

从商业谈判到优化方法

— 来自社会生活理念的帮助

Bingsheng He

Department of Mathematics, Nanjing University

Email: hebma@nju.edu.cn

Homepage: math.nju.edu.cn/~hebma

南方科技大学 2015年6月2日

主要内容

- 许多优化问题等价于一个 min-max 问题（或称鞍点问题）。min-max 问题往往有经济方面的背景。
- 求解 min-max 问题, 犹如买卖双方博弈, 一方要支付尽可能小, 另一方要收益尽可能大. min-max 问题的解, 相当于买卖双方的一个双赢策略.
- 以商业谈判的视角, 辅以变分不等式为工具的数学解释, 讨论求解 min-max 问题的迭代方法, 说明社会生活理念对求解数学问题会有有益的启示和帮助.
- 展示根据这些简单理念设计的方法, 在国际应用数学界引起“过望”的反响. 说明简单的方法才有生命力, 才会被工程界拿去应用.

1 序言—从线性规划谈起

我们从一组线性规划谈起. 设某饲养场要购进一批饲料.

要求

- 含有 m 种营养素, 第 i -种营养素总量不少于 $b_i (i = 1, \dots, m)$ 千克.

市场

- 市场上共有 n 种自然饲料 (米糠, 大豆, 玉米, 鱼粉 等等), 第 j 种自然饲料每公斤的价格为 $c_j (j = 1, \dots, n)$ 元, 第 i 种营养素的含量为 a_{ij} .

	米糠	大豆	玉米	鱼粉	
$V - A$	a_{11}	a_{12}	a_{13}	a_{14}	2000
$V - B$	a_{21}	a_{22}	a_{23}	a_{24}	5000
自然饲料单价	c_1	c_2	c_3	c_4	

$$\sum_{i=1}^m a_{ij} \leq 1.$$

为尽可能节省支出, 饲养场该如何采购?

- 设饲养场将采购这 n 种自然饲料的量分别为 x_1, \dots, x_n 公斤. 总花费是

$$c_1x_1 + c_2x_2 + \dots + c_nx_n.$$

- 饲养场采购的自然饲料中第 i 种营养素的总量必须不小于 b_i . 也就是说

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n \geq b_i.$$

饲养场满足营养要求的最小支出采购计划是线性规划问题:

$$\begin{aligned}
 \min \quad & c_1x_1 + c_2x_2 + \dots + c_nx_n \\
 \text{s. t.} \quad & a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \geq b_1, \\
 & a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \geq b_2, \\
 & \dots\dots\dots \\
 & a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \geq b_m, \\
 & x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0.
 \end{aligned} \tag{1.1}$$

通常, 我们称这个采购问题为原始线性规划问题 (Primal Problem) .

对应于上述原始线性规划问题, 有一个与之相应的对偶线性规划问题.

- 设有一家营养素公司, 它能供应饲养场所需的各种营养素. 他们声称营养素价格是低廉的.
- 买了我们的营养素去配制含同量营养成分的自然饲料, 价格也不会高于市场上自然饲料的价格.

为了最大赢利, 营养素公司该怎样去确定营养素的价格?

- 设营养素公司对这 m 种营养素将设定的价格分别为 y_1, \dots, y_m . 他们能获取的总收入是

$$b_1y_1 + b_2y_2 + \dots + b_my_m.$$

- 用营养素合成一公斤第 j 种自然饲料的合成价格必须不大于市场自然饲料的价格 c_j .

$$a_{1j}y_1 + a_{2j}y_2 + \dots + a_{mj}y_m \leq c_j.$$

营养素公司实现最大盈利的价格方案是线性规划问题

$$\begin{aligned} \max \quad & b_1 y_1 + b_2 y_2 + \cdots + b_m y_m \\ \text{s. t.} \quad & a_{11} y_1 + a_{21} y_2 + \cdots + a_{m1} y_m \leq c_1 \\ & a_{12} y_1 + a_{22} y_2 + \cdots + a_{m2} y_m \leq c_2 \\ & \cdots \\ & a_{1n} y_1 + a_{2n} y_2 + \cdots + a_{mn} y_m \leq c_n \\ & y_1, y_2, \dots, y_m \geq 0. \end{aligned} \tag{1.2}$$

线性规划 (1.2) 称为问题 (1.1) 的对偶问题 (Dual Problem) .

- 线性规划中, 满足约束条件的解叫做可行解.
- **Primal Problem** 目标函数值最小的可行解称为原始问题的最优解.
- **Dual Problem** 目标函数值最大的可行解称为对偶问题的最优解.

采用矩阵一向量的表达式,

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}, \quad c = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix},$$

和

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}.$$

原始-对偶问题 (1.1) 和 (1.2) 就可以分别写成

$$\begin{array}{ll} \min & c^T x \\ \text{(Primal)} & Ax \geq b \\ & x \geq 0. \end{array} \quad \begin{array}{ll} \max & b^T y \\ \text{(Dual)} & A^T y \leq c \\ & y \geq 0. \end{array} \quad (1.3)$$

饲养场一定要实现采购计划

营养素公司恪守价格承诺

(原始可行)

$$Ax \geq b$$

$$x \geq 0.$$

(对偶可行)

$$A^T y \leq c$$

$$y \geq 0.$$

$x \geq 0, c - A^T y \geq 0$		$y \geq 0, Ax - b \geq 0$
$x^T (c - A^T y) \geq 0$		$y^T (Ax - b) \geq 0$
$c^T x \geq x^T A^T y$	$x^T A^T y = y^T Ax$	$y^T Ax \geq y^T b$
	$c^T x \geq b^T y$	

在大家都理性的前提下 $c^T x \geq b^T y$

Lagrange 函数

线性规划问题的 Lagrange 函数是定义在 $\mathfrak{R}_+^n \times \mathfrak{R}_+^m$ 上的

$$L(x, y) = c^T x - y^T (Ax - b). \quad (1.4)$$

鞍点

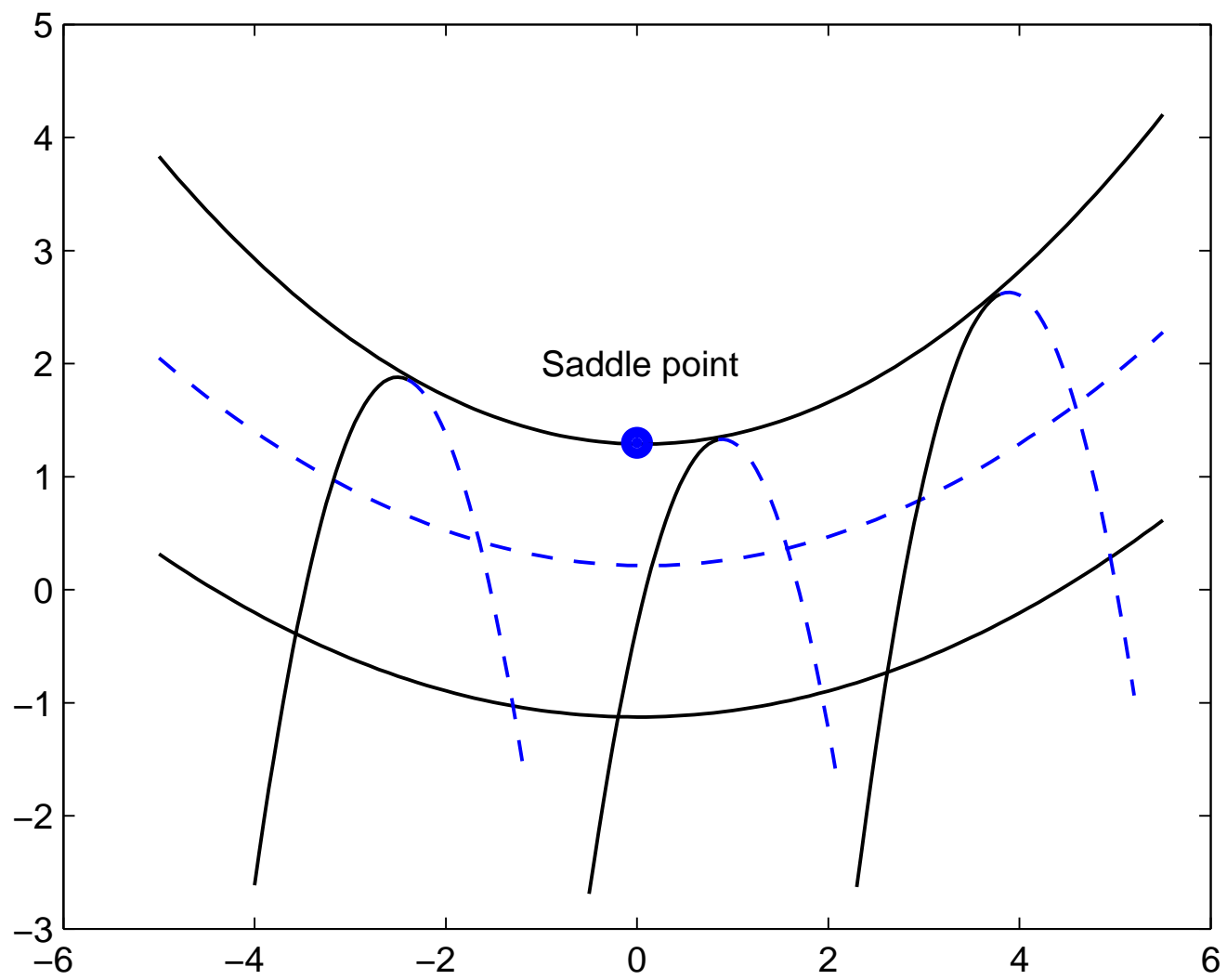
如果 $x^* \in \mathfrak{R}_+^n$ 和 $y^* \in \mathfrak{R}_+^m$ 满足

