

# 条件梯度法求解非线性分裂可行问题

宇振盛, 王子伦

上海理工大学理学院, 上海  
Email: zhsh-yu@163.com, 1257578347@qq.com

收稿日期: 2020年9月1日; 录用日期: 2020年9月19日; 发布日期: 2020年9月27日

---

## 摘要

本文研究了分裂可行问题的条件梯度算法, 该算法将求解迭代方向转化成求解一线性子问题, 并以线搜索得到的步长作为凸因子, 当前方向与上一步迭代点的凸组合作为新的迭代点。算法在迭代的更新步中不使用投影, 并且得到的解有较好的稀疏性和低秩性。我们获得了算法的收敛性并给出数值实验对比分析了本文的算法与相关算法在同一算例下的表现情况, 得到了良好的结果。

## 关键词

非线性分裂可行问题, 条件梯度法, 替代函数, 稀疏约束集, IMRT问题

---

# Solving Nonlinear Split Feasibility Problem via Conditional Gradient Method

Zhensheng Yu, Zilun Wang

College of Science, University of Shanghai for Science and Technology, Shanghai  
Email: zhsh-yu@163.com, 1257578347@qq.com

Received: Sep. 1<sup>st</sup>, 2020; accepted: Sep. 19<sup>th</sup>, 2020; published: Sep. 27<sup>th</sup>, 2020

---

## Abstract

In this paper, we consider a conditional gradient algorithm for the split feasibility problem. We get the iterative direction by solving a linear subproblem, use the step size obtained by the line search as the convex factor, and get the new iteration point by the convex combination of the current direction and the previous step iteration point. The algorithm does not use projection in the update step. And the obtained solution has good properties of sparsity and low rank. The numerical experiment compares the performance of the proposed algorithm with other algorithms under the same calculation example.

## Keywords

Nonlinear Split Feasibility, Conditional Gradient Method, Surrogate Function, Sparse Constraint, IMRT

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

本文主要研究欧几里得空间中的单集分裂可行问题(Nonlinear Split Feasibility Problem), 即找一个  $x^* \in R^n$  使其满足

$$x^* \in C \text{ 并且 } Ax^* \in Q \quad (1.1)$$

其中  $C \subseteq R_n$  与  $Q \subseteq R_m$  都为非空闭凸集,  $A$  代表矩阵或者线性变换。该问题由 Censor 等人[1]提出, 它在放射治疗(IMRT)计划领域有着重要的作用[2] [3], 在工程技术领域, 例如信号和图像恢复, 基因调控网络推论以及用于处理高维稀疏数据的 LASSO 和 Dantzig Selector 方法[4] [5] [6]中也有广泛应用。

循环投影法是最先被提出用于求解分裂可行问题的一种迭代算法, 其迭代公式为:

$$x^{k+1} = A^{-1} (I + AA^T)^{-1} (AP_C(x^k) + AA^T P_Q(Ax^k)) \quad (1.2)$$

这里,  $P_C$  和  $P_Q$  分别表示向非空闭凸集  $C$  和  $Q$  上的正交投影。该方法显然不是很理想, 因为在每次迭代中都需要计算矩阵的逆。当将该方法应用于大规模的实际问题时, 求解逆矩阵非常困难, 这也可能导致算法花费太长时间。因此, Byrne 给出了著名的 CQ 算法[7]。在之后的分裂可行问题算法的发展中, CQ 算法得到更多的关注。假设  $C$  和  $Q$  分别是欧几里得或希尔伯特空间  $H_1$  和  $H_2$  上的闭合凸子集, 并且  $A$  是有界线性算子。CQ 算法的迭代格式如下

$$x^{k+1} = P_C(x^k - \theta A^*(I - P_Q)Ax^k) \quad (1.3)$$

学者们还提出了一些其他的方法[8] [9] [10] [11], 将分裂可行问题等价于不动点问题或变分不等式问题进行求解, 因此应用于其等效问题的算法可以作适应性调整并应用于分裂可行问题, 例如 Krasnosel'skii-Mann (KM)算法[12] [13]在解决分裂可行问题方面取得了重大突破。Zhao 将 KM 迭代形式应用于扰动 CQ 算法。由于 KM 迭代算法具有较弱的收敛性, 因此作为一种特殊形式, CQ 算法也具有较弱的收敛性。Xu [14]改进了上述算法, 并很好地解决了这一缺点, 改进算法具有较强的收敛性。

考虑对分裂可行问题的一种情况: 在欧式空间中存在集合  $C_i \subseteq R_n$ ,  $Q_j \subseteq R_m$  其中  $i=1,2,\dots,p$ ,  $j=1,2,\dots,r$  并满足(1.1)式条件, 这个问题称为多集分裂可行性问题(MSSFP), 更进一步, 我们设矢量值函数或非线性函数  $F: R_n \rightarrow R_m$  代替  $A$ 。本文中我们考虑双集合非线性分裂可行性问题, 即

$$x^* \in C \text{ and } F(x^*) \in Q \quad (1.4)$$

在 2005 年, Censor 等人[2]提出了一种投影梯度算法, 最小化一个使用最大最小法(Majorize-Minimize Method)构建的目标函数如下

$$f(x) = \frac{1}{2} \sum_i v_i \text{dist}(x, C_i)^2 + \frac{1}{2} \sum_j w_j \text{dist}[H(x), Q_j]^2 \quad (1.5)$$

