|_2,$$

并假设所有可行点位于半径 R 的球中。通过添加一个冗余线性约束，可把 $(x, s) = (0, s_0)$ 放在 Phase I 中心路径上，其中

$$s_0 = mGR + F.$$

若 Phase I 最优值为 $\bar{p}^* \neq 0$ ，判定可行或不可行所需 Newton 步数至多具有量级

$$O\left(\sqrt{m} \log \frac{m^2 GR}{|\bar{p}^*|}\right). \quad (11.35)$$

难度参数

$|\bar{p}^*|$ 越小，系统越接近“严格可行与不可行”的分界，判定越困难。这与 §11.4 的数值现象一致。

Phase I 与 Phase II 的总复杂度

组合分析使用带两个冗余约束的 Phase I:

$$f_0(x) \leq M, \quad \mathbf{a}^\top \mathbf{x} \leq 1.$$

若问题可行, Phase I 在 $s = 0$ 处进入 Phase II 中心路径。

Phase I

$$N_{\text{I}} = O\left(\sqrt{m+2} \log \frac{(m+1)(m+2)GR}{|\bar{p}^*|}\right).$$

找到严格可行点或证实不可行。

Phase II

$$N_{\text{II}} = O\left(\sqrt{m+1} \log \frac{(m+1)(M-p^*)}{\varepsilon}\right).$$

把间隙降至 ε 。

总 Newton 步数由 $N_{\text{I}} + N_{\text{II}}$ 控制; 问题接近可行性边界时, Phase I 项会变大。

Boyd & Vandenberghe, *Convex Optimization*, §11.5.5, equations (11.36)–(11.37)

复杂度结论的适用范围

理论给出的保证

- ▣ 每次中心化步数有与 t 无关的统一上界;
- ▣ 精度代价按 $\log(1/\varepsilon)$ 增长;
- ▣ 恰当理论参数下得到 $\sqrt{m}$ 量级。

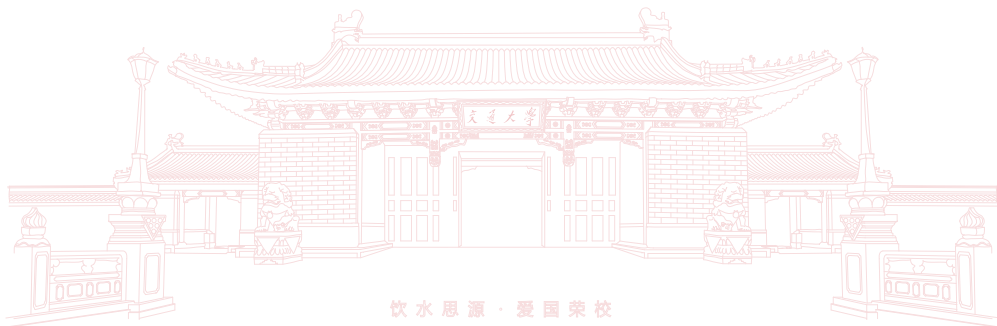
数值实践中的现象

- ▣ 最坏界远高于实际步数;
- ▣ Newton 步数常随规模增长很慢;
- ▣ 单步线性代数往往决定总时间。

自协调改写是否在具体实现中更快，需要结合模型结构判断；其确定优势是能够获得严格复杂度证书。

06

广义不等式与锥规划



广义不等式问题

考虑

$$\begin{aligned} \min_{\mathbf{x}} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \preceq_{K_i} 0, \quad i = 1, \dots, q, \\ & A\mathbf{x} = \mathbf{b}, \end{aligned} \tag{11.38}$$

其中 $K_i \subseteq \mathbb{R}^{k_i}$ 为 proper cone, f_i 对 K_i 凸。

KKT 条件

$$\begin{aligned} A\mathbf{x}^* &= \mathbf{b}, & f_i(\mathbf{x}^*) &\preceq_{K_i} 0, \\ \boldsymbol{\lambda}_i^* &\succeq_{K_i^*} 0, \quad \nabla f_0(\mathbf{x}^*) + \sum_i Df_i(\mathbf{x}^*)^\top \boldsymbol{\lambda}_i^* + A^\top \boldsymbol{\nu}^* &= 0, \\ (\boldsymbol{\lambda}_i^*)^\top f_i(\mathbf{x}^*) &= 0. \end{aligned} \tag{11.39}$$

标量不等式是 $K_i = \mathbb{R}_+$ 的特殊情形; SOCP 与 SDP 由其他锥产生。

广义对数函数

proper cone K 上的广义对数 $\psi : \text{int } K \rightarrow \mathbb{R}$ 满足

$$\psi(s\mathbf{x}) = \psi(\mathbf{x}) + \theta \log s, \quad s > 0,$$

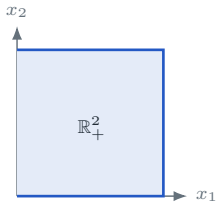
其中 $\theta > 0$ 为 degree。定义还要求 ψ 闭、严格凹且二阶连续可微；由对数齐次性可推出

$$\nabla \psi(\mathbf{x}) \in \text{int } K^*, \quad \mathbf{x}^\top \nabla \psi(\mathbf{x}) = \theta. \tag{11.40}$$