$$L_{y \in \mathfrak{R}_+^m}(x^*, y) \leq L(x^*, y^*) \leq L_{x \in \mathfrak{R}_+^n}(x, y^*), \quad (1.5)$$

则称 (x^*, y^*) 是 Lagrange 函数 $L(x, y)$ 在 $\mathfrak{R}_+^n \times \mathfrak{R}_+^m$ 上的一个鞍点.

求饲养场的最优采购策略 x^* 和营养素公司最优价格策略 y^* , 相当于求 Lagrange 函数 (1.4) 在 $\mathfrak{R}_+^n \times \mathfrak{R}_+^m$ 上的一个鞍点.

换句话说, 要求得一对 (x^*, y^*) , 使其满足 (1.5).



2 优化问题与鞍点问题

2.1 凸优化与变分不等式（解的性质）

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

$$\min\{\theta(x) \mid x \in \Omega\} \quad (2.1)$$

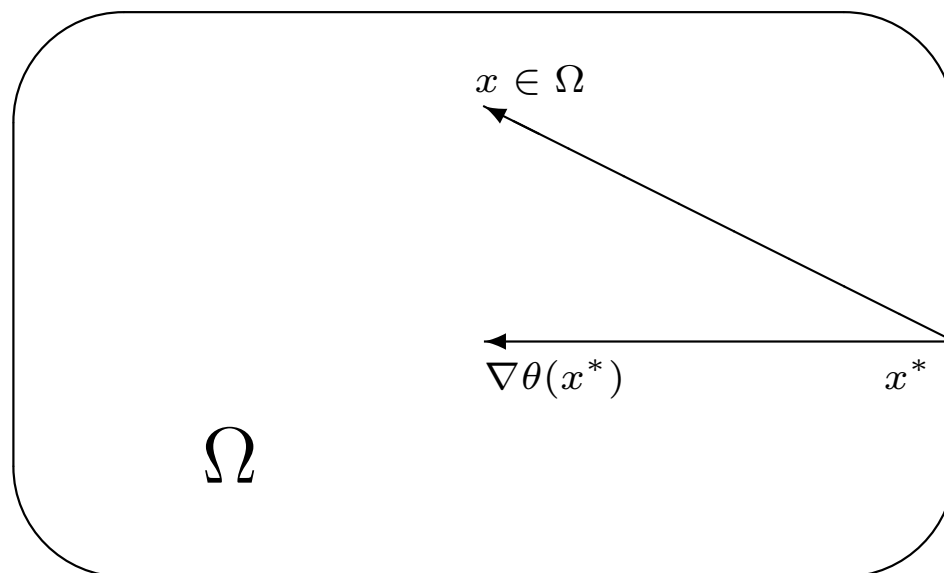
的最优性条件.

- 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

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

x^* 是最优解（瞎子爬山）： $x^* \in \Omega$ 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 (2.1) 就归结为求

$$x^* \in \Omega, (x - x^*)^T f(x^*) \geq 0, \forall x \in \Omega. \quad (2.2)$$

上述形式称为一个变分不等式, 记为 $\mathbf{VI}(f, \Omega)$.

2.2 极大-极小问题与相应的变分不等式

考虑如下一般形式的 $\min - \max$ 问题

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \Phi(x, y) := \theta_1(x) - y^T Ax - \theta_2(y), \quad (2.3)$$

其中 $\theta_1(x) : \mathfrak{R}^n \rightarrow \mathfrak{R}$, $\theta_2(y) : \mathfrak{R}^m \rightarrow \mathfrak{R}$ 是凸函数（不一定可微）。
 $A \in \mathfrak{R}^{m \times n}$, $\mathcal{X} \subset \mathfrak{R}^n$, $\mathcal{Y} \subset \mathfrak{R}^m$ 是闭凸集。

- 线性约束的凸优化问题

$$\min\{\theta(x) \mid Ax = b \text{ (or } \geq b), x \in \mathcal{X}\} \quad (2.4)$$

的 Lagrange 函数可以写成

$$L(x, y) = \theta(x) - y^T (Ax - b). \quad (2.5)$$

记 $\theta_1(x) = \theta(x)$, $\theta_2(y) = -b^T y$. 鞍点问题 (2.5) 是问题 (2.3) 的一个特例.

- 科学计算中的许多问题（例如图像处理中离散 ROF 模型 [11]）的数学形式也是形如 (2.3) 的一个 $\min - \max$ 问题.

min – max 问题相当于一个变分不等式

假如 (x^*, y^*) 是 (2.3) 的一个解 (函数 $\Phi(x, y)$ 在 $\mathcal{X} \times \mathcal{Y}$ 上的鞍点), 则有

$$\Phi_{y \in \mathcal{Y}}(x^*, y) \leq \Phi(x^*, y^*) \leq \Phi_{x \in \mathcal{X}}(x, y^*). \quad (2.6)$$

这时, 由 (2.6) 的右边的不等式

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

可以得到

$$x^* \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^*) + (x - x^*)^T (-A^T y^*) \geq 0, \quad \forall x \in \mathcal{X}. \quad (2.7a)$$

同样的道理, 从 (2.6) 左边的不等式

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

可以推得变分不等式

$$y^* \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^*) + (y - y^*)^T (Ax^*) \geq 0, \quad \forall y \in \mathcal{Y}. \quad (2.7b)$$

因此, 求 $\Phi(x, y)$ (2.3) 在 $\mathcal{X} \times \mathcal{Y}$ 上的鞍点等价于求 (混合) 变分不等式的解

$$u^* \in \Omega, \quad \theta(u) - \theta(u^*) + (u - u^*)^T F(u^*) \geq 0, \quad \forall u \in \Omega, \quad (2.8a)$$

where

$$\theta(u) = \theta_1(x) + \theta_2(y), \quad u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad F(u) = \begin{pmatrix} -A^T y \\ Ax \end{pmatrix} \quad (2.8b)$$

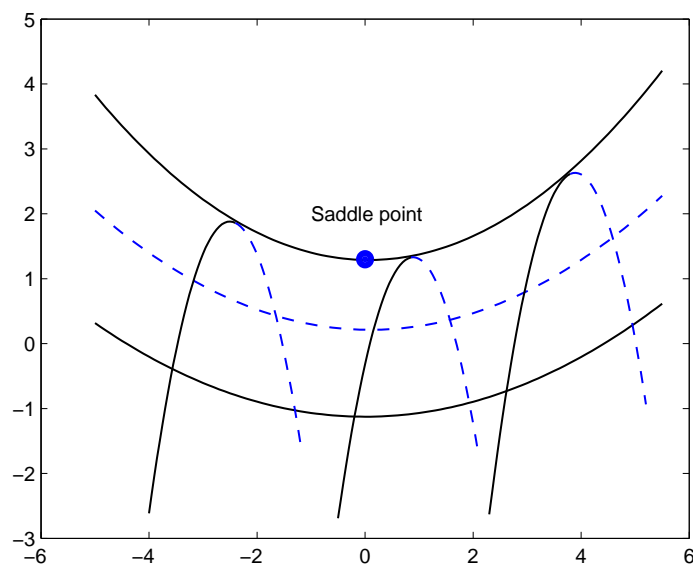
and

$$\Omega = \mathcal{X} \times \mathcal{Y}.$$

说 F 是单调的, 如果算子 F 满足

$$(u - v)^T (F(u) - F(v)) \geq 0.$$

如果变分不等式 (2.8) 中 θ 是凸函数, F 是单调算子的, 则称 (2.8) 为单调变分不等式.