其中  $dist(x, y)$  表示向量  $x$  和  $y$  之间的欧氏距离。权重系数或者凸系数  $v_i$  和  $w_j$  为正数且满足  $\sum v_i + \sum w_j = 1$ 。算法的迭代公式为:

$$x^{k+1} = P_{\Omega} \left\{ x^k + s \left( \sum_{i=1}^l \alpha_i (P_{C_i}(x^k) - x^k) + \sum_{j=1}^r \beta_j A^T (P_{Q_j}(Ax^k) - Ax^k) \right) \right\} \quad (1.6)$$

所以问题转化为最小化问题  $\min f(x)$ 。我们的目标是找到一个解  $x^*$ ，使在此点的目标函数的梯度满足  $\nabla f(x^*) < \varepsilon$ 。根据上面的替代函数，我们可以推导出非线性分裂可行问题的目标函数和梯度函数分别为

$$f(x) = \frac{1}{2} \|x - P_C(x)\|^2 + \frac{1}{2} \|F(x) - P_Q(F(x))\|^2 \quad (1.7)$$

$$\nabla f(x) = (x - P_C x) + \nabla F(x)^T (F(x) - P_Q(F(x))) \quad (1.8)$$

其中  $\nabla F(x): R^m \rightarrow R^n$  是映射  $F$  在  $x$  点处的雅可比矩阵。

李志宝[15]提出了一种自适应投影类型方法，其文章算法基于投影梯度法加上自适应的技术求解非线性分裂可行问题。作者基于一阶信息使用了新的步长搜索，迭代公式为  $x^{k+1} = P_{\Omega} [x^k - \beta_{k+1} \nabla p(x^k)]$ ，其中步长  $\beta_{k+1}$  满足

$$\beta_{k+1} \|\nabla f(x^k) - \nabla f(x^{k+1})\|^2 \leq (2 - \delta)(x^k - x^{k+1})^T (\nabla f(x^k) - \nabla f(x^{k+1}))$$

2014 年 A. Gibali [16]提出了一种求解非线性分裂可行问题的信赖域方法也展现出了良好的收敛性与数值实验结果。

本文，我们给出非线性分裂可行问题的一个条件梯度算法，我们考虑等价的无约束优化问题

$$\min f(x) = \frac{1}{2} \|x - P_C(x)\|^2 + \frac{1}{2} \|F(x) - P_Q(F(x))\|^2$$

该算法将求解迭代方向转化成求解一线性子问题，并以线搜索得到的步长作为凸因子，当前方向与上一步迭代点的凸组合作为新的迭代点。算法在迭代的更新步中不使用投影，并且得到的解有较好的稀疏性和低秩性。我们获得了算法的收敛性并给出数值实验对比分析  $x$  了本文的算法与相关算法在同一算例下的表现情况，得到了良好的结果。

## 2. 预备知识

### 2.1. 欧式空间上的投影的性质

设向量  $x$  为  $R_n$  空间的任意向量，内积定义为  $\langle \cdot, \cdot \rangle$ ， $x$  的范数定义为  $\|x\| = \sqrt{\langle x, x \rangle}$ 。向量  $x$  向  $C$  集上的投影用  $P_C$  表示，定义式为

$$P_C x = \arg \min_{y \in C} \|y - x\|, \quad \forall y \in C \quad (2.1)$$

**定理 2.1** [17]  $P_C$  与  $I - P_C$  是非扩张映射，有以下结论成立:

$$\|P_C x - P_C y\| \leq \|x - y\| \quad (2.2)$$

$$\|(I - P_C)(x - y)\| \leq \|x - y\| \quad (2.3)$$

其中  $x, y \in C$ 。

若算子  $T$  满足  $FixT = \{x \in D : x = Tx\}$ ，则  $T$  是一非扩张算子，称作不动点映射。我们定义

$$T_C \text{ 与 } T_Q : T_C := \{x \in C : x = P_C x\}, \quad T_Q := \{H(x) = y \in Q : y = P_Q y\} \quad (2.4)$$

**引理 2.1** 若迭代点  $x^k$  满足

$$x^k \in T_C \cap T_Q \tag{2.5}$$

则说明  $x^k$  为最优解。

**证明:** 若  $x = x^k \in T_C \cap T_Q$ , 则  $x^k = P_C x^k$  且  $y^k = F(x^k) = P_Q(F(x^k))$ , 所以

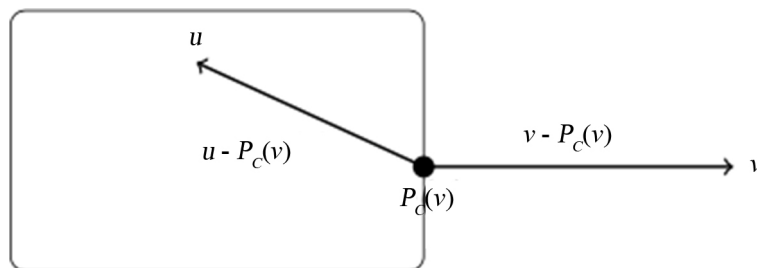
$$f(x^k) = \frac{1}{2} \|x^k - P_C(x^k)\|^2 + \frac{1}{2} \|F(x^k) - P_Q(F(x^k))\|^2 = 0$$

显然  $f(x) \geq 0$ , 所以当  $x = x^k$  时, 目标函数达到最小值 0。根据投影定理, 当向量与其投影之间的差为 0 时, 我们可以得出结论: 向量  $x^k$  在集合  $C$  的内部或边界上, 并且  $F(x^k)$  也同时位于集合  $Q$  的内部或边界上, 即上述两向量满足非线性分裂可行性问题的约束, 因此获得了一个解, 这意味着向量  $x^k$  为问题的最优解。

**引理 2.2** 设  $C$  是  $R^n$  上的闭凸集, 则存在

$$(v - P_C(v))^T (u - P_C(v)) \leq 0 \tag{2.6}$$

此不等式关系以二维为例如图 1 所示如下:



**Figure 1.** Lemma 2.2 schematic diagram of projection inequality

**图 1.** 引理 2.2 投影不等式示意图

向范数球  $N_D = \{z \in H : \|z - z_0\| \leq r\}$  上投影的解析表达式为

$$P_C x = \begin{cases} x, & x \in C \\ z_0 + \frac{r(x - z_0)}{\|x - z_0\|}, & x \notin C \end{cases}$$

**推论 1:** 假设  $K$  是空间  $X$  中的闭凸子集。定义任何满足  $x \in X$  的唯一元素为  $\pi_K(x) \in K$ , 有

$$x - \pi_K(x) = \min_{y \in K} \|x - y\| \tag{2.7}$$

其性质为如下变分不等式

$$\langle \pi_K(x) - x, \pi_K(x) - y \rangle \leq 0, \quad \forall y \in K \tag{2.8}$$

**命题 1:** 假设  $K$  是一个非空闭凸集, 而  $d_K^2$  是在  $X$  上由  $d_K^2(x) = \inf_{y \in K} \|x - y\|^2$  定义的函数。  $d_K^2$  是连续可微的, 其梯度为

$$\nabla d_K^2(x) = 2(x - \pi_K(x)) \tag{2.9}$$

定义等式约束集:

$$\mathcal{E} = \left\{ x \in R^n : (\alpha_1, \alpha_2, \dots, \alpha_r)^T x = (\beta_1, \beta_2, \dots, \beta_r) \right\} \rightarrow \left\{ x \in R^n : \alpha_i^T x = \beta_i \right\}$$

与不等式约束集:

$$\mathcal{J} = \{x \in R^n : (\alpha_1, \alpha_2, \dots, \alpha_r)^T x \leq (\beta_1, \beta_2, \dots, \beta_r)\} \rightarrow \{x \in R^n : \alpha_i^T x \leq \beta_i\}$$

向量  $x$  向集合  $\mathcal{E}$  与集合  $\mathcal{J}$  的投影解析表达式为:

$$P_{\mathcal{E}}(x) = x + (b_i - a_i^T x) \frac{a_i}{\|a_i\|^2}$$

$$P_{\mathcal{J}}(x) = x + \min\{0, b_i - a_i^T x\} \frac{a_i}{\|a_i\|^2}$$

在本文中, 我们还将考虑稀疏约束下的非线性分裂可行问题[18]

$$S_r := \{x \in R^n \mid \|x\|_0 \leq r, 0 \leq r \leq n\}$$

定义向量  $x$  的指标集  $I(x) = \{i_1, i_2, \dots, i_r\} \subseteq \{1, 2, \dots, n\}$ , 例如  $x = \{x_{i_1}, x_{i_2}, \dots, x_{i_r}\}$ , 则其指标集为  $I(x) = \{I_1, I_2, \dots, I_{2n}\}$ ,  $x = \{x_{I_1}, x_{I_2}, \dots, x_{I_{2n}}\}$ , 且有

$$\max_{i \in I(x)} \|x_i\| \leq \min_{i \in I(x)} \|x_i\|$$

定义  $x$  在稀疏集  $S_r$  上的投影为  $P_{S_r}(x)$ , 根据投影的性质, 记  $M_s(\|x\|)$  为向量  $x$  中最大的分量, 我们有:

$$P_{C_s}(x) = \begin{cases} x_i & x_i > M_s(\|x\|) \\ 0 \text{ 或 } x_i & x_i = M_s(\|x\|) \\ 0 & x_i < M_s(\|x\|) \end{cases}$$

由于 0 范数约束问题是一个 NP 难问题, 我们使用范数和范数的最优凸近似[19]在问题(1.1)中, 如果集合  $C$  是一稀疏集合  $C_s = \{x \in R^n \mid \|x\|_0 \leq r\}$ , 则其可以与下述问题等价

$$\min_x f(x)$$

s.t.  $\|x\|_1 \leq \mu$

## 2.2. 条件梯度法

Marguerite Frank 和 Philip Wolfe [20]在 1956 年首次提出条件梯度法, 方法的思想是在每一次迭代中通过求解一个线性的子问题得到最优的下降方向, 在更新步时与上一步的迭代点凸组合得到下一步的迭代点。条件梯度法由于其非常好的稀疏性与低秩性, 在大数据与机器学习热门的近几年里收获了极大的关注[21][22], 其算法描述如下:

$$\max_{\lambda} h(\lambda)$$

s.t.  $\lambda \in Q$

在当前迭代点  $x^k$  处, 该算法考虑目标函数的线性化, 并求解出该线性函数的最小值(定义域相同)。通过 LO oracle 解决线性子问题, 也可根据不同的问题适当调整子问题的处理方法, 从而大大提高了整个算法的灵活性。

以下两个图是条件梯度法的迭代过程的示意图。图 2 左边表示集合  $C$  中的迭代过程。每个迭代方向都来自当前迭代向量, 沿着  $C$  集合的边界向量的方向。步长可以控制每个迭代向量与边界向量之间的距离以实现更合适的下降。图 2 右侧图是在分裂可行问题下的, 我们的解集是图中的集  $C$  和集  $Q$  的交集。

