

凸优化和单调变分不等式的收缩算法

第一讲：变分不等式能作为 多种问题的统一表述模式

Variational inequality is a uniform
approach for different problems

南京大学数学系 何炳生

hebma@nju.edu.cn

变分不等式是一种统一的问题表述模式. 管理科学与统计计算中存在大量凸优化问题. 信号处理, 图像恢复, 矩阵完整化, 机器学习等信息技术领域中也许多问题可以归结为（或松弛成）一个凸优化问题. 凸优化的一阶必要性条件就是一个单调变分不等式. 在变分不等式的框架下研究凸优化的求解方法, 就像微积分中用导数求一元函数的极值, 常常会带来很大的方便.

除了通常的最优化问题以外, 互补问题是约束为正卦限的变分不等式. 经济活动中的空间价格平衡, 保护资源—保障供给中的调控手段, 用经济手段解决交通疏导等问题, 都可以用变分不等式（或其特殊形式互补问题）来描述. 这一讲叙述以下一些常见问题与变分不等式的关系.

- 凸优化与单调变分不等式
- 商品流通、保护资源、保障供给中的变分不等式
- 交通疏导中的变分不等式
- 广义线性规划问题的线性变分不等式
- 最短距离和问题的线性变分不等式
- 极小化最大特征值之变分不等式

1 凸优化与单调变分不等式

连续优化方法中一些代表性的数学模型

- 简单约束优化问题 $\min \{f(x) \mid x \in \Omega\}.$
- 线性约束优化问题 $\min\{\theta(x) \mid Ax = b, x \in \mathcal{X}\}.$
- 结构型优化问题 $\min\{\theta_1(x)+\theta_2(y) \mid Ax + By = b, x \in \mathcal{X}, y \in \mathcal{Y}\}.$
- 非线性互补问题 $x \geq 0, F(x) \geq 0, x^T F(x) = 0.$
- 变分不等式 $x \in \Omega, (x' - x)^T F(x) \geq 0, \forall x' \in \Omega.$

我们只讲述求解凸优化和单调变分不等式的收缩算法. 在讨论求解方法的安排上, 我们从一般问题(非线性互补问题和变分不等式) 的求解方法(L2– L3)开始; 然后利用优化问题的性质, 讨论线性约束优化问题的求解方法(L4– L7).

L8 讲述求解简单约束优化问题的只用导数信息的收缩算法. L9 讨论线性约束优化问题对偶变量的收缩算法. L10 是只用梯度的线性约束凸优化的求解. L11– L14 讨论结构型优化问题的收缩算法. L15– L18 讨论多个可分离算子的结构型优化问题的收缩算法. 最后, L19– L20 对主要算法的收敛速率做了论述.

1.1 可微凸优化等价于一个特殊的单调变分不等式

设 Ω 是 \mathbb{R}^n 中的非空闭凸集, 我们先来讨论可微凸优化问题

$$\min\{\theta(x) \mid x \in \Omega\} \quad (1.1)$$

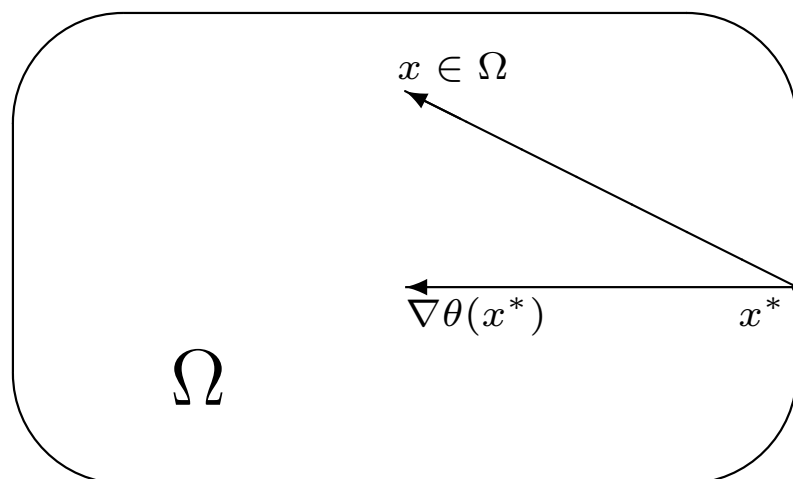
的最优性条件.

- 如果某一点 x^* 是最优点, 它必须属于 Ω ,
- 并且从这点出发的所有可行方向都不是下降方向.

我们用 $\nabla\theta(x)$ 表示 $\theta(x)$ 的梯度, 并记

- $Sd(x) = \{s \in \mathbb{R}^n \mid s^T \nabla\theta(x) < 0\}$, 为点 x 处的下降方向集;
- $Sf(x) = \{s \in \mathbb{R}^n \mid s = x' - x, x' \in \Omega\}$, 为点 x 处的可行方向集.

x^* 是最优解: $x^* \in \Omega$ 且 $Sf(x^*) \cap Sd(x^*) = \emptyset$



可微优化与变分不等式的关系

可以写成 $x^* \in \Omega, (x - x^*)^T \nabla \theta(x^*) \geq 0, \forall x \in \Omega$

将 $\nabla \theta(x)$ 写成 $f(x)$, 可微凸优化问题 (1.1) 就归结为求

$$x \in \Omega, (x' - x)^T f(x) \geq 0, \quad \forall x' \in \Omega. \quad (1.2)$$

以上形式就是一个变分不等式.

1.2 线性约束的凸优化问题

我们考虑一般的线性约束凸优化问题

$$\min\{\theta(x) \mid Ax = b, x \in \mathcal{X}\}, \quad (1.3)$$

其中 $\theta : \mathbb{R}^n \rightarrow \mathbb{R}$ 是凸函数, $\mathcal{X} \subset \mathbb{R}^n$ 是闭凸集, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$.

设 λ 是 Lagrange 乘子, 问题 (1.3) 的 Lagrange 函数

$$L(x, \lambda) = \theta(x) - \lambda^T (Ax - b)$$

定义在 $\mathcal{X} \times \mathbb{R}^m$ 上, 它的鞍点 $(x^*, \lambda^*) \in \mathcal{X} \times \mathbb{R}^m$ 满足

$$L_{\lambda \in \mathbb{R}^m}(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L_{x \in \mathcal{X}}(x, \lambda^*)$$

为了简便, 我们假设 $\theta(x)$ 是可微的, 并且记 $\nabla \theta(x) = f(x)$, 鞍点问题等价于

$$\begin{cases} x^* \in \mathcal{X}, & (x - x^*)^T (f(x^*) - A^T \lambda^*) \geq 0, \quad \forall x \in \mathcal{X}, \\ \lambda^* \in \mathbb{R}^m, & Ax^* = b. \end{cases} \quad (1.4)$$

上述最优性条件的第二部分也同样可以表示成

$$\lambda^* \in \Re^m, \quad (\lambda - \lambda^*)^T (Ax^* - b) \geq 0, \quad \forall \lambda \in \Re^m.$$

如果记

$$u = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(u) = \begin{pmatrix} f(x) - A^T \lambda \\ Ax - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \Re^m,$$

那么, (1.4) 可以写成

$$u^* \in \Omega, \quad (u - u^*)^T F(u^*) \geq 0, \quad \forall u \in \Omega.$$

如果 $\theta(x)$ 是非光滑的, 记 $F_a(u) = \begin{pmatrix} -A^T \lambda \\ Ax - b \end{pmatrix}$, 那么, 求 Lagrange 函数的

鞍点就等价于求混合变分不等式

$$u^* \in \Omega, \quad \theta(x) - \theta(x^*) + (u - u^*)^T F_a(u^*) \geq 0, \quad \forall u \in \Omega$$

的解 u^* . 记 $F_a(u) = Mu + q$, 则 M 是斜对称矩阵, $F_a(u)$ 是仿射单调算子.