变分不等式 (2.8) 中的

$$F(u) = \begin{pmatrix} -A^T y \\ Ax \end{pmatrix} = \begin{pmatrix} 0 & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = Mu,$$

其中 M 是反对称矩阵 ($M^T = -M$), 因此

$$(u - v)^T (F(u) - F(v)) \equiv 0.$$

所以 F 是单调的.

- 我们记 (2.8) 的解集为 Ω^* . 导出 min – max 问题 (2.3) 和变分不等式 (2.8) 的等价性, 并不是说要通过求解变分不等式来求解 min – max 问题, 而是要在变分不等式的框架下分析算法.
- 从经济博弈的观点, 对 min – max 问题 (2.3) 的求解方法进行讨论.
- 虽然不是所有的 min – max 问题都是买卖问题, 我们仍然可以看作买卖两方, 一方要支付尽可能小, 另一方要收益尽可能大. x 和 y 分别是买卖两方的决策变量. 迭代的过程犹如一个谈判的过程, 最后要得到双赢的决策变量 $u^* = (x^*, y^*) \in \Omega^*$.

3 变分不等式的邻近点算法—PPA 算法

这一节仅从数学的角度考虑变分不等式 (2.8) 的邻近点算法 (**Proximal Point Algorithm**, 简称 **PPA** 算法) [9, 10].

3.1 邻近点算法

PPA 算法是一种产生的迭代序列越来越靠近解集的方法.

对给定的 $u^k \in \mathfrak{R}^{m+n}$ 和 $r > 0$, 邻近点算法的第 k -次迭代中的子问题是

$$(PPA) \quad u \in \Omega, \quad \theta(v) - \theta(u) + (v - u)^T \{F(u) + r(u - u^k)\} \geq 0, \quad \forall v \in \Omega. \quad (3.1)$$

设 u^{k+1} 是 (3.1) 的解, 那么我们有 $u^{k+1} \in \Omega$,

$$\theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T \{F(u^{k+1}) + r(u^{k+1} - u^k)\} \geq 0, \quad \forall u \in \Omega, \quad (3.2)$$

以 u^* 替代上式中的 u , 得到

$$(u^{k+1} - u^*)^T r(u^k - u^{k+1}) \geq \{(u^{k+1} - u^*)^T F(u^{k+1}) + \theta(u^{k+1}) - \theta(u^*)\}. \quad (3.3)$$

由 F 是仿射算子 ($F(u) = Mu$) 且矩阵 M 是反对称的, 可得

$$(u^{k+1} - u^*)^T F(u^{k+1}) = (u^{k+1} - u^*)^T F(u^*).$$

因为 $u^{k+1} \in \Omega$ 并且 u^* 是 $\text{VI}(\Omega, F)$ 的解, 根据变分不等式 (2.8) 的定义, 有

$$\theta(u^{k+1}) - \theta(u^*) + (u^{k+1} - u^*)^T F(u^*) \geq 0.$$

根据 (3.3) 和上面两个关系式得到下面的关键不等式

$$(u^{k+1} - u^*)^T (u^k - u^{k+1}) \geq 0, \quad \forall u^* \in \Omega^*. \quad (3.4)$$

由 (3.4) 可以推出,

$$\begin{aligned}\|u^k - u^*\|^2 &= \|(u^{k+1} - u^*) + (u^k - u^{k+1})\|^2 \\ &\geq \|u^{k+1} - u^*\|^2 + \|u^k - u^{k+1}\|^2.\end{aligned}$$

因而

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - \|u^k - u^{k+1}\|^2. \quad (3.5)$$

-
- 序列 $\{u^k\}$ 满足的以上关系式, 称作序列 $\{u^k\}$ 在欧氏模下对解集 Ω^* 是 Fejér 单调的. 邻近算法的收敛性可见 [9] 和 [10].
 - 如果 $r = 0$, 子问题 (3.1) 就成了原问题 (2.8). 对与鞍点问题等价的变分不等式, 邻近点算法 (3.1) 只有抽象的理论意义, 一般在实际问题的求解中很少应用.
-

假如我们将 (3.1) 中的常数 $r > 0$ 换成一个对称正定矩阵 G , (3.2) 就变成

$$\theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T \{F(u^{k+1}) + G(u^{k+1} - u^k)\} \geq 0, \quad \forall u \in \Omega. \quad (3.6)$$

由同样的道理可以导出

$$(u^{k+1} - u^*)^T G(u^k - u^{k+1}) \geq 0, \quad \forall u^* \in \Omega^*.$$

最后会有

$$\|u^{k+1} - u^*\|_G^2 \leq \|u^k - u^*\|_G^2 - \|u^k - u^{k+1}\|_G^2. \quad (3.7)$$

-
- 由 (3.6) 生成新的迭代点 u^{k+1} 的算法称为 G -模下的 PPA 算法. 由 PPA 算法 (3.6) 生成的序列 $\{u^k\}$ 具有的性质 (3.7) 称为 G -模下 Fejér 单调的.
 - 后面我们将会看到, 对与鞍点问题等价的变分不等式, 通过适当选配的正定矩阵 G , 邻近点算法 (3.6) 是容易实现的.
-

Assumptions:

对给定的 $r, s > 0$ 和 $p \in \mathfrak{R}^n, q \in \mathfrak{R}^m$, 下述问题

$$\min\{\theta_1(x) + \frac{r}{2}\|x - p\|^2 \mid x \in \mathcal{X}\}, \quad \min\{\theta_2(y) + \frac{s}{2}\|y - q\|^2 \mid y \in \mathcal{Y}\}.$$

是容易求解的. 这样的 x, y -子问题的解 \tilde{x} 和 \tilde{y} 就会有

$$\tilde{x} \in \mathcal{X}, \quad (\theta_1(x) - \theta_1(\tilde{x})) + (x - \tilde{x})^T \{r(\tilde{x} - p)\} \geq 0, \quad \forall x \in \mathcal{X},$$

和

$$\tilde{y} \in \mathcal{Y}, \quad (\theta_2(y) - \theta_2(\tilde{y})) + (y - \tilde{y})^T \{s(\tilde{y} - q)\} \geq 0, \quad \forall y \in \mathcal{Y}.$$

4 以商业谈判的视角看求解方法

我们要求解的是一个 $\min - \max$ 问题

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \Phi(x, y) := \theta_1(x) - y^T Ax - \theta_2(y).$$

这个问题等价于一个变分不等式

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

采用 PPA 算法, 对给定的 u^k , 我们要求一个 u^{k+1} , 使得它满足

$$\theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T \{F(u^{k+1}) + G(u^{k+1} - u^k)\} \geq 0, \quad \forall u \in \Omega,$$

其中 G 是一个对称正定矩阵. 这样, 就会有

$$\|u^{k+1} - u^*\|_G^2 \leq \|u^k - u^*\|_G^2 - \|u^k - u^{k+1}\|_G^2.$$

下面, 我们从商业谈判的视角来看如何实现这种算法。

4.1 双方只顾自己利益的策略—PDG Method

我们把 $u^k = (x^k, y^k)$ 中的 x^k 和 y^k 看作是双方谈判到昨天各自的决策变量. 希望今天谈判得到的决策变量 (x^{k+1}, y^{k+1}) 在某种度量意义下更靠近最优的双赢决策变量 (x^*, y^*) . 谈判往往是这样进行的:

买方在知悉卖方昨天的决策变量 y^k , 同时考虑与自己昨天的决策变量 x^k 差距不要太大前提下, 追求支付尽可能小. 因此买方的问题是

$$x^{k+1} = \arg \min \{ \Phi(x, y^k) + \frac{r}{2} \|x - x^k\|^2 \mid x \in \mathcal{X} \}. \quad (4.1a)$$

类似地, 卖方在考虑新的一天的策略时, 在顾及了买方昨天的决策变量 x^k , 同时考虑与自己昨天的决策变量 y^k 差距不要太大的前提下, 追求收益尽可能大. 因此卖方通过

$$y^{k+1} = \arg \max \{ \Phi(x^k, y) - \frac{s}{2} \|y - y^k\|^2 \mid y \in \mathcal{Y} \}, \quad (4.1b)$$

得到他新的决策变量.