使用 FW 方法解决分裂可行问题是非常行之有效的。

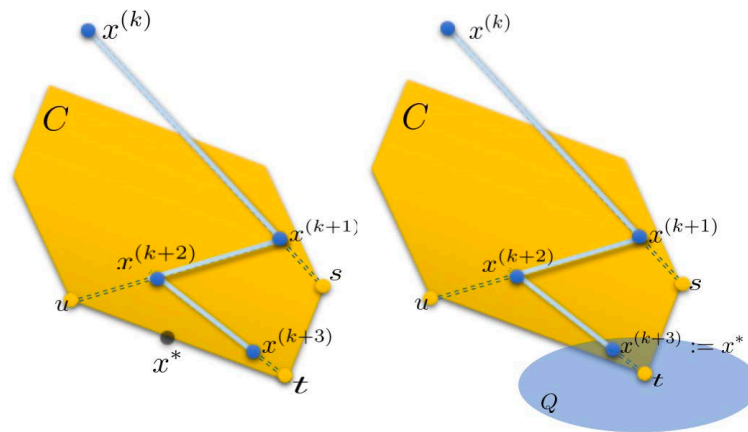


Figure 2. Conditional gradient method iteration diagram  
图 2. 条件梯度法迭代示意图

算法 1——原始条件梯度法

步 0: 设置初始点  $x^0 \in Q$ ;

步 1: 若  $\nabla f(x^{(k)}) = 0$  停止, 否则计算  $s^{(k)} := \arg \max \langle s, \nabla f(x^{(k)}) \rangle$  其中  $s \in D$ ;

步 2: 更新迭代点  $x^{(k+1)} := (1-\gamma)x^{(k)} + \gamma s^{(k)}$  其中  $\gamma = 2/k+2 \in [0,1)$ ,  $k = k+1$ , 返回步 1。

### 2.3. 对偶间隙

定义目标函数的对偶间隙[23] [24]

$$\mathcal{G}(x) = \max \langle x - s, \nabla f(x) \rangle, s \in C \tag{2.10}$$

对偶间隙是作为算法每次迭代的副产品, 无需进行额外的计算。同时我们记在  $x^k$  处的下降方向为  $d_k^{FW} = s^k - x^k$ 。

定义 2.1 (下降方向)若目标函数在集合  $C$  上是连续可微的, 则  $d_k^{FW}$  为下降方向的充要条件是  $\mathcal{G}(x^k) < 0$ , 并且当  $\mathcal{G}(x^k) = 0$  时, 目标函数达到最小值,  $x^k$  为最优解。

证明: 由于

$$s^{(k)} := \arg \max \langle s, \nabla f(x^{(k)}) \rangle$$

有  $s^k$  与  $x^k$  处梯度内积最小, 所以

$$\begin{aligned} \langle s^k, \nabla f(x^k) \rangle &\leq \langle x^k, \nabla f(x^k) \rangle \\ \langle s^k, \nabla f(x^k) \rangle - \langle x^k, \nabla f(x^k) \rangle &\leq 0 \\ \langle s^k - x^k, \nabla f(x^k) \rangle &\leq 0 \\ \nabla f(x^k)^\top d_k^{FW} &\leq 0 \end{aligned}$$

因此, 可以看出,  $d_k^{FW}$  是下降方向, 当且仅当  $x^k$  是最优解时, 等号成立。

引理 2.4 为一收敛序列, 并且有  $\lim_{i \rightarrow \infty} \mathcal{G}(x^i) = 0$ 。

证明略。

### 3. 算法

以下是条件梯度法解决非线性分裂可行问题的算法步骤:

---

#### 算法 2——带线搜索条件梯度法

---

**步 0:** 设置初始点  $x^0$ ,  $k=0, \beta \in (0,1), \delta \in (0,0.5), \epsilon < 10^{-6}$

**步 1:** 若  $\|\mathcal{G}(x^k)\| < \epsilon$  终止算法并输出  $x^k$ , 否则计算  $s^{(k)} := \arg \max \langle s, \nabla f(x^{(k)}) \rangle$  其中  $s \in C$

**步 2:** 计算下降方向  $d_k^{FW} = s^k - x^k$ , 并求出一个正整数  $m$ , 满足

$$f(x^k + \lambda d_k^{FW}) \leq f(x^k) + \delta \beta^m \nabla f(x^k)^T d_k^{FW}$$

得到线搜索步长  $\lambda_k = \beta^m$

**步 3:** 更新迭代点  $x^k = x^k + \lambda_k d_k^{FW}, k = k+1$ , 返回步 1

---

**引理 3.1** 对于  $\forall x, x^i \in C$ ,  $\langle \nabla f(x^i), x^* \rangle \leq \langle \nabla f(x^i), x \rangle$ 。

**证明:** 目标函数  $f(x)$  在  $x^*$  处与任意  $x^r \in C$  的泰勒展开式为

$$f(x^*) = f(x^i) + \nabla f(x^i)(x^* - x^i) + \mathcal{O}(x) = 0$$

与

$$f(x^r) = f(x^i) + \nabla f(x^i)(x^r - x^i) + \mathcal{O}(x) = \epsilon_r$$

其中  $\epsilon_r > 0$ , 因此我们有

$$\nabla f(x^i)(x^* - x^i) < \nabla f(x^i)(x^r - x^i)$$

所以

$$\langle \nabla f(x^i), x^* \rangle \leq \langle \nabla f(x^i), x^r \rangle, \forall x^r, x^i \in C$$

证毕。我们自然推导出以下引理

**引理 3.2**  $f(x^k) - f(x^{k-1}) < 0$

**证明**

$$\begin{aligned} f(x^k) - f(x^{k-1}) &= f(x^i) + \nabla f(x^i)(x^k - x^i) - [f(x^i) + \nabla f(x^i)(x^{k-1} - x^i)] \\ &= \nabla f(x^i)(x^k - x^i) - \nabla f(x^i)(x^{k-1} - x^i) \\ &= \langle \nabla f(x^i), x^k \rangle - \langle \nabla f(x^i), x^{k-1} \rangle \\ &= \nabla f(x^i)(x^k - x^{k-1}) = \lambda_{k-1} \nabla f(x^i) d_{k-1}^{FW} \leq 0 \end{aligned}$$