1.3 可分离结构的线性约束凸优化问题

可分离结构的线性约束凸优化问题是指

$$\min \{ \theta_1(x) + \theta_2(y) \mid Ax + By = b, x \in \mathcal{X}, y \in \mathcal{Y} \} \quad (1.5)$$

其中 $\theta_1(x)$ 和 $\theta_2(y)$ 是可微凸函数. 设 λ 是 Lagrange 乘子, 上述问题的 Lagrange 函数

$$L(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T (Ax + By - b)$$

定义在 $\mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m$ 上, 设 $f(x) = \nabla \theta_1(x)$, $g(y) = \nabla \theta_2(y)$. 这个可分离结构凸优化问题的一阶最优性条件是

$$(x^*, y^*, \lambda^*) \in \Omega, \quad \begin{cases} (x - x^*)^T (f(x^*) - A^T \lambda^*) \geq 0, \\ (y - y^*)^T (g(y^*) - B^T \lambda^*) \geq 0, \\ (\lambda - \lambda^*)^T (Ax^* + By^* - b) \geq 0, \end{cases} \quad \forall (x, y, \lambda) \in \Omega, \quad (1.6)$$

其中

$$\Omega = \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m.$$

利用

$$u = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad F(u) = \begin{pmatrix} f(x) - A^T \lambda \\ g(y) - B^T \lambda \\ Ax + By - b \end{pmatrix} \quad \text{和} \quad F_a(u) = \begin{pmatrix} -A^T \lambda \\ -B^T \lambda \\ Ax + By - b \end{pmatrix},$$

这个可分离凸优化问题的一阶最优性条件可以写成

$$u^* \in \Omega, \quad (u - u^*)^T F(u^*) \geq 0, \quad \forall u \in \Omega.$$

当 $\theta_1(x), \theta_2(y)$ 是不可微凸函数时, 问题 (1.5) 等价于混合变分不等式

$$u^* \in \Omega, \quad \begin{pmatrix} (\theta_1(x) - \theta_1(x^*))^+ \\ \theta_2(y) - \theta_2(y^*) \end{pmatrix} + (u - u^*)^T F_a(u^*) \geq 0, \quad \forall u \in \Omega.$$

我们在后面的章节中展示, 在变分不等式框架下研究凸优化求解方法, 无论在算法设计和收敛性证明方面, 会变得相当简单和方便.

1.4 非线性互补问题是一类特殊的变分不等式

凸优化问题 (1.1) 中, 如果 $\Omega = \mathfrak{R}_+^n$ (实 n -维空间中的非负卦限) 并且 $\theta(x)$ 是可微的, 根据 §1.1 中的分析, x^* 是最优解的充要条件是

$$x^* \geq 0, \quad (x - x^*)^T \nabla \theta(x^*) \geq 0, \quad \forall x \geq 0.$$

设 F 是 $\mathfrak{R}^n \rightarrow \mathfrak{R}^n$ 的一个算子. 非负卦限 \mathfrak{R}_+^n 上变分不等式的形式是

$$\mathbf{VI}(\mathfrak{R}_+^n, F): \quad x \geq 0, \quad (x' - x)^T F(x) \geq 0, \quad \forall x' \geq 0.$$

非线性互补问题是最优化理论与方法中一类很重要的问题. 它的数学形式是

$$(\mathbf{NCP}) \quad x \geq 0, \quad F(x) \geq 0, \quad x^T F(x) = 0. \quad (1.7)$$

事实上, **NCP** 是 $\Omega = \mathfrak{R}_+^n$ 的一类变分不等式. 因此, 在 $\Omega = \mathfrak{R}_+^n$ 时, 可微凸优化问题 (1.1) 与一个互补问题等价.

证明

为帮助阅读理解, 我们给出**NCP**与 $\text{VI}(\mathcal{R}_+^n, F)$ 等价的具体证明.

如果 x 是 **NCP** 的解, 那么有 $x \geq 0$ 和 $F(x) \geq 0$.

对于任意的 $x' \geq 0$ 有 $(x')^T F(x) \geq 0$.

又因 $x^T F(x) = 0$, 得

$$\begin{aligned} (x' - x)^T F(x) \\ &= (x')^T F(x) \\ &\geq 0. \end{aligned}$$

所以 x 是 $\text{VI}(R_+^n, F)$ 的一个解.

反过来, 如果 x 是 $\text{VI}(R_+^n, F)$ 的一个解, 则 $x \geq 0$.

将 $x' = 0$ 和 $x' = 2x$ 代入 $(x' - x)^T F(x) \geq 0$, 我们得到 $\mp x^T F(x) \geq 0$. 因此 $x^T F(x) = 0$.

要证明 x 是 **NCP** 的解, 只剩下 $F(x) \geq 0$ 需要证明, 对此采用反证法. 如果 $F(x)$ 的某个分量 $F_j(x) < 0$, 我们取 x' , 使得

$$x'_i = \begin{cases} x_i, & \text{if } i \neq j \\ x_j + 1, & \text{if } i = j \end{cases}$$

这样的 $x' \geq 0$. 但 $(x' - x)^T F(x) = F_j(x) < 0$, 这与 x 是 **VI** 的解矛盾.

1.5 最小一模问题与等价的变分不等式

最小一模问题 (Least absolute deviations) 的数学形式是

$$\min \|Ax - b\|_1, \quad (1.8)$$

其中 $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. 与最小二乘相比, 它提供了一种更稳健的数据拟合模型. 与最小二乘不同, 它是一个非光滑凸优化问题. 这类问题被广泛应用于统计学和经济研究, 一些最新的文献都有提及. 例如:

- T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer, second edition, 2009. §10.6.
- J. M. Wooldridge. Introductory Econometrics: A Modern Approach. South Western College Publications, fourth edition, 2009. §9.6.