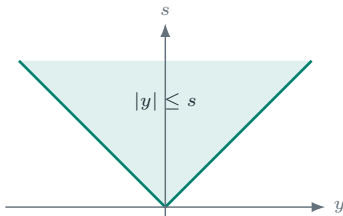
锥 K	广义对数 ψ	degree
$\mathbb{R}_+^k$	$\sum_{j=1}^k \log x_j$	k
二阶锥	$\log(s^2 - \ \mathbf{y}\ _2^2)$	2
$\mathbb{S}_+^p$	$\log \det X$	p

后续复杂度分析还需验证相应障碍 $-\psi$ 的自协调性；这不是“广义对数”定义本身的一部分。

三类锥的几何



逐分量非负



二阶锥截面，坐标等比例

$X \succeq 0$
全部特征值非负

半正定锥

广义障碍在锥边界处发散，并使用锥的对偶结构生成对偶变量。

广义障碍与中心路径

令 ψ_i 是 K_i 的广义对数, degree 为 θ_i 。定义

$$\phi(\mathbf{x}) = -\sum_{i=1}^q \psi_i(-f_i(\mathbf{x})), \quad \text{dom } \phi = \{\mathbf{x} \mid f_i(\mathbf{x}) \prec_{K_i} 0\}.$$

中心点最小化

$$\min_{A\mathbf{x}=\mathbf{b}} t f_0(\mathbf{x}) + \phi(\mathbf{x})$$

并满足

$$t \nabla f_0(\mathbf{x}^*(t)) + \sum_i Df_i(\mathbf{x}^*(t))^{\top} \nabla \psi_i(-f_i(\mathbf{x}^*(t))) + A^{\top} \hat{\mathbf{v}} = 0. \quad (11.41)$$

与标量情形的对应

$\mathbb{R}_+^m$ 上取 $\psi(s) = \sum_i \log s_i$, 上式退化为普通对数障碍中心路径。

广义中心点的对偶间隙

定义

$$\lambda_i^*(t) = \frac{1}{t} \nabla \psi_i(-f_i(\mathbf{x}^*(t))), \quad \nu^*(t) = \frac{\hat{\nu}}{t}. \quad (11.42)$$

由 $\mathbf{x}^\top \nabla \psi_i(\mathbf{x}) = \theta_i$,

$$(\lambda_i^*(t))^\top f_i(\mathbf{x}^*(t)) = -\frac{\theta_i}{t}. \quad (11.43)$$

令

$$\theta = \sum_{i=1}^q \theta_i,$$

则原始一对偶间隙为

$$\boxed{\theta/t}.$$



Boyd & Vandenberghe, *Convex Optimization*, §11.6.1, equations (11.42)–(11.43)

例 11.8: SOCP 障碍

SOCP

$$\min \mathbf{f}^\top \mathbf{x} \quad \text{subject to} \quad \|A_i \mathbf{x} + \mathbf{b}_i\|_2 \leq \mathbf{c}_i^\top \mathbf{x} + d_i, \quad i = 1, \dots, q$$

的障碍为

$$\phi(\mathbf{x}) = - \sum_{i=1}^q \log \left[(\mathbf{c}_i^\top \mathbf{x} + d_i)^2 - \|A_i \mathbf{x} + \mathbf{b}_i\|_2^2 \right]. \quad (11.45)$$

每个二阶锥的 degree 为 2, 故

$$\theta = 2q, \quad f_0(\mathbf{x}^*(t)) - p^* \leq \frac{2q}{t}.$$

严格内部

Newton 线搜索必须维持

$$\|A_i \mathbf{x} + \mathbf{b}_i\|_2 < \mathbf{c}_i^\top \mathbf{x} + d_i.$$

应用例 6: 应急设施选址

为三个地点

$$\mathbf{a}_1 = (0, 0), \quad \mathbf{a}_2 = (2, 0), \quad \mathbf{a}_3 = (0, 2)$$

选择设施位置 $\mathbf{x} \in \mathbb{R}^2$, 最小化最远距离:

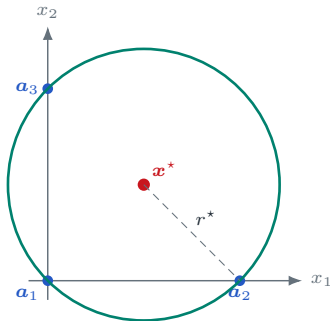
$$\min_{\mathbf{x}, r} r \quad \text{subject to} \quad \|\mathbf{x} - \mathbf{a}_i\|_2 \leq r, \quad i = 1, 2, 3.$$

这是含三个二阶锥约束的 SOCP。障碍为

$$-\sum_{i=1}^3 \log(r^2 - \|\mathbf{x} - \mathbf{a}_i\|_2^2).$$

最优解

$$\mathbf{x}^* = (1, 1), \quad r^* = \sqrt{2}.$$



横纵坐标单位长度一致。

自编设施选址例; SOCP 障碍参见 Boyd & Vandenberghe, Example 11.8

例 11.9: SDP 障碍

考虑

$$\min_{\mathbf{x}} \mathbf{c}^\top \mathbf{x} \quad \text{subject to} \quad F(\mathbf{x}) = F_0 + \sum_{j=1}^n x_j F_j \preceq 0,$$

其中 $F_j \in \mathbb{S}^p$ 。对半正定锥使用 $\log \det$:

$$\phi(\mathbf{x}) = -\log \det(-F(\mathbf{x})).$$

梯度

$$\frac{\partial \phi}{\partial x_j} = -\operatorname{tr}(F(\mathbf{x})^{-1} F_j) = \operatorname{tr}((-F)^{-1} F_j).$$