其中  $\lambda_{k-1} \in (0,1)$ 。当且仅当  $x^i$  是最优解时, 等号成立。

## 4. 数值实验

### 4.1. 数值结果

论文应采在本节中我们将通过如下的例子对算法 2 进行数值实验, 并与其他算法进行对比。

考虑非线性分裂可行问题:

$$F(x) := [x_1, x_2, 2x_1^2 + 2x_2^2 + 1]$$

非空闭凸集  $C_{NL}, C_S$  与  $Q_{NL}$  分别是:

(问题 1)  $C_{NL} := \{x \in R^2 \mid \|x\|_2 \leq 1\}$ ,  $Q_{NL} := \{y \in R^3 \mid \|y-d\|_2 \leq 1\}$

(问题 2)  $C_S := \{x \in R^2 \mid \|x\|_0 \leq 1\}$ ,  $Q_{NL} := \{y \in R^3 \mid \|y-d\|_2 \leq 1\}$

其中  $d = [0, 1.8, 3]^T$ 。

本文的数值实验是使用 MATLAB2018a 在同一台计算机上进行的。图 3 是针对上述问题, 三种算法的迭代路径示意图。对于问题 1, 本文比较了最大值最小值法(MM), 传统投影梯度算法(Projected Gradient Method)和本文提出的算法 2 在不同初始点的迭代次数与计算时间的差异。从表 1 可以看出, 与其他两种算法相比, MM 法需要更多的迭代才能到达求解区域。与算法 2 相比, PGM 的迭代路径明显更“曲折”。图 4 中迭代次数与目标函数值的比较图。可以看出, 算法 2, PGM 和 MM 方法可以使目标函数值在先前的迭代中迅速减小到较小的值, 但是 MM 需要更多的迭代次数, PGM 在迭代的上一步中, 函数值将反弹, 算法 2 迅速收敛到 0, 从而获得良好的结果。表 1 列出了一些特殊的初

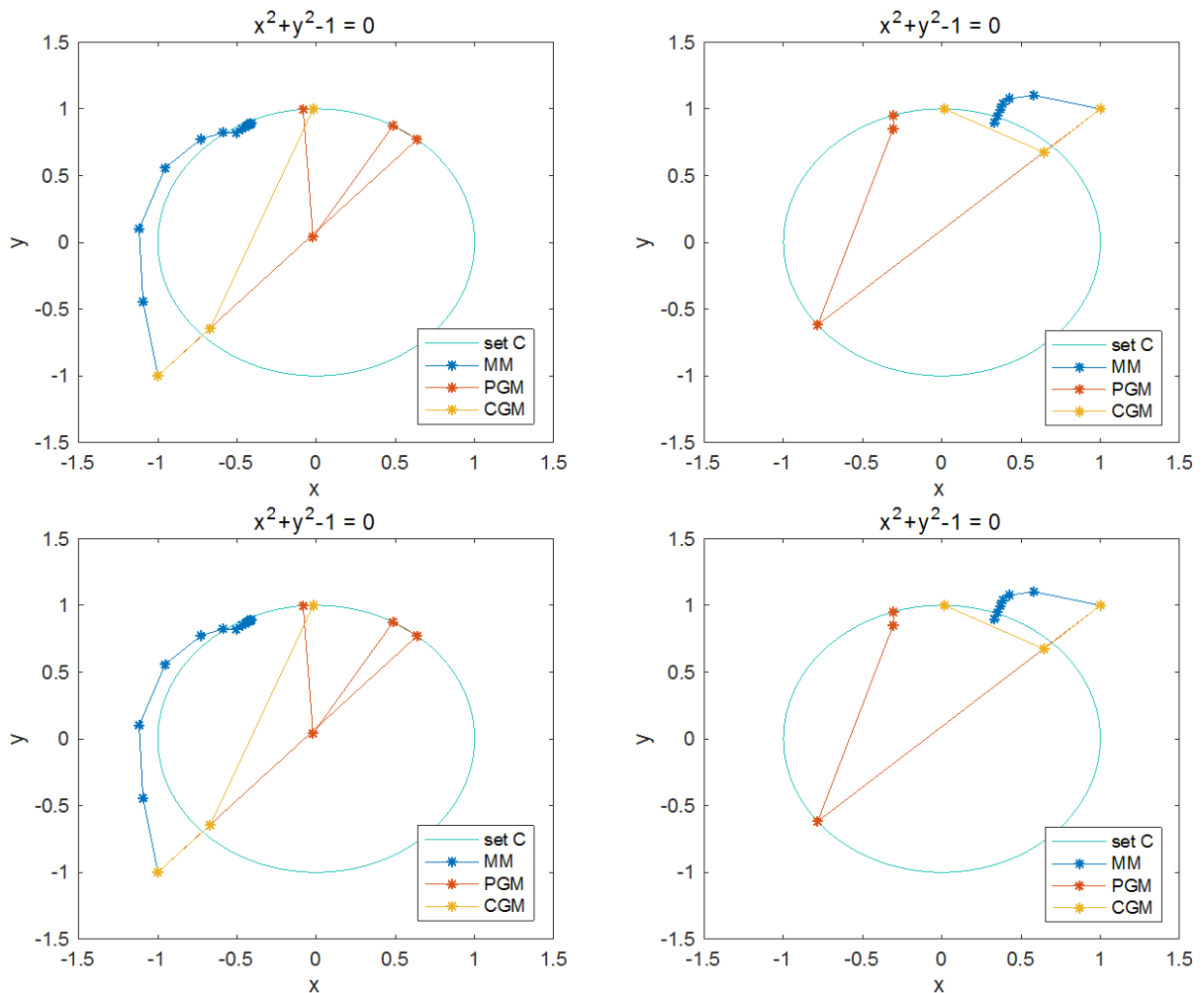


Figure 3. Comparison of iterative paths of the three algorithms under different initial points