我们用 e 表示每个分量都是 1 的 m -维向量. 容易验证

$$\|d\|_1 = \max\{y^T d \mid y \in B_\infty\}, \quad B_\infty = \{y \in \mathbb{R}^m \mid -e \leq y \leq e\}.$$

因此, 问题 (1.8) 就是一个形式为

$$\min_{x \in \mathbb{R}^n} \max_{y \in B_\infty} y^T (Ax - b).$$

的 min-max 问题. 它的等价形式是下面的线性变分不等式:

$$x^* \in \mathbb{R}^n, y^* \in B_\infty, \quad \begin{cases} (x - x^*)^T (A^T y^*) \geq 0, & \forall x \in \mathbb{R}^n, \\ (y - y^*)^T (-Ax^* + b) \geq 0, & \forall y \in B_\infty. \end{cases}$$

可以写成更简单紧凑的形式

$$u^* \in \Omega, \quad (u - u^*)^T (Mu^* + q) \geq 0, \quad \forall u \in \Omega$$

其中

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad M = \begin{pmatrix} 0 & A^T \\ -A & 0 \end{pmatrix}, \quad q = \begin{pmatrix} 0 \\ b \end{pmatrix}, \quad \Omega = \mathbb{R}^n \times B_\infty.$$

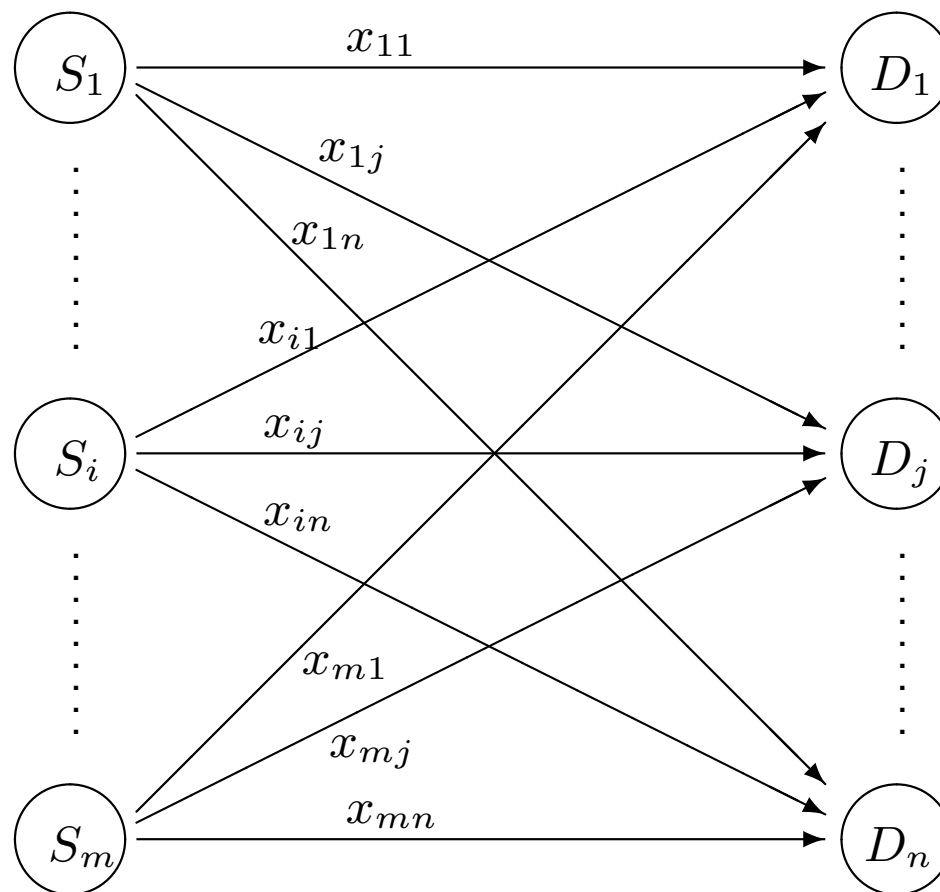
2 保护资源、保障供给中的互补问题

我们以下面的例子来说明一个经济平衡模型：

假设某种商品（例如煤）由 m 个资源地生产和 n 个需求地消费. 它由经营者们从资源地采购运到需求地销售. 经营者会根据贪婪原理找到他们的最优经营方案. 记

$$s_i = \sum_{j=1}^n x_{ij}$$

$$d_j = \sum_{i=1}^m x_{ij}$$



一些记号

S_i : 该种商品的第 i 个资源地;

D_j : 该种商品的第 j 个需求地;

x_{ij} : 从 S_i 到 D_j 的交易量;

s_i : 经营者们在资源地 S_i 的采购总量, $s_i = \sum_{j=1}^n x_{ij}$;

d_j : 经营者在需求地 D_j 的销售总量, $d_j = \sum_{i=1}^m x_{ij}$;

h_i^s : 经营者在资源地 S_i 处的采购价, 是 S_i 处采购量的函数;

h_j^d : 经营者在需求地 D_j 处的销售价, 是 D_j 处到货量的函数;

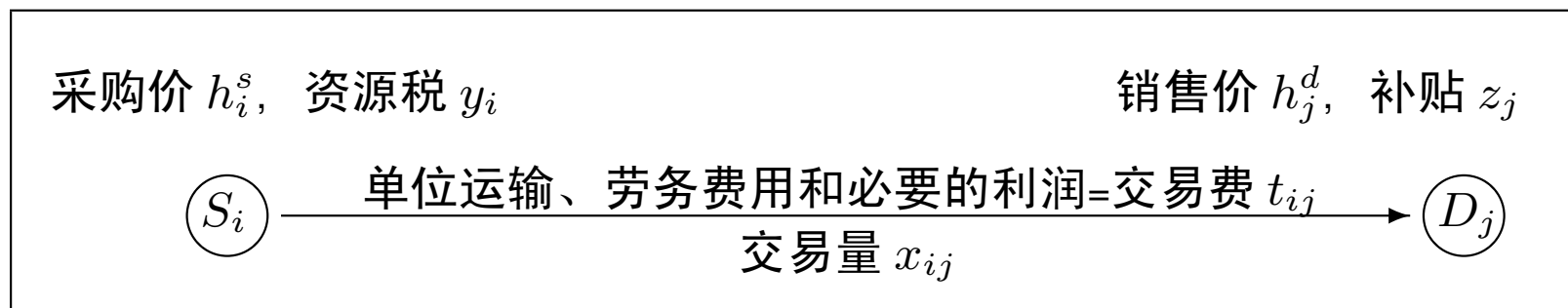
t_{ij} : 从 S_i 到 D_j 的交易费用 (包括运输费用等交易成本);

y_i : 政府为避免资源过度开采而在资源地 S_i 向经营者征收的资源税;

z_j : 政府为保障供给而在需求地 D_j 给经营者的经营补贴.

2.1 经营者追求利益最大化之互补问题

用下图表示第 i 个资源地 S_i 和第 j 个需求地 D_j 之间的采购—销售关系.