这里 degree 等于矩阵阶数 p , 而不是标量变量数 n 。

对偶点与间隙

$$Z^*(t) = \frac{1}{t} (-F(\mathbf{x}^*(t)))^{-1} \succ 0, \quad \text{gap} = \frac{p}{t}.$$

数值例 7: 二维半正定设计

考虑

$$\min_{x_1, x_2} x_1 + x_2 \quad \text{subject to} \quad \begin{bmatrix} x_1 & 0 \\ 0 & x_2 \end{bmatrix} - \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix} \succeq 0.$$

半正定条件等价于

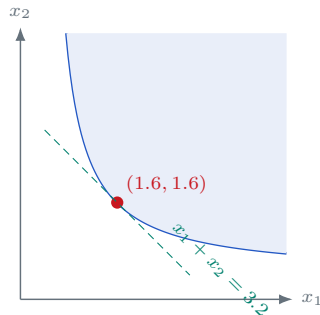
$$x_1 \geq 1, \quad x_2 \geq 1, \quad (x_1 - 1)(x_2 - 1) \geq 0.36.$$

对称性给出

$$x_1^* = x_2^* = 1.6, \quad p^* = 3.2.$$

严格内部的障碍为

$$-\log((x_1 - 1)(x_2 - 1) - 0.36).$$



自编 SDP 数值例; 障碍与对偶点参见 Boyd & Vandenberghe, Example 11.9

广义障碍法与 Phase I

普通障碍法中的 m 全部替换为总 degree θ :

$$\text{停止条件 } \frac{\theta}{t} < \varepsilon, \quad \text{外层次数 } \left\lceil \frac{\log(\theta/(t^{(0)}\varepsilon))}{\log \mu} \right\rceil.$$

广义 Phase I 选择 $e_i \succ_{K_i} 0$, 求解

$$\begin{aligned} \min_{x,s} \quad & s \\ \text{subject to} \quad & f_i(x) \preceq_{K_i} s e_i, \quad i = 1, \dots, q, \\ & Ax = b. \end{aligned}$$

- ▣ 最优值为负: 得到严格锥可行点。
- ▣ 最优值为正: 对偶点给出广义不等式系统的不可行证书。
- ▣ 最优值接近零: 问题接近严格可行性的边界。

教材数值例：SOCP 与 SDP

§11.6.3 对随机严格原始一对偶可行的 SOCP 与 SDP 使用障碍法。

小规模 SOCP

$n = 50$, $q = 50$, 每个 $A_i \in \mathbb{R}^{5 \times 50}$ 。初始间隙 100, 取不同 μ :

- ▣ 每次中心化所需 Newton 步近似恒定;
- ▣ $\mu \geq 10$ 时总步数约几十次。

小规模与系列 SDP

小例中 $x \in \mathbb{R}^{100}$, 矩阵阶数 $p = 100$ 。系列例

$$\min \mathbf{1}^\top x \quad \text{subject to} \quad A + \text{diag}(x) \succeq 0$$

把 n 从 10 增至 1000, 平均 Newton 步仅缓慢增长。

三类问题 LP、SOCP、SDP 的外层曲线相似; 差别主要体现在单步线性代数。

广义不等式的复杂度

若 $tf_0 + \phi$ 闭且自协调、次水平集有界，则

$$\begin{aligned} \mu t f_0(\mathbf{x}^*(t)) + \phi(\mathbf{x}^*(t)) - \mu t f_0(\mathbf{x}^*(\mu t)) - \phi(\mathbf{x}^*(\mu t)) \\ \leq \theta(\mu - 1 - \log \mu). \end{aligned} \tag{11.48}$$

所以每次中心化的 Newton 步数不超过

$$\frac{\theta(\mu - 1 - \log \mu)}{\gamma} + c.$$

统一结论

普通不等式的复杂度理论保持不变，只需将约束数 m 替换成锥障碍的总 degree θ 。

SOCP 每个锥贡献 2； $p \times p$ PSD 锥贡献 p 。

对偶锥上的广义对数

令 ψ 是 K 上 degree 为 θ 的广义对数。定义其对偶广义对数

$$\bar{\psi}(\mathbf{y}) = \inf_{\mathbf{x} \succ_{K^*} 0} \left(\mathbf{y}^\top \mathbf{x} - \psi(\mathbf{x}) \right), \quad \mathbf{y} \in \text{int } K^*. \tag{11.49}$$

得到 Fenchel 型不等式

$$\bar{\psi}(\mathbf{y}) + \psi(\mathbf{x}) \leq \mathbf{x}^\top \mathbf{y}, \tag{11.50}$$