图 3. 不同初始点下三种算法的迭代路径比较图

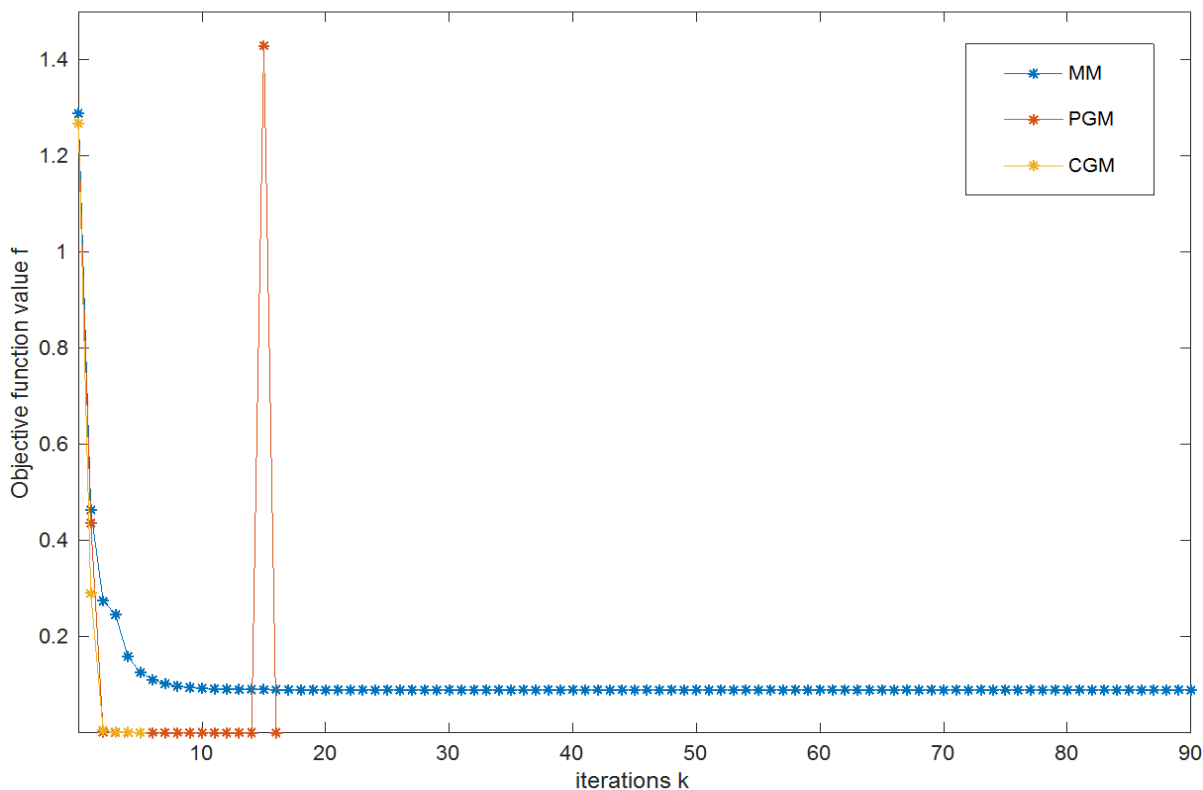


始点对三种算法的迭代次数, 迭代时间和最终目标函数值的影响。可以看出, 不同的初始点对 MM 方法有很大的影响, PGM 也显示出一定的波动性。在这种情况下, 本文提出的算法 2 具有良好的性能, 影响最小。

**Table 1.** Comparison table of solving efficiency of three methods for more special initial points

**表 1.** 较特殊初始点的三种方法求解效率比较表

问题	初始点	MM			PGM			CGM		
		迭代数	时间	$\ f(x)\ $	迭代数	时间	$\ f(x)\ $	迭代数	时间	$\ f(x)\ $
1	$(\pi, -\log 6)$	8	0.020s	0	30	0.020s	0	4	0.080s	0
	$(10^8, 10^9)$	>500	NaN	NaN	5	0.020s	0	10	0.050s	0
	$(0, 0)$	80	0.110s	0	1	0.000s	0	1	0.020s	0
2	$(\pi, -\log 6)$	Inf	Inf	NaN	4	0.050s	0	2	0.100s	0
	$(10^8, 10^9)$	Inf	Inf	NaN	6	0.020s	0	2	0.070s	0
	$(0, 0)$	116	0.110s	9.1e-06	2	0.010s	0	1	0.050s	0



**Figure 4.** The change of the objective function value of different algorithms in the iterative process

**图 4.** 不同算法在迭代过程中目标函数值的变化情况

#### 4.2. 不同步长选取策略

算法中步长的选取方式有所不同[25], 本文主要考虑如下求解步长的方法

- 1) 经典条件梯度法步长:  $\lambda_{\omega} = 2/k + 2$
- 2) 简单平均步长:  $\lambda_{\alpha} = 1/k + 1$
- 3) 常数步长:  $\lambda_c = \hat{\lambda}$