如果 $(h_i^s + y_i + t_{ij}) \geq (h_j^d + z_j)$, 没有人愿意做亏本买卖, 故 $x_{ij} = 0$;

反之, 根据**贪婪原理**, 经营者会尽可能增大经营量 x_{ij} , (通常这会导致采购价的上涨和销售价的下降) 直到

$$(h_i^s + y_i + t_{ij}) = (h_j^d + z_j).$$

用数学语言描述就是:

$$h_i^s + y_i + t_{ij} \begin{cases} \geq h_j^d + z_j, & \text{如果 } x_{ij} = 0, \\ = h_j^d + z_j, & \text{如果 } x_{ij} > 0. \end{cases}$$

- 对给定的资源税率 y_1, y_2, \dots, y_m 和补贴标准 z_1, z_2, \dots, z_n (政策),

经营者们会根据“贪婪原理”找到他们的最优经营方案 x_{ij} (对策).

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix},$$

如果记

$$F(X) = \begin{pmatrix} F_{11} & F_{12} & \dots & F_{1n} \\ F_{21} & F_{22} & \dots & F_{2n} \\ \vdots & \vdots & & \vdots \\ F_{m1} & F_{m2} & \dots & F_{mn} \end{pmatrix},$$

其中

$$F_{ij}(X) = \{h_i^s(s_i) + y_i + t_{ij}\} - \{h_j^d(d_j) + z_j\},$$

$$s_i = \sum_{j=1}^n x_{ij}, \quad d_j = \sum_{i=1}^m x_{ij}.$$

那么, 由经营者的最优经营方案形成的非负 $m \times n$ 矩阵 X 和 $F(X)$ 中, 同样下标的元素中最多只能有一个大于零. 换句话说, 经营者解决的问题是互补问题

$$X \geq 0, \quad F(X) \geq 0, \quad \text{Trace}(X^T F(X)) = 0.$$

上式中的 $\text{Trace}(\cdot)$ 表示矩阵的“迹”——矩阵对角元的和.

- 我们能观测到的是问题解中相应的

$$s_i(y, z) = \sum_{j=1}^n x_{ij}(y, z) \quad \text{and} \quad d_j(y, z) = \sum_{i=1}^m x_{ij}(y, z)$$

它们是由政策向量 (y, z) 决定的.

2.2 政府部门制定最优政策之互补问题

政府部门的职责—保护资源和保障供给

- 坚持可持续发展及环境保护，防止资源被过度开采， $s_i \leq s_i^{\max}$;
- 保障基本供给，从而保证社会稳定和民生， $d_j \geq d_j^{\min}$.

手段(政策):

- 在资源过度开采的产地征收资源税， y_1, \dots, y_m ;
- 供应紧张的需求地给经营者提供补贴， z_1, \dots, z_n .

经营者们会根据政策给出对策. 政府的任务是给出‘最好’的政策.

政府部门需要的最优政策是一个“黑箱”条件下的互补问题:

- 保护了资源, 同时又让经济尽可能繁荣;

$$y \geq 0, \quad s^{\max} - s(y, z) \geq 0, \quad y^T (s^{\max} - s(y, z)) = 0,$$

- 保障了供给, 同时又尽可能节约财政支出.

$$z \geq 0, \quad d(y, z) - d^{\min} \geq 0, \quad z^T (d(y, z) - d^{\min}) = 0.$$

职能部门的互补问题的紧凑形式是:

$$u \geq 0, \quad F(u) \geq 0, \quad u^T F(u) = 0,$$

其中

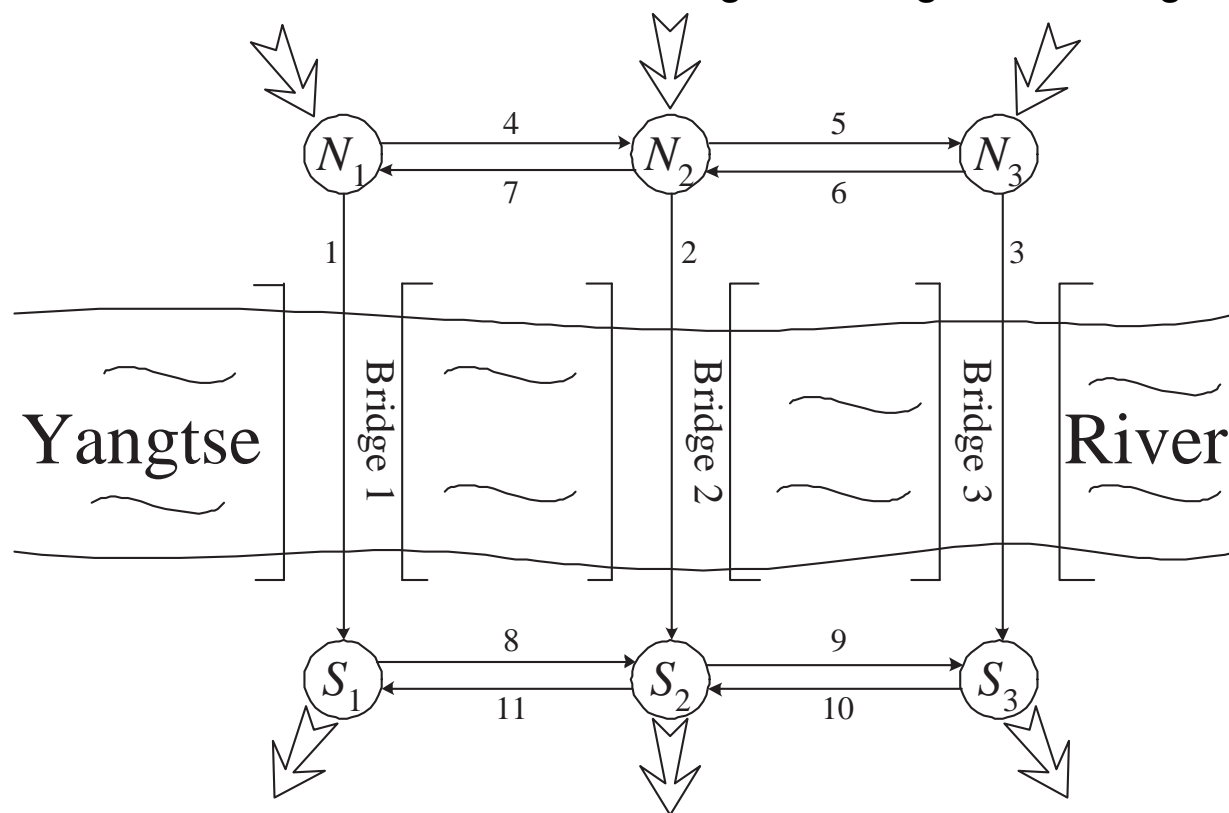
$$u = \begin{pmatrix} y \\ z \end{pmatrix}, \quad F(u) = \begin{pmatrix} s^{\max} - s(u) \\ d(u) - d^{\min} \end{pmatrix},$$

$s(u)$ 和 $d(u)$ 是 u 的函数.