根据子问题 (4.1a) 和 (4.1b) 的最优性条件, 利用 (2.8b) 中给的记号, 由 (4.1) 产生的 $u^{k+1} = (x^{k+1}, y^{k+1})$ 就是变分不等式

$$u^{k+1} \in \Omega, \theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T \{F(u^{k+1}) + Q(u^{k+1} - u^k)\} \geq 0, \forall u \in \Omega, \quad (4.2)$$

的解, 其中

$$Q = \begin{pmatrix} rI_n & A^T \\ -A & sI_m \end{pmatrix}.$$

- 买卖双方虽然都考虑到自己的新决策变量不能离自己昨天的决策变量太远, 但在各自追求自己的利益最大化的时候, 只顾及了对方昨天的决策变量. 我们称这种算法为原始-对偶梯度方法(**Primal-Dual Gradient Method**).
- 用这种方法产生的序列 $\{u^k\} = \{(x^k, y^k)\}$, 是不可能向任何一个双赢决策变量 $u^* = (x^*, y^*) \in \Omega^*$ 收敛的.

可以用一个简单的例子就可以说明

考虑以下的一对原始—对偶线性规划问题：

$$\begin{array}{ll} \min & x \\ \text{(Primal)} & \text{s. t. } x \geq 1 \\ & x \geq 0. \end{array} \qquad \begin{array}{ll} \max & y \\ \text{(Dual)} & \text{s. t. } y \leq 1 \\ & y \geq 0. \end{array}$$

这对问题的最优解分别是 $x^* = 1$ 和 $y^* = 1$. 线性规划问题的 Lagrange 函数

$$L(x, y) = x - y(x - 1) \tag{4.3}$$

定义在 $R_+ \times R_+$ 上. $(x^*, y^*) = (1, 1)$ 也是 Lagrange 函数在 $R_+ \times R_+$ 上的唯一鞍点.

取 $r = s = 1$, 用 (4.1) 方法求解, 如图 1 所示, 若以 $(x^0, y^0) = (0, 0)$ 为初始点, 迭代序列 $\{u^k\}$ 在 11 个点上循环, 永远不会收敛到解点 $(1, 1)$.

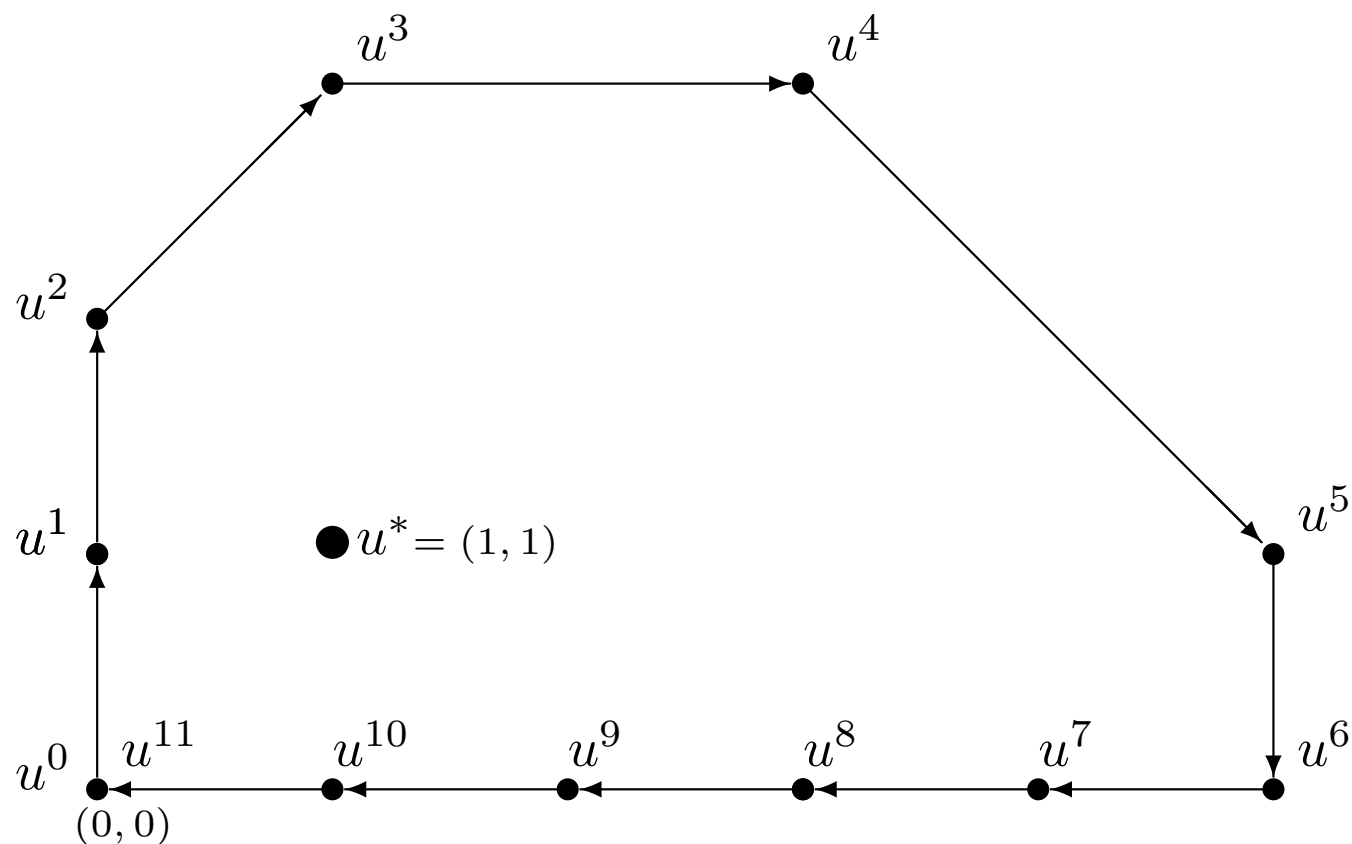


图 1. 原始-对偶梯度方法产生的迭代序列

4.2 前后出牌的策略—PDHG Method

把 x^k 和 y^k 看作是昨天双方谈判的决策变量. 现在考虑买卖双方先后出牌的策略. 不妨设买方上午出牌, 卖方下午出牌.

买方知悉卖方昨天的决策变量 y^k , 同时考虑今天的决策变量 x^{k+1} 不能与昨天自己的决策变量 x^k 差距太大. 因此, 买方通过求解

$$x^{k+1} = \arg \min \{ \Phi(x, y^k) + \frac{r}{2} \|x - x^k\|^2 \mid x \in \mathcal{X} \}, \quad (4.4a)$$

得到他的支付尽可能小的新决策变量 x^{k+1} .

卖方在考虑今天的决策变量时, 知悉了买方今天上午给出了新的决策变量 x^{k+1} , 他依追求收益最大化的原则新决策变量 y^{k+1} 时, 同时考虑与自己昨天的决策变量 y^k 差距不要太大. 因此, 他的新决策变量 y^{k+1} 通过

$$y^{k+1} = \arg \max \{ \Phi(x^{k+1}, y) - \frac{s}{2} \|y - y^k\|^2 \mid y \in \mathcal{Y} \} \quad (4.4b)$$

求得.

该方法称为原始-对偶混合梯度法(Primal-Dual Hybrid Gradient Method—PDHG).

由子问题 (4.4a) 和 (4.4b) 的最优性条件, 并利用 (2.8b) 中给的记号, 原始-对偶混合梯度法 (4.4) 产生的 $u^{k+1} = (x^{k+1}, y^{k+1})$ 就是

$$u^{k+1} \in \Omega, \theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T \{F(u^{k+1}) + Q(u^{k+1} - u^k)\} \geq 0, \forall u \in \Omega, \quad (4.5)$$

的解, 其中

$$Q = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix}.$$

这种先后出牌的策略, 并不公平, 方法也是不可能收敛的.

对第 4.1 节中线性规划问题, 用 PDHG 求解, 我们还是取 $r = s = 1$, 并用 (4.4) 求 Lagrange 函数 (4.3) 的鞍点. 以 $(x^0, y^0) = (0, 0)$ 为初始点, 如图 2 所示, 从 $u^1 = (0, 1)$ 开始, 迭代序列 $\{u^k\}$ 在 6 个点上循环, 永远不会收敛到解点 $(1, 1)$.