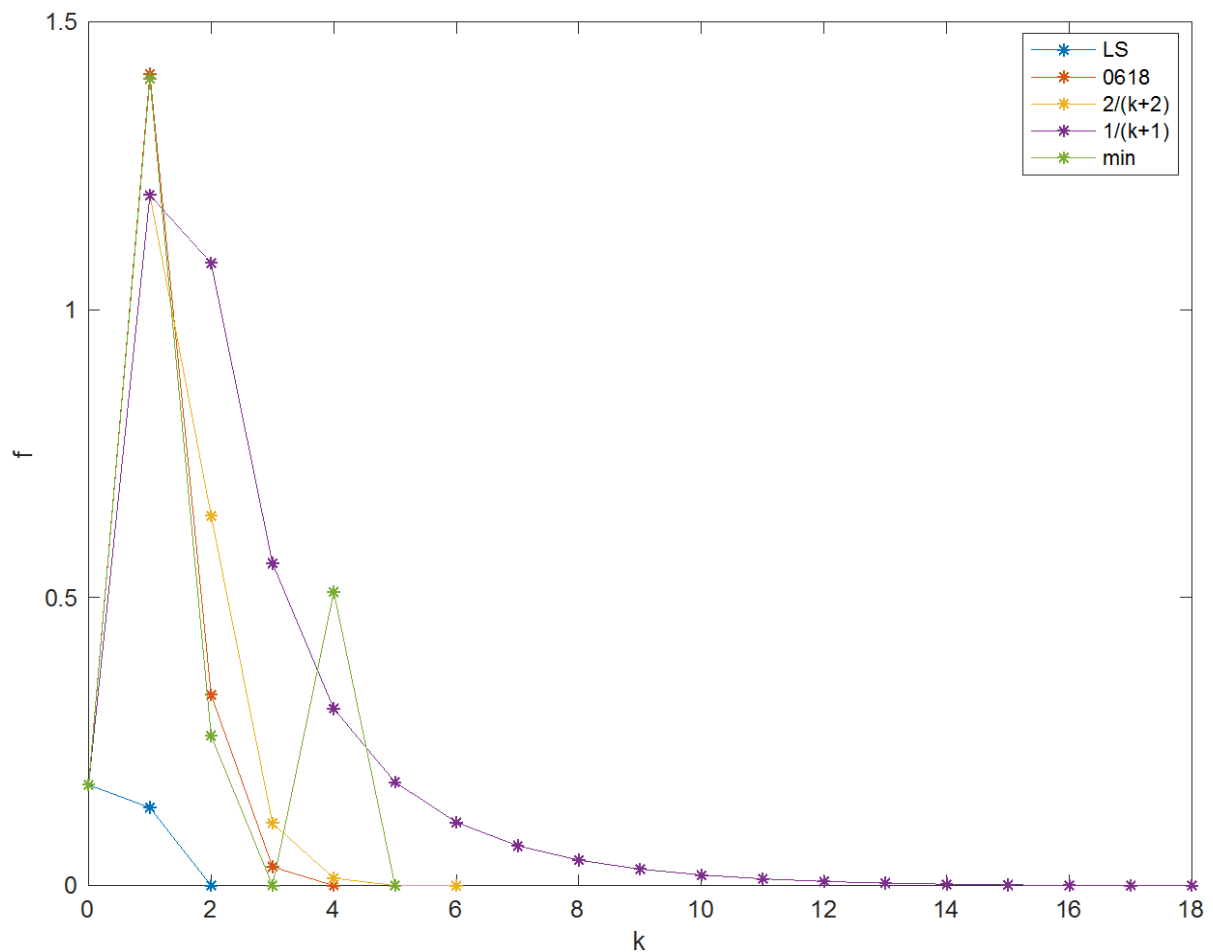
4) 线搜索步长(算法 2):  $\lambda_{LS}$

5) 最小步长选取法:  $\lambda_{\min} = \left\{ 1, \frac{(x^k - s^k)^T \nabla f(x^k)}{2 \|x^k - s^k\|^2} \right\}$

我们分析了五种步长选择策略在非线性分裂可行问题中的表现情况, 不同步长下的迭代次数比较表如表 2 所示, 图 5 则对比了迭代过程中目标函数的变化过程情况。

**Table 2.** Comparison table of various algorithm values under different step length  
**表 2.** 不同步长下算法各项数值的比较表

步长	初始点 $x_0 = rand(2,1)$		
	迭代数	时间	$\ f(x)\ $ & $\ \nabla f(x)\ $
$\lambda_w$	6	0.1400s	0/0
$\lambda_a$	18	1.2400s	0/0
$\lambda_c$	4	0.1100s	0/0
$\lambda_{LS}$	2	0.0060s	0/0
$\lambda_{\min}$	5	0.1500s	0/0



**Figure 5.** Schematic diagram of the iterative process under asynchronous length  
**图 5.** 不同步长下迭代过程示意图

## 5. 结论

本文主要研究利用带线搜索的条件梯度法求解非线性分裂可行问题, 根据理论和数值实验的结果, 算法在迭代次数、运算时间和实现的可行性方面, 优于其他投影类算法, 并且灵活性更强。本文仅使用 MATLAB 中的优化工具箱, 没有针对本算例使用特定的方法求解, 在求解迭代方向时, 可以运用不同的方法, 以实现对不同问题的更有效解决方案。对于步长的线搜索, 本文使用 Armijo 线搜索。在以后的研究中, 对于更复杂的高纬度问题, 可尝试使用非单调线搜索来避免迭代向量非常接近集合的交点时出现的折线步太多的情况。