黑箱问题, 是指 $F(u)$ 是 u 的函数, 有确定的关系, 但没有函数表达式.

3 交通疏导问题之互补问题

设某跨江城市有三座长江大桥, 分别为 Bridge-1, Bridge-2 和 Bridge-3.



不失一般性, 我们可以将 N_1, N_2, N_3 看作由北向南的车辆在江北的出发地, 同时把 S_1, S_2, S_3 看作它们在江南的集散地. 出发地之间和集散地之间都有

道路联结. 所考虑的交通问题可以看作一个带‘有向边’ (oriented link) 的网络.

从 N_i 到 S_j , $i = 1, 2, 3$; $j = 1, 2, 3$. 共有 9 个 O/D 对 (O/D Pair) .

将这些 O/D 对编号并记为 $\omega_1, \dots, \omega_9$, 对每个 O/D 对的运输任务, 驾驶员可选择分别经过 Bridge-1, Bridge-2, Bridge-3 的 3 条不同路(路由边联结而成) 行走.

用 P_{ω_j} 表示联结 O/D 对 ω_j 的路的集合. 将联结 9 个 O/D 对的 27 条路(path) 编号并记为 p_1, \dots, p_{27} . 关于进一步的网络分析可以参考 [11].

O/D 对 ω	P_{ω}	Links on the path		
$\omega_1 = (N_1, S_1)$	$P_{\omega_1} = \{p_1, p_2, p_3\}$	$p_1 = \{1\}$	$p_2 = \{4, 2, 11\}$	$p_3 = \{4, 5, 3, 10, 11\}$
$\omega_2 = (N_1, S_2)$	$P_{\omega_2} = \{p_4, p_5, p_6\}$	$p_4 = \{1, 8\}$	$p_5 = \{4, 2\}$	$p_6 = \{4, 5, 3, 10\}$
$\omega_3 = (N_1, S_3)$	$P_{\omega_3} = \{p_7, p_8, p_9\}$	$p_7 = \{1, 8, 9\}$	$p_8 = \{4, 2, 9\}$	$p_9 = \{4, 5, 3\}$
$\omega_4 = (N_2, S_1)$	$P_{\omega_4} = \{p_{10}, p_{11}, p_{12}\}$	$p_{10} = \{7, 1\}$	$p_{11} = \{2, 11\}$	$p_{12} = \{5, 3, 10, 11\}$
$\omega_5 = (N_2, S_2)$	$P_{\omega_5} = \{p_{13}, p_{14}, p_{15}\}$	$p_{13} = \{7, 1, 8\}$	$p_{14} = \{2\}$	$p_{15} = \{5, 3, 10\}$
$\omega_6 = (N_2, S_3)$	$P_{\omega_6} = \{p_{16}, p_{17}, p_{18}\}$	$p_{16} = \{7, 1, 8, 9\}$	$p_{17} = \{2, 9\}$	$p_{18} = \{5, 3\}$
$\omega_7 = (N_3, S_1)$	$P_{\omega_7} = \{p_{19}, p_{20}, p_{21}\}$	$p_{19} = \{6, 7, 1\}$	$p_{20} = \{6, 2, 11\}$	$p_{21} = \{3, 10, 11\}$
$\omega_8 = (N_3, S_2)$	$P_{\omega_8} = \{p_{22}, p_{23}, p_{24}\}$	$p_{22} = \{6, 7, 1, 8\}$	$p_{23} = \{6, 2\}$	$p_{24} = \{3, 10\}$
$\omega_9 = (N_3, S_3)$	$P_{\omega_9} = \{p_{25}, p_{26}, p_{27}\}$	$p_{25} = \{6, 7, 1, 8, 9\}$	$p_{26} = \{6, 2, 9\}$	$p_{27} = \{3\}$

上述网络分析是为道路使用者服务的. 我们只对管理部门的问题感兴趣, 假

设只收过桥费. 管理部门想制定一个适当的收费标准合理控制桥上流量.

驾驶员基于 Wardrop 原理的最优出行方案 — 最小费用路径

对给定的大桥收费 $y = (y_1, y_2, y_3)$, 驾驶员会找到他们的最优出行方案.

管理部门要求通过大桥合理收费控制桥上的流量

- $0 \leq y \in \Re^3$: 桥上的收费向量;
- $f(y) \in \Re^3$: 桥上的流量, 它是收费 y 的函数;
- $0 < b \in \Re^3$: 管理部门希望控制的桥上的流量上界.

管理部门要求解的数学问题是

$$y \geq 0, \quad F(y) = b - f(y) \geq 0, \quad y^T F(y) = 0.$$

同样, 流量 $f(y)$ 确是收费 y 的函数, 但没有表达式. 只能对给定的自变量, 观测相应的函数值, 使用只用函数值的方法. 而这种观测, 往往代价不菲. 我们从事投影收缩算法研究, 一度着眼点是要得到效率高一些的、应对社会计算问题的只用函数值和少用函数值的方法.

4 广义线性规划及鞍点问题之变分不等式

4.1 广义线性规划及之变分不等式

线性规划的标准形式是 $\min\{c^T x \mid Ax = b, x \geq 0\}$. 其中 $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$. 在实际经济问题中, 向量 b 一般表示需求量, c 表示价格. 我们允许 b 和 c 都在一定范围之内变动, 考虑更一般的问题

$$\min\{\max_{\eta \in C} \eta^T x \mid Ax \in B, x \in D\} \quad (4.1)$$

其中 $C, D \subset \mathbb{R}^n$, $B \subset \mathbb{R}^m$ 是闭凸集. 这样的问题我们称之为广义线性规划. 引进辅助变量 y 和 Lagrange 乘子 λ , 得到广义线性规划 (4.1) 的 Lagrange 函数

$$L(x, y, \lambda, \eta) = \eta^T x - \lambda^T (Ax - y),$$

它定义在 $(D \times B) \times (\mathbb{R}^m \times C)$ 上. 广义线性规划等价与一个 min-max 问题

$$L_{\lambda \in \mathbb{R}^m, \eta \in C} L(x^*, y^*, \lambda, \eta) \leq L(x^*, y^*, \lambda^*, \eta^*) \leq L_{x \in D, y \in B} L(x, y, \lambda^*, \eta^*).$$

设 $(x^*, y^*, \lambda^*, \eta^*)$ 是上述 min-max 问题的解, 则有

$$\begin{cases} x^* \in D, & (x - x^*)^T (-A^T \lambda^* + \eta^*) \geq 0, & \forall x \in D \\ y^* \in B, & (y - y^*)^T (\lambda^*) \geq 0, & \forall y \in B \\ \lambda^* \in \mathbb{R}^m, & (\lambda - \lambda^*)^T (Ax^* - y^*) \geq 0, & \forall \lambda \in \mathbb{R}^m \\ \eta^* \in C, & (\eta - \eta^*)^T (-x^*) \geq 0, & \forall \eta \in C. \end{cases}$$