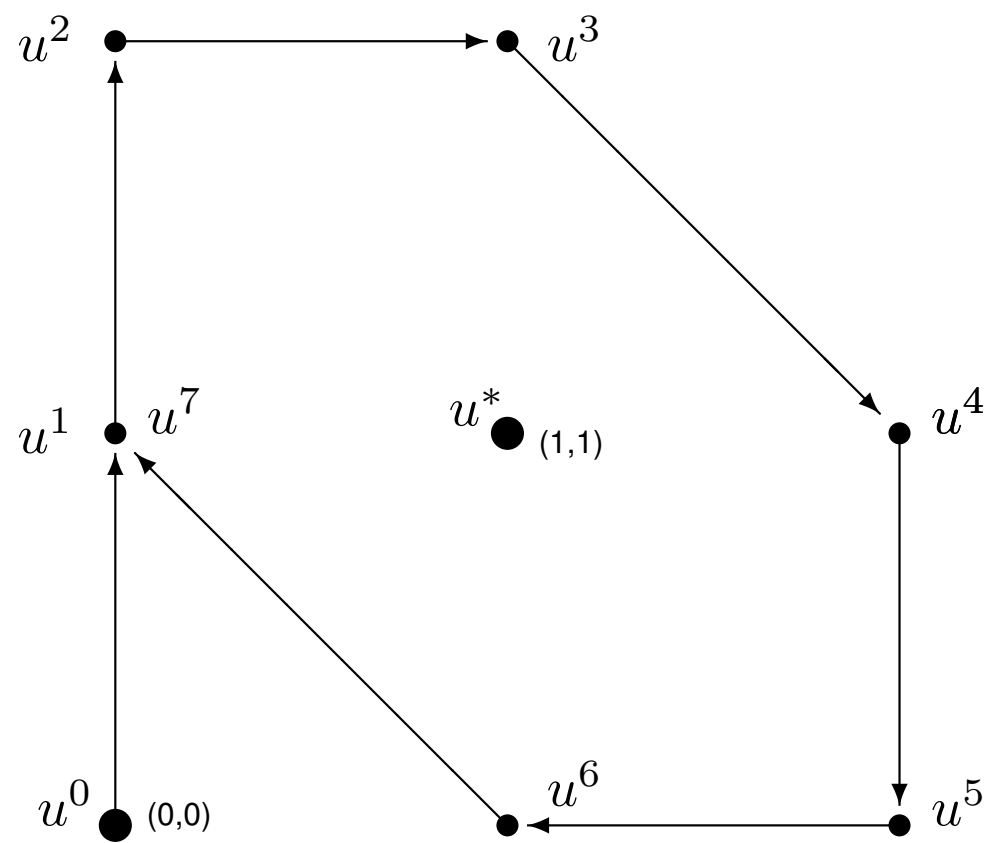


图 2. 原始-对偶混合梯度方法产生的迭代序列

4.3 按需定制的邻近点算法—Customized PPA

把 x^k 和 y^k 看作是昨天双方谈判的决策变量. 方法 (4.1) 和方法 (4.4), 都可以看作买方上午出牌, 卖方下午出牌.

- 方法 (4.1) 中, 卖方在下午出牌时用的仍是买方昨天提供的信息.
- 方法 (4.4) 中, 卖方在下午出牌时采用了买方上午更新的信息.

因此, 从 (4.1) 到 (4.4), 差别仅在于卖方的策略由

$$y^{k+1} = \arg \max \{ \Phi(x^k, y) - \frac{s}{2} \|y - y^k\|^2 \mid y \in \mathcal{Y} \}$$

变化到

$$y^{k+1} = \arg \max \{ \Phi(x^{k+1}, y) - \frac{s}{2} \|y - y^k\|^2 \mid y \in \mathcal{Y} \}.$$

注意到这时变分不等式 (4.2) 与 (4.5) 中的矩阵 Q 的左下角子块从 $-A$ 变化到一个 $m \times n$ 的零矩阵.

$$-A \quad \Rightarrow \quad 0_{m \times n}.$$

如果进一步将 0 矩阵变化到矩阵 A , 那么相应的 Q 就成了一个对称矩阵.

$$\begin{pmatrix} rI_n & A^T \\ -A & sI_m \end{pmatrix} \Rightarrow \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \Rightarrow \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}.$$

由于 $\begin{pmatrix} rI_n & A^T \\ -A & sI_m \end{pmatrix}$ 变成 $\begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix}$,
 是由 x^k 变到 $x^k + (x^{k+1} - x^k)$ 导致的,

所以, 考虑买方上午出牌, 卖方下午出牌的游戏规则时,

要使得 $\begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix}$ 变成 $\begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}$,
 只要将 $x^k + (x^{k+1} - x^k)$ 变成 $x^k + 2(x^{k+1} - x^k)$.

换句话说, 我们只要将 (4.4) 改造成

$$x^{k+1} = \arg \min \{ \Phi(x, y^k) + \frac{r}{2} \|x - x^k\|^2 \mid x \in \mathcal{X} \}, \quad (4.6a)$$

和

$$y^{k+1} = \arg \max \{ \Phi([2x^{k+1} - x^k], y) - \frac{s}{2} \|y - y^k\|^2 \mid y \in \mathcal{Y} \}. \quad (4.6b)$$

这种方法可以解释为后者利用了前者的最新信息, 在考虑自己的策略的时候, 应该让前者有一个更大的前进量.

由子问题 (4.6a) 和 (4.6b) 的最优性条件, $u^{k+1} = (x^{k+1}, y^{k+1})$ 是变分不等式

$$u^{k+1} \in \Omega, \theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T \{ F(u^{k+1}) + G(u^{k+1} - u^k) \} \geq 0, \forall u \in \Omega, \quad (4.7)$$

的解, 其中

$$G = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}.$$

G 是对称矩阵, 当 $rs > \|A^T A\|$ 时, G 是正定的.

- 第 3 节的分析告诉我们, u^{k+1} 是变分不等式问题 (2.8) 在 G -模下 PPA (邻近点) 算法的解. 由 (4.6) (其等价形式 (4.7)) 产生的序列 $\{u^k\}$ 具有收缩性质 (3.7) 并最终收敛于问题 (2.8) 的一个解点.
- 这一节的 PPA 方法, 根据算法的收敛性需要, 人为地构造了一个正定矩阵 G . 我们称这个方法为按需定制的 PPA 算法(Customized PPA) [8].

按需定制的 PPA (Customized PPA) 算法产生的迭代序列收敛于问题的解点, 理论能证明, 计算实践中也得到证实.

我们还可以拿第 4.1 节中线性规划的例子来说明. 采用按需定制的 PPA 求解, 同样取 $r = s = 1$, 以 $(x^0, y^0) = (0, 0)$ 为初始点, 收敛的示意如图 3 所示.

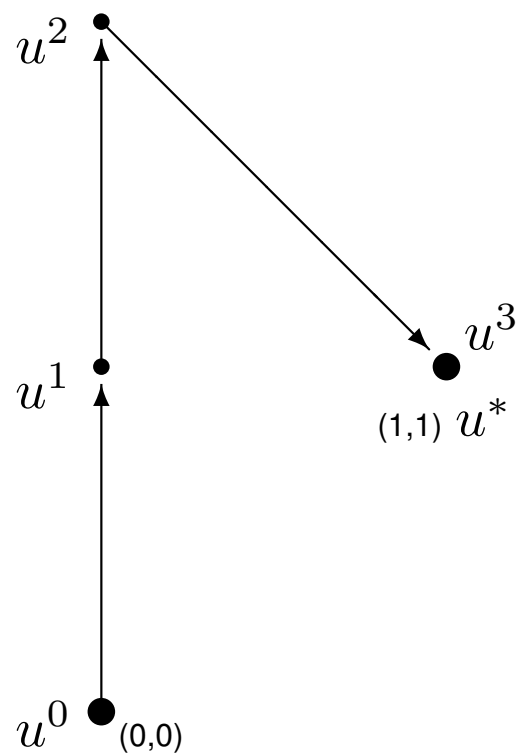


图 3. 应需定制 PPA 方法产生的迭代序列

这种方法在图像重构上得到广泛的应用, 读者可以参考文献 [1] 和 [2].

4.4 加速的定制邻近点算法—Extended Customized PPA

- 应需定制的邻近点算法 (4.6)使得在 G -模意义下 u^{k+1} 比 u^k 更靠近解集.
- 在商业谈判中,买卖双方在经过一天的艰苦谈判,取得比前一天更接近双赢策略,晚上,双方一起出去喝喝茶,在友好的气氛下把当天的成果再一起向前推进一步.
- 用算法来描述,就是

$$u^{k+1} := u^k - \gamma(u^k - u^{k+1}), \quad \gamma \in (1, 2), \quad (4.8)$$

上式右端的 u^{k+1} 是由 (4.6) 提供的.