且等号成立当且仅当 $\nabla \psi(\mathbf{x}) = \mathbf{y}$ 。

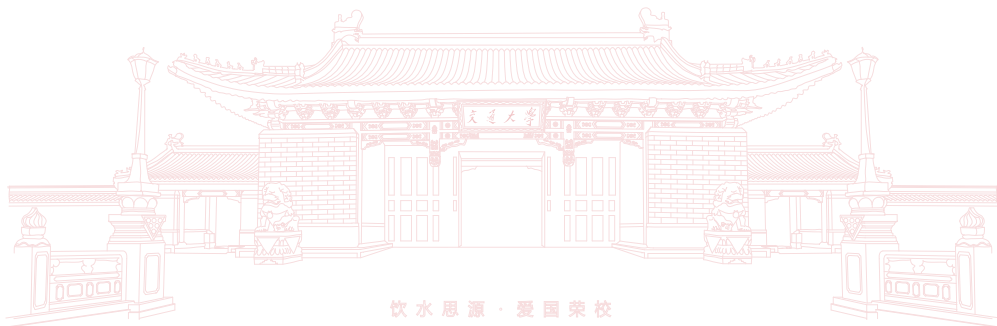
锥	对偶广义对数
二阶锥	原 $\log(s^2 - \ \mathbf{y}\ ^2)$ 加常数
半正定锥	$\log \det Y + p$, 仍为 log-det 形式

该不等式是证明广义中心化上界 (11.48) 的关键。

Boyd & Vandenberghe, *Convex Optimization*, §11.6.4, equations (11.49)–(11.52), Examples 11.12–11.13

07

原始-对偶内点法



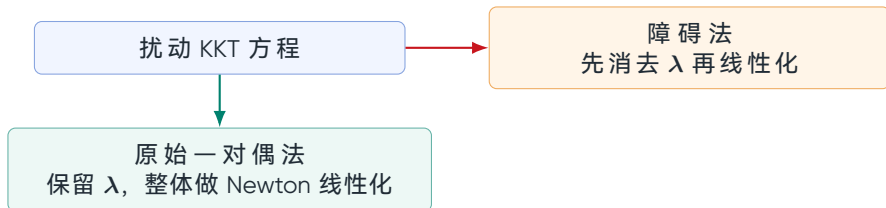
障碍法与原始一对偶方法

障碍法

- ▣ 外层更新 t , 内层求中心点;
- ▣ 每个内层点保持原始可行;
- ▣ 中心化结束时构造对偶点;
- ▣ 适合解释中心路径与复杂度。

原始一对偶方法

- ▣ 只有一个迭代循环;
- ▣ 同时更新 x, λ, ν ;
- ▣ 中间点可以原始或对偶不可行;
- ▣ 高精度时通常更快。



三个 KKT 残差

令

$$\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}) \quad \cdots \quad f_m(\mathbf{x})]^\top, \quad D\mathbf{f}(\mathbf{x}) = \begin{bmatrix} \nabla f_1(\mathbf{x})^\top \\ \vdots \\ \nabla f_m(\mathbf{x})^\top \end{bmatrix}.$$

扰动 KKT 残差写成

$$\mathbf{r}_t(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{bmatrix} r_{\text{dual}} \\ r_{\text{cent}} \\ r_{\text{pri}} \end{bmatrix} = \begin{bmatrix} \nabla f_0(\mathbf{x}) + D\mathbf{f}(\mathbf{x})^\top \boldsymbol{\lambda} + A^\top \boldsymbol{\nu} \\ -\text{diag}(\boldsymbol{\lambda}) \mathbf{f}(\mathbf{x}) - (1/t) \mathbf{1} \\ A\mathbf{x} - \mathbf{b} \end{bmatrix}. \quad (11.53)$$

对偶残差
驻点条件

中心性残差
扰动互补条件

原始残差
等式可行性

Boyd & Vandenberghe, *Convex Optimization*, §11.7.1, equation (11.53)

原始一对偶搜索方向

对固定 t , 在 $f_i(x) < 0$ 、 $\lambda \succ 0$ 的点线性化 $r_t = 0$:

$$\begin{bmatrix} \nabla^2 f_0 + \sum_i \lambda_i \nabla^2 f_i & Df^\top & A^\top \\ -\text{diag}(\lambda)Df & -\text{diag}(\mathbf{f}) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta \nu \end{bmatrix} = - \begin{bmatrix} r_{\text{dual}} \\ r_{\text{cent}} \\ r_{\text{pri}} \end{bmatrix}. \quad (11.54)$$

若 $r_{\text{pri}} = 0$, 则

$$A\Delta x = 0,$$

方向同时依赖当前原始变量与乘子。边界松弛和乘子共同决定曲率与尺度。

所以任何步长都保持线性等式可行。

实际实现通常先消去 $\Delta \lambda$, 再求具有 KKT 结构的降维系统。

Boyd & Vandenberghe, *Convex Optimization*, §11.7.1, equation (11.54)

与障碍法 Newton 方向的比较

消去 $\Delta\lambda$ 后, 原始一对偶方向满足

$$\begin{bmatrix} H_{\text{pd}} & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_{\text{pd}} \\ \Delta \boldsymbol{\nu}_{\text{pd}} \end{bmatrix} = - \begin{bmatrix} \nabla f_0 + \sum_i \frac{\nabla f_i}{-t f_i} + A^\top \boldsymbol{\nu} \\ r_{\text{pri}} \end{bmatrix},$$

其中

$$H_{\text{pd}} = \nabla^2 f_0 + \sum_i \lambda_i \nabla^2 f_i + \sum_i \frac{\lambda_i}{-f_i} \nabla f_i \nabla f_i^\top. \quad (11.56)$$

方向何时一致

当

$$-\lambda_i f_i(\mathbf{x}) = 1/t, \quad \forall i,$$

即当前点满足扰动互补条件时, $H_{\text{pd}} = (1/t)H_{\text{bar}}$, 两种原始方向相同。

代理对偶间隙

中间迭代点未必原始、对偶可行，无法直接计算真正对偶间隙。定义

$$\widehat{\eta}(x, \lambda) = -f(x)^\top \lambda, \quad f(x) \prec 0, \quad \lambda \succeq 0. \quad (11.59)$$