更紧凑的形式可以写成

$$w^* \in \Omega, \quad (w - w^*)^T (Mw^* + q) \geq 0, \quad \forall w \in \Omega$$

其中

$$w = \begin{pmatrix} x \\ y \\ \lambda \\ \eta \end{pmatrix}, \quad M = \begin{pmatrix} 0 & 0 & -A^T & I \\ 0 & 0 & I & 0 \\ A & -I & 0 & 0 \\ -I & 0 & 0 & 0 \end{pmatrix}, \quad q = 0$$

和

$$\Omega = D \times B \times \mathbb{R}^m \times C.$$

4.2 一般鞍点问题之变分不等式

用全变差极小处理图像去模糊 [1], 经离散化以后, 问题的数学模型是

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \Phi(x, y) := \theta_1(x) + y^T A x - \theta_2(y) \quad (4.2)$$

其中 $\mathcal{X} \subset \mathbb{R}^n$, $\mathcal{Y} \subset \mathbb{R}^m$ 是闭凸集, $A \in \mathbb{R}^{m \times n}$. $\theta_1(x) : \mathbb{R}^n \rightarrow \mathbb{R}$, $\theta_2(y) : \mathbb{R}^m \rightarrow \mathbb{R}$ 为凸函数. 如果 $(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}$ 是问题 (4.2) 的解, 则有

$$\Phi_{y \in \mathcal{Y}}(x^*, y) \leq \Phi(x^*, y^*) \leq \Phi_{x \in \mathcal{X}}(x, y^*).$$

换句话说, (x^*, y^*) 是 $\Phi(x, y)$ 在 $\mathcal{X} \times \mathcal{Y}$ 上的鞍点. 因此, 问题 (4.2) 能够转换成等价的变分不等式: 求 $(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}$, 使得

$$\begin{pmatrix} x - x^* \\ y - y^* \end{pmatrix}^T \begin{pmatrix} f(x^*) + A^T y^* \\ g(y^*) - A x^* \end{pmatrix} \geq 0, \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}, \quad (4.3)$$

其中 $f(x) \in \partial\theta_1(x)$, $g(y) \in \partial\theta_2(y)$. 用记号

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad F(u) = \begin{pmatrix} f(x) + A^T y \\ -Ax + g(y) \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y}.$$

问题 (4.3) 就是变分不等式

$$u^* \in \Omega, \quad (u - u^*)^T F(u^*) \geq 0, \quad \forall u \in \Omega. \quad (4.4)$$

由于 $\theta_1(x)$ 和 $\theta_2(y)$ 是凸函数, 容易验证, 这里的 $F(u)$ 是单调的. 关心这类问题求解方法的可以参阅第四讲与第五讲. 当 $\theta_1(x)$, $\theta_2(y)$ 是非光滑凸函数时, 问题 (4.2) 也可以表述成下面的混合变分不等式问题:

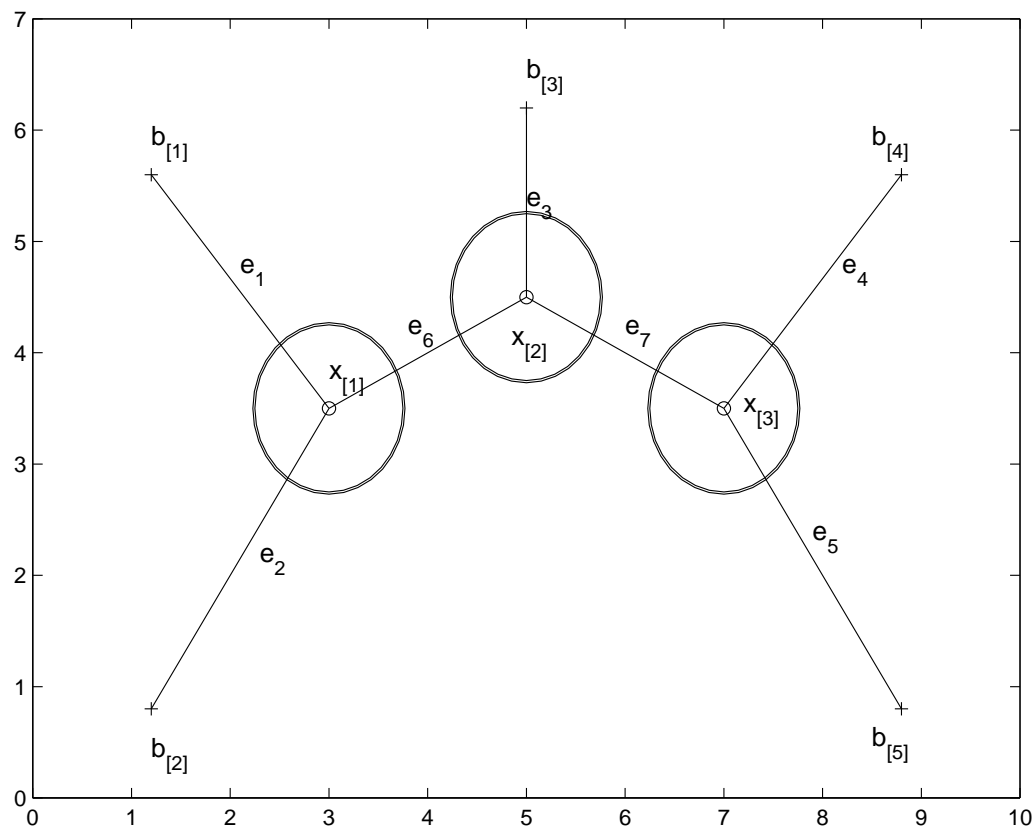
$$\theta(u) - \theta(u^*) + \begin{pmatrix} x - x^* \\ y - y^* \end{pmatrix}^T \begin{pmatrix} A^T y^* \\ -Ax^* \end{pmatrix} \geq 0, \quad \forall u \in \Omega.$$

其中 $\theta(u) = \theta_1(x) + \theta_2(y)$.

5 最短距离和问题之变分不等式

有些典型的非光滑凸优化问题可以化成结构相当简单的变分不等式. 这一节我们以最短距离和问题为例加以说明.

假设 $b_{[1]}, \dots, b_{[5]}$ 是确定的村镇. 现在要用一个如右图的网络把它们连接起来. $x_{[1]}, x_{[2]}, x_{[3]}$ 是待选的连接点, 它们必须在划定的范围内. 如何确定 $x_{[1]}, x_{[2]}, x_{[3]}$ 的位置, 使得网络线路长度最短.