- 这种方法,常常会提高收敛速度 [6, 8]. 笔者主页上讲义 [3] 的第 4 讲附有不多这样的例子并附有相应的 Matlab 程序.

4.5 Application for matrix completion problems

Let M be a given $m \times n$ matrix, Π is the elements indices set of M ,

$$\Pi = \{(ij) \mid i \in \{1, 2, \dots, m\}, j \in \{1, 2, \dots, n\}\}.$$

The mathematical form of the matrix completion problem is relaxed to

$$\min\{\|X\|_* \mid X_{ij} = M_{ij}, (ij) \in \Pi\}, \quad (4.9)$$

where $\|\cdot\|_*$ is the nuclear norm—the sum of the singular values of a given matrix.

The problem (4.9) is a convex optimization of form (2.4). The matrix A in (2.4) for the linear constraints

$$X_{ij} = M_{ij}, (ij) \in \Pi,$$

is a projection matrix, and thus $\|A^T A\| = 1$.

We let $Z \in \Re^{n \times n}$ as the Lagrangian multiplier to the constraints $X_\Omega = M_\Omega$.

For given (X^k, Z^k) , applying (4.6) to produce (X^{k+1}, Z^{k+1}) :

1. Producing Z^{k+1} by

$$z_{\Omega}^{k+1} = Z_{\Omega}^k - \frac{1}{s}(X_{\Omega}^k - M_{\Omega}). \quad (4.10)$$

2. Finding X^{k+1} by

$$X^{k+1} = \arg \min \left\{ \|X\|_* + \frac{r}{2} \left\| X - \frac{1}{r} [X^k - (2\tilde{Z}_{\Omega}^k - Z_{\Omega}^k)] \right\|_F^2 \right\}. \quad (4.11)$$

Then, the new iterate is given by

$$X^{k+1} := X^k - \gamma(X^k - X^{k+1}), \quad Z^{k+1} := Z^k - \gamma(Z^k - z^{k+1}).$$

Test examples

The test examples is taken from

- ◇ J. F. Cai, E. J. Candès and Z. W. Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optim. **20**, 1956-1982, 2010.

Code for Creating the test examples of Matrix Completion

```

%% Creating the test examples of the matrix Completion problem      %(1)
clear all;  clc                                                    %(2)
maxIt=100;          tol = 1e-4;                                     %(3)
r=0.005;           s=1.01/r;          gamma=1.5;                  %(4)
    n=200;          ra = 10;          oversampling = 5;           %(5)
% n=1000;    ra=100;    oversampling = 3; %% Iteration No. 31     %(6)
% n=1000;    ra=50;    oversampling = 4; %% Iteration No. 36     %(7)
% n=1000;    ra=10;    oversampling = 6; %% Iteration No. 78     %(8)
%% Generating the test problem                                     %(9)
rs = randseed;          randn('state',rs);                         %(10)
M=randn(n,ra)*randn(ra,n);          %% The matrix will be completed %(11)
df =ra*(n*2-ra);          %% The freedom of the matrix            %(12)
mo=oversampling;                                                %(13)
m =min(mo*df,round(.99*n*n));          %% No. of the known elements  %(14)
Omega= randsample(n^2,m);          %% Define the subset Omega    %(15)
fprintf('Matrix: n=%4d  Rank(M)=%3d  Oversampling=%2d \n',n,ra,mo);%(16)

```

Code: Customized PPA for Matrix Completion Problem

```

function PPAE(n,r,s,M,Omega,maxIt,tol,gamma)    % Initial Process %%(1)
X=zeros(n);      Y=zeros(n);      YT=zeros(n);    %(2)
nM0=norm(M(Omega),'fro');    eps=1;    VioKKT=1;    k=0;    tic;    %(3)
%% Minimum nuclear norm solution by PPA method    %(4)
while (eps > tol && k<= maxIt)    %(5)
    if mod(k,5)==0    %(6)
        fprintf(' It=%3d |X-M|/|M|=%9.2e VioKKT=%9.2e\n',k,eps,VioKKT); end; %(7)
        k=k+1;      X0=X;      Y0=Y;    %(8)
        YT(Omega)=Y0(Omega)-(X0(Omega)-M(Omega))/s;      EY=Y-YT;    %(9)
        A = X0 + (YT*2-Y0)/r;      [U,D,V]=svd(A,0);    %(10)
        D=D-eye(n)/r;      D=max(D,0);      XT=(U*D)*V';      EX=X-XT;    %(11)
        DXM=XT(Omega)-M(Omega);      eps = norm(DXM,'fro')/nM0;    %(12)
        VioKKT = max( max(max(abs(EX)))*r, max(max(abs(EY))) );    %(13)
        if (eps <= tol)    gamma=1;    end;    %(14)
        X = X0 - EX*gamma;    %(15)
        Y(Omega) = Y0(Omega) - EY(Omega)*gamma;    %(16)
    end;    %(17)
    fprintf(' It=%3d |X-M|/|M|=%9.2e VioKKT=%9.2e \n',k,eps,VioKKT);    %(18)
    RelEr=norm((X-M),'fro')/norm(M,'fro');    toc;    %(19)
    fprintf(' Relative error = %9.2e Rank(X)=%3d \n',RelEr,rank(X));    %(20)
    fprintf(' Violation of KKT Condition = %9.2e \n',VioKKT);    %(21)

```

Numerical Results: using SVD in Matlab

Unknown $n \times n$ matrix M				Computational Results		
n	$\text{rank}(ra)$	m/d_{ra}	m/n^2	#iters	times(Sec.)	relative error
1000	10	6	0.12	76	841.59	9.38E-5
1000	50	4	0.39	37	406.24	1.21E-4
1000	100	3	0.58	31	362.58	1.50E-4

Numerical Results: Using SVD in PROPACK

Unknown $n \times n$ matrix M				Computational Results			
n	$\text{rank}(ra)$	m/d_{ra}	m/n^2	#iters	times(Sec.)	relative error	Cai
1000	10	6	0.12	76	30.99	9.30E-5	117
1000	50	4	0.39	36	40.25	1.29E-4	114
1000	100	3	0.58	30	42.45	1.50E-4	119

♣ The paper by J. F. Cai, E. J. Candès and Z. W. Shen is the first publication in SIAM J. Optimization for matrix completion problem. For the same accuracy, their method needs **117**, **114** and **129** iterations (See the first three examples in Table 5.1 of this paper, pp. 1974).

5 Customized PPA 的应用与引用

5.1 Application in Image Science

Frame of VI is recognized by some Researcher in Image Science

**Diagonal preconditioning for first order primal-dual algorithms
in convex optimization***

Thomas Pock

Institute for Computer Graphics and Vision
Graz University of Technology
`pock@icg.tugraz.at`

Antonin Chambolle

CMAP & CNRS
École Polytechnique
`antonin.chambolle@cmap.polytechnique.fr`

- A. Chambolle and T. Pock, A first-order primal-dual algorithms for convex problem with applications to imaging, J. Math. Imaging Vision, **40**, 120-145, 2011.

preconditioned algorithm. In very recent work [10], it has been shown that the iterates (2) can be written in form of a proximal point algorithm [14], which greatly simplifies the convergence analysis.

From the optimality conditions of the iterates (4) and the convexity of G and F^* it follows that for any $(x, y) \in X \times Y$ the iterates x^{k+1} and y^{k+1} satisfy

$$\left\langle \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}, F \begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} + M \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} \right\rangle \geq 0, \quad (5)$$

where

$$F \begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \begin{pmatrix} \partial G(x^{k+1}) + K^T y^{k+1} \\ \partial F^*(y^{k+1}) - K x^{k+1} \end{pmatrix},$$

and

$$M = \begin{bmatrix} T^{-1} & -K^T \\ -\theta K & \Sigma^{-1} \end{bmatrix}. \quad (6)$$

It is easy to check, that the variational inequality (5) now takes the form of a proximal point algorithm [10, 14, 16].

- [9] L. Ford and D. Fulkerson. *Flows in Networks*. Princeton University Press, Princeton, New Jersey, 1962.
- [10] B. He and X. Yuan. Convergence analysis of primal-dual algorithms for total variation image restoration. Technical report, Nanjing University, China, 2010.
- [11] J. Lellmann, D. Breitenreicher, and C. Schnörr. Fast and exact primal-dual iterations for variational problems in computer vision. In K. Daniilidis, P. Maragos, and N. Paragios, editors, *European Conference on Computer Vision (ECCV)*, volume 6312 of *LNCS*, pages 494–505. Springer Berlin / Heidelberg, 2010.

我们的论文正式发表在

- B. S. He and X. M. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective, *SIAM J. Imag. Sci.*, 5, 119-149, 2012.

5.2 Goldstein, Esser and Baraniuk 的引用

ADAPTIVE PRIMAL-DUAL HYBRID GRADIENT METHODS FOR SADDLE-POINT PROBLEMS

TOM GOLDSTEIN, ERNIE ESSER, RICHARD BARANIUK

ABSTRACT. The Primal-Dual hybrid gradient (PDHG) method is a powerful optimization scheme that breaks complex problems into simple sub-steps. Unfortunately, PDHG methods require the user to choose stepsize parameters, and the speed of convergence is highly sensitive to this choice. We introduce new adaptive PDHG schemes that automatically tune the stepsize parameters for fast convergence without user inputs. We prove rigorous convergence results for our methods, and identify the conditions required for convergence. We also develop practical implementations of adaptive schemes that formally satisfy the convergence requirements. Numerical experiments show that adaptive PDHG methods have advantages over non-adaptive implementations in terms of both efficiency and simplicity for the user.

arXiv:1305.0546v1 [math.NA] 2 May 2013

作者在这篇文章中采用 1. 自调比的思想. 2. 变分不等式的框架,

4. RESIDUAL BALANCING

When choosing the stepsize for PDHG, there is a tradeoff between the primal and dual residuals. Choosing a large value of τ_k creates a very powerful minimization step in the primal variables and a slow maximization step in the dual variables, resulting in very small primal residuals at the cost of large dual residuals. Choosing τ_k to be small, on the other hand, results in small dual residuals at the cost of large primal errors.

Ideally, one would like to choose stepsizes so that the larger of P_k and D_k is as small as possible. If we assume the primal/dual residuals decrease/increase monotonically with τ_k , then $\max\{P_k, D_k\}$ is minimized when both residuals are equal in magnitude. This suggests that τ_k be chosen to “balance” the primal and dual residual – i.e., the primal and dual residuals should be roughly the same size, up to some scaling to account for units. **This principle has been suggested for other iterative methods (see [26, 27]).**

作者提到, 这里的自调比原理是在 [26, 27] 中的迭代方法提出的, [26] 是我们的文章, Boyd 等在 [27] 中介绍我们的自调比原理.

6. CONVERGENCE THEORY

6.1. **Variational Inequality Formulation.** For notational simplicity, we define the vector quantities

$$(29) \quad u_k = \begin{pmatrix} x_k \\ y_k \end{pmatrix}, \quad R(u) = \begin{pmatrix} P(x, y) \\ D(x, y) \end{pmatrix},$$

and the matrices

$$(30) \quad M_k = \begin{pmatrix} \tau_k^{-1} I & -A^T \\ -A & \sigma_k^{-1} I \end{pmatrix}, \quad H_k = \begin{pmatrix} \tau_k^{-1} I & 0 \\ 0 & \sigma_k^{-1} I \end{pmatrix}.$$

This notation was first suggested to simplify PDHG by He and Yuan [7].

作者指出, 这个表达方式首先是由He & Yuan 提出来简化 PDHG 的。

Following [7], it is possible to formulate each step of PDHG as a solution to the variational inequality

$$(31) \quad 0 \in R(u_{k+1}) + M_k(u_{k+1} - u_k).$$

Also, the optimality conditions (5) and (6) can be written succinctly as

$$(32) \quad 0 \in R(u^*).$$

Note that R is monotone (see [28]), meaning that

$$\langle u - \hat{u}, R(u) - R(\hat{u}) \rangle \geq 0, \quad \forall u, \hat{u}.$$

这里 Following 的 [7] 就是 He & Yuan 2012 年发表在 SIAM J. Image Science 上的文章。

This first lemma has been adapted from He and Yuan [7]. It shows that the stability condition **C1** forces the matrix M_k to be positive definite. When this condition is satisfied, the operator $\langle u, M_k u \rangle = \|u\|_{M_k}$ is a proper norm and can be used in our convergence analysis.

作者说到, Lemma 1 从我们的论文 [7] Adapt 而来。

- [7] B. He and X. Yuan, “Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective,” *SIAM J. Img. Sci.*, vol. 5, pp. 119–149, Jan. 2012.
- [26] B. He, H. Yang, and S. Wang, “Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities,” *Journal of Optimization Theory and Applications*, vol. 106, no. 2, pp. 337–356, 2000.
- [27] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers,” *Foundations and Trends in Machine Learning*, 2010.

6 简单的方法与过望的引用

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

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

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

$$\min\{\theta_1(x) + \theta_2(y) + \theta_3(z) \mid Ax + By + Cz = b, x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\}$$

6.1 投影收缩算法: Jordan 课题组的引用

UC Berkeley 计算机科学系 Jordan 教授课题组的文章.

Structured Prediction via the Extragradient Method

Ben Taskar

Computer Science

UC Berkeley, Berkeley, CA 94720

taskar@cs.berkeley.edu

Simon Lacoste-Julien

Computer Science

UC Berkeley, Berkeley, CA 94720

slacoste@cs.berkeley.edu

Michael I. Jordan

Computer Science and Statistics

UC Berkeley, Berkeley, CA 94720

jordan@cs.berkeley.edu

Michael I. Jordan 教授是美国科学院和工程院院士.

采用我们同样的表述方式

Let us define a residual error function for Eq. (10):

$$e(\mathbf{u}, \beta) = \mathbf{u} - P_{\mathcal{U}}(\mathbf{u} - \beta T(\mathbf{u})).$$

Note that for $\beta > 0$, the zeros of $e(\mathbf{u}, \beta)$ are precisely the solutions of Eq. (10):

$$e(\mathbf{u}, \beta) = 0 \quad \Leftrightarrow \quad \mathbf{u} \in \mathcal{U}^*. \quad (11)$$

直接引用我们的定理证明收敛性

Theorem A.1 (Theorem 2.1 in [11]) Let $\{\beta_k\}$ be a sequence such that $\inf_k \{\beta_k\} = \beta_{\min} > 0$ and let $c_0 > 0$ be a constant. If the sequence $\{\mathbf{u}^k\}$ satisfies:

$$\|\mathbf{u}^{k+1} - \mathbf{u}^*\|^2 \leq \|\mathbf{u}^k - \mathbf{u}^*\|^2 - c_0 \|e(\mathbf{u}^k, \beta_k)\|^2, \quad \forall \mathbf{u}^* \in \mathcal{U}^*, \quad (13)$$

then $\{\mathbf{u}^k\}$ converges to a solution point of Eq. (10).

Jordan 等作者这篇论文的部分参考文献:

- [10] F. Guerriero and P. Tseng. Implementation and test of auction methods for solving generalized network flow problems with separable convex cost. *Journal of Optimization Theory and Applications*, 115(1):113–144, October 2002.
- [11] B.S. He and L. Z. Liao. Improvements of some projection methods for monotone nonlinear variational inequalities. *JOTA*, 112:111:128, 2002.
- [12] M. Jerrum and A. Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM J. Comput.*, 22, 1993.

Jordan 课题组发表在 Machine Learning 上的文章

Jordan 课题组在 Machine Learning 上论文的首页

Journal of Machine Learning Research 7 (2006) 1627–1653

Submitted 10/05; Published 7/06

Structured Prediction, Dual Extragradient and Bregman Projections

Ben Taskar

Simon Lacoste-Julien

Computer Science

University of California

Berkeley, CA 94720, USA

Michael I. Jordan

Computer Science and Statistics

University of California

Berkeley, CA 94720, USA

TASKAR@CS.BERKELEY.EDU

SLACOSTE@CS.BERKELEY.EDU

JORDAN@CS.BERKELEY.EDU

STRUCTURED PREDICTION, DUAL EXTRAGRADIENT AND BREGMAN PROJECTIONS

where $\|\mathbf{F}\|_2$ is the largest singular value of the matrix \mathbf{F} . In practice, various simple heuristics can be considered for setting the stepsize, including search procedures based on optimizing the gap merit function (see, e.g., He and Liao, 2002).

这里实际上是考虑：采用的步长与方向都使得新的迭代点离解集的距离越来越近。关于这类算法, 作者建议参阅我们的论文。

Jordan 课题组的这篇论文的部分参考文献:

Trevor Hastie, Saharon Rosset, Robert Tibshirani, and J. Zhu. The entire regularization path for the support vector machine. *Journal of Machine Learning Research*, 5:1391–1415, 2004.