若 $r_{\text{pri}} = r_{\text{dual}} = 0$ ，则

$$\widehat{\eta} = f_0(x) - g(\lambda, \nu),$$

它就是真正的原始一对偶间隙。

与当前点对应的中心参数是

$$t_{\text{current}} = \frac{m}{\widehat{\eta}}.$$

算法取更大的目标值

$$t = \mu \frac{m}{\widehat{\eta}}.$$

终止判断

需要同时检查 $\|r_{\text{pri}}\|_2$ 、 $\|r_{\text{dual}}\|_2$ 与 $\widehat{\eta}$ ；只看代理间隙并不足够。

Boyd & Vandenberghe, *Convex Optimization*, §11.7.2, surrogate duality gap (11.59)

原始一对偶内点算法

Algorithm 11.2

给定 $f(x) \prec 0$ 、 $\lambda \succ 0$ 、 $\mu > 1$ 、容差 $\varepsilon_{\text{feas}}, \varepsilon > 0$:

1. 计算 $\hat{\eta} = -f(x)^\top \lambda$, 取

$$t = \mu m / \hat{\eta}.$$

2. 解线性系统 (11.54), 得到 $(\Delta x, \Delta \lambda, \Delta \nu)$ 。

3. 通过回溯线搜索选择步长 s , 更新

$$(x, \lambda, \nu) \leftarrow (x, \lambda, \nu) + s(\Delta x, \Delta \lambda, \Delta \nu).$$

当

$$\|r_{\text{pri}}\|_2 \leq \varepsilon_{\text{feas}}, \quad \|r_{\text{dual}}\|_2 \leq \varepsilon_{\text{feas}}, \quad \hat{\eta} \leq \varepsilon$$

时停止。

原始一对偶线搜索

先保证乘子正:

$$s_{\max} = \min \left\{ 1, \min_{\Delta \lambda_i < 0} \left(-\frac{\lambda_i}{\Delta \lambda_i} \right) \right\}, \quad s \leftarrow 0.99 s_{\max}.$$

原始一对偶线搜索

先保证乘子正:

$$s_{\max} = \min \left\{ 1, \min_{\Delta \lambda_i < 0} \left(-\frac{\lambda_i}{\Delta \lambda_i} \right) \right\}, \quad s \leftarrow 0.99 s_{\max}.$$

再反复令 $s \leftarrow \beta s$, 直到

$$f(x + s\Delta x) < 0.$$

原始一对偶线搜索

先保证乘子正:

$$s_{\max} = \min \left\{ 1, \min_{\Delta\lambda_i < 0} \left(-\frac{\lambda_i}{\Delta\lambda_i} \right) \right\}, \quad s \leftarrow 0.99s_{\max}.$$

再反复令 $s \leftarrow \beta s$, 直到

$$f(\mathbf{x} + s\Delta\mathbf{x}) < 0.$$

最后继续回溯, 直到

$$\|r_t(\mathbf{y} + s\Delta\mathbf{y})\|_2 \leq (1 - \alpha s)\|r_t(\mathbf{y})\|_2.$$

乘子正性

不等式严格内部

总 KKT 残差充分下降

数值例 8: 两变量标准 LP 的中心对

考虑

$$\min_{x_1, x_2} x_1 + 2x_2 \quad \text{subject to} \quad x_1 + x_2 = 1, \quad \mathbf{x} \succeq 0.$$

KKT 扰动条件为

$$\mathbf{c} + A^\top \nu - \boldsymbol{\lambda} = 0, \quad x_1 + x_2 = 1, \quad x_i \lambda_i = 1/t.$$

解得下列中心原始一对偶对:

t	x_1	x_2	λ_1	λ_2	gap = $2/t$
1	0.6180	0.3820	1.6180	2.6180	2
10	0.9099	0.0901	0.1099	1.1099	0.2

当 $t \rightarrow \infty$, $\mathbf{x}^*(t) \rightarrow (1, 0)$, 第二个非负约束活跃, 其乘子趋向 1。

自编数值例; 扰动 KKT 结构参见 Boyd & Vandenberghe, §§11.2.2 and 11.7

教材例：小规模 LP 与 GP

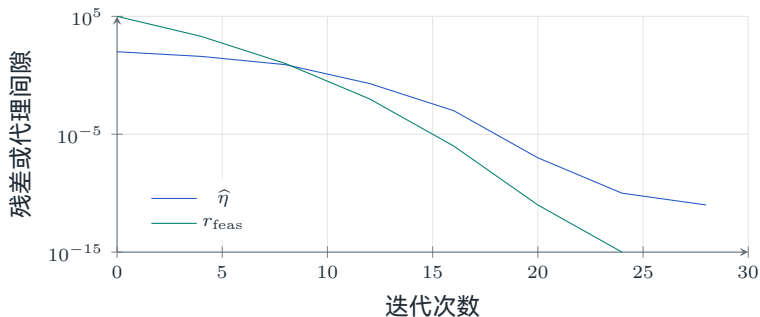
教材用与 §11.3.2 相同的 LP、GP 数据测试原始-对偶方法：

$$\mu = 10, \quad \alpha = 0.01, \quad \beta = 0.5, \quad \varepsilon = 10^{-8}.$$