5.1 欧氏模下的最短距离和问题

欧氏模下的最短距离和问题的数学模型是

$$\min_{x[j] \in X_j} \left\{ \begin{array}{l} \|x_{[1]} - b_{[1]}\|_2 + \|x_{[1]} - b_{[2]}\|_2 + \|x_{[2]} - b_{[3]}\|_2 \\ + \|x_{[3]} - b_{[4]}\|_2 + \|x_{[3]} - b_{[5]}\|_2 \\ + \|x_{[1]} - x_{[2]}\|_2 + \|x_{[2]} - x_{[3]}\|_2 \end{array} \right\}. \quad (5.1)$$

注意到这里是距离和问题, 不是距离的平方和问题. 此类问题是一个非光滑凸优化问题.

将问题转化为 min-max 问题

注意到对任意的 $d \in \mathbb{R}^2$, 有

$$\|d\|_2 = \max_{\xi \in B_2} \xi^T d, \quad (5.2)$$

其中

$$B_2 = \{\xi \in \Re^2 \mid \|\xi\|_2 \leq 1\}.$$

利用 (5.2), 上述欧氏模意义下的最短距离和问题可以化为min-max问题

$$\min_{x_{[i]} \in X_i} \max_{z_{[j]} \in B_2} \left\{ \begin{aligned} & z_{[1]}^T(x_{[1]} - b_{[1]}) + z_{[2]}^T(x_{[1]} - b_{[2]}) + z_{[3]}^T(x_{[2]} - b_{[3]}) \\ & + z_{[4]}^T(x_{[3]} - b_{[4]}) + z_{[5]}^T(x_{[3]} - b_{[5]}) \\ & + z_{[6]}^T(x_{[1]} - x_{[2]}) + z_{[7]}^T(x_{[2]} - x_{[3]}) \end{aligned} \right\}$$

它的紧凑形式为

$$\min_{x \in \mathcal{X}} \max_{z \in \mathcal{B}_2} z^T(Ax - b)$$

其中

$$\mathcal{X} = X_1 \times X_2 \times X_3, \quad \mathcal{B}_2 = B_2 \times B_2 \times \cdots \times B_2.$$

$$x = \begin{pmatrix} x_{[1]} \\ x_{[2]} \\ x_{[3]} \end{pmatrix}, \quad z = \begin{pmatrix} z_{[1]} \\ z_{[2]} \\ \vdots \\ z_{[7]} \end{pmatrix}. \quad (5.3)$$

分块矩阵 A 和向量 b 的结构分别是

$$A = \begin{pmatrix} I_2 & & & & & \\ & I_2 & & & & \\ & & I_2 & & & \\ & & & I_2 & & \\ & & & & I_2 & \\ I_2 & -I_2 & & & & \\ & & I_2 & -I_2 & & \end{pmatrix}, \quad b = \begin{pmatrix} b_{[1]} \\ b_{[2]} \\ b_{[3]} \\ b_{[4]} \\ b_{[5]} \\ 0 \\ 0 \end{pmatrix}.$$

设 $(x^*, z^*) \in \mathcal{X} \times \mathcal{B}_2$ 是 min-max 问题的解,

则对所有的 $x \in \mathcal{X}$ 和 $z \in \mathcal{B}$, 有

$$z^T(Ax^* - b) \leq z^{*T}(Ax^* - b) \leq z^{*T}(Ax - b)$$

它的等价形式是下面的线性变分不等式:

$$x^* \in \mathcal{X}, z^* \in \mathcal{B}_2, \quad \begin{cases} (x - x^*)^T(A^T z^*) \geq 0, & \forall x \in \mathcal{X}, \\ (z - z^*)^T(-Ax^* + b) \geq 0, & \forall z \in \mathcal{B}_2. \end{cases}$$

可以写成更简单紧凑的形式

$$u^* \in \Omega, \quad (u - u^*)^T(Mu^* + q) \geq 0, \quad \forall u \in \Omega$$

其中

$$u = \begin{pmatrix} x \\ z \end{pmatrix}, \quad M = \begin{pmatrix} 0 & A^T \\ -A & 0 \end{pmatrix}, \quad q = \begin{pmatrix} 0 \\ b \end{pmatrix}, \quad \Omega = \mathcal{X} \times \mathcal{B}_2.$$

5.2 l_1 -模下的最短距离和问题

图示问题的 l_1 -模最短距离的数学形式是

$$\min_{x_{[j]} \in X_j} \left\{ \begin{array}{lll} \|x_{[1]} - b_{[1]}\|_1 & + \|x_{[1]} - b_{[2]}\|_1 & + \|x_{[2]} - b_{[3]}\|_1 \\ & + \|x_{[3]} - b_{[4]}\|_1 & + \|x_{[3]} - b_{[5]}\|_1 \\ & + \|x_{[1]} - x_{[2]}\|_1 & + \|x_{[2]} - x_{[3]}\|_1 \end{array} \right\}. \quad (5.4)$$

由于对任意的 $d \in \Re^2$, 有

$$\|d\|_1 = \max_{\xi \in B_\infty} \xi^T d,$$

其中

$$B_{\infty} = \{\xi \in R^2 \mid \|\xi\|_{\infty} \leq 1\}.$$

上述 l_1 -模意义下的最短距离和问题化为 min-max 问题. 它的紧凑形式为

$$\min_{x \in \mathcal{X}} \max_{z \in B_{\infty}} z^T (Ax - b)$$

与欧氏模下的距离和问题一样, 它等价于变分不等式

$$u^* \in \Omega, \quad (u - u^*)^T (Mu^* + q) \geq 0, \quad \forall u \in \Omega$$

矩阵 M 和向量 q 都不变, 所不同的只是这时集合

$$\Omega = \mathcal{X} \times B_{\infty}, \quad B_{\infty} = B_{\infty} \times B_{\infty} \times \cdots \times B_{\infty}.$$

5.3 l_∞ -模下的最短距离和问题

图示问题的 l_∞ -模最短距离的数学形式是

$$\min_{x_{[j]} \in X_j} \left\{ \begin{array}{lll} \|x_{[1]} - b_{[1]}\|_\infty & + \|x_{[1]} - b_{[2]}\|_\infty & + \|x_{[2]} - b_{[3]}\|_\infty \\ & + \|x_{[3]} - b_{[4]}\|_\infty & + \|x_{[3]} - b_{[5]}\|_\infty \\ & + \|x_{[1]} - x_{[2]}\|_\infty & + \|x_{[2]} - x_{[3]}\|_\infty \end{array} \right\}. \quad (5.5)$$

由于对任意的 $d \in \Re^2$, 有

$$\|d\|_\infty = \max_{\xi \in B_1} \xi^T d,$$

其中

$$B_1 = \{\xi \in R^2 \mid \|\xi\|_1 \leq 1\}.$$

与欧氏模下的距离和问题一样, 它也可以表示成一个 min-max 问题, 并化成等价的变分不等式

$$u^* \in \Omega, \quad (u - u^*)^T (Mu^* + q) \geq 0, \quad \forall u \in \Omega$$

矩阵 M 和向量 q 都不变, 所不同的只是这时集合

$$\Omega = \mathcal{X} \times \mathcal{B}_1, \quad \mathcal{B}_1 = B_1 \times B_1 \times \cdots \times B_1.$$

将处理欧氏模问题时的 \mathcal{B}_2 换成了 \mathcal{B}_1 . 我们会在第三讲具体介绍这类最短距离和问题的求解方法.