Bingsheng He and Li-Zhi Liao. Improvements of some projection methods for monotone nonlinear variational inequalities. *Journal of Optimization Theory and Applications*, 112:111–128, 2002.

Mark Jerrum and Alistair Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM Journal on Computing*, 22:1087–1116, 1993.

6.2 乘子交替方向法(ADMM): Boyd 课题组的引用

交替方向法. 最优化中采用分裂算法, 就像大工程必须分包。从招收第一批博士生开始, 在收缩算法框架下研究交替方向法就是我们的一个主要课题。交替方向法就是把大问题分解成容易求解的一系列小问题去做。对该方法中影响收敛速度的参数选择, 10 多年前发表的自调比准则, 被 Boyd 教授在文章中称为一个简单而有效的公式。

Foundations and Trends® in
Machine Learning
Vol. 3, No. 1 (2010) 1–122
© 2011 S. Boyd, N. Parikh, E. Chu, B. Peleato
and J. Eckstein
DOI: 10.1561/22000000016

now
the essence of knowledge

Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers

By **Stephen Boyd**, Neal Parikh, Eric Chu,
Borja Peleato and Jonathan Eckstein

Stephen Boyd 是 Stanford 大学电子工程系教授, 美国工程院院士.
2006 年世界数学家大会 45 分钟邀请报告人.

Boyd 等在这篇论文的第20 页中说到:

A simple scheme that often works well is (see, *e.g.*, [96, 169]):

$$\rho^{k+1} := \begin{cases} \tau^{\text{incr}} \rho^k & \text{if } \|r^k\|_2 > \mu \|s^k\|_2 \\ \rho^k / \tau^{\text{decr}} & \text{if } \|s^k\|_2 > \mu \|r^k\|_2 \\ \rho^k & \text{otherwise,} \end{cases} \quad (3.13)$$

where $\mu > 1$, $\tau^{\text{incr}} > 1$, and $\tau^{\text{decr}} > 1$ are parameters. Typical choices might be $\mu = 10$ and $\tau^{\text{incr}} = \tau^{\text{decr}} = 2$. The idea behind this penalty parameter update is to try to keep the primal and dual residual norms within a factor of μ of one another as they both converge to zero.

Boyd 等这篇论文的部分参考文献:

- [95] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, second ed., 2009.
- [96] B. S. He, H. Yang, and S. L. Wang, “Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities,” *Journal of Optimization Theory and Applications*, vol. 106, no. 2, no. 2, pp. 337–356, 2000.
- [97] M. R. Hestenes, “Multiplier and gradient methods,” *Journal of Optimization Theory and Applications*, vol. 4, pp. 302–320, 1969.

- [169] S. L. Wang and L. Z. Liao, “Decomposition method with a variable parameter for a class of monotone variational inequality problems,” *Journal of Optimization Theory and Applications*, vol. 109, no. 2, no. 2, pp. 415–429, 2001.

交替方向法广受重视, 我们完成了它（包括线性化策略）的 $O(1/t)$ 收敛速率证明

6.3 多个算子的分裂收缩算法：Osher 课题组的引用

- E. Esser, M. Möller, S. Osher, G. Sapiro and J. Xin, A convex model for non-negative matrix factorization and dimensionality reduction on physical space, IEEE Trans. Imag. Process., 21(7), 3239-3252, 2012.

A convex model for non-negative matrix factorization and dimensionality reduction on physical space

Ernie Esser, Michael Möller, Stanley Osher, Guillermo Sapiro, Jack Xin

Stanley Osher 是 UCLA 数学系教授, 美国科学院院士, 2010 年世界数学家大会一小时邀请报告人.

他们考虑的是降维问题, 归结成三个可分离算子的线性约束凸优化

$$\min_{T \geq 0, V_j \in D_j, e \in E} \zeta \sum_i \max_j(T_{i,j}) + \langle R_w \sigma C_w, T \rangle \quad (15)$$

such that $YT - X_s = V - X_s \text{diag}(e).$

其中说到, 问题(15) 稍微复杂一些, 需要用允许将函数分裂成两部分以上的方法处理

Since the convex functional for the extended model (15) is slightly more complicated, it is convenient to use a variant of ADMM that allows the functional to be split into more than two parts. The method proposed by He, Tao and Yuan in [29] is appropriate for this application. Again introduce a

对这个应用, 由 He, Tao, Yuan 在引文 [29] 中提出的方法是适用的。

使用 [29] 中提出的 ADMM 类方法, 可以通过迭代下面的子问题 (取参数 $\delta > 0$ 和 $\mu > 2$) 得到增广 Lagrangian 的一个鞍点。

Using the ADMM-like method in [29], a saddle point of the augmented Lagrangian can be found by iteratively solving the following subproblems with parameters $\delta > 0$ and $\mu > 2$,

由于用不同的方法求解扩展了的模型, 这里需要一个额外的参数 μ , 根据 [29], 对这个应用, 参数 μ 必须大于 2. 我们取 $\mu = 2.01$.

alternating minimization refinement step. Due to the different algorithm used to solve the extended model, there is an additional numerical parameter μ , which for this application must be greater than two according to [29]. We set μ equal to 2.01. There are also model parameters r_j and γ for modeling

Osher 课题组论文的部分参考文献:

- [28] E. Candes, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis,” 2009, http://arxiv.org/PS_cache/arxiv/pdf/0912/0912.3599v1.pdf.
- [29] B. He, M. Tao, and X. Yuan, “A splitting method for separate convex programming with linking linear constraints,” Tech. Rep., 2010.
- [30] Y. Sun, C. Ridge, F. del Rio, A.J. Shaka, and J. Xin, “Postprocessing and sparse blind source separation of positive and partially overlapped data,” Tech. Rep., 2010.

我们的论文 [29] 在 2013 年录用, 今年会在 **IMA Numerical Analysis** 发表

- B.S. He, M. Tao and X.M. Yuan, A splitting method for separable convex programming.

Conclusions

- Customized PPA belongs to the class of Proximal Point Algorithms.
- The discussed first order splitting contraction methods are only appropriate for some structured convex optimization in practical applications.

感想：数学之美，不是纯数学的专利。为应用服务的最优化方法研究，同样可以追求简单与统一。简单，他人才会看懂使用；统一，自己才有美的享受。

✘ 追求简单与统一，是我研究工作欲罢不能的原因 ✘

✘ 感谢导师的培养，朋友的支持，还有经常提出问题一起讨论的学生 ✘

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] T. Goldstein, E. Esser and R. Baraniuk, Adaptive primal-dual hybrid gradient method for saddle-point problems, arXiv: 1305.0546v1 [math.NA] 2 May 2013.
- [3] B.S. He, Lectures of Contraction Methods for Convex Optimization and Monotone Variational Inequalities, <http://math.nju.edu.cn/~hebma>
- [4] B. S. He and Y. Shen, On the convergence rate of customized proximal point algorithm for convex optimization and saddle-point problem (in Chinese). *Sic Sin Math*, **42**, 515-525, 2012. doi: 10.1360/012011-1049
- [5] B.S. He, H. Yang and S.L. Wang, Alternating direction method with self- adaptive penalty parameters for monotone variational inequalities. *JOTA*, 106: 337 – 356, 2000.
- [6] B.S. He and X. M. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective, *SIAM Journal on Imaging*

Science, **5**, 119-149, 2012.

- [7] B.S. He and X.M. Yuan, On the $O(1/n)$ Convergence Rate of the Douglas-Rachford Alternating Direction Method, SIAM J. Numer. Anal. 50, 700-709, 2012.
- [8] B. S. He, X. M. Yuan and W.X. Zhang, A customized proximal point algorithm for convex minimization with linear constraints, Comput. Optim. Appl., 56: 559-572, 2013.
- [9] B. Martinet, Regularisation, d'inéquations variationnelles par approximations succesives, Rev. Francaise d'Inform. Recherche Oper., **4**, 154-159, 1970.
- [10] R.T. Rockafellar, Monotone operators and the proximal point algorithm, SIAM J. Cont. Optim. **14**, 877-898, 1976.
- [11] L. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D, 60 (1992), 227-238.
- [12] M. Zhu and T. F. Chan, An efficient primal-dual hybrid gradient algorithm for total variation image restoration, CAM Reports 08-34, UCLA, 2008.



Thank you very much for your attention !