起点只要求不等式严格满足，并取

$$\lambda_i^{(0)} = -1/f_i(\mathbf{x}^{(0)}),$$

使初始代理间隙为 100。



两类问题都在约二十余次迭代内把残差和间隙降至很小。

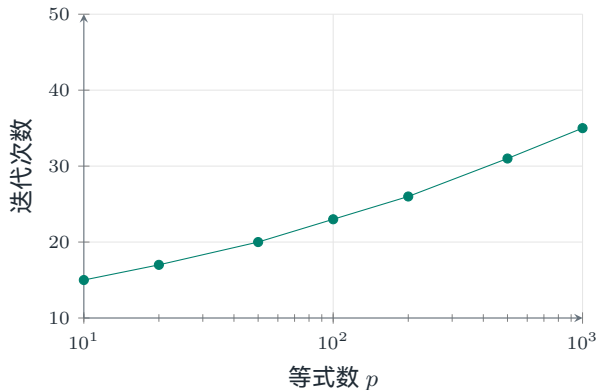
Boyd & Vandenberghe, *Convex Optimization*, §11.7.4, Small LP and GP; Figures 11.21–11.22

教材例：LP 规模增长

对 $n = 2p$ 的标准 LP 系列，教材从

$$\mathbf{x}^{(0)} = \mathbf{1}, \quad \boldsymbol{\lambda}^{(0)} = \mathbf{1}, \quad \nu^{(0)} = 0$$

启动原始一对偶方法，以 $\varepsilon = 10^{-8}$ 终止。



迭代数从约 15 增至 35，随规模近似对数增长。

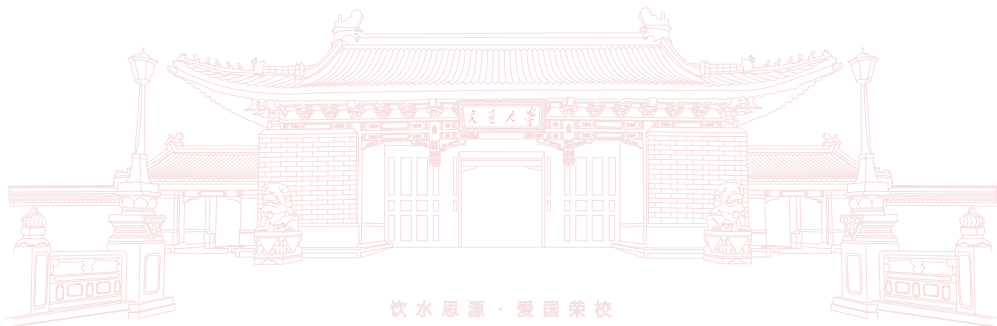
比较

虽然从不可行点起步、终止精度更高，原始一对偶方法的迭代数仍与障碍法处于同一数量级。

Boyd & Vandenberghe, *Convex Optimization*, §11.7.4, A family of LPs; Figure 11.23

08

实现



饮水思源 · 爱国荣校

KKT 线性系统决定单步成本

障碍法中心化 Newton 方程统一写成

$$\begin{bmatrix} H & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x} \\ \nu_{\text{nt}} \end{bmatrix} = - \begin{bmatrix} \mathbf{g} \\ 0 \end{bmatrix}, \quad (11.60)$$

其中

$$H = t \nabla^2 f_0 + \sum_i \frac{\nabla f_i \nabla f_i^\top}{f_i^2} + \sum_i \frac{\nabla^2 f_i}{-f_i},$$
$$\mathbf{g} = t \nabla f_0 + \sum_i \frac{\nabla f_i}{-f_i}.$$

实现目标

避免形成和分解一个无结构的稠密 $(n + p) \times (n + p)$ 矩阵。应利用消元、稀疏、带状、块对角或对角加低秩结构。

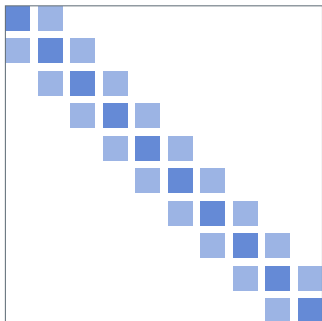
原始一对偶 Newton 方程具有相同的 KKT 骨架，以下结构同样适用。

Boyd & Vandenberghe, *Convex Optimization*, §11.8, equation (11.60)

稀疏与对角加低秩

稀疏问题

若每个 f_i 只依赖少量变量，则梯度、Hessian 和 A 都稀疏。使用稀疏对称不定分解或消元树，可显著减少填充。



可分目标和少量线性约束

若 $\nabla^2 f_0 = D$ 为对角矩阵， f_i 仿射，则

$$H = D + UCU^\top.$$

可用矩阵求逆引理把 n 维系统化为秩数维系统。

几条稠密约束通常只产生少量稠密行列，仍可在稀疏分解中专门处理。

Boyd & Vandenberghe, *Convex Optimization*, §11.8, Sparse problems; separable objective and few linear constraints

标准形式 LP 的消元

对

$$\min \mathbf{c}^\top \mathbf{x} \quad \text{subject to} \quad A\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \succeq 0,$$

中心化目标为 $t\mathbf{c}^\top \mathbf{x} - \sum_j \log x_j$ 。令 $X = \text{diag}(\mathbf{x})$, Newton 系统是

$$\begin{bmatrix} X^{-2} & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x} \\ \boldsymbol{\nu} \end{bmatrix} = \begin{bmatrix} -t\mathbf{c} + X^{-1}\mathbf{1} \\ 0 \end{bmatrix}.$$

先消去 $\Delta \mathbf{x}$:

$$\Delta \mathbf{x} = -tX^2\mathbf{c} + \mathbf{x} - X^2A^\top\boldsymbol{\nu},$$

再解 Schur 补系统

$$\boxed{AX^2A^\top\boldsymbol{\nu} = -tAX^2\mathbf{c} + \mathbf{b}}.$$

A 满行秩时 $AX^2A^\top \succ 0$, 可用 Cholesky; 若 A 稀疏, Schur 补通常也可稀疏处理。

数值例 9: 标准 LP 的一次消元

取

$$A = \begin{bmatrix} 1 & 1 \end{bmatrix}, \quad \mathbf{b} = 1, \quad \mathbf{c} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix}, \quad t = 1.$$

此时

$$X^2 = \text{diag}(0.36, 0.16), \quad AX^2A^\top = 0.52,$$

且

$$-tAX^2\mathbf{c} + \mathbf{b} = -(0.36 + 0.32) + 1 = 0.32.$$

故

$$\nu_{\text{nt}} = 0.61538,$$

$$\Delta\mathbf{x} = -X^2\mathbf{c} + \mathbf{x} - X^2A^\top\nu_{\text{nt}} = \begin{bmatrix} 0.01846 \\ -0.01846 \end{bmatrix}.$$

检查 $A\Delta\mathbf{x} = 0$: Newton 方向自动保持 $x_1 + x_2 = 1$ 。

ℓ_1 逼近转化为 LP

$$\min_x \|Ax - b\|_1$$

引入 $y \in \mathbb{R}^m$:

$$\begin{aligned} \min_{x,y} \quad & \mathbf{1}^\top y \\ \text{subject to} \quad & Ax - b \preceq y, \\ & -(Ax - b) \preceq y. \end{aligned}$$

令

$$D_1 = \text{diag}(b - Ax + y)^{-2}, \quad D_2 = \text{diag}(-b + Ax + y)^{-2}.$$

消去 Δy 后, Δx 满足

$$A^\top D A \Delta x = -A^\top g, \tag{11.61}$$

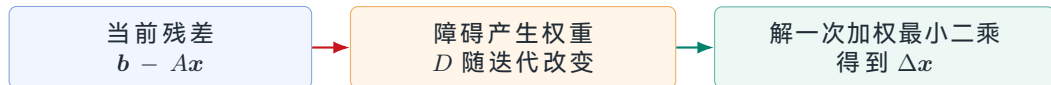
其中

$$D = 4D_1 D_2 (D_1 + D_2)^{-1} = 2 \left(\text{diag}(y)^2 + \text{diag}(b - Ax)^2 \right)^{-1}.$$

ℓ_1 内点步是加权最小二乘

方程 (11.61) 正是加权最小二乘的正规方程：

$$\min_{\Delta x} \left\| D^{1/2} (A \Delta x + D^{-1} g) \right\|_2^2.$$



- ▣ 若 A 是稀疏矩阵、卷积算子或快速变换，结构可以在每次加权最小二乘中复用。
- ▣ ℓ_1 拟合的成本可理解为“少量权重变化的最小二乘求解”。

应用例 10: 稳健传感器标定

五次标定观测为

$$a = (1, 2, 3, 4, 5), \quad b = (1.1, 1.9, 3.2, 3.9, 9.0).$$

最后一个观测含明显离群误差。比较

$$\min_{\theta} \|a\theta - b\|_2^2 \quad \text{与} \quad \min_{\theta} \|a\theta - b\|_1.$$

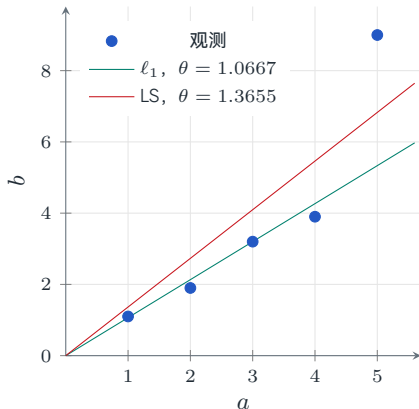
两种斜率

$$\theta_{\text{LS}} = \frac{a^\top b}{a^\top a} = \frac{75.1}{55} = 1.3655, \quad \theta_{\ell_1} = 1.0667.$$

ℓ_1 解是比值 b_i/a_i 关于权重 a_i 的加权中位数，受离群点影响较小。

ℓ_1 的内点步反复求解

$$a^\top D a \Delta \theta = -a^\top g,$$



自编标定例；实现结构参见 Boyd & Vandenberghe, §11.8.2

SDP 的 Newton 系统

对

$$\min c^\top x \quad \text{subject to} \quad F(x) = \sum_{i=1}^n x_i F_i + G \preceq 0,$$

令 $S = -F(x) \succ 0$ 。中心化目标为

$$tc^\top x - \log \det S.$$

Newton 方程 $H\Delta x = -g$ 中

$$H_{ij} = \text{tr}(S^{-1} F_i S^{-1} F_j), \quad g_i = tc_i + \text{tr}(S^{-1} F_i).$$

稠密情形的主要量级

形成和求解系统的量级由

$$\max\{np^3, n^2p^2, n^3\}$$

控制。共同块对角结构可以逐块利用；一般稀疏 F_i 并不保证 Hessian 稀疏。

应用例 11: 协方差上界设计

设两个传感器的误差耦合矩阵为

$$B = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}.$$

设计对角预算 $\text{diag}(\mathbf{x})$ 使

$$\text{diag}(\mathbf{x}) - B \succeq 0$$

并最小化 $x_1 + x_2$ 。这是前述最优解 (1.6, 1.6) 的 SDP。

令

$$S = \text{diag}(\mathbf{x}) - B.$$

在严格可行点 $\mathbf{x} = (2, 2)$,

$$S^{-1} = \frac{1}{0.64} \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix} = \begin{bmatrix} 1.5625 & 0.9375 \\ 0.9375 & 1.5625 \end{bmatrix}.$$

由于 $F_i = -E_{ii}$,

$$H_{ij} = (S_{ij}^{-1})^2.$$

所以

$$H = \begin{bmatrix} 2.4414 & 0.8789 \\ 0.8789 & 2.4414 \end{bmatrix}.$$

一个矩阵逆同时给出整个 Hessian。

网络速率优化

固定每条流的路径, 令 $A \in \{0, 1\}^{L \times n}$ 为“链路一流”关联矩阵:

$$A_{\ell j} = 1 \iff \text{流 } j \text{ 经过链路 } \ell.$$

问题为

$$\max_{\mathbf{x}} \sum_{j=1}^n U_j(x_j) \quad \text{subject to} \quad A\mathbf{x} \preceq \mathbf{c}, \quad \mathbf{x} \succeq 0. \quad (11.62)$$

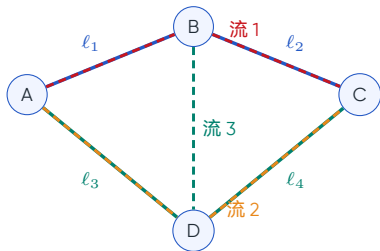
中心化 Hessian 具有

$$H = D_0 + A^\top D_1 A + D_2,$$

其中

$$D_0 = -t \text{diag}(U_j''(x_j)), \quad D_1 = \text{diag}(\mathbf{c} - A\mathbf{x})^{-2}, \quad D_2 = \text{diag}(\mathbf{x})^{-2}.$$

网络结构决定 Hessian 稀疏性



$H_{ij} \neq 0 \iff$ 流 i, j 共享某条链路.

短路径、每条链路承载少量流时, H 稀疏。也可用矩阵求逆引理转为 $L \times L$ 系统; 此时两个链路共享某条流才产生非零耦合。

应在 $n \times n$ 的“流耦合系统”和 $L \times L$ 的“链路耦合系统”之间选择更小、更稀疏者。

数值例 12: 三条流、两条链路

取

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad c = \begin{bmatrix} 3 \\ 2.5 \end{bmatrix}, \quad U_j(x_j) = \log x_j.$$

问题

$$\max_{x \succ 0} \log x_1 + \log x_2 + \log x_3 \quad \text{subject to} \quad x_1 + x_2 \leq 3, \quad x_2 + x_3 \leq 2.5.$$

两个容量约束在最优点均活跃。令 $x_1 = 3 - x_2$ 、 $x_3 = 2.5 - x_2$ ，一阶条件为

$$-\frac{1}{3 - x_2} + \frac{1}{x_2} - \frac{1}{2.5 - x_2} = 0.$$

数值解

$$\mathbf{x}^* = (2.0946, 0.9054, 1.5946).$$

共享两条链路的流 2 获得较小速率；稀疏矩阵 A 直接编码这种竞争关系。

选择原始系统还是 Schur 补

结构	优先考虑	原因
$p \ll n$, H^{-1} 易算	$p \times p$ Schur 补	消去 Δx , 系统明显降维
$n \ll p$ 或 Schur 补变稠密	原始 KKT 系统	避免形成稠密 $AH^{-1}A^T$
带状/块对角 Hessian	结构化直接分解	保持局部耦合, 减少填充
网络或图模型	比较流系统与链路系统	选择维数更小、图更稀疏的一侧
SDP 共同块结构	按块形成逆与 Hessian	矩阵运算可并行, 避免大稠密块

线性方程的数学等价形式很多; 数值实现应根据维数、稀疏模式和条件数选择。

第 11 章的结构

中心路径与证书

- ▣ 对数障碍把不等式变为光滑目标;
- ▣ 中心点同时生成对偶可行点;
- ▣ m/t 或 θ/t 给出可验证误差界。

初始化与复杂度

Phase I 找严格可行点或证实不可行。自协调性给出每次中心化的统一界和 $O(\sqrt{m} \log(1/\epsilon))$ 型最坏复杂度。

Lecture 07: 统计估计

最大似然、maximum entropy、协方差估计与实验设计将产生 log-sum-exp、entropy、log-det 及 conic models; 本讲的 primal-dual certificates 与 KKT 求解器为这些模型提供计算基础。

锥与原始一对偶方法

普通障碍扩展到正交锥、二阶锥与半正定锥。原始一对偶方法同时更新原始变量与乘子, 通常更适合高精度计算。

实现

实际时间主要花在线性系统。LP 消元、 ℓ_1 加权最小二乘、SDP 块结构和网络稀疏图决定单步成本。



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

自然科学研究院
INSTITUTE OF NATURAL SCIENCES

Fin