6 极小化最大特征值之变分不等式

设 $A_i, i = 1, \dots, m$ 和 B 是给定的 $n \times n$ 对称矩阵, $c \in \Re^m$. 考虑问题

$$\min_{x \in \mathcal{X}} \{ \max \lambda_{\max} [\mathcal{A}(x) + B] + c^T x \} \quad (6.1)$$

其中 $\mathcal{X} \subset \Re^m$,

$$\mathcal{A}(x) = \sum_{i=1}^m x_i A_i. \quad (6.2)$$

这是经典半定规划中的对偶问题[9]. 对给定的 $n \times n$ 矩阵 $A = (a_{ij})$, 矩阵 A 的迹记为 $\mathbf{Tr}(A)$, 其定义为 (矩阵对角元的和)

$$\mathbf{Tr}(A) = \sum_{j=1}^n a_{jj}.$$

熟知, 矩阵的特征值的和等于矩阵的迹. 对由 (6.2) 定义的 $\mathcal{A}(x)$ 和给定的 $n \times n$ 矩阵 Y , 我们用

$$\mathcal{A}^*(Y) = \begin{pmatrix} \mathbf{Tr}(A_1 Y) \\ \vdots \\ \mathbf{Tr}(A_m Y) \end{pmatrix}$$

定义 $\mathcal{A}^*(Y) \in \Re^m$. 在一些应用问题中, 集合 \mathcal{X} 往往是个单纯形, 即 $\mathcal{X} = \{x \in \Re^m \mid \sum_{i=1}^m x_i = 1, x_i \geq 0\}$. 注意到相似变换不改变矩阵特征

值这一事实, 可以得到

$$\lambda_{\max}(A) = \max\{\mathbf{Tr}(YA) \mid \mathbf{Tr}(Y) = 1, Y \succeq 0\}.$$

所以, 问题 (6.1) 等价于下面的 $\min - \max$ 问题:

$$\min_{x \in \mathcal{X}} \max_{Y \succeq 0} \left\{ \sum_{i=1}^m (x_i \mathbf{Tr}(YA_i)) + \mathbf{Tr}(YB) + c^T x \mid \mathbf{Tr}(Y) = 1 \right\}. \quad (6.3)$$

设 (x^*, Y^*) 是问题 (6.3) 的解, 那么有 $\lambda^* \in \Re$, 使得

$$\Phi_{y \in \mathcal{S}_+^n, \lambda \in \Re}(x^*, Y, \lambda) \leq \Phi(x^*, Y^*, \lambda^*) \leq \Phi_{x \in \mathcal{X}}(x, Y^*, \lambda^*),$$

其中

$$\Phi(x, Y, \lambda) = \sum_{i=1}^m (x_i \mathbf{Tr}(YA_i)) + \mathbf{Tr}(YB) + c^T x + \lambda(\mathbf{Tr}(Y) - 1),$$

是问题 (6.3) 的 Lagrange 函数. 问题就转化为求函数 $\Phi(x, Y, \lambda)$ 在 $\mathcal{X} \times (\mathcal{S}_+^n \times \Re)$ 上的鞍点 (x^*, Y^*, λ^*) . 因此, 问题 (6.1) 可以转换成下面

的变分不等式: 求 $(x^*, Y^*, \lambda^*) \in \mathcal{X} \times \mathcal{S}_+^n \times \mathfrak{R}$, 使得

$$\begin{cases} x^* \in \mathcal{X}, & (x - x^*)^T \{\mathcal{A}^*(Y^*) + c\} \geq 0, & \forall x \in \mathcal{X}, \\ Y^* \succeq 0, & \langle Y - Y^*, -\mathcal{A}(x^*) - \lambda^* I - B \rangle \geq 0, & \forall Y \succeq 0, \\ \lambda^* \in \mathfrak{R}, & (\lambda - \lambda^*)(\text{Tr}(Y^*) - 1) \geq 0, & \forall \lambda \in \mathfrak{R}. \end{cases}$$

上述变分不等式的紧致形式就是

$$u^* \in \Omega, \quad (u - u^*)^T F(u^*) \geq 0, \quad \forall u \in \Omega,$$

其中

$$u = \begin{pmatrix} x \\ Y \\ \lambda \end{pmatrix}, \quad F(u) = \begin{pmatrix} \mathcal{A}^*(Y) + c \\ -\mathcal{A}(x) - \lambda I - B \\ \text{Tr}(Y) - 1 \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{S}_+^n \times \mathfrak{R}.$$

变分不等式是描述数学模型的强有力的统一工具

当然, 这并不是说, 上面提及的问题都要化成变分不等式去做.

References

- [1] A. Chambolle and T. Pock, A first-order primal-dual algorithms for convex problem with applications to imaging, *J. Math. Imaging Vison*, **40**, 120-145, 2011.
- [2] P.T. Harker, J.S.Pang, Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications, *Mathematical Programming*, **48**, 161-220, 1990.
- [3] B.S. He, A projection and contraction method for a class of linear complementarity problems and its application in convex quadratic programming, *Applied Mathematics and Optimization*, **25**, 247-262, 1992.
- [4] B.S. He, A class of projection and contraction methods for monotone variational inequalities, *Applied Mathematics and Optimization*, **35**, 69-76, 1997.
- [5] B.S He and L-Z Liao, Improvements of some projection methods for monotone nonlinear variational inequalities, *Journal of Optimization Theory and Applications*, **112**, 111-128, 2002
- [6] B.S. He, X.M. Yuan and J.J.Z. Zhang, Comparison of two kinds of prediction-correction methods for monotone variational inequalities, *Computational Optimization and Applications*, **27**, 247-267, 2004
- [7] B. S. He and M.-H. Xu, A general framework of contraction methods for monotone variational inequalities, *Pacific J. Optimization*, **4**, 195-212, 2008.
- [8] B.S. He, W. Xu , H. Yang, and X.M. Yuan, Solving resource protection and supply guarantee problems in economic equilibrium, *Netw. Spat. Econ.*, **11**,. 127-138, 2011.
- [9] L. Vandenberghe and S. Boyd, Semidefinite programming, *SIAM Review*, **38**, 49-95, 1995.
- [10] 何炳生, 论求解单调变分不等式的一些投影收缩算法, 《*计算数学*》, **18**, 54-60, 1996.
- [11] 何炳生, 邵虎, 徐明华, 大桥流量分配调控中的隐式互补问题. *中国科技论文在线优秀论文集*, 第三辑, 1-10, 2006.
- [12] 何炳生, 徐薇, 杨海, 袁晓明, 经济平衡中的一类保护资源何保障供给问题, *中国科技论文在线优秀论文集*, 第三辑, 11—19, 2006.