## 参考文献

- [1] Censor, Y. and Elfving, T. (1994) A Multiprojection Algorithm Using Bregman Projections in a Product Space. *Numerical Algorithms*, **8**, 221-239. <https://doi.org/10.1007/BF02142692>
- [2] Censor, Y., Elfving, T., Kopf, N. and Bortfeld, T. (2005) The Multiple-Sets Split Feasibility Problem and Its Applications for Inverse Problems. *Inverse Problems*, **21**, 2071-2084. <https://doi.org/10.1088/0266-5611/21/6/017>
- [3] Censor, Y., Bortfeld, T., Martin, B. and Trofimov, A. (2006) A Unified Approach for Inversion Problems in Intensity-Modulated Radiation Therapy. *Physics in Medicine & Biology*, **51**, 2353-2365. <https://doi.org/10.1088/0031-9155/51/10/001>
- [4] Byrne, C. (2003) A Unified Treatment of Some Iterative Algorithms in Signal Processing and Image Reconstruction. *Inverse Problems*, **20**, 103-120. <https://doi.org/10.1088/0266-5611/20/1/006>
- [5] Xu, H.-K. (2014) Properties and Iterative Methods for the Lasso and Its Variants. *Chinese Annals of Mathematics, Series B*, **35**, 501-518. <https://doi.org/10.1007/s11401-014-0829-9>
- [6] He, H.J. and Xu, H.-K. (2017) Splitting Methods for Split Feasibility Problems with Application to Dantzig Selectors. *Inverse Problems*, **33**, Article ID: 055003. <https://doi.org/10.1088/1361-6420/aa5ec5>
- [7] Byrne, C. (2002) Iterative Oblique Projection onto Convex Sets and the Split Feasibility Problem. *Inverse Problems*, **18**, 441-453. <https://doi.org/10.1088/0266-5611/18/2/310>
- [8] Guélat, J. and Marcotte, P. (1986) Some Comments on Wolfe's "Away Step". *Mathematical Programming*, **35**, 110-119. <https://doi.org/10.1007/BF01589445>
- [9] Shang, M. and Zhang, C. and Xiu, N.H. (2014) Minimal Zero Norm Solutions of Linear Complementarity Problems. *Journal of Optimization Theory and Applications*, **163**, 795-814. <https://doi.org/10.1007/s10957-014-0549-z>
- [10] Byrne, C., Censor, Y., Gibali, A. and Reich, S. (2012) The Split Common Null Point Problem. *Journal of Nonlinear and Convex Analysis*, **13**, 759-775.
- [11] Masad, E. and Reich, S. (2007) A Note on the Multiple-Set Split Convex Feasibility Problem in Hilbert Space. *Journal of Nonlinear and Convex Analysis*, **8**, 367-371.
- [12] Krasnoselskii, M.A. (1955) Two Remarks about the Method of Successive Approximations. *Uspekhi Matematicheskikh Nauk*, **10**, 123-127.
- [13] Robert, M.W. (1953) Mean Value Methods in Iteration. *Proceedings of the American Mathematical Society*, **4**, 506-510. <https://doi.org/10.1090/S0002-9939-1953-0054846-3>
- [14] Xu, H.-K. (2006) A Variable Krasnosel'skii-Mann Algorithm and the Multiple-Set Split Feasibility Problem. *Inverse Problems*, **22**, 2021. <https://doi.org/10.1088/0266-5611/22/6/007>
- [15] Li, Z.B., Han, D.R. and Zhang, W.X. (2013) A Self-Adaptive Projection-Type Method for Nonlinear Multiple-Sets Split Feasibility Problem. *Inverse Problems in Science and Engineering*, **21**, 155-170. <https://doi.org/10.1080/17415977.2012.677445>
- [16] Gibali, A., Küfer, K.H. and Süß, P. (2014) Successive Linear Programming Approach for Solving the Nonlinear Split Feasibility Problem. *Journal of Nonlinear and Convex Analysis*, **15**, 345-353.
- [17] Xu, J., Chi, E.C., Yang, M. and Lange, K. (2018) A Majorization-Minimization Algorithm for Split Feasibility Problems. *Computational Optimization and Applications*, **71**, 795-828. <https://doi.org/10.1007/s10589-018-0025-z>
- [18] 畅含笑, 孙军, 屈彪. 带稀疏约束的分裂可行问题的算法[J]. *应用数学进展*, 2016, 5(2): 269-275.
- [19] Censor, Y., Gibali, A. and Reich, S. (2012) Algorithms for the Split Variational Inequality Problem. *Numerical Algorithms*, **59**, 301-323. <https://doi.org/10.1007/s11075-011-9490-5>
- [20] Frank, M. and Wolfe, P. (1956) An Algorithm for Quadratic Programming. *Naval Research Logistics Quarterly*, **3**,

- 
- 95-110. <https://doi.org/10.1002/nav.3800030109>
- [21] Freund, R.M. and Grigas, P. (2016) New Analysis and Results for the Frank-Wolfe Method. *Mathematical Programming*, **155**, 199-230. <https://doi.org/10.1007/s10107-014-0841-6>
- [22] Goncalves, M.L.N. and Melo, J.G. (2017) A Newton Conditional Gradient Method for Constrained Nonlinear Systems. *Journal of Computational and Applied Mathematics*, **311**, 473-483. <https://doi.org/10.1016/j.cam.2016.08.009>
- [23] Jaggi, M. (2013) Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization. *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, 17-19 January 2013, 427-435.
- [24] Hearn, D.W. (1982) The Gap Function of a Convex Program. *Operations Research Letters*, **1**, 67-71. [https://doi.org/10.1016/0167-6377\(82\)90049-9](https://doi.org/10.1016/0167-6377(82)90049-9)
- [25] Bauschke, H.H. and Borwein, J.M. (1996) On Projection Algorithms for Solving Convex Feasibility Problems. *SIAM Review*, **38**, 367-426. <https://doi.org/10.1137/S0036144593251